SCIENCE AND TECHNOLOGY METRICS

BY

Dr. Ronald N. Kostoff
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217
Phone: 703-696-4198
Fax: 703-696-3098
Internet: kostofr@onr.navy.mil

(*THE VIEWS IN THIS REPORT ARE SOLELY THOSE OF THE AUTHOR AND DO NOT REPRESENT THE VIEWS OF THE DEPARTMENT OF THE NAVY*)

I. ABSTRACT

This document describes the rationale for, and implementation of, the expanded use of the proper metrics in the evaluation of science and technology (S&T). The document starts with an Executive Overview and Conclusions regarding the application of metrics to the entire S&T development cycle, including its key role in setting incentives for S&T development. Then, after describing how the evolution of S&T has influenced the present burgeoning interest in quantitative S&T metrics, this monograph defines different types of S&T metrics, followed by the main principles of high quality metrics-based S&T evaluations. After a broad overview of quantitative approaches to research assessment, the document focuses on the main approaches of bibliometrics and econometrics, including a novel section on bibliometric collaboration indicators. It then describes the bibliometrics-related family of approaches known as co-occurrence phenomena, describes a network modeling approach to quantifying research impacts, and ends the main text body with a description of a metrics-based expert systems approach for supporting research assessment.

There are a substantial number of Appendices that make the present document essentially a self-contained monograph. Appendix 12 contains extensive data describing the infrastructure of the S&T metrics literature (including the seminal documents in S&T metrics), and it is followed by a very extensive Bibliography that contains over 7500 key references in S&T metrics. The Bibliography includes both those specific references identified in the body of this document's text, and suggestions for further reading in this broad technical area.

KEYWORDS: science and technology; metrics; research assessment; bibliometrics; scientometrics; cost-benefit; econometrics; co-occurrence; network modeling; research impact; expert systems; rate of return; citation analysis; co-word; co-citation, discovery, innovation.

TABLE OF CONTENTS

I. ABSTRACT
I-A. EXECUTIVE OVERVIEW AND CONCLUSIONS.
II. INTRODUCTION/ BACKGROUND
III. DEFINITIONS/ PRINCIPLES OF HIGH QUALITY METRICS
IV. SCIENCE AND TECHNOLOGY METRICS
V. APPENDICES
APPENDIX 1: METRICS IN SUPPORT OF PEER REVIEW
1-A: PEER REVIEW: THE APPROPRIATE GPRA METRIC FOR RESEARCH
1-B: METRICS FOR PEER REVIEW OF BASIC AND APPLIED RESEARCH
1-C: METRICS FOR PEER REVIEW OF ADVANCED TECHNOLOGY DEVELOPMENT
APPENDIX 2: THE UNDER-REPORTING OF RESEARCH IMPACT
APPENDIX 3: UTILITY OF CITATION ANALYSES
3-A: CHARACTERISTICS OF HIGHLY-CITED AND POORLY-CITED PAPERS
3-B: CITATION ANALYSIS OF RESEARCH PERFORMER QUALITY
3-C: CITATION DIFFERENTIALS
APPENDIX 4: DISPLAY OF BIBLIOMETRICS RESULTS
APPENDIX 5-A: CITATION NORMALIZATION APPROACHES
APPENDIX 5-B: CITATION ANALYSIS CROSS-FIELD NORMALIZATION: A NEW PARADIGM
APPENDIX 5-C: IS CITATION NORMALIZATION REALISTIC?
APPENDIX 5-D: CAB – CITATION-ASSISTED BACKGROUND
APPENDIX 6: THE PIED PIPER EFFECT: A SPECIFIC EXAMPLE
APPENDIX 7: EXAMPLES OF S&T BIBLIOMETRICS STUDIES
7-A: FULLERENES RESEARCH
7-B: AIRCRAFT
7-C: ANALYTICAL CHEMISTRY
7-D: ELECTRIC POWER SOURCES
7-E: ELECTROCHEMICAL POWER SOURCES
7-F: NONLINEAR DYNAMICS
7-G: FRACTALS
7-H: CITATION MINING-GRANULAR SYSTEM DYNAMICS
7-I: CITATION MINING-MACROMOLECULAR MASS SPECTROMETRY
APPENDIX 8: SCIENCE AND TECHNOLOGY TRANSITIONS
APPENDIX 9-A: NETWORK MODELING FOR DIRECT/INDIRECT IMPACTS
APPENDIX 9-B: NETWORK MODELING FOR ROADMAPS
APPENDIX 10: EXPERT NETWORKS
APPENDIX 11: POTENTIAL USE OF ENTROPY IN RESEARCH EVALUATION
APPENDIX 12: INFRASTRUCTURE OF S&T METRICS TECHNICAL LITERATURE
VI. BIBLIOGRAPHY

I-A. EXECUTIVE OVERVIEW AND CONCLUSIONS

I-A-1. Introduction

The products of science and technology (S&T) underpin modern economies and defense capabilities. Government and industry provide the bulk of resources for S&T development, with government supplying the majority of basic science resources, and industry contributing substantial resources to more advanced technology development. In both organizations, S&T accountability procedures have become more requested, more visible, more frequent, and more formal. Questions persist about the most credible methods for insuring accountability to satisfy a variety of stakeholders.

Peer review, the expert judgment by research specialists, has been the traditional method used for S&T accountability [Kostoff, 2004q]. Performance metrics (the counting of research activity, outputs, impacts, and quantification of outcomes) tend to be advocated by S&T decision-makers who may not be technical specialists, but want independent credible measures of S&T quality and progress that could support resource allocation decisions. The consensus of most of the S&T community is that peer review is the preferred approach to be used for S&T accountability (evaluation/ assessment), strongly supported by the use of 'appropriate metrics' [Kostoff, 1997a]. However, the selection of 'appropriate metrics' remains an outstanding issue. The present document aims to provide some insight to the role of metrics in the S&T accountability process, and the criteria for selection of metrics most appropriate to the problems being addressed. ***In particular, because S&T metrics can serve as S&T development incentives, the present document highlights the positive and negative intended and unintended consequences for S&T that could result from incorrect selection of S&T metrics.***

The remainder of this Executive Overview describes

- S&T Accountability
- Effects of S&T expenditures
    - Structure
    - Flow
- Attributes of S&T Metrics
    - Qualitative/ Quantitative Metrics
    - Prospective/ Retrospective Metrics
- Impact of Metrics Selection on Strategic Management
- Unintended Negative Consequences from Metrics Selection
- Re-Balancing Quantitative and Qualitative Metrics

I-A-2. S&T Accountability

What is S&T accountability, how is it performed, and how does it relate to metrics?

The S&T enterprise can be viewed from a decision-consequences perspective as having two major components: 1) expenditure of S&T funds, and 2) the S&T-related effects resulting from

those expenditures. S&T accountability is the **identification and assessment/ evaluation** of the S&T-related effects resulting from the S&T expenditures. S&T accountability is performed through evaluations/ assessments of the expenditures and resulting effects by a combination of 1) experts in the relevant S&T disciplines and 2) experts in technical and mission areas impacted by the S&T under evaluation. Metrics are the instruments that enable the identification and assessment/ evaluation of the S&T-related effects. The challenge is to identify the suite of metrics instruments that will enable credible accountability without being overly burdensome, unwieldy, or expensive.

I-A-3. Effects of S&T Expenditures

The effects of S&T expenditures can be classified into two major categories: 1) structure, and 2) flow. The structure represents characteristic features of the S&T being conducted (e.g., merit, approach, team, risk, status), while the flow can be conceptualized as the flux of product from the S&T being conducted (e.g., activity, output, impact, outcome). The challenge mentioned above translates into selecting metrics (aka evaluation criteria) that will provide adequate resolution into the nature of the structural and flow effects of the S&T expenditures.

I-A-3-i. Structure

The structure category contains all the non-flow characteristic features of the S&T resulting from the S&T expenditures. How many, and what types of, evaluation criteria are required to provide adequate insight to the structural characteristics of the project/ program being reviewed? Large numbers of criteria become unwieldy operationally, provide excessive resolution, and mask/ dilute the major insights and findings from the review. Very small numbers of criteria provide inadequate insight/ resolution to the project's/ program's structure to identify and understand any specific structural problems that exist, and are inadequate for program/ project management purposes.

A minimum set of evaluation criteria for the structure category that balances adequate insight/ resolution with operational flexibility consists of the following five criteria: 'merit', 'approach', 'team quality', 'risk', 'status' [Kostoff, 1997n]. These criteria are differentiated chronologically by the S&T development cycle stage in which they first exert influence on the decision-making process (planning -→ portfolio selection -→ review -→ transition), as follows.

'Merit' addresses the importance of the S&T being reviewed to both the larger S&T community and the sponsoring organization's mission, specifically, whether the appropriate overall objectives (in the context of the sponsoring organization's mission) are being pursued by the project/ program under review. The focus is on S&T and mission end goals, not on approach. 'Merit' exerts influence on the decision-making process starting at the earliest stages of S&T planning, and continues to exert influence on the portfolio selection, review, and transition stages. Examples of 'merit' metrics could include research merit, mission relevance, etc.

'Approach' addresses the conduct of the S&T project/ program, specifically whether the conduct will lead to attainment of the specified S&T project/ program goals and objectives. 'Approach' exerts influence on the decision-making process at the portfolio selection stage, and continues to

exert influence on the review and transition stages. Examples of 'approach' metrics could include balance between experiment and theory, balance between resources and objectives, state-of-the-art of instrumentation, coordination with other organizations, etc.

'Team Quality' addresses the competence of the people who manage and perform the S&T. 'Team Quality' exerts influence on the decision-making process at the portfolio selection stage, and continues to exert influence on the review and transition stages. Examples of 'team quality' metrics could include team publication quality, citations, awards, honors, etc.

'Risk' addresses the degree of certainty that the S&T project/ program will achieve its stated goals and objectives, and has some relation to the quality of S&T performers and the approach selected. 'Risk' exerts influence on the decision-making process at the portfolio selection stage, and continues to exert influence on the review and transition stages. Examples of 'risk' metrics could include probability of achieving S&T objectives, probability of impacting long-range mission, probability of successful demonstration, etc.

'Status' addresses the progress that has been made on the S&T development, and has some relation to the quality of S&T performers and approach, and to risk. 'Status' exerts influence on the decision-making process at the review stage, and continues to exert influence on the transition stages. Examples of 'status' metrics could include technology readiness level, objectives completed, technical milestones completed, etc.

I-A-3-ii. Flow

The flow category contains all the S&T product-related effects resulting from the S&T expenditures. These product-related effects can be classified into four categories (activity, output, impact, outcome), differentiated by their temporal distance from the time the S&T funds were expended.

'Activity' reflects the S&T infrastructure generated from the initial S&T expenditures. It starts immediately after the portfolio selection stage, and continues through all successive stages. The 'activity' is under direct control of the S&T resources recipient. Examples of 'activity' metrics could include numbers and types/ quality of people conducting the S&T, numbers and types/ quality of equipment used in the S&T, and numbers and types/ quality of facilities used for the S&T. There is some overlap between the 'team quality' criterion used for structure evaluation in the review and transition stages, and the people quality component of 'activity'.

'Output' reflects the initial products from the S&T under review. It starts well after the portfolio selection stage, continues through the review and transition stages, and may continue even after transition due to long lag times. The 'output' is under direct control of the S&T resources recipient. Examples of 'output' metrics could include numbers and quality of journal papers, numbers and quality of patents, numbers and quality of presentations, etc.

'Impact' reflects the influence of the S&T under review on the external S&T and potential user communities. It starts typically years after the initiation of 'output', and can continue years after transition (decades in some cases). The 'impact' is not under the control of the S&T resources

recipient, but rather under external control, typically (but not exclusively) by other members of the S&T community. Examples of 'impact' metrics could include numbers and quality of paper and patent citations, numbers and quality of awards/ honors, numbers and quality of downstream development plans altered due to S&T outputs, etc.

'Outcome' reflects the far downstream effects of the S&T under review on the larger scale societal goals. It starts well after transition, perhaps even decades afterwards, and can continue for many years/ decades. The 'outcome' is not under the control of the S&T resources recipient, but rather is impacted by changing user interests, environmental, political, financial, legal, international, and other non-technical considerations. Examples of 'outcome' metrics could include lives saved, cost savings, increased capability, improved rate of return, improved quality of life, etc.

'Activity', 'output', and 'impact' are (mathematically) products of quantity times unit quality. Thus, if publication outputs are being evaluated, not only are the numbers of publications important, but the quality of each publication is important as well. These three flow criteria can be separated mathematically into their quantity and quality components for simple estimations, but for credible S&T evaluation, the quantity-quality product is required. Since 'outcomes' tend to be fewer in number than the above three flow quantities, but larger in magnitude of effect, separation of 'outcomes' into quantity and quality components is not useful. Detailed analyses by experts are required for credible 'outcome' results.

The categories of structure and flow effects (potential evaluation criteria categories) resulting from S&T expenditures have been defined, and some metrics examples have been provided. The question now arises as to the intrinsic properties of these metrics, and how these properties affect operational use of the metrics.

I-A-4. Attributes of Metrics for S&T Evaluation

S&T metrics have two fundamental intrinsic characteristics that span the 'objectivity-time' continuum. The 'objectivity' characteristic ranges from objective (quantitative, machine-supplied data) to subjective (qualitative, human-supplied data). The temporal characteristic ranges from retrospective (looking backward in time) to prospective (looking forward in time). Each of these two intrinsic characteristics will be discussed in more detail.

I-A-4-i. Qualitative/ Quantitative Metrics

The two fundamental approaches to S&T evaluation, peer review and performance metrics, use two intrinsically different types of metrics. Peer review uses qualitative (subjective) metrics, and performance metrics uses quantitiative (objective) metrics. Both types of evaluation also use metrics that are a hybrid of qualitative and quantitative. Purely qualitative metrics use data supplied by humans. These subjective types of data are typically judgments of items (e.g., manuscript quality, level of project risk, degree of project innovation, level of project technological readiness, quality of researchers, etc). Purely quantitative metrics use data supplied by machine, with minimal human assumptions. These objective types of data are typically counts of items (e.g., numbers of papers, numbers of patents, numbers of transitions,

numbers of researchers, revenues generated, etc).

Hybrid metrics use data supplied by machine, supplemented by substantial human judgment on which machine data is selected for analysis and how the machine data is aggregated to quantify the metric. The people who perform the data selection and aggregation to quantify the hybrid metrics require substantial knowledge of the underlying S&T, and perhaps business, marketing, and application data as well, depending on the specific hybrid metrics selected. This is contrasted with the simple counting of papers, patents, citations used for the purely quantitative metrics, where many assumptions or much judgment are not required from the analyst, nor is any understanding about the underlying S&T required by the analyst, These objective/ subjective hybrid metrics are typically outcome-related (cost-benefit ratios, rates of return, cost savings, or their national security/ medical equivalents).

The subjective qualitative metrics applied to S&T evaluation today tend to have the following characteristics:

- More complex in concept than simple item counts
- More expensive to obtain
- More manually intensive, and less amenable to automation
- More training required for implementation and interpretation
- Less consistency across projects

The objective quantitative metrics used in S&T evaluation today have their origins in ***industrial-age production measures***. Quantitative metrics based on past data tend to involve quantity of S&T productivity counts. These types of productivity metrics are (relative to the subjective qualitative metrics):

- Simpler in concept
- Relatively inexpensive to obtain
- Easily amenable to automation
- Implemented and interpreted with minimal training

The criteria categories defined for structure (merit, approach, team, risk, status) tend to be qualitative metrics. The criteria categories defined for flow (activity, output, impact, outcome) tend to be 1) quantitative for the counting component of activity, output, and impact, 2) qualitative for the non-counting components of these criteria categories, and 3) hybrid for the outcomes. For both types of metrics, one important selection consideration today is minimal disruption to the organization's operations.

Both quantitative and qualitative metrics have different levels of certainty and credibility, depending on whether they use past, present, or future data. The relation between time perspective, credibility, and application will now be examined.

I-A-4-ii. Prospective/ Retrospective/ Present Metrics Utilization

Prospective use of metrics involves prediction/ estimation of the metrics' values at future points

in time. The uncertainty/ credibility associated with the metrics' values increases with the length of prediction/ estimation time. As an example, a cost-benefit estimate of market implementation in 2020 of products resulting from S&T performed today would be a prospective hybrid metric, with substantial uncertainty. As another example, the impact on quality of life in 2020 of S&T performed today would be a prospective qualitative metric, also with substantial uncertainty.

Conversely, retrospective use of metrics involves tabulation of the metrics' values from past points in time. Retrospective tabulation is an inherently more certain and credible process. As an example, the cumulative number of citations received over the past decade by papers published in the mid-1990s would be a retrospective quantitative metric, with a high degree of certainty. As another example, the impact on quality of life in 2000 of S&T performed in the 1960s would be a retrospective qualitative metric, again with relatively high certainty.

Finally, 'present' metrics involves specification of the metrics' values at the present time. As an example, the quality of the approach of an ongoing S&T project would be a present qualitative metric. As another example, the specific performance status today of a fighter aircraft prototype under development would be a present quantitative metric.

The rationale for, and value of, using metrics retrospectively, in the present, or prospectively depends on the intended application. Retrospective use of metrics tends to be valuable for:

- Generating lessons learned from past development
- Marketing based on actual achievements
- Identifying management environments conducive to successful development
- Rewarding personnel involved in successful development
- Accountability based on past performance

However, retrospective use of most quantitative metrics (e.g., number of citations recorded, number of awards received, amount of revenue generated) and qualitative metrics (e.g., quality of demonstrated impact on S&T, quality of awards, quality of life enhancement demonstrated) is of limited value for some S&T management purposes. These include program modifications (directions, budgets, personnel) to correct real-time performance problems, new program selection based on potential impact and payoff, and marketing based on potential payoff.

In particular, the availability of <u>impact</u> or especially <u>outcome</u> data resulting from S&T program execution typically occurs too far downstream from the S&T program initiation to influence future program execution (research direction, budgets, personnel). For example, paper citation data would not be available for credible evaluation purposes until at least six (or preferably more) years after an S&T project had been initiated, given the reality of publication delays for the initial published papers and for the subsequent citing papers. Market implementation data would not be available for one or two decades after S&T project initiation (for most technologies). These long time intervals between S&T program initiation and the availability of data for evaluation purposes preclude the use of this retrospective data to impact the original S&T program decision-makers or influence the S&T program's direction in a timely manner.

However, in special cases, use of short-term retrospective metrics (e.g., number and quality of

papers recently published, researchers recently hired) could provide timely data to partially influence program execution decisions. More importantly for this type of application (influence present program execution decisions) would be the use of 'present' metrics from recent peer reviews (e.g., research team quality, research approach quality, progress status, technology readiness distribution and associated quality of distribution bands, etc). Having this current data would insure that actions taken to correct problems (with the S&T project) identified by the evaluation/ metrics would be applied to the people, allocations, and budgets responsible for those problems.

Prospective use of quantitative and qualitative metrics (e.g., estimated impact on S&T, estimated sales revenue streams, estimated operational cost savings, estimated quality of life enhancement, estimated increase in military capabilities) is quite valuable for some of the applications unavailable to retrospective use of metrics, including:

- new program selection based on potential impact and payoff, and
- marketing based on potential payoff, especially marketing at early stages of the S&T development

Unfortunately, the data generated prospectively are far more uncertain than the retrospective data. Prospective S&T metrics data should be generated by researchers with a thorough understanding of the S&T at all phases of its proposed evolution trajectory from the present to its future estimation point, if such data are to have credibility.

If selected and applied properly, metrics can be of substantial benefit for strategic management (and marketing) at all stages of the S&T development cycle shown above. But what is the relation between selection of S&T metrics and strategic management of S&T development?

I-A-5. Impact of Metrics Selection on Strategic Management

In many research project/ program evaluations, 'productivity' (in the broader context of including all the 'flow' categories defined previously) assumes a central role, and in a very real sense is where the 'rubber meets the road'. Not only are the **numbers** of 'activities', 'outputs', 'impacts', and 'outcomes' important for determining 'productivity', but the **quality** of these 'productivity' items is equally or more important. Unfortunately, *there is a severe imbalance today between the use of retrospective quantitative and qualitative indicators in the reporting of S&T ' roductivity'*. Due to the simplicity and other advantages of obtaining the retrospective activity/ output/ impact quantitative vs qualitative metrics data shown above, much of S&T 'productivity' reported today is quantitative alone. This can have many **negative unintended consequences**. The following sections relate these consequences to the types of metrics used, and how the metrics can be selected to both **minimize the negative unintended consequences** and **promote positive intended consequences**.

In practice, one strong reason for the selection of the simple retrospective quantitative 'productivity' metrics is to make minimal time demands on sponsor organization Program Officers and field Research Performers. To accomplish this arbitrary (but understandable) objective of minimal intrusion on the organization's operations, the data available from ordinary

organizational business operations becomes the major source of data to populate the metrics. The available data from organizational business operations thus serves as the pro forma driver for determining the metrics to be used, which in turn determines the objectives whose S&T progress will be gauged by the metrics. **This is the reverse of what would be desired from strategic management of S&T**:

- Set objectives for desired outputs and outcomes,
- Define metrics that would gauge S&T progress toward meeting these objectives,
- Determine the data required to populate these metrics.

What are the consequences to S&T development of available organizational business data determining the metrics selected for evaluation?

I-A-6. Unintended Negative Consequences from Metrics Selection

For data gathering in physical, environmental, engineering, and life sciences applications, care is taken to insure that the measuring instruments have minimal impact on the state of the system being measured. Except for the fundamental limitations on measurement precision imposed by Heisenberg's Uncertainty Principle, which becomes of concern only at very small scales, these instruments are becoming more able to exert minimal influence on states of systems being measured.

For the S&T development cycle, the situation is intrinsically different. The metrics employed have the potential to influence the S&T development trajectory. Additionally, they have the potential to serve as incentives and thereby distort the development results and objectives severely, sometimes in very unintended directions. In particular, if production-based 'productivity' metrics are perceived by the S&T sponsors and performers to be the dominant form used for S&T evaluation, the incentives for S&T sponsors and performers alike are to:

- Alter the types of S&T performed,
- Alter the types of S&T documents produced

to maximize output **quantity**. These distorted incentives lead to ***negative unintended consequences***. Weingart [2005] summarizes a few of these ***negative unintended consequences*** as follows:

- Increase publication counts by fragmenting articles. An upcoming publication by the author confirms this phenomenon of 'paper inflation' for a specific technical discipline.
- Propose conservative but safe research projects. The objective here is to minimize the risk of failure, and insure the continual supply of publications.
- Increasing publication quantity at the expense of quality. This is especially true when quality metrics are not included in the measurement suite.
- Increasing bias toward short-term performance as opposed to long-term research capacity.
- Increasing bias toward conventional approaches.

Perhaps the most serious negative impact of expanded use of conventional production-based 'productivity' metrics in the S&T development cycle would be the strong ***negative incentives*** provided for radical discovery and innovation, **counter to the recommendations in the recent National Innovation Initiative Report of the Council on Competitiveness [NIIR, 2004] to strongly promote this type of radical discovery and innovation**. As shown in a recent report by the present author on radical discovery and innovation [Kostoff, 2005a], much of truly radical discovery and innovation will involve cross-discipline extrapolation of concepts. Unfortunately, very strong negative incentives exist for cross-disciplinary or inter-disciplinary research [Kostoff, 2002g]:

- Much time is required for the performers to learn multiple disciplines or new disciplines, leaving less time for publishing, and reducing publication (and patent) outputs;
- Much time is required for coordinating and synchronizing research across disciplines, subtracting time that could be devoted to generating publications and other outputs;
- Journal review of trans-discipline manuscript submission is much more difficult, resulting in higher manuscript rejection rates, and reducing publication outputs;
- Grants are more difficult to obtain because of the trans-disciplinary review problem, reducing metrics based on research support funds obtained;
- All these effects impact tenure and honors/ awards negatively, reducing metrics based on achievements.

In the same vein, an insightful Washington Post article of 10 April 2005 by Rick Weiss decried the decline of non-applied curiosity-driven research. Mr. Weiss attributes the reason for the decline to 'deliverables' (e.g., specific products, profits, outcomes) becoming the dominant force driving research agendas.

When the three perspectives of this section (Weingart, Kostoff, Weiss) are integrated, a clear picture emerges. Most, if not all, quantitative metrics that are perceived by the research performers to influence research funding and research rewards will, on average, steer the research away from curiosity-driven multi/ inter-disciplinary high-risk research characteristic of radical discovery and toward mono-disciplinary low-risk low-discovery research. When meeting quantitative milestone and deliverable targets becomes the de facto research goal, any diversions that cost time and involve risk will of necessity be eliminated.

What can be done to counter these negative incentives for radical discovery and innovation?

I-A-7. Re-Balancing Quantitative and Qualitative Metrics

To correct today's de facto imbalance of quantitative to qualitative retrospective 'productivity' metrics, qualitative metrics should be added to the suite of criteria used for S&T evaluation. These metrics could include (but not be limited to): Innovation Potential, Creativity, Discovery Potential, Originality, Level of Risk, Probability of Success, Potential for Mission Impact, Research Merit, Research Approach Quality, Potential for Transition, Program Executability, Team Quality, Technology Readiness Level, Exploitation of External S&T, Leveraging of External S&T.

(It should be noted that, for inclusion of more qualitative metrics in the suite of evaluation instruments/metrics, there is no guarantee that the present desire for minimal disruption of research sponsors and performers during the evaluation process will be achieved. Additionally, for inclusion of either quantitative or qualitative metrics that have been determined starting from objectives and goals rather than available organizational business operations data, there is also no guarantee that the present desire for minimal disruption of research sponsors and performers during the evaluation process will be achieved.)

A general rule for metrics selection to insure some minimum balance between quantitative and qualitative productivity metrics is that ***every purely quantitative productivity metric should be accompanied by one or more qualitative metrics***. Thus, if one output displayed is number of journal papers produced, the quality of those papers should be added (see [Kostoff, 1997n-Attachment 19] for a cost-efficient method to obtain this data using existing journal review procedures). If one output is number of transitions produced, the quality and potential impact of those transitions should be added (see [Kostoff, 2004o] for methods of obtaining this data at different levels of accuracy). If one output is number of researchers developed, the quality of those researchers should be added. The necessity for the performers to now maximize the quantity-quality product, rather than to maximize the quantity only, will lead to different more desirable types of research.

Thus, selection of appropriate metrics to use for the S&T development cycle will involve a tradeoff among 1) providing positive incentives for meeting organizational and national objectives; 2) cost savings and improved quality due to increased accountability; and 3) the full costs of implementation. It is highly recommended that, before implementing specific metrics for application to any part of the S&T development cycle, an organization should identify and evaluate the ***intended and unintended*** consequences of the specific metrics' implementation, and identify the impact of these consequences on the organization's core mission.

II. INTRODUCTION/ BACKGROUND

II-A. Introduction

This document describes the rationale for, and implementation of, the expanded use of proper metrics in the evaluation of science and technology (S&T). The present section of this document (section II) describes the evolution of S&T, especially research, from a rich man's pastime to a major government enterprise. This historical background is necessary to provide the context for the present burgeoning interest in quantitative research metrics. Specifically, the background section describes:

- the linkages between research and technology advances;
- the reasons for the decline of industrial research and the concomitant growth of government research;
- the parallel increase of both research accountability and the use of quantitative measures in the research accountability process;
- the problems of relating these quantitative research metrics to research value; and
- the lack of a systematic approach to tracking and collecting this raw research benefit data, and the subsequent under-reporting of the impact of research.

The next section (section III) defines research metrics, and then categorizes the types of research metrics with the following generic taxonomy:

- Direct S&T Metrics - Input/ Output/ Productivity
- S&T Metametrics - Near-Term - Impact
- S&T Metametrics - Long-Term - Impact/ Outcome

Section III also lists the principles of high quality metrics-based R&D evaluations. These principles address:

- the commitment of the evaluating organization's senior management to high-quality metrics-based S&T evaluations
- the assessment manager's motivation to perform a technically credible assessment
- the role and competency of technical experts in a metrics-based S&T evaluation
- criteria for metric selection
- THE NECESSITY FOR EVERY S&T METRIC, AND ASSOCIATED DATA,

PRESENTED IN A STUDY OR BRIEFING TO HAVE A DECISION FOCUS, TO CONTRIBUTE TO THE ANSWER OF A QUESTION WHICH IN TURN WOULD BE THE BASIS OF A RECOMMENDATION FOR FUTURE ACTION

- reliability or repeatibility
- normalization and standardization across different science and technology areas
- global data awareness
- cost consciousness
- maintenance of high ethical standards throughout the process

Section IV, Science and Technology Metrics is, with the exception of the massive Bibliography, the longest and most detailed section of this monograph. After a broad overview of quantitative approaches to research assessment, this section focuses on the main approaches of bibliometrics and econometrics, then describes the bibliometrics-related family of approaches known as co-occurrence phenomena, then describes a network modeling approach to quantifying research impacts, and ends with a metrics-based expert systems approach for supporting research assessment.

Section V contains a substantial number of Appendices that make the present document essentially a self-contained monograph. Finally, section VI contains a very extensive Bibliography of key references in S&T metrics. It includes both those specific references identified in the body of this document's text, and suggestions for further reading in this broad technical area.

II-B. Background

Basic research provides the underpinnings for many of the technological advances of recent history, although there are examples where technology-driven needs (technology traction) motivate basic research as well. The evidence from many diverse retrospective studies, such as TRACES, Hindsight and DARPA accomplishments [IITRI, 1968; DOD, 1969; IDA, 1991; Kostoff, 1997n], strongly confirms the chains of strong linkages from basic research to technological innovations. Intuition then concludes that the economic benefits of these technological successes are attributable to their basic research origins. Unfortunately, the intuitive linkages between the cost of basic research and its eventual payoffs have been difficult to translate into convincing analytical arguments using present economic approaches.

From the private sector's perspective, basic research is very difficult to justify without substantial tax and other economic incentives. With non-negligible discount rates and long time spans between the research costs and eventual payoffs (for most, not all, research), benefit-cost ratios of most basic research computed using microeconomic analysis tend to be very small. In addition, the assumption that the organization conducting the basic research is the one that will receive the eventual (albeit small in a discounted sense) payoff may not be valid, in many cases.

For these economic reasons, industrial sponsorship of basic research throughout the world has had a long decline.

Historically, basic research has evolved from a rich man's pastime [SCIENCE, 1998] to industrial support to mainly government sponsorship. In the twentieth century, specifically during the period between the two World Wars, research funds were very limited worldwide. In many technical disciplines, European research surpassed that of the the United States. World War II changed this relationship, since the resources of Europe and Asia had to be devoted to conducting the War and rebuilding in its aftermath. The U.S. became the dominant industrial and government sponsor of basic research.

After WWII, U.S. companies had no serious competition in the world for two decades. Europe and the Pacific Rim had been destroyed by the war, and large U.S. companies gained both expansion and substantial profits due to lack of competition. They established corporate research centers as affordable luxuries for the following diverse reasons:

- for public relations purposes;
- because of liberal tax policies;
- as a method to attract and screen potentially bright new employees;
- as a vehicle to obtain rapidly expanding Federal research dollars;
- as a way of maintaining a window on the technological advances of their domestic and foreign competitors; and
- to develop new ideas that might eventually pay off for themselves.

As Europe and Asia recovered and became strong corporate competitors, the profitability and size of U.S. companies became more endangered. Many companies could no longer afford the luxury of basic research with its long and uncertain payoff horizon, and they closed their non-profitable research centers. Those that remained open focused their research to contribute more to short term profitability. Companies that have become absorbed in the recent trend toward deregulation and competition have shifted their basic research to the more focused side of the spectrum, since the relatively stable and secure regulated income that allowed such fundamental research no longer exists. The point here is that pure economics of increased domestic and world-wide competition drove U. S. domestic industry out of basic research, and the same competition drove much of foreign industry out of basic research as well.

After WWII, basic research support from the U.S. Federal government increased sharply. There are many reasons for this, the foremost being the recognition that basic research fuels the engines of innovation, and it is the government's role to insure the continuity of this fuel supply. In addition, the U.S. economy was expanding, and money was available for basic research without the need for detailed expository justification of its benefits. As the European and Asian economies rebounded after the War, their government-sponsored research increased as well.

Increasing global competition has had further impacts on the intrinsic structure of basic research. As the U. S. federal debt increased dramatically over much of the 1980s and 1990s, competition for federal funds became more severe. Basic research, with its long-term payoff horizon, now has to compete strongly with medicare, welfare, and other service provision and development programs. In Europe and Asia, basic research has undergone a similar transformation, with more of a strategic focus to the research.

In the U. S., the combination of a strong economy and weak inflation in the 1990s (and the rebounding economy of recent years) has kept interest rates low, and has shielded federal funds recipients from the full consequences of the large debt and other economic dislocations. In the research arena, NSF and NIH research budgets have increased, DOE and DOD budgets have not increased in real terms. Projections for overall Federal research funding, as reported in the media, are optimistic at present. Whether this stable overall support for research can be maintained indefinitely is, in the author's opinion, questionable. For a federal debt of five trillion dollars, even a one percent rise in interest rates would have a $50 billion dollar yearly impact on the federal budget, and would place all federal funds recipients in much greater jeapordy. A doubling of interest rates or worse, as occurred in the late 1970s/ early 1980s could have disasterous consequences for all federal recipients, especially those with long-horizon benefits such as research.

In this environment of scarce government funds, accountability of all government programs has increased substantially. For research to compete strongly for federal funds, the benefits from research need to receive full accounting and be articulated clearly. The implementation of the Government Performance and Results Act of 1993 (GPRA) [GPRA, 1993; Kostoff, 2004i], with its strong reliance on the use of metrics in S&T accountability, has begun to place even more emphasis on this research accounting requirement (See Appendix 1-A for an article in Science [Kostoff, 1997a] containing a summary description of the GPRA, and potential problems arising from strong reliance on S&T metrics).

There are two major characteristics of this increased accountability, whether from GPRA or other oversight sources: more detailed programmatic information is requested by the program assessors, and more quantified information is requested. What has motivated this dramatic increase in data requests? The upsurge in computer availability over the past decade has enabled large quantities of detailed information to be stored, tracked, and interpreted, and has driven the request for the large volumes of detailed program information. The request for increased quantitative information also derives from the increased computer capabilities for handling and analyzing large amounts of this type of data. In addition, there is substantial motivation from the assessors to have simple quantitative indicators that could drive the resource allocation process, and substantiate and justify the resource allocation decisions that are generated, rather than use the more complex and expensive and subjective qualitative peer review evaluation processes.

There are, however, substantial problems with the application of metrics to allocation decisions on proposed or continuing research. When a research unit is being evaluated, typically to ascertain whether its budget should be modified and/ or new research should be supported, there are three fundamental questions which are asked implicitly or directly: 1) What has been the

breadth of long-term impacts of research performed in the past; 2) What have been the accomplishments and impacts of research performed recently, and what are the estimated future societal impacts of this research; 3) What is the projected knowledge to be gained from proposed research, what types of benefits could be obtained if successful, and what is the confidence level that these different types of near and long-term payoffs will be realized?

The simple research output metrics, such as published papers and patents, can be easily quantified in the short term. However, they are intermediate measures only. The long-term benefit measures amenable to quantification, such as some societal outcomes or economic payoffs, cannot be generated in the short term. Because the research oversight organizations want valid performance metrics applicable to existing research (see question 2 in the previous paragraph), the question arises whether credible short term proxies for long-term research impacts and outcomes can be defined. Considerable research and correlation studies are necessary to produce credible answers to this question.

One final issue with appropriate use of S&T metrics concerns the systematic collection and tabulation of data required for their generation. The present informal and unstructured system for tracking and disseminating research products and downstream impacts has many deficiencies, resulting in a gross under-reporting of the broad range of research products, benefits and outcomes. Historically, there has been no central mechanism for documenting impacts, and no collective will among the federal agencies to expend the resources necessary. Thus, there exists a dual deficiency with respect to quantitative determination of research benefits. Not only are there deficiencies and limitations of how the metrics results are interpreted to translate to impacts and benefits, but there are major deficiencies in the tracking and collection of the raw data itself. Appendix 2 addresses this problem in more detail, and provides some potential solutions.

III. DEFINITIONS/ PRINCIPLES OF HIGH QUALITY METRICS

TABLE OF CONTENTS FOR SECTION III

III-A. DEFINITIONS

III-A-1. Overview
III-A-2. Taxonomy of S&T Metrics
III-A-2-i. Overview
III-A-2-i-a. Output vs Outcome Metrics
III-A-2-i-b. Normalized vs Un-normalized Metrics
III-A-2-ii. S&T Metrics Categories
III-A-2-ii-a. Direct S&T Metrics - Input/ Output/ Productivity
III-A-2-ii-b. S&T Metametrics - Near-Term - Impact
III-A-2-ii-c. S&T Metametrics - Long-Term - Impact/ Outcome

III-B. PRINCIPLES OF HIGH QUALITY METRICS-BASED S&T EVALUATIONS

III-B-1. Overview
III-B-2. Principles
III-B-2-a. Senior Management Commitment
III-B-2-b. Assessment Manager Motivation
III-B-2-c. Competency of Technical Evaluators
III-B-2-d. Criteria for Metric Selection
III-B-2-e. Relevance of Metric to Future Action
III-B-2-f. Reliability of Evaluation
III-B-2-g. Normalization Across Technical Disciplines
III-B-2-h Global Data Awareness
III-B-2-i. Cost of Metrics-based Evaluations
III-B-2-j. Maintenance of High Ethical Standards

III-A-1. Overview

The dictionary definition of a metric is a 'standard of measurement'. In physical science, a metric is used to quantify physical and tangible items (mass, size, etc.). For science and technology evaluation, metrics have a different meaning and application. Metrics selected for S&T evaluation derive from the intrinsic unique features of S&T products and outputs, and can include economic, financial, and other research environmental and management metrics.

What is the purpose of using S&T metrics? Metrics are not an end in themselves; they are a means to an end. Like any S&T management decision aid, their ultimate purpose is to support maximum acceleration of S&T progress efficiently, consistent with the mission of the sponsor's organization. Metrics support this objective by quantifying progress toward the S&T targets. Ideally, goals and objectives would be developed iteratively with their metrics to generate a final organizational strategic plan that explicitly or implicitly presents the metrics in parallel with the strategic goals and objectives. Any strategic or tactical S&T development plan whose strategic or tactical goals and objectives are not expressible in terms of quantifiable metrics should, except for extreme circumstances, be viewed with suspicion.

The author has assessed many strategic plans in many organizatios throughout the Federal government. In essentially every instance, goals that had no associated metrics were public relations creations, and were completely useless operationally. Additionally, perhaps the most valuable exercise from a strategic management perspective that the author has observed has been the transformation of strategic plans from metrics-free to metrics-bound. The organic understanding gained when re-structuring and re-framing the organization's goals to make them amenable to quantification is extremely beneficial, and can provide substantial insight to strengthening the organization's uni'ue mission in the context of related and parent organizations.

For basic research in particular, the goal is increased knowledge and understanding. These goals are ethereal multi-dimensional multi-faceted quantities, not amenable to direct measurements using today's technology. What can be measured directly are the various expressions and manifestations and embodiments of this knowledge, such as numbers of papers/ patents/ speeches. Because of the intrinsic complexity of knowledge, none of these relatively simplistic measures can serve as a valid stand-alone proxy metric for knowledge. Trying to portray knowledge through its metrics is analogous to portraying a scene through a portrait. Each brush stroke adds to the accuracy with which the scene is portrayed, but many brush strokes are necessary for even a moderately accurate reflection of the scene. With S&T metrics, combinations of metrics along with expert interpretation of their meaning are required to understand more fully both the output and short and long-term impacts of the knowledge generated from the S&T. But what are the different types of metrics that can be used for S&T?

III-A-2. Taxonomy of S&T Metrics

III-A-2-i. Overview

III-A-2-i-a. Output vs Outcome Metrics

There are a variety of S&T metrics commonly used. The simplest metrics (input/ output) relate to the time frame at or near when the research is performed, and the more complex metrics (impact/ outcome) relate to time frames further downstream. Consider the analogy of the research process to the nuclear fission process to help understand the intrinsic differentiation between these types of metrics.

In nuclear fission, neutrons interact with fissile material. The nucleus is fissioned (split) into energetic fission fragments and several neutrons, and other forms of radiation are generated as well. Under critical mass conditions, these fission-produced neutrons have further interactions with fissile and fertile material, generating more neutrons, more fission fragments, more radiation, and breeding more fissile material. The fissile material generated can then be either consumed in situ or separated out for future use, and the energy/ power from the fission reactions can be transferred to power converters to provide electricity and/ or heat. Additionally, either fission products, or neutron-irradiated stable target materials, can be used as beneficial radioactive isotopes (for food irradiation, nuclear medicine diagnostics, etc.).

Assume the fission process is analogous to the research process. The primary products of the fission process, fission fragments and neutrons and radiation, are the analogs of the primary products of the research process, papers and patents and students. These primary products in both cases are simple quantities, the results of a relatively few interactions that are easily trackable. The primary metrics of the fission process are the distribution functions which effectively 'count' the primary fission products, and the primary metrics of the research process are the distribution functions which count the primary research products of papers and patents.

In the fission example, the primary products, while important in describing the efficiency and other details of the focused fission process, serve as an intermediary. The main interest is in the downstream impacts and influence resulting from the fission process. Unlike the primary products, these downstream 'products' result from many more and complex interactions that are far less easy to track than the primary products. Parameters other than technical (e.g., geopolitical, economic, financial) influence the final deployment of downstream products. For the civilian use of nuclear power, metrics are generated to describe these downstream 'outcomes', such as electricity supplied, fossil fuel saved, bacteria destroyed by food irradiation, lives saved by early detection with radioisotopes, etc. These downstream metrics represent intrinsically more complex and abstract phenomena than the primary metrics, and are in many cases much more difficult to quantify than the primary metrics.

In the research analog, the primary products, while important in describing the efficiency of short-term outputs, also serve as an intermediary. Again, the main interest is in the longer term impacts and influence resulting from the research process. In parallel, the longer term impacts and outcomes of research are influenced by diverse environmental parameters (geopolitical, economic, financial). Metrics can be generated analogously to describe these downstream outcomes, such as improved performance military systems, safer civilian aircraft, lower cost automobiles, more effective drugs, etc., with these downstream metrics also being intrinsically more complex and difficult to quantify than the primary metrics.

III-A-2-i-b. Normalized vs Un-normalized Metrics

One major difference between S&T metrics and physical science metrics revolves around their use in practice. Consider the fission analogy again, this time focusing on the power output of a fission reactor. What types of metrics can be employed to quantify this process?

The simplest metric would quantify the absolute un-normalized value of the power output. This metric would offer some small amount of information, but would be of limited use in practice. It offers no information about the input resources required to achieve the measured power level, and therefore gives no indication of the efficiency of the conversion process.

The next level of complexity metric would provide an efficiency measure, the power output divided by the power input. By itself, this metric still offers limited information, since there is no comparison with the efficiencies of competitive systems or processes. When this metric's quantitative value is compared with efficiencies of other systems, then information useful for decision-making becomes possible.

However, in physical systems, while comparative use of metrics allows critical choices to be made on the basis of performance, it still has limitations. As Appendix 5-B shows in more detail for the specific metric example of citation analysis, comparing power output among different engines gives no indication of actual performance relative to ultimate performance. It provides no understanding as to how much potential improvement is possible with a given engine's performance, and therefore is of no help in advancing the technology of engines. The solution used by the engineering community is to compare a given engine's efficiency with its theoretical ultimate efficiency. Since the Carnot efficiency indicates the highest efficiency an engine can achieve when operating between two temperatures, a valuable use of efficiency metrics becomes the comparison of an engine's efficiency with that of its Carnot efficiency. This allows performance standards and development targets to be set for engines, and converts the metric from an interesting relative indicator to a serious tool for management control.

Consider now how S&T metrics are used in practice, relative to the analogous physical sciences use presented. For illustrative purposes, consider the metric of paper citations, although the conclusions will apply to most other S&T metrics. Most citation studies present one of two metric uses: 1) Absolute numbers of citations to papers from an individual/ group/ organization, and/ or 2) Comparison of these absolute numbers of citations with citations from competing individuals/ groups/ organizations. Only in the rarest of circumstances are the numbers of citations normalized to some input parameter, such as the funding received by the project represented by the paper being cited, or the funding received by a group whose paper citations are being examined. And nowhere has the author seen an analogous comparison of citations received to potential citations possible, the Carnot efficiency analog for citations. Appendix 5-B addresses this issue in more detail, and presents one possible approach to obtaining this effective Carnot efficiency for citations.

The present limitations in understanding ultimate performance values for S&T metrics translate into limitations in their use as management and performance targets. While S&T metrics appropriately normalized for technical discipline and other environmental parameters can be used (cautiously) for comparative purposes, they require much more theoretical development

before their full potential as useful measures of S&T impact and performance can be realized. In addition, more understanding of ultimate performance values for S&T metrics would support a more powerful use of these metrics; namely, their use as management performance targets and controls. This is especially true for those metrics which could be classified more as management performance metrics than output or impact metrics, such as the collaboration metrics addressed later.

The taxonomy below divides the research metrics into two generic classes, primary metrics and metametrics, and then subdivides the metametrics into short-term and long-term. The short-term metametrics are typically straightforward operations on the primary metrics, and in some sense still serve as intermediate quantities. The long-term metametrics in many cases bypass the primary products/ metrics, and deal mainly with gross resource inputs and net long-term outputs. This is analogous again to the fission example, where the long-term metametric of civilian power supplied from a reactor neglects the fission product/ neutron distribution details, and deals directly with resource inputs and power outputs.

III-A-2-ii. S&T Metrics Categories

III-A-2-ii-a. Direct S&T Metrics - Input/ Output/ Productivity

The major components of research measured directly include input/ activity (e.g., number of people working on research, amount of resources devoted to research) and output/ productivity (e.g., papers, papers per resource unit, patents, speeches). These quantities are mostly measured in or near the time frame during which the research is performed. Most of even these relatively simple measures need two aspects for credibility and utility; a magnitude component and a quality component. For example, it is important to know not only that a research group published ten papers in a year from a $1 million per annum program, but also to know the caliber of journals in which those papers were published. Another important characteristic of output metrics is that the output/ productivity data that quantifies these metrics is under the control of the performer.

Obtaining the magnitude component of most of these metrics is relatively straightforward. It is a simple counting process, and with many of the comprehensive databases and algorithmic capabilities available today, it becomes a rapid efficient process. Obtaining the quality component is more complex and time intensive, since it is a highly subjective process which requires substantial judgement on the part of the assessors.

The above discussion has focused on individual primary metrics. However, as stated in the overview to the present section, because of the multi-faceted nature of research, combinations of metrics are required to provide a more complete picture of the research product. These different metrics can be presented to decision-makers separately, which can be confusing and time-consuming if large numbers of primary metrics are presented, or they can be aggregated. In this way, figures of merit can be generated which combine the different primary metrics into a single primary megametric [Geisler, 1996]. Provision of this megametric to management, along with the combination and prioritization rules, allows the research product to be estimated simply and rapidly, and potential problem areas to be pinpointed rapidly.

III-A-2-ii-b. S&T Metametrics - Near-Term - Impact

The metrics in this category are derived from operations performed on the direct or primary metrics described above. These near-term metametrics tend to reflect S&T impact based on the primary metrics, and tend to be generated/ measured at points in time moderately after the research has been performed. Not only are these measures still relatively simple, but the types of impacts they measure are simple and relatively near-term. The impacts tend to be on other research or early technology development. Again, most of these measures need the two aspects for credibility mentioned above; a magnitude component and a quality component. For example, it is important to know not only that a research group received 100 citations to their papers in a given year, but also to know both numbers of citations relative to other similar papers and the caliber of papers/ authors citing the primary papers. Obtaining the magnitude component is still a relatively time efficient process, but obtaining the quality component can be very time intensive. Contrary to the output or productivity metrics, the data that quantifies these impact metrics is, to a large extent, not under the control of the performer.

A similar argument to the one in the preceding section can be made for the need to combine individual metametrics into one, or a few, megametrics. In fact, there are benefits to combining individual primary and metametrics into one, or a few, megametrics. For example, assume that a project's output and near-term impact are characterized by twenty primary metrics and near-term metametrics, and assume that these metrics are not monolithic in their message. While examination of each of the metrics may be of interest to the analyst, a weighted impact figure of merit which reflected the organization's priorities would be very useful to managers and decision-makers [Geisler, 1996]. If such a figure of merit indicated a potential problem with the research's net impact, then, with modern display technology, the individual metric components of the figure of merit could be rapidly displayed and the causes of the problem could be investigated at a lower level of detail.

III-A-2-ii-c. S&T Metametrics - Long-Term - Impact/ Outcome

The metrics in this category tend to integrate out and incorporate the primary productivity measures and the intermediate impact measures. These outcome metrics also tend to include highly uncertain data, and tend to require complex and far-ranging data difficult to obtain. For example, a cost-benefit metric for a research program performed in the past would require an understanding of the breadth of influence which the research program had, and might require very subjective methods for generating benefit data (e.g., value of lives saved, value of more comfortable living). This analysis might not use any of the short term primary or metametrics (papers, citations, students graduated), but would focus directly on market-based metrics (expenditures, sales, revenues). Or, it could include valuation of some shorter-term metrics, such as quantifying economic benefit attached to training 10,000 Ph.Ds. A projected cost-benefit metric for research being proposed or performed would require in addition estimates of highly uncertain future cost and benefit data, and environmental economic and financial data such as discount rates. As in the previous section, a readily deconvolveable figure of merit that integrated long-term metametrics, or combinations of the different types of metrics, would be a very valuable tool for management's use.

III-B. PRINCIPLES OF HIGH QUALITY METRICS-BASED S&T EVALUATIONS

III-B-1. Overview

As shown by the Bibliography to this paper, there are hundreds of documents that describe S&T metrics, and substantially less that describe their credible applications to the evaluation of S&T. One major problem in reading these documents is the inability to ascertain the quality of the application, or assessment. There is no Consumer Reports, or Good Housekeeping Seal of Approval, which provides independent tests of the quality of a metrics-based S&T evaluation. Unlike the physical and engineering sciences, there are no primary physical reference standards against which one can benchmark the assessment product.

Most of the S&T metrics literature focus has effectively been on metrics as an end in themselves. Relatively few studies have been done on the issues and principles underlying S&T metrics, and even fewer studies have addressed how metrics can be used to support S&T evaluations in real-world applications. This conclusion was confirmed most graphically by a recent metrics literature survey conducted by the author. Most of the documents retrieved described the generation of a multitude of metrics of large data aggregates, with no indication of the relevance of these metrics to any questions or decisions supporting S&T evaluations.

The foundation of this problem is the strong dichotomy between the researchers who publish metrics studies in the literature, and the managers who use metrics to support budgetary allocation and other management decisions. Most of the people who employ metrics for management purposes do not document them in the literature. Most of the principle and concept and (potential) application papers in the metrics literature are written by people who have never used or applied metrics for management decision-making purposes. In addition, many of the researchers who perform metrics studies focus on single approaches or single approach applications, in order to promote the concepts that they have developed. The managers who use metrics, conversely, have very eclectic requirements. They need suites of metrics, or suites of metrics combined with other evaluation approaches, in order to perform comprehensive multi-faceted S&T evaluations. Thus, there is a serious schism between the incentives and products of the metrics researchers (suppliers) and the incentives and requirements of the metrics users (customers).

Consequently, there are two major gaps in the literature on S&T metrics. First, there are few relevant papers published. Second, most of the concept and principle and (potential) application papers that do exist bear little relation to the reality of what is required to quantitatively support science and technology assessments and evaluations for decision-making. Because of the deficiency of metrics studies relevant to S&T applications, it is difficult to extract the conditions for high quality metrics-based evaluations solely from the open literature. Drastic alterations in this overall situation are required if metrics are going to support the GPRA requirements in any credible manner.

Despite these severe deficiencies identified, more specific requirements, or underlying principles, necessary for a high quality metrics-based S&T evaluation can be formulated. The

author's experience, based on examining the S&T metrics literature, evaluating many types of S&T programs and projects and proposals with the use of metrics in concert with other techniques, and developing different types of metrics, leads to the following conclusions about the factors critical to high-quality metric-based S&T evaluations.

III-B-2. Principles

III-B-2-a. Senior Management Commitment

The most important factor in a high-quality metrics-based S&T evaluation is the serious commitment of the evaluating organization's senior management to high-quality metrics-based S&T evaluations, and the associated emplacement of rewards and incentives to encourage such evaluations.

III-B-2-b. Assessment Manager Motivation

The second most important factor is the assessment manager's motivation to perform a technically credible assessment. The manager:

1) sets the boundary conditions and constraints on the assessment's scope;

2) selects the final metrics used from a myriad of potential choices;

3) selects the methodologies for how these metrics will be combined/ integrated/ interpreted, and

4) selects the experts who will perform the interpretation.

In particular, if the evaluation manager does not follow, either consciously or subconsciously, the highest standards in selecting these experts, the evaluation's final conclusions could be substantially determined even before the evaluation process begins.

III-B-2-c. Statement of Objectives

The third most important factor is the transmission of a clear, unambiguous statement of the metrics-based evaluations objectives (and conduct) and potential impact/consequences to all participants at the initiation of the process. Participants are usually more motivated to contribute when they understand the importance of the evaluation to the achievement of the organizations goals, and understand in particular how they and the organization will be potentially impacted by the evaluations outcome.

Clear objectives and goals tend to derive from the seamless integration of evaluation processes in general into the organization's business operations. Evaluation processes should not be incorporated in the management tools as an afterthought, as is the case in practice today, but should be part of the organization's front-end design. This allows optimal matching between data generating/ gathering and evaluation requirements, not the present procedure of force fitting evaluation criteria and processes to whatever data is produced from non-evaluation

requirements. When the evaluation processes are integrated with the organizations strategic management, the objectives drive the metrics which in turn determine what data should be gathered. Ad hoc evaluation processes tend to let the available data drive the metrics and the quantifiable goals.

III-B-2-d. Competency of Technical Evaluators

The fourth most important factor is the role and competency of technical experts in a metrics-based S&T evaluation. Metrics should not be used as a stand-alone diagnostic instrument. Analogous to a medical exam, even quantitative metric results from suites of instruments require expert interpretation to be placed into proper context and gain credibility. The metrics results should contribute to, and be subordinate to, an effective peer review of the technical area being examined [Kostoff, 1997a]. Thus, this third critical factor consists of the evaluation experts' competence and objectivity. Each expert should be technically competent in his subject area, and the competence of the total evaluation team should cover the multiple research and technology areas critically related to the science or technology area of present interest. In addition, the team's focus should not be limited to disciplines related only to the present technology area (which tends to reinforce the status quo and provide conclusions along very narrow lines), but should be broadened to disciplines and technologies which have the potential to impact the overall evaluation's highest-level objectives (which would be more likely to provide equitable consideration to revolutionary new paradigms).

III-B-2-e. Criteria for Metric Selection

The fifth most important factor is criteria for metric selection. These criteria and the resultant metrics will depend on:

- the interests of the audience for the evaluation,
- the nature of the benefits and impacts,
- the availability and quality of the underlying data,
- the accuracy and quality of results desired,
- the complementary metrics available and suites of metrics desired for the complete analysis,
- the status of algorithms and analysis techniques, and
- the capabilities of the evaluation team.

III-B-2-f. Relevance of Metric to Future Action

A factor of equal importance to criteria is one that has been violated in every metrics briefing the author has attended spanning many government agencies, industrial organizations, and academic institutions.

EVERY S&T METRIC, AND ASSOCIATED DATA, PRESENTED IN A STUDY OR BRIEFING SHOULD HAVE A DECISION FOCUS; IT SHOULD CONTRIBUTE TO THE ANSWER OF A QUESTION WHICH IN TURN WOULD BE THE BASIS OF A RECOMMENDATION FOR FUTURE ACTION.

Metrics and associated data that do not perform this function become an end in themselves, offer no insight to the central focus of the study or briefing, and provide no contribution to decision-making. They dilute the theme of the study, and, over time, tend to devalue the worth of metrics in credible research evaluations. Because of the political popularity and subsequent proliferation of S&T metrics, the widespread availability of data, and the ease with which this data can be electronically gathered/ aggregated/ displayed, most S&T metrics briefings and studies are immersed in data geared to impress rather than inform.

III-B-2-g. Reliability of Evaluation

Another factor of equal importance is reliability or repeatibility. To what degree would a metrics-based evaluation be replicated if a completely different team were involved in selection, analysis, and interpretation of the metrics data? If each evaluation team were to generate different metrics, and particularly far different interpretations of metrics, for the same topic, then what meaning or credibility or value can be assigned to any metrics-based evaluation? To minimize repeatibility problems, a diverse segment of the competent technical community should be involved in the construction and execution of the the evaluation.

III-B-2-h. Metrics Integration

The eighth most important factor is the seamless integration of metrics in particular, and evaluation processes in general, into the organization's business operations. Evaluation processes should not be incorporated in the management tools as an afterthought, as is the case in practice today, but should be part of the organization's front-end design. This allows optimal matching between data generating/ gathering and evaluation requirements, not the present procedure of force fitting metrics and evaluation processes to whatever data is produced from non-evaluation requirements.

III-B-2-i. Normalization Across Technical Disciplines

For evaluations which will be used as a basis for comparison of science and technology programs or projects, the ninth most important factor is normalization and standardization across different science and technology areas. For science and technology areas which have some similarity, use of common experts (on the evaluation teams) with broad backgrounds which overlap the disciplines can provide some degree of standardization. For very disparate science and technology areas, some allowances need to be made for the relative strategic value of each discipline to the organization, and arbitrary corrections applied for benefit estimation differences and biases. Even in this case of disparate disciplines, some normalization is possible by having some common team members with broad backgrounds contributing to the evaluations for diverse programs and projects. However, normalization of the metrics for each science or technology area's unique characteristics is a fundamental requirement. Because credible normalization requires substantial time and judgement, it tends to be an operational area where quality is sacrificed for expediency.

III-B-2-j. Global Data Awareness

A tenth factor of equal importance is data awareness [Kostoff, 2003a]. In all of the decision aids, placement of the technology of interest in the larger context of technology development and availability world-wide is an absolute necessity. This tends to be a central deficiency of most management decision aids. Lack of S&T documentation, inaccessibility of S&T that is documented, inability to retrieve S&T documents due to poor retrieval methods, inability to extract information from large retrievals, and general lack of interest and will in global data awareness, mitigate against attaining comprehensive global data awareness.

III-B-2-k. Cost of Metrics-based Evaluations

An eleventh critical factor for quality metrics-based evaluations is cost. The true total costs of developing a high quality evaluation using credible suites of metrics, sophisticated normalization techniques, and diverse experts for analyses and interpretation can be considerable, but tend to be understated. For high quality evaluations, where sufficient expertise is represented on the evaluation team, the major contributor to total costs is the time of all the individuals involved in normalizing and interpreting the data. With high quality personnel involved in the evaluation process, time costs are high, and the total evaluation costs can be non-negligible. Especially when a metrics-based evaluation is performed in tandem to a qualitative peer-review process [Kostoff, 1997a], the real costs of these experts could be substantial. Costs should not be neglected in designing a high quality metrics-based S&T evaluation process.

III-B-2-j. Maintenance of High Ethical Standards

The final critical factor, and perhaps the foundational factor, in high quality metrics-based evaluations is the maintenance of high ethical standards throughout the process. There is a plethora of potential ethical issues, including technical fraud, technical misconduct, betraying confidential information, and unduly profiting from access to privileged information, because there is an inherent bias/ conflict of interest in the process when real experts are desired to design, analyze, and interpret a metrics-based evaluation. The evaluation managers need to be vigilant for undue signs of distortion aimed at personal gain.

IV. SCIENCE AND TECHNOLOGY METRICS

TABLE OF CONTENTS FOR SECTION IV

IV-A. OVERVIEW

IV-B. BIBLIOMETRICS

IV-B-1. Overview
IV-B-2. Bibliometric Axioms
IV-B-3. Generic Bibliometric Uses
IV-B-4. Generic Bibliometric Analysis Approaches
IV-B-5. Problems with Bibliometrics
IV-B-5-i. Personal Example
IV-B-5-ii. Limited Federal and Industrial Use of Bibliometrics
IV-B-5-iii. Normalization Problems and Approaches
IV-B-5-iv. Problems with Incomplete References
IV-B-5-v. Collective Distortions: The Pied Piper Effect
IV-B-6. Examples of Bibliometric Studies
IV-B-6-i. Overview of Different Bibliometric Study Types Performed
IV-B-6-i-a. Macroscale Bibliometric Studies
IV-B-6-i-b. Microscale Bibliometrics Studies
IV-B-6-ii. Specific Bibliometric Studies with Different Indicators
IV-B-6-ii-a. Publications
IV-B-6-ii-b. Publication Citations
IV-B-6-ii-c. Patents and Patent Citations
IV-B-6-ii-d. Combinations of Publications/ Patents/ Citations
IV-B-6-ii-e Science and Technology Transitions
IV-B-6-ii-f. Collaboration Indicators
IV-B-6-ii-f-1. Collaboration Indicators for Vertical Integration
IV-B-6-ii-g. Other Indicators
IV-B-6-ii-h. Multiple Indicators
IV-B-6-ii-i. Indicators Integrated with other Techniques

IV-C. COST-BENEFIT/ ECONOMIC ANALYSES
IV-C-1. Background
IV-C-2. Classical Microlevel Application
IV-C-3. Macrolevel Analyses
IV-C-4. Production Function Approach
IV-C-5. Macrolevel Marginal Cost-Benefit Application
IV-C-6. Specific Cost-Benefit Studies with Different Approaches

IV-D. COST-EFFICIENCY

IV-E. CO-OCCURRENCE PHENOMENA
IV-E-1. Background
IV-E-2. Overview Summary
IV-E-3. Co-citation Analysis
IV-E-4. Co-nomination and Co-classification Analyses
IV-E-5. Co-word Analysis
IV-E-5-i. Database Tomography
IV-E-6. Specific Co-occurrence Studies with Different Indicators

IV-F. NETWORK MODELING FOR DIRECT/INDIRECT IMPACTS
IV-F-1. Background
IV-F-2. Summary of Methodology

IV-G. EXPERT NETWORKS

IV-H. THE METRICS OF SCIENCE AND TECHNOLOGY

IV-I. S&T METRICS - SUMMARY AND CONCLUSIONS

IV-A. OVERVIEW

This section addresses some critical issues in the applicability of quantitative performance measures to the assessment of S&T, with emphasis on basic research. The strengths and weaknesses of metrics applied as S&T performance measures are examined. The remainder of this section provides an overview of the quantitative approaches used in S&T assessment.

Quantitative approaches to research assessment focus on the numerics associated with the performance and outcomes of research. The main approaches used are bibliometrics and econometrics such as cost-benefit and production function analysis. This section focuses on these three main approaches, then

- describes the bibliometrics-related family of approaches known as co-occurrence phenomena,
- describes a network modeling approach to quantifying research impacts, and
- ends with an expert systems approach for supporting research assessment.

Studies reported in the literature tend not to adhere strictly to the metrics taxonomy presented above. In particular, bibliometrics analyses tend to report mixtures of primary and short-term metametrics without addressing the significances of the differences. In order to allow an easy mapping from the present document into results reported in the literature, the literature approaches and groupings will be retained, but any problems associated with combining the different types of metrics improperly will be discussed where necessary.

IV-B. BIBLIOMETRICS

IV-B-1. Overview

This section overviews the scope and breadth of bibliometrics studies performed. It

- starts with examples of bibliometric indicators (IV-B-1),
- presents fundamental axioms that underly the utilization and validity of bibliometric analysis (IV-B-2),
- describes the four generic uses of bibliometric analyses (IV-B-3),
- summarizes the four major steps in any bibliometrics analysis (IV-B-4),
- illuminates a broad range of conceptual and operational problems with bibliometrics analyses (IV-B-5),
- overviews briefly the types of bibliometric applications that have been performed (IV-B-6), and
- ends with moderate descriptions of specific bibliometric studies performed using a wide variety of indicators.

Bibliometrics, especially evaluative bibliometrics, uses counts of publications, patents, citations and other potentially informative items to develop science and technology performance indicators. It includes both the direct or primary metrics and the near-term metametrics defined in the section III taxonomy. The choice of important bibliometric indicators to use for research performance measurement may not be straightforward. A 1993 study surveyed about 4,000 researchers to identify appropriate bibliometric indicators for their particular disciplines [Australia, 1993]. The

respondents were grouped in major discipline categories across a broad spectrum of research areas. While the major discipline categories agreed on the importance of publications in refereed journals as a performance indicator, there was not agreement about the relative values of the remaining 19 indicators provided to the respondents. For the respondents in total, the important performance indicators were:

1. Publications (publication of research results in refereed journals);

2. Peer Reviewed Books (research results published as commercial books reviewed by peers);

3. Keynote Addresses (invitations to deliver keynote addresses, or present refereed papers and other refereed presentations at major conferences related to one's profession);

4. Conference Proceedings (publication of research results in refereed conference proceedings);

5. Citation Impact (publication of research results in journals weighted by citation impact);

6. Chapters in Books (research results published as chapters in commercial books reviewed by peers);

7. Competitive Grants (ability to attract competitive, peer reviewed grants from the ARC, NH&MRC, rural R&D corporations and similar government agencies).

These bibliometric indicators can be used as part of an analytical process to measure scientific and technological accomplishment. Because of the volume of documented scientific and technological accomplishments being produced (5,000 scientific papers published in refereed scientific journals every working day worldwide; 1,000 new patent documents issued every working day worldwide), use of computerized analyses incorporating quantitative indicators is necessary to understand the implications of this technical output [Narin, 1994].

IV-B-2. Bibliometric Axioms

Narin states three axioms that underly the utilization and validity of bibliometric analysis. The first axiom is activity measurement: that counts of patents and papers provide valid indicators of R&D activity in the subject areas of those patents or papers, and at the institution from which they originate. This axiom has degrees of validity which can vary significantly across authors, technical disciplines, and organizations. Cultural historical reasons, classification issues, corporate proprietary issues, and myriad other causes can and do contribute to open source literature having substantial gaps in documented information of existing and past activity in specific technical fields. The more that the open source literature of a specific technical discipline can serve as a representative sample of the total literature in this discipline, the more valid is this axiom.

The second axiom is impact measurement: that the number of times those patents or papers are cited in subsequent patents or papers provides valid indicators of the impact or importance of the cited patents and papers. However, there could be weightings applied to the raw count data, depending on the perceived importance of the journals containing the citing papers. Also, the impacts would be on

allied research fields or technologies, not necessarily long-term impacts on the originating organization's mission. Finally, as discussed later in this section, and in more detail in Appendix 3, there are many reasons for including (or excluding) specific documents in a paper's references. Therefore, the number of citations received by a given document may not be a unique indicator of the document's impact or importance. Substantial expert interpretation is required before conclusions can be drawn as to the importance or impact of a particular document on the technical field.

The third axiom is linkage measurement: that the citations from papers to papers, from patents to patents and from patents to papers provide indicators of intellectual linkages between the organizations which are producing the patents and papers, and knowledge linkage between their subject areas [Narin, 1994]. Again, there are many reasons documents are cited other than valid intellectual linkage, and expert analyses are required before specific conclusions can be drawn.

IV-B-3. Generic Bibliometric Uses

Bibliometrics (and other S&T metrics) have been used for a variety of purposes, including::

- S&T marketing; S&T assessment and diagnostics;
- S&T management; and
- resource allocation.

Specific uses of bibliometrics can be categorized into four levels of aggregation [Narin, 1994]:

1. policy (evaluation of national or regional technical performance);

2. strategy (evaluation of the scientific performance of universities or the technological performance of companies);

3. tactics (tracing and tracking R&D activity in specific scientific and technological areas or problems);

4. conventional (identifying specific activities and specific people engaged in research and development).

Policy questions deal with the analysis of very large numbers of papers and patents, often hundreds of thousands at a time, to characterize the scientific and technological output of nations and regions. Strategic analyses tend to deal with thousands to tens of thousands of papers or patents at a time, numbers that characterize the publication or patent output of universities and companies. Tactical analyses tend to deal with hundreds to thousands of papers or patents, and deal typically with activity within a specific subject area. Finally, conventional information retrieval tends to deal with identifying individual papers, patents, and clusters of interest to an individual scientist or engineer or research manager working on a specific research project [Narin, 1994].

IV-B-4. Generic Bibliometric Analysis Approaches

The first, and major, step in the performance of a high quality bibliometric analysis in any of the

above four levels of aggregation is acceptance by the potential user of the above three axioms to validate the credibility of the bibliometric approach. Once this hurdle has been passed, the second step is to select the suite of bibliometric indicators most appropriate to achieving the objectives of the study, and in parallel, select the highest quality and reliability raw indicator products (data and databases). The third step is to apply analyses of the highest statistical precision and accuracy to these indicators [Braun, 1989, 1990, 1993]. The fourth step, which determines the credibility and utility of the results, is the interpretation and visual display of the results. The results of the most stringent analyses will be relatively worthless if they are not placed in the larger evaluation context and if they are not displayed in a concise and lucid form. See Appendix 4 for a more detailed discussion of indicator display issues

IV-B-5. Problems with Bibliometrics

IV-B-5-i. Personal Example

Generating the bibliometric raw data and performing computer manipulations on this data are relatively straightforward processes. Normalizing and interpreting and assigning meaning to this data lies at the source of the difficulties with bibliometrics. A personal anecdote partially illustrates this point.

A few years ago, the author was asked to be part of a team that reviewed a component of a large Federal agency laboratory. Identification of the agency and laboratory is not important for this discussion. The team judged the work of the component to be excellent, but the number of papers produced relative to the component's funding was extremely small. Since the agency was trying to improve publication output of its laboratories, the team recommended that the component try to increase its publications.

A couple of years later, the team revisited the laboratory component. This time, the publication record was much improved. However, had the quality of research improved? No, the quality was excellent in the first review and remained excellent in the second review. Had the quantity of research increased? No; in fact, one could probably make the argument that there was less research produced, since research time had to be sacrificed in writing the extra papers. Were the users more satisfied? No, since in either case the direct users were getting the quantity and quality research product they wanted, and were converting it to technology.

There appeared to be three main benefits of emphasis on publication. First, there was increased dissemination of the laboratory's results to the larger research community, which theoretically could have been of value to the community not familiar with the laboratory's work. The agency improved its bibliometric statistics, which it could then display as an example of increasing research productivity. In addition, there was probably some enhancement of the laboratory's and researchers' prestige (and subsequent marketing) due to the increased recognition in the published literature.

The main point to be derived from the above anecdote is that the fundamental bibliometric unit, the published paper in a peer reviewed journal, is not research; it is a documentation of research. While its contents are important in disseminating the research results and evaluating the quality and quantity of research produced, the documentation counts need to be associated with many more

caveats and to be supported by much interpretation before they can become useful in a research evaluation.

In addition, there is a more serious problem with the published peer-reviewed research paper as presently structured for the tracking of intellectual heritage or impact. The typical paper focuses, in priority order, on research approach, research product, and intellectual heritage (references). This focus derives from performer priorities, not sponsor tracking priorities. The completeness of the references, the adequacy of the references, and the relative importance of each reference, are governed by the performer's subjectivity and the limited space available for the paper. Thus, the present structure and design of the research paper is not the optimal structure required for research impact tracking, and contributes to an under-reporting of the impact of research. This limitation is more than an academic issue; it could have consequences on the reporting of research products and impacts required under the Government Performance and Results Act of 1993. For a more detailed discussion of this under-reporting phenomenon, see Appendix 2.

IV-B-5-ii. Limited Federal and Industrial Use of Bibliometrics

A comprehensive review of bibliometrics [White, 1989] shows the sparsity of bibliometric studies for research impact evaluation reported by the Federal government. The reason for this is due in part to the following problems with publication and citation counts [King, 1987; Oberski, 1988; OTA, 1986]:

1) Publication counts:

a. indicates quantity of output, not quality;

b. non-journal methods of communication ignored;

c. publication practices vary across fields, journals, employing institutions;

d. choice of a suitable, inclusive database is problematical;

e. undesirable publishing practices (artificially inflated numbers of co-authors, artificially shorter papers) increasing.

2) Citations:

a. intellectual link between citing source and reference article may not always exist;

b. incorrect work may be highly cited;

c. methodological papers among most highly cited;

d. self-citation may artificially inflate citation rates;

e. citations lost in automated searches due to spelling differences and inconsistencies;

f. Science Citation Index (SCI) changes over time;

g. SCI biased in favor of English language journals;

h. same problems as publication counts.

In response to Cawkell's [1977] claims that 'citation anomalies have little effect-they are like random noise in the presence of strong repetitive signals,' MacRoberts [1989] stated the Federal concerns about bibliometrics eloquently: "When only a fraction of influences are cited, when what is cited is a biased sample of what is used, when influences from the informal level of scientific communication are excluded, when citations are not all the same type, and so on, the 'signal' may be repetitive, but it is also weak, distorted, fragmented, incoherent, filtered, and noisy".

Another reason for limited Federal use can be inferred from Narin [1976], where studies on the publication and citation distribution functions for individuals are reviewed. The conclusion drawn, from studies such as those of Lotka, Shockley, De Solla Price, and Cole and Cole, is that **very few of the active researchers are producing the heavily cited papers**. How motivated are funding agencies to report these hyperbolic productivity distributions for different programs in the open literature, especially since many questions exist as to the accuracy and completeness of the bibliometric indicators? This conclusion raises the further question of the role actually played by the less productive researchers (as measured by publication and citation counts): is the productivity of the elite somehow dependent on the output of the less influential, or is the role of the less productive members that of maintaining the stability of the research infrastructure and educating future generations of researchers?

IV-B-5-iii. Normalization Problems and Approaches

Another problem with bibliometrics is cross-discipline comparisons of outputs. For example, how should the paper or citation output of a program in Solid-State Physics be compared to that of Shallow Water Acoustics. What types of normalizations are required to allow comparisons among these different types of programs and fields. Is there a threshold for disaggregation below which the normalization factors apply to all the subfields. For example, can the normalization factor for Acoustics be applied to a program in High Frequency Shallow Water Acoustics, or can the normalization factor for Shallow Water Acoustics be applied to the program in High Frequency Shallow Water Acoustics? The author has addressed these issues in more detail in a recent paper using normalization domains of decreasingly smaller extent [Kostoff, 2005j], and the technique and conclusions are summarized in Appendix 5-C,.

While many researchers and organizations have been concerned about this issue, a group centered at the Library of the Hungarian Academy of Sciences has been addressing the problem of output comparisons, including cross-discipline comparisons, in detail for many years. The normalization solutions they propose are excerpted from a 1993 publication [Schubert, 1993], and are presented in Appendix 5-A. In addition, the author has generated a new approach (citation efficiency) for comparing citation rates across different disciplines [Kostoff, 1997i], and excerpts are contained in Appendix 5-B.

IV-B-5-iv. Problems with Incomplete References

In a comprehensive survey of problems with citation analysis, MacRoberts and MacRoberts [1996] list many deficiencies with citation analysis. In particular, they read papers in technical fields with which they were familiar, and compared the influence evident (to them) in the text with what was contained in the bibliography. They found that approximately 30% o f the influence was cited. Their paper is one of the few cases where this type of validation study has been performed. However, even this innovative study illuminates the difficulties of establishing reference standards for bibliometrics analyses; the benchmark as to what references should have been cited was an arbitrary judgement made by the authors. This issue of relative reference completeness is discussed in somewhat more detail in Appendix 3. The author has recently generated a methodical approach for insuring that the seminal background papers in any discipline are retrieved [Kostoff, 2005g], and this approach is summarized in Appendix 5-D.

IV-B-5-v. Collective Distortions: The Pied Piper Effect

One of the main concerns with using citations as a stand-alone measure of quality and impact has been the potential bimodal interpretation of the numerical results. A paper could receive high citations because of its high quality, or because the citers disagree with it. However, there is a third interpretation that further precludes citations being utilized in stand-alone mode, which the author has termed the "Pied Piper" effect.

Assume there is a present-day mainstream (characterized by high citations) approach in a specific field of research; for example, the chemical/ radiation/ surgical approach to treating cancer (See Appendix 6 for a more detailed example of the "Pied Piper Effect"). Assume that in, say, fifty years a cure for cancer is discovered, and the curative approach has nothing to do with today's mainstream highly-cited research. In fact, assume it turns out that today's highly-cited mainstream approach was completely orthogonal or even antithetical to the correct approach, and that one of the alternative lowly-cited approaches existing today provided the foundation for the eventual cure. Then what meaning can be ascribed to those research papers in cancer today that define the mainstream approach; i.e., they are highly cited for supposedly positive reasons?

In this case, a paper's high citations are a measure of the extent to which the paper's author(s) has persuaded the research community that the research direction contained in his paper is the correct one. The citations are not a measure of the intrinsic correctness of the research direction. In fact, the citations may reflect the desire of a closed research community (the author and the citers) to persuade a larger community (which could include politicians and other resource allocators) that the research direction is the correct one. The citations become the operational mechanism by which the established infrastructure is able to protect its intellectual and capital investments and exclude other competitive approaches which could threaten the integrity of that infrastructure. Citations become the vehicle by which scientific monopoly is established and perpetuated.

This is the "Pied Piper Effect". *The large number of citations in the above medical example becomes a measure of the extent of the problem, the extent of the diversion from the correct path, not*

the extent of progress toward the solution. The "Pied Piper Effect" is a key reason why, especially in the case of revolutionary research, citations and other quantitative measures must be part of and subordinate to a broadly constituted peer review in any credible evaluation and assessment of research impact and quality [Kostoff, 1997a].

Since citation analysis has had substantial usage in the literature as a key approach for estimating research impact and quality, it will receive a disproportionate share of attention in the present document. Appendix 3 is an excerpt from a paper by the author describing different uses and purposes for citation analysis. The appendix includes uses of citations for: bookmarks, intellectual heritage, tracking of research impact, and self-serving purposes. It also shows the limitations of citations as a stand-alone measure of impact or quality.

IV-B-6. Examples of Bibliometric Studies

IV-B-6-i. Overview of Different Bibliometric Study Types Performed

Bibliometric studies have been performed over a wide range of levels, from analysis of a performer or even selected documents produced by a performer to analysis of national output or total discipline output. There is a belief in the bibliometrics community that the analyses become more valid as the domain of interest increases in size. The supposedly wide range of fluctuations of results across small units integrates out when these units are aggregated (a 'Law of Large Numbers' effect), and theoretically the larger domain unit analyses are the most credible.

However, the author has performed many bibliometric analyses of small units. If these types of studies are restricted to pinpointing problem areas for further investigation, and if time and effort are invested in obtaining quality data for the analysis, very useful results can be obtained. For those readers interested in a source focused on this broad range of bibliometric analyses, the journal Scientometrics is a very good starting point.

IV-B-6-i-a. Macroscale Bibliometric Studies

Macroscale bibliometric studies characterize science activity at the national [e.g., Hicks, 1986; Braun, 1989], international, and discipline level. The biennial Science and Engineering Indicators report [NSF, 1996] tabulates data on characteristics of personnel in science, funds spent, publications and citations by country and field, and many other bibliometric indicators. Another study at the national level was aimed at evaluating the comparative international standing of British science [Martin, 1990]. Using publication counts and citation counts, the authors evaluated scientific output of different countries by technical discipline as a function of time. A study similar in concept was published recently [King, 2004]. It drew conclusions about national capabilities in research based on country aggregate bibliometrics. In a short note commenting on [King, 2004], the present author concluded that the country aggregate results could be misleading for some applications, and comparisons for specific critical technologies were far more important [Kostoff, 2004g]. All the above studies use comparative metrics only; they compare productivity metrics of one group to another. They do not relate metric values to some desirable or theoretical limiting value. If all groups, for example, are underperforming, this fact will not be captured by the types of metrics employed.

There is little evidence that the results from such studies have much influence on policy or decision-making; i.e., the allocation of resources. As Martin et al point out in their conclusions, there is potential benefit for a country to understand its position vis-a-vis that of its competitors in different science areas, in order to be able to exploit opportunities which may arise in those areas. However, which indicators are appropriate and how they should impact allocation decisions are open questions.

IV-B-6-i-b. Microscale Bibliometrics Studies

There have been numerous microscale bibliometric studies reported in the literature [e.g., Frame, 1983; McAllister, 1983; Mullins, 1987, 1988; Moed, 1988; Irvine, 1989; Van Raan, 1989; Luukkonen, 1990a, 1990b, 1992]. With the notable exception of the NIH [OTA, 1986], few Federal agencies report use of microscale bibliometric studies to evaluate programs and influence research planning in the published literature. The NIH bibliometric-based evaluations included the effectiveness of various research support mechanisms and training programs, the publication performance of the different institutes, the responsiveness of the research programs to their congressional mandate, and the comparative productivity of NIH-sponsored research and similar international programs.

Publication Citation Analysis

Two papers in the late 1980s [Narin, 1987b, 1989] described determination of whether significant relationships existed among major cancer research events, funding mechanisms, and performer locations; compared the quality of research supported by large grants and small grants from the National Institute of Dental Research; evaluated patterns of publication of the NIH intramural programs as a measure of the research performance of NIH; and evaluated quality of research as a function of size of the extramural funding institution. Most of the NIH studies focused on aggregated comparison studies (large grants vs small, large schools vs small schools, domestic vs foreign, etc).

Patent Citation Analysis

Patent citation analysis has the potential to provide insight to the conversion of science to technology [Carpenter, 1981, 1982, 1983; Narin, 1984; Wallmark, 1986; Collins, 1988; Narin, 1988a, 1988b, 1988c; Van Vianen, 1990; Narin, 1991, 1992]. Much of the Federal government support of the development of patent citation analysis was by the NSF [e.g., Carpenter, 1980; Narin, 1987a], although **there is little published evidence now of widespread Federal use of this capability**. Some studies have focused on utilization of patent citation analysis for corporate intelligence and planning purposes (Narin, 1990, 1992a, 1992b). Some of the data presented verify further Lotka's Productivity Law, where relatively few people in a laboratory are producing large numbers of patents. In the example presented in Narin [1992b], the patents of the most productive inventor are highly cited, further demonstrating his key importance. Narin concludes that highly productive research labs are built around a small number of highly productive, key individuals.

An ongoing study of citations to scientific papers from the front pages of U.S. patents has potentially

important implications for science and technology policy. Some results showed that, for different countries that file patents with the U.S. patent system, each country's patents in the U.S. cite their own scientific papers three times as often as would be expected, after normalizing out the size of each country's science [Narin, 1994]. To end this discussion of patent citation analysis on a cautionary note, courtesy of Pavitt [1991], it is not yet clear to what extent the 'other publications', cited in patents, reproduce basic or applied research, from universities or corporate laboratories. In addition, a high proportion [Pavitt's estimation] of technology is not patented, because it is kept secret, because it is tacit and non-codifiable art, or because - as in the case of software technology - it is very difficult to protect through patents. Finally, while patent citations can be used to track the science conversion process or the technical influence trajectory, the value of the magnitude of the metric is still limited through lack of comparison with theoretically achieveable targets.

Research Product Dissemination

Despite these limitations, bibliometrics may have utility in providing insight into research product dissemination. For example, in a series of presentations to large Federally-funded laboratories [Kostoff, 1992b], the following suite of bibliometric studies was proposed:

1. Examine distribution of disciplines in co-authored papers, to see whether the multidisciplinary strengths of the lab are being utilized fully;

2. Examine distribution of organizations in co-authored papers, to determine the extent of lab collaboration with universities/ industry/ other labs and countries;

3. Examine nature (basic/ applied/ discipline/ quality) of citing journals, other citing media (patents), citing author disciplines, citing author organizations, to ascertain whether lab's products are reaching the intended customer(s);

4. Determine whether the lab has its share of high impact (heavily cited) papers and patents, viewed by some analysts as a requirement for technical leadership;

5. Determine which countries are citing the lab's papers and patents, to see whether there is foreign exploitation of technology and in which disciplines;

6. Identify papers and patents cited by the lab's papers and patents, to ascertain degree of lab's exploitation of foreign and other domestic technology.

While it was also recommended that the lab compare its output (papers/ citations normalized over disciplines) with that of other similar institutions, this quantitative comparison should be approached with great caution. A comparative bibliometric analysis of 53 laboratories [Miller, 1992] clustered the labs into six types (Regulation and Control, Project Management, Science Frontier, Service, Devices, Survey), and stated that "comparisons of scientific impacts should be made only with laboratories that are comparable in their primary task and research outputs". The report concluded further that:

1. Bibliometric indicators and scientific publications are not the only outputs that should be

measured, but the other types of outputs differ for different labs;

2. Bibliometric indicators are not equally valid across different types of laboratories;

3. Bibliometric indicators are less useful for the evaluation of research laboratories involved in closed publication markets.

In addition, studies were performed [Kostoff, 1992c] to track the dissemination of information from accelerated research programs. Key papers (P1) resulting from these programs were identified, then the citing papers for these key papers (P2) were identified, then the next generation of citing papers (P3) which cited P2 were identified, and so on. The breadth of disciplines impacted by the key papers (P1) can be identified from the succeeding generations of citing papers. The type of analysis performed provided more of a qualitative than quantitative estimation of breadth of impact. Preliminary results show that some very fundamental papers impact across a wide spectrum of disciplines, while some high quality but more narrowly focused research papers impact one main discipline very strongly through succeeding generations of citations. Because of the large amounts of data required for a complete analysis, especially where highly cited papers and their descendents are concerned, present efforts focus on methods to reduce data requirements and retain a credible analysis.

IV-B-6-ii. Specific Bibliometric Studies with Different Indicators

In this section, a number of bibliometric studies which examine different indicators or combinations of indicators, are described in moderate levels of detail.

IV-B-6-ii-a. Publications

Computer-Mediated Communication and Publication Productivity Among Faculty

This study [Cohen, 1996] investigated whether faculty who use computer mediated communication (CMC) achieve greater scholarly productivity as measured by publications and a higher incidence in the following prestige factors: receipt of awards; service on a regional or national committee of a professional organization; service on an editorial board of a refereed journal; service as a principal investigator on an externally funded project; or performance of other research on an externally funded project. It also investigated whether faculty who use CMC at less research- oriented institutions realize disproportional benefit from their use of CMC. Data were collected in Fall 1994. A positive relationship was found between the frequency of use of CMC and publications, including coauthored publications. CMC users also had a higher incidence of prestige factors. In addition to statistically significant relationships between CMC use and productivity measures, faculty judged CMC to be of some utility to their productivity. Nevertheless, there did not appear to be a "democratizing effect" which would yield disproportionate benefit to those from less research-oriented institutions.

Research Volume Published

This study [Towe, 1995] measures an important component of the research output of Australian

economics and econometrics teaching departments, namely, the number of pages published, during the period 1988-93, in journals listed by the Journal of Economic Literature. Based on page counts it is found that department rankings are similar over a broad range of journal groupings. It is also found that the median numbers of pages published by each of the groups of senior lecturers, associate professors and professors are quite small, indicating that within these groups research output is highly concentrated among a few active publishers.

Describing and Explaining Research Productivity

This study [Ramsden, 1994] describes results from a study of academic productivity in Australian higher education. It estimates the output (in terms of quantity of publications) of individual staff and academic departments across different subject areas and types of institution. Concerning research productivity, Australian academics resemble their colleagues in other countries: the average is low, while the range of variation is high. Most papers are produced by few academic staff. Several potential correlates of productivity, including level of research activity, subject area, institutional type, gender, age, early interest in research, and satisfaction with the promotions system, are examined. A model linking departmental context to personal research performance through departmental and personal research activity is developed and tested. The results support the view that structural factors (such as how academic departments are managed and led) combine with personal variables (such as intrinsic interest in the subject matter of one's discipline) to determine levels of productivity. There is also evidence that research and teaching do not form a single dimension of academic performance.

Effects of Resource Concentration and Group Size on Research Performance

One study [Johnston, 1994] reports the results of a study commissioned by the Australian National Board of Employment, Education and Training, which examines in detail the effect of resource concentration on research performance, and the basis for critical mass, economies of scale, critical time and risk strategy hypotheses. The widespread introduction of policies of resource concentration around the world are found to have been based on little examined assumptions, and in operation to be at times counter-productive. In general relationships between group size and productivity are found to be linear, though there does appear to be evidence for an optimal size of 5-8. Detailed results and policy implications of these findings are presented.

In a previous series of studies aimed at investigating the dependence of per-capita research output (R) of an interacting group of research workers on the size of the group, it was shown that the per-capita research output of various research groups and institutes in U. S. A., U. K., Pakistan and Bangladesh shows an initial approximately linear rise, followed by one or more mixima, the first one being at group size of 6 to 8 persons. In the present study [Qurashi, 1993], the author presents a fine analysis of the reported data for (a) physics departments of U. K. universities (in 1985-86) and (b) mathematics departments of two universities in Greece (from 1975 to 1984), using close sampling-intervals of DELTAN = 2 and 3 for group-sizes. The results of this reanalysis show that the data for U. K. physics departments exhibits a series of peaks of per-capita research output (R) at N = 11, 19, 25, 36, 46, etc., which compare well with the corresponding maxima already found in the 1977 per- capita output of National Cancer Institute, U. S. A., at N = 7, 15, 26, 34 and 44. Comparison of these two yields the following mean positions for the five peaks viz N = 9 +/- 2, 17

+/- 2, 26 +/- 0, 35 +/- 1 and 45 +/- 1. These appear to be close to multiples of 8.5, indicating the possibility that a sub-group of 8 to 9 persons could be forming a basic unit of interaction in these particular research groups. The data from the mathematics departments of two Greek universities, which falls in the range of N = 20 to N = 44, also shows two maxima, of per-capita output at N = 27 and 34.5 (and possibly one at about 18), which fit in well with the pattern described above. It appears likely that the above concept could open up new avenues in management practices. Accordingly, further studies are in hand on the relevant characteristics of the output of various institutes and, if possible, a fuller study of size and nature of the sub- groups noted above.

Normalization Bias

The bibliometric indicators currently used to assess scientific production have a serious flaw: a notable bias is produced when different subfields are compared. In this study [Schwartz, 1996], the authors demonstrate the existence of this bias using the impact factor (IF) indicator. The impact factor is related to the quality of a published article, but only when each specific subfield is taken separately: only 15.6% of the subfields we studied were found to have homogeneous means. The bias involved can be very misleading when bibliometric estimators are used as a basis for assigning research funds. To improve this situation, the authors propose a new estimator, the RPU, based on a normalization of the impact factor that minimizes bias and permits comparison among subfields. The RPU of a journal is calculated with the formula: RPU=10(1-exp (-IF/x)), where IF is the impact factor of the journal and x the mean IF for the subfield in which the journal belongs. The RPU retains the advantages of the impact factor: simplicity of calculation, immediacy and objectivity, and increases homogeneous subfields from 15.6% to 93.7%.

A Quantitative Bibliometric Study of the Formation of a Field.

A quantitative technique is illustrated which uses publication statistics from a bibliography of citations in the area of weak interactions to provide a view of trends and patterns in the development of the field during the period from 1950 to 1960 [White, 1986]. An overview is given of what the physicists working in weak interactions during this period were doing as indicated by an analysis of the subjects of their papers. The dominant problems and concerns are discussed. Focus is then turned to the events surrounding the emergence of the tau/theta particle puzzle, the discovery of parity nonconservation, and the resolution offered by the V-A theory. Displaying the data from the citation index in unusual ways highlights dominant issues of the period, especially the close relationship between theory and experiment in the latter half of the decade.

IV-B-6-ii-b. Publication Citations

Citation Issues

The first study [Wang, 1996] identifies several aspects of citing behavior (reasons for citing, criteria used in decision making, and mete-level documentation concerns) by directly questioning researchers about decisions to cite or not to cite specific documents. An important finding is the existence of meta-level concerns which may indicate documentation styles which influence a decision to cite a document in addition to situational factors related to its actual use during research. It reports the preliminary results of the citing decisions in an empirical, longitudinal study of

document use by academic economists and graduate students during several phases of their research projects.

The goal of another study [Liu, 1993] was to obtain insights into the citation process focusing on scientists' citing motivation. Different from most citation studies, the research findings were derived from directly questioning Chinese physicists. This exploratory study revealed that the number of citations (termed as citation output) a scientist cited in a publication was not directly associated with the essentiality of these citations (termed as citation essentiality). Instead, citation output was related to an external factor, while citation essentiality was related to a number of internal motivations. As a result, a citation relationship model was established. The study shows that an author's citing behavior is unique, personal and complex. Further investigations are needed to articulate the nature and norms of this more-private-than-public process.

Another study on citation comprehensiveness [Lichbach, 1992] surveys nearly two hundred scholarly works that use mathematical methods, which include stochastic models, difference and differential equation models, expected utility models, and various types of game theoretic models, to study domestic political conflict (DPC), which includes terrorism, guerrilla wars and insurrections. A citation count reveals that the DPC articles surveyed here cite less than three quarters of an article from within their own DPC modelling tradition and cite less than two articles from any DPC modelling tradition. The only exceptions to the rule that "nobody cites nobody else" am the stochastic and expected utility modelers. The author concludes conclude that the "field" of formal models of DPC hardly exists: few authors read other authors, few articles cite other articles, few models build on other models. Several suggestions aimed at promoting greater accumulation in formal models of DPC are offered.

Relationships Between Cited and Citing Articles

It is assumed that a paper which cites an earlier document shares a subject relationship with that particular document. In order to determine if this assumption is valid, a study was conducted by analysing 1000 articles from the Science Citation Index(R) and Social Sciences Citation Index(R) [Ali, 1993]. These articles were selected in ten different disciplines by using a purposive sampling technique. Various Spearman's Correlation Coefficient tests were computed to find out if a subject relationship existed between the Articles which have the same keywords in their titles (Parent Articles and Related Records). Through the analysis, the hypothesis has been verified showing that there is a relationship between the articles which are citing the same references. This was determined by co-occurrences of the same keywords among the shared references. However, there are some unique differences in the science and the social science disciplines that exist in these two databases.

A somewhat different perspective was obtained in another study using a different approach [Harter, 1993]. This study examined directly the assumption that the act of referencing another author's work in a scholarly or research paper is usually assumed to signal a direct semantic relationship between the citing and cited work. The purpose of the research was to investigate the semantic relationship between citing and cited documents for a sample of document pairs in three journals in library and information science: Library Journal, College and Research Libraries, and Journal of the American Society for Information Science. A macroanalysis, based on a comparison of the Library of Congress class numbers assigned citing and cited documents, and a microanalysis, based on a comparison of

descriptors assigned citing and cited documents by three indexing and abstracting journals, ERIC, LISA, and Library Literature, were conducted. Both analyses suggested that the subject similarity among pairs of cited and citing documents is typically very small, supporting a subjective, psychological view of relevance and a trial-and-error, heuristic understanding of the information search and research processes. The results of the study have implications for collection development, for an understanding of psychological relevance, and for the results of doing information retrieval using cited references. Several intriguing methodological questions are raised for future research, including the role of indexing depth, specificity, and quality on the measurement of document similarity.

Citation Problems

Five core library science journals were examined to study the accuracy of citations in library literature [Pandit, 1993]. A total of 1,094 references from 131 articles were verified directly by comparing the published citation with the original publication. In 193 references, 223 errors were detected. A review of citations at manuscript stage was also carried out for one of the journals. The results of the study show that library and information professionals, in spite of their awareness of difficulties posed by inaccurate citations, are prone to making such mistakes themselves. The study emphasizes a need for greater awareness among LIS professionals of keeping their citations error free, and suggests other aspects of the subject for further study.

Another study examined ethnic bias in citation practices [Greenwald, 1994]. Recent experimental findings of subtle forms of prejudice prompted a search for a similar phenomenon outside the laboratory. In Study 1, with a sample of more than 12 000 citations by North American social scientists, names of both citing and cited authors were classified as Jewish, nonJewish, or other. Author's name category was associated with 41 per cent greater odds of citing an author from the same name category. Study 2 included over 17 000 citations from a much narrower research domain (prejudice research), and found a similar (40 per cent) surplus in odds of citing an author of the author's own ethnic name category. Further analyses failed to support two hypotheses - differential assortment of researchers by ethnicity to research topics, and selective citation of acquaintances' works - that were plausible alternatives to the hypothesis that the observed citation discrimination revealed implicit (unconsciously operating) prejudicial attitudes. The authors conjectured that, given the sociopolitically liberal reputation of social scientists (and of prejudice researchers especially), it seems unlikely that the observed bias in citations reflected conscious prejudicial attitudes.

A study on highly cited papers describes examples of influential and/or highly cited papers that were initially rejected by one or more scientific journals [Campanario, 1995]. The work reported in eight of the papers eventually earned Nobel prizes for their authors; six papers later became the most cited of the journals in which they were published. Also described are influential and highly cited scientific books whose authors encountered problems in publishing them. These case studies suggest that, although rejection may subsequently result in an improved manuscript, on other occasions referees may simply have failed to appreciate a paper's importance. Many of these rejected papers also reported unexpected findings or discoveries that challenged conventional models or interpretations.

Research Citation Impact

In the opinion of the authors of a study on citations in mathematics [Korevaar, 1996], many mathematicians are not convinced that citation counts do in fact provide useful information in the field of mathematics. According to these mathematicians, citation and publication habits differ completely from scholarly fields such as chemistry or physics. Therefore, it is impossible to derive valid information regarding research performance from citation counts. The aim of the present study was to obtain more insight into the significance of citation-based indicators in the field of mathematics. In particular, to what extent do citation-scores mirror the opinions of experts concerning the quality of a paper or a journal? A survey was conducted to answer this question. Top journals, as qualified by experts, receive significantly higher citation rates than good journals. These good journals, in sum, have significantly higher scores than journals with the qualification less good. Top publications, recorded in the ISI database, receive on the average 15 times more citations than the mean score within the field of mathematics as a whole. In conclusion, the experts' views on top publications or top journals correspond very well to bibliometric indicators based on citation counts.

Another study [Plomp, 1994] examined the highly cited papers of professors as an indicator of a research group's scientific performance. In the first part of the study, the citations in 1986 and 1987 of 3938 papers published in 1985 by 324 research groups in the faculties of science and of medicine of eight universities in the Netherlands were analyzed. Because of the large statistical spread of (1) the number of short-term citations of papers cited equally frequently over a long period, and (2) the number of citations over a long period of papers by the same author, short-term citation scores appear to be an unreliable indicator of a research group's contribution to science. In the second part of the study an alternative approach is presented, based on a subdivision of the 3938 papers in papers authored by professors with 0-2, 3-8, or greater-than-or-equal-to 9 highly cited papers (HCPs, greater-than-or-equal-to 25 citations) to their name. Very large citation score differences were found for the three categories. For example: for papers first-authored by a professor, the average number of citations per person in 1986 and 1987 for 1985 papers was for 161 professors with greater-than-or-equal-to 9 HCPs a factor 14 larger than for 575 professors with only 0-2 HCPs; for papers co-authored by professors, this factor was 6.6. These findings justify the conclusion that the number of HCPs scored by the professors (and other senior scientists) during their entire career is a much more reliable predictor of the performance of a research group than the number of short-term citations of the articles published by the group within a short period. A research group's contribution to science is primarily determined by the individual scientific talents of its members.

A third study in this section [Eom, 1993] identified the most influential contributors in the DSS area in the U.S., examined their contributions, and reviewed the institutional publishing records at the leading U.S. universities which are actively publishing DSS research. To measure the influence/contributions of leading universities and contributors, the authors used the bibliographic citations of the publications on the specific DSS applications. The critical assumption of this study was that "bibliographic citations are an acceptable surrogate for the actual influence of various information sources." (M.J. Culnan, Management Science 32, 2, feb 1986, 156-172) This study identified thirty-two leading U.S. universities with eighty-one of their affiliated members and twenty three most influential researchers. Among the leading U.S. universities identified, two universities are truly outstanding: The University of Texas-Austin and MIT. Regardless of any types of yardsticks which may be applied to measure their contributions, these two universities may be recognized as centers of excellent DSS research in the U.S.A. in terms of the number of research

publications, the number of total citation frequencies, and the number of active researchers in the DSS related areas.

A fourth study [Anglin, 1991] focused on the patterns of communication in the field of instructional technology and examined the reference lists provided with each article or review in three journals for a period of five years to determine: (1) who the most cited authors in the field are; (2) whether invisible colleges exist in the field; and (3) if invisible colleges do exist, who the participants are in each invisible college. The journals studied were the Journal of Instructional Development (JID, Spring 1985-Fall 1990); Educational Communication and Technology Journal (ECTJ, Spring 1985-Summer 1990); and Performance Improvement Quarterly (PIQ, Spring 1988-Fall 1990).

The name of each author, co-author, or editor of works cited was entered in a database together with the name of the journal, date of citation, and volume and issue numbers of the journal. The number of citations per author was recorded, and individuals were included in the study if they had been cited a minimum of five times. From the 12,220 citations entered in the database for all three journals, 386 individuals were selected. The highest numbers of citations reported were 83 (R. M.Gagne), 76 (R. D. Tennyson), and 43 (R. Kaufman). The results of a hierarchical cluster analysis among frequently cited individuals identified 53 homogeneous groups. For many of the groups dominant individuals could also be identified. The results of the study support the conclusion that there are 'many' invisible colleges in the field, and that the groups of frequently cited individuals do significantly influence the development of the field and the practice of industrial design.
The final study in this section [Adams, 1996] examined the available United States data on academic research and development (R&D) expenditures and the number of papers published and the number of citations to these papers as possible measures of "output" of this enterprise. The authors examined these numbers for science and engineering as a whole, for five selected major fields, and at the individual university field level. The published data in Science and Engineering Indicators imply sharply diminishing returns to academic R&D using published papers as an "output" measure. These data are quite problematic. Using a newer set of data on papers and citations, based on an "expanding" set of journals and the newly released Bureau of Economic Analysis R&D deflators, changes the picture drastically, eliminating the appearance of diminishing returns but raising the question of why the input prices of academic R&D are rising so much faster than either the gross domestic product deflator or the implicit R&D deflator in industry. A production function analysis of such data at the individual field level follows. It indicates significant diminishing returns to "own" R&D, with the R&D coefficients hovering around 0.5 for estimates with paper numbers as the dependent variable and around 0.6 if total citations are used as the dependent variable. When scientists and engineers are substituted in place of R&D as the right-hand side variables, the coefficient on papers rises from 0.5 to 0.8, and the coefficient on citations rises from 0.6 to 0.9, indicating systematic measurement problems with R&D as the sole input into the production of scientific output. But allowing for individual university field effects drives these numbers down significantly below unity. Because in the aggregate both paper numbers and citations are growing as fast or faster than R&D, this finding can be interpreted as leaving a major, yet unmeasured, role for the contribution of spillovers from other fields, other universities, and other countries.

IV-B-6-ii-c. Patents and Patent Citations

Patent citations, especially to research papers cited by the patents, provide some indication of

science to technology conversion. Probably the most consistent organization producing studies of different aspects of patent citations over the past decade has been CHI, Inc. The first few studies described summarize key aspects of CHI's work over this period.

CHI Efforts

In the first study [Kitti, 1983], quantitative indicators for foreign technological presence in the United States were reported on the basis of data derived from the front pages of U.S. patents issued from 1971-1980. It was noted that the percent of foreign-owned and -invented patents in the U.S. patent system increased from 26 percent in 1971 to 38 percent in 1980. The areas with the greatest increases were those where there had been recent influxes of foreign products--for example, motorcycles, radios and televisions, and primary metals. It was found that the percent of citations given by foreign-owned and -invented patents in the U.S. to foreign origin patents in the U.S. system was two and one-half times as large as those given by U.S.-owned and -invented patents to foreign origin patents. In addition, approximately one-fourth of all U.S. patents from 1971-1980 were owned by multi-national corporations. It was suggested that research be undertaken to address the relationship between these indicators and various economic and trade statistics.

A subsequent analysis [Narin, 1986] of Japanese-invented patents appearing in the U.S. patent system over the 10-year period 1975-84, showed that the share of U.S. patents with Japanese inventors increased from 8.8% of all U.S. patents in 1975 to 16.5% in 1984, while the share of patents with U.S. inventors decreased from 64.9% to 57.1%. Japanese inventors obtained 8% more U.S. patents while U.S. inventors obtained 8% fewer, and the rest of the world's inventors remained approximately constant: in the U.S. patent system, the increase in Japanese share was entirely at the expense of the United States. The Japanese patents were shown to be quite concentrated in relatively high-technology classes related, especially, to those areas of consumer products where there is a major Japanese presence, including electronics, photography, and automotive technology. There was also a growing Japanese presence in the pharmaceutical area. When looked at from the point-of-view of citation analysis--that is, considering highly cited patents to be patents of particular technical impact and quality--the Japanese performance was just as impressive. Among the most highly cited few percent of U.S. patents, the Japanese have 30 to 50% more patents than expected, and the Japanese inventors are patenting in the most highly cited 1% of patents--the areas in which the Japanese have substantial numbers of these very highly cited patents are automotive technology, semiconductor electronics, photocopying and photography, and pharmaceuticals and pharmaceutical chemistry. The implication of all of the above is that the Japanese position in patented technology is strong, growing and based on high quality, high impact technology which has been invented by Japanese inventors.

The third study's research [Narin, 1985] formulated a series of quantitative indicators of corporate technological strength, using data from U.S. patents and U.S. patent citations. These indicators were generated for 18 U.S. pharmaceutical companies. The research then examined the extent of correlation between peer judgement of research performance, literature-based indicators of research publication, corporate financial performance, and the various patent and patent citation indicators. The findings implied that not only are counts of patents an excellent indicator of overall corporate technological strength, but also that the occurrence of highly-cited, high-impact patents may be a particularly good indicator of corporate growth.

The final two studies reported describe the Tech-Line database and some results from studies with the database. TECH-LINE CD provides technology indicators to complement existing financial data [CHI, 1996]. With TECH-LINE data, financial analysts, corporate analysts and economists can determine an organization's technological strength and trends, and technologically rank and compare companies within an industry for long-term investment strategies. TECH-LINE's company profiles allow an analyst to compare a company's technological strength to its financial performance. TECH-LINE measures technological strength, activity, and position for over 1,000 public and private companies, universities and government agencies worldwide which received the most U.S. patents in the last five years. TECH-LINE's company indicators are based on 500,000 U.S. patents and nearly 4,000,000 patent citations. TECH-LINE is important because technology is the major force driving industrial companies, and any comprehensive assessment of a technological company must include an analysis of its technological strength. Companies with high technological strengths are likely to prosper, while companies with obsolete technologies are likely to decline. TECH-LINE indicators are designed to complement financial indicators, so that technological excellence can be used as an explicit measure of value of an individual company, or region, or industry, or nation. Each organization's strength in TECH-LINE is profiled both overall and within 57 SIC product groupings.

A basic description of patent citation cycles is provided for 1,100 major companies and organizations covered by the TECH-LINE database [Narin, 1993]. The average U.S. patent has five to six "references cited-U.S. patent documents." The properties of these patent citations are shown to vary widely from one technology to another. For example, patents in Office Computing and Accounting, a relatively hot area, are cited almost three times as frequently as patents in Organic Chemicals, a less active area of patenting. Similarly, technology cycle times vary widely-from five to six years in fast moving electronics areas to twelve to fifteen years in some of the slow moving areas of mechanical technology. Citations to earlier patents peak at patents three to five years old, rather similar to the peak citation time for scientific literature. Since these citation peaks and cycle times are relatively short, and represent the difference between current art and prior art, this indicates, in one sense, that the technological lifetime of an invention may be much shorter than its legal and commercial life times.

Geographic Boundary Flows

The extent to which new technological knowledge flows across institutional and national boundaries is a question of great importance for public policy and the modeling of economic growth. In this study [Jaffe, 1993a], the authors develop a model of the process generating subsequent citations to patents as a lens for viewing knowledge diffusion. They find that the probability of patent citation over time after a patent is granted fits well to a double-exponential function that can be interpreted as the mixture of diffusion and obsolescense functions. The results indicate that diffusion is geographically localized. Controlling for other factors, within-country citations are more numerous and come more quickly than those that cross country boundaries.

A related study [Jaffe, 1993b] compares the geographic location of patent citations with that of the cited patents, as evidence of the extent to which knowledge spillovers are geographically localized. We find that citations to domestic patents are more likely to be domestic, and more likely to come from the same state and SMSA as the cited patents, compared with a "control frequency" reflecting

the pre-existing concentration of related research activity. These effects are particularly significant at the local (SMSA) level. Localization fades over time, but only very slowly. There is no evidence that more "basic" inventions diffuse more rapidly than others.

Technological Niches

This study [Almeida, 1997] examined the innovative ability of small firms in the semiconductor industry regarding their exploration of technological diversity and their integration within local knowledge networks. Through the analysis of patent data, the authors compared the innovative activity of start-up firms and larger firms. They found that small firms explore new technological areas by innovating in less 'crowded' areas. The analysis of patent citation data revealed that small firms are tied into regional knowledge networks to a greater extent than large firms. These findings point to the role of entrepreneurial firms in the exploration of new technological spaces and in the diffusion of their accumulated knowledge through local small firm networks.

Another study [Podolny, 1995] considered what factors determine whether an innovation becomes a foundation for future technological developments rather than a "dead end." The authors introduced the concept of the technological niche, which includes a focal innovation, the innovations on which the focal innovation builds, the innovations that build upon the focal innovation, and the technological ties among the innovations within the niche. Using patents and patent citations to measure characteristics of innovation niches within tile semiconductor industry, the authors showed that the size of the niche and the status of the actors within the niche have a positive effect on the likelihood that subsequent innovations will build upon the focal innovation. Competitive intensity within the niche has a negative effect on this likelihood.

In a subsequent study [Podolny, 1996], the conception of an organization-specific niche is defined by two properties: crowding and status. The authors hypothesize that crowding suppresses an organization's life chances and that status enhances life chances, especially for those organizations in uncrowded niches. They operationalize this conception of the niche using patents and patent citations, and they find support for these hypotheses in an examination of technological competition in the worldwide semiconductor industry. In the conclusion, they compare these findings to the earlier research and highlight some of the particular advantages of this conception of the niche.

Defense Technology Transfers

Although technology is considered to be a strategic asset for an organization, interplay in technology among organizations is necessary. Technology may be considered a bank which organizations both contribute to and draw from. Such interactions among organizations in technology follow different patterns. This study [Chakrabarti, 1993] presented some preliminary results from a study that aimed at addressing this issue. By using patent-citation data, this study showed how the benefits to participating firms change with industry type, organization class, country of origin, etc.

A follow-up study [Chakrabarti, 1994] investigated the pattern of transfer of technology between defence firms and other organizations. Using eight large defence contractors, Boeing, General Dynamics, Grumman, Lockheed, Martin Marietta, McDonnell Douglas, Raytheon and United Technologies, as sample, the authors analysed their patents. They were particularly interested in the

pattern of citations. By using patents as the 'tracer' in the technology interaction, the authors were able to characterize the pattern, nature and effectiveness of the technology interactions between the defence and non-defence sectors. In the exchange of technological information between a firm and other organizations, the authors defined technology input to a firm X as the citations of patents made by firm X. Similarly, technology output of firm X was defined as the number of citations received by its patents from other patents of other organizations. Once the authors knew the identity of the organizations, they could observe the technology exchange between the defence and the non-defence sectors, between the US defence firms and foreign firms. The intensity and efficiency of transfer of technology were computed from these data.

IV-B-6-ii-d. Combinations of Publications/ Patents/ Citations

The purpose of the first study presented [Reisher, 1984] was to determine the degree to which the NIDR Dental Research Institutes and Centers achieved selected program objectives relating to resource utilization and recruitment, multidisciplinary research and collaborations with other institutions. A bibliometric comparison of papers from the Centers with papers published under investigator-initiated R01 grants was undertaken to test eight hypotheses on the following topics: frequency of publication; impact of publication; type and number of support; multiple authorship; multidisciplinarity; width of utilization, and scientist background. Some conclusions of this preliminary study were that the Centers' scientists are of similar productivity and quality as the R01 investigators as measured by the number of papers per scientist per year, and by the number of citations per paper.

The second study [Nederhof, 1993] involved a comparison of bibliometrics results with peer review. The research performance of research units in economics was evaluated by simultaneous efforts of peers and bibliometricians, with extensive interactive comparison of results afterwards. The authors studied trends in productivity and impact of six economics research groups in the period 1980-1988. These groups participate in a large (above one million pounds) research programme of a national Research Council. Research performance of the groups was compared to the world average by means of the Journal Citation Score method. In order to investigate the influence of one key scientist (the "star effect"), the authors applied a sensitivity analysis to the performance of the research groups by elimination of the papers (and subsequent citations) of one key member. Furthermore, to provide insight into the fields to which a group directs its work, and the fields in which a group has its most important contributions, comparisons were made of publishing and citing journal packets. Similarly, citations to the work of the research groups were analysed for country of origin. The authors compared the results of the bibliometric part of this study with those of a simultaneous peer review study. The bibliometric study yielded clear and meaningful results, notwithstanding the increasingly applied nature of the research groups. Results from peer review and bibliometric studies appear to be complementary and mutually supportive. The participants of the bibliometrics peer review "confrontation" meeting regarded the exercise as most valuable, with lessons for the Research Council both for the future of research programmes and the form of evaluation used for large awards.

The final study reported in this section involved examination of publication and citation rates [McGinnis, 1982]. The careers of 557 biochemists were studied in order to answer the following questions: Who gets postdoctoral training and why? How does such training affect subsequent

employment opportunities? Does postdoctoral training increase later research productivity? Results showed that predoctoral research productivity had no effect on who gets postdoctoral training or where one gets it. Getting postdoctoral training does not seem to affect one's chances of getting a prestigious job, but where the training occurred has a major impact on the prestige of subsequent jobs. In contrast, having had postdoctoral training seems to result in substantial increases in later citation rates, but where the training occurred makes little difference in citation rates. The modest effect of postdoctoral training on publication rates disappears when employment sector is held constant.

Appendix 7 contains selected examples of bibliometrics studies performed for a variety of science and technology disciplines by the author's group.

IV-B-6-ii-e. Science and Technology Transitions

In practice, one of the most widely used metrics for gauging the progress of science and technology is transition metrics. These are metrics that incorporate

- the number of transitions (across development levels) per unit of time,
- the potential impact or benefit eventually resulting from these transitions, and
- the probability that each transition will eventually achieve the potential impact

A more detailed analysis of transition metrics is contained in Appendix 8.

IV-B-6-ii-f. Collaboration Indicators

Collaboration among researchers has been increasing steadily for decades. This collaboration has covered different:

- disciplines;
- development categories;
- institutions;
- geographical regions;
- countries, etc.

There is a belief that collaboration improves the quality of the final research product by bringing different perspectives to bear on solving the problem. In particular, approaches used to solve problems in one field may be extrapolated or modified to solve conceptually similar problems in other fields. A 1997 article in The Washington Post on innovation examined research performed on teams of collaborators. The summary findings were that innovation was enhanced when the teams consisted of researchers from disparate disciplines, and that innovation was not enhanced over that of individual investigators when the teams consisted of individuals from similar backgrounds/ disciplines.

There have been a number of studies examining the impact of collaboration on quality, innovation, technology transfer, and other quantities. While collaboration can be viewed as a metric, it is more of an intermediate or proxy or management metric as opposed to a definitive quality metric such as citations or awards or cost/ benefit. Similar to the output/ impact metrics discussed previously, the collaboration metric suffers from lack of theoretical understanding as to what ultimate values should be, and therefore its use is limited as a management target or control.

For example, in the illustrative case of vertical integration at the end of this section, what should be the management targets for the appropriate mix of basic research/ applied research/ early technology development/ advanced technology development in a given vertical structure, or in a group of vertical structures? Without this target or control, what meaning can one assign to a specific vertical integration metric? Nevertheless, because of the growing importance of collaborations, it will be treated here as a separate S&T metric. In particular, the last study reported in this section [Kostoff, 1997c] describes how collaboration can help accelerate the conversion of science to technology. Associated commentary following the study summary describes potential metrics for quantifying the effects of vertically integrated program management on quality and transitionability of the science and technology product. It should be noted that the collaboration process (interdisciplinary research) has many associated disincentives relative to mono-discipline research, as stated in the Introduction. A 2002 article [Kostoff, 2002g] addresses these disincentives in detail.

University-Industry Collaboration

The first two projects reported deal with university/ industry collaboration. The first study [Tornquist, 1996] investigates the assumption that scientific research taking place in universities "trickles down" to industry. Publication characteristics are used to examine the collaboration and utilization behavior of scientists employed in the computer equipment and aircraft industries. The data indicate that these industries are using research generated by university scientists and that collaboration between sectors is occurring. Four sets of factors (article, firm, industry , and university characteristics) are used to explain research utilization and publication practices. Logistical regression results confirm that university/ firm proximity is associated with increased collaboration and that collaborative relationships promote firm utilization of university research. These results indicate that university policymakers should consider ways to encourage collaborative relationships between sectors to promote information transfer. Further, the result linking proximity and collaboration suggests support for academic scientific activities should be encouraged at the local level.

Previous studies on collaborative research emphasize industry/ university collaboration conducted in a subset of academic disciplines associated with applied engineering. These studies focus on motivations, mechanisms, financial costs and financial benefits of collaborative research while paying little attention to the impact of collaborative research on academic productivity. The purpose of the second study reported on university/ industry collaboration [Landry, 1996] is to attempt to compensate for some of these shortcomings. First, the authors present a survey which includes responses from academic researchers of all the scientific disciplines. Second, the study takes into account and compares the collaborative relationships between university researchers, between university researchers and industry, and between university researchers and other institutions,

especially government agencies, local governments and organized interest groups. And third, the authors assess the impact of these collaborative activities on the academic productivity of the university researchers.

The results of this study show that collaboration, whether it be undertaken with universities, industries or institutions, may indeed increase researchers' productivity. The authors find this to be true whether or not such relationships begin early in a researcher's career. They also find this to be true whether or not the collaborators have an intellectual symmetry. The effect of collaboration on productivity varies according to both the scientists' geographical closeness to their partners and on their field of research. It was found that collaboration between researchers and industry had significantly more impact on productivity than collaborations between researchers and their peers or researchers and other institutions. Scientists in humanities were found to produce less materials in collaboration than scientists in other fields. And, scientists involved in collaboration aimed mostly at producing patented and unpatented products, scientific instruments, software and artistic production were also found to produce less. In sum, given that collaboration contributes to the increase of scientific productivity, the authors conclude that government decision makers and university administrators should encourage researchers to forge collaborative relationships.

Biomedical Collaboration

The third project reported in this section [Zucker, 1996] concerns collaboration of 'star' scientists with other researchers. The authors found that the most productive ("star") bioscientists had intellectual human capital of extraordinary scientific and pecuniary value for some 10-15 years after Cohen and Boyer's 1973 founding discovery for biotechnology [Cohen, S., Chang, A., Boyer, H. & Helling, R. (1973) Proc. Natl. Acad. Sci. USA 70, 3240-3244]. This extraordinary value was due to the union of still scarce knowledge of the new research techniques and genius and vision to apply them in novel, valuable ways. As in other sciences, star bioscientists were very protective of their techniques, ideas, and discoveries in the early years of the revolution, tending to collaborate more within their own institution, which slowed diffusion to other scientists. Close, bench-level working ties between stars and firm scientists were needed to accomplish commercialization of the breakthroughs. Where and when star scientists were actively producing publications is a key predictor of where and when commercial firms began to use biotechnology. The extent of collaboration by a firm's scientists with stars is a powerful predictor of its success: for an average firm, 5 articles coauthored by an academic star and the firm's scientists result in about 5 more products in development, 3.5 more products on the market, and 860 more employees. Articles by stars collaborating with or employed by firms have significantly higher rates of citation than other articles by the same or other stars. The U.S. scientific and economic infrastructure has been particularly effective in fostering and commercializing the bioscientific revolution, These results provide insight to the process by which scientific breakthroughs become economic growth and consider implications for policy.

Another study [Bordons, 1996] also examined collaboration in biomedical research. Bibliometric indicators were used to analyse international, domestic and local collaboration in publications of Spanish authors in three Biomedical subfields: Neurosciences, Gastroenterology and Cardiovascular System as covered by the SCI database. Team size, visibility and basic-applied level of research were analysed according to collaboration scope. International collaboration was linked to higher

visibility documents. Cluster analysis of the most productive authors and centres provides a description of collaboration habits and actors in the three subfields. A positive correlation was found between productivity and international and domestic collaboration at the author level.

International Collaboration

This project [Luukkonen, 1993] provided further evidence of the value of international collaboration. A growing science policy interest in international scientific collaboration has brought about a multitude of studies which attempt to measure the extent of international scientific collaboration between countries and to explore intercountry collaborative networks. This study attempts in particular to clarify the methodology that is being used or can be used for this purpose and discusses the adequacy of the methods. The study concludes that, in an analysis of collaborative links, it is essential to use both absolute and relative measures. The latter normalize differences in country size. Each yields a different type of information. Absolute measures yield an answer to questions such as which countries are central in the international network of science, whether collaborative links reveal a centre - periphery relationship, and which countries are the most important collaborative partners of another country. Relative measures provide answers to questions of the intensity of collaborative links.

The next study [Melin, 1998] examines international collaboration patterns of selected European countries. The collaborative pattern of all Nordic universities, as well as a few universities in the UK and the Netherlands, is analyzed using institutionally cc-authored articles retrieved from Science Citation Index.(TM) The study shows that there are no major differences between universities of various size when it comes to the proportion of articles with internal, national, or international co-authorships. There are some country variations, but within each country, the differences among the universities are small, if any. When cc-authorships were fractionalized according to the number of times a given university occurs among the addresses of an article, there were still no significant differences between universities of varying size. Since external collaboration, whether it is national or international, accounts for more than half of all articles produced by the universities, one is inclined to conclude that the universities function as a kind of cosmopolitan hotel housing nodes of scientific networks that are becoming increasingly international.

Economic Impacts of Collaborative Research

The next two studies reported examine the economic impacts of collaborative research. American companies have embraced collaborative research ventures as an organizational form conducive for carrying out critical, advanced research programs. This is evidenced, in part, by the rapid growth in consortium research since the passage of the US National Cooperative Research Act of 1984. However, there is a conspicuous absence of detailed case studies that document the returns to member companies involved in collaborative research ventures. This void is due to the perception, both on the part of consortium managers and member companies, that such an evaluation would lack rigour and be too cumbersome to undertake. This study [Link, 1997] presents a general methodology for evaluating the returns to collaborative research membership, and illustrates it by summarizing an analysis of the private returns to the corporate members of a cooperative research venture called SEMATECH.

The second economics-related collaboration study examines defense procurement [Hartley, 1993]. International collaboration in the development and production of defence equipment is said to reduce procurement costs and improve export prospects. However, critics argue that joint ventures cost more than national programmes, are more prone to cost escalation and take longer to complete. These claims are evaluated by comparing collaborative and national military aircraft using a variety of performance indicators. The evidence suggests that for military aircraft collaboration leads to cost savings and greater scales of output, with only limited support for the view that joint projects take longer to develop. There is little evidence that collaborative projects perform better in export markets than their national rivals.

A Dissenting View on Collaboration

The next study in this section presents a somewhat different view on the value of collaborative research [Avkiran, 1997]. The study reports an empirical comparison of quality of collaborative research with the quality of individual research. Quality of a paper is measured by the citation rate over the four years following the year of publication. Papers published in fourteen Finance journals between 1987-1991 are sampled. The study author finds there is no significant difference between the quality of collaborative and individual research. He recommends that decision-makers should hesitate in interpreting collaborative research as a definitive sign of ability to produce better research.

IV-B-6-ii-f-1. Collaboration Indicators for Vertical Integration

The one study in this section [Kostoff, 1997c] focuses on the value of collaboration for accelerating the conversion of science to technology. The study shows that, as the technology marketplace has become global, the efficient and timely transfer of technology has assumed paramount importance. Delays in commercializing technologies can translate into surrendering substantial market shares to national or international competitors. The study also asserts that there is very little in the literature addressing the problem of how science, especially fundamental science, gets converted eventually to technology, and how the efficiency (minimization of time and other resource utilization) of this process can be improved. The study then provides examples of how different types of collaboration can help address some of these problems.

The study starts by defining the two major variants of retrospective studies which have examined the science-technology evolution process. One type starts with a successful technology or system and works backwards to identify the critical R&D events which led to the end product. The other type starts with initial research grants and traces evolution forward to identify impacts. The tracing backwards approach is favored for two reasons: 1) the data are easier to obtain, since forward tracking is essentially non-existent for evolving research; and 2) the sponsors have little interest in examining research that may have gone nowhere.

Many examples of retrospective studies are presented and discussed. In particular, in the 1960s, a study named Project Hindsight was sponsored by the Department of Defense [DOD, 1969]. Hindsight examined twenty successful military systems, and identified the critical R&D events which led to the successful systems. Hindsight examined characteristics of these critical R&D events to see whether any general principles could be extracted. While there were problems with some of the constraints placed on the Hindsight study, nevertheless, some valuable conclusions emerged. In particular, a major conclusion related to the science-technology conversion process was

that the results of research were most likely to be used when the researcher was intimately aware of the needs of the applications engineer.

From the author's viewpoint, Project Hindsight, with all of its limitations [Kostoff, 1994d, 1997n], produced very relevant findings for the science-technology conversion problem. A conceptual principle for accelerating the science-technology conversion can be abstracted from the Hindsight results, and it is important to separate the conceptual principle from the implementations of the principle. In this manner, one does not become bound by the limitations of any particular implementation. This principle, termed by the author as Heightened Dual Awareness (HDA) [Kostoff, 1997c], states that in order for the science-technology conversion to be accelerated, at least two necessary conditions must be fulfilled: 1) the researcher must be intimately aware of the needs of the applications engineer; 2) the potential user of the research, or transitionee, must be aware of the progress and results of the research. In addition, if third parties are involved in the conversion and development process, such as vendors, their awareness of both ends of the conversion cycle must be maintained as well. To the degree that each of these requirements is not fulfilled, the science-technology conversion will be retarded and delayed.

In the study, a number of laboratory examples illustrate the most straightforward application of the HDA principle. The researchers and developers are physically contiguous, and in many cases are the same person. Thus, the dual awareness is readily effected by the intrinsic structure of the physical environment, and complex management structures are not necessary to enhance dual awareness. However, it is also shown that the HDA principle as a major driver of eventual utility is not limited to the performer and potential user; it is applicable to the research sponsor environment as well. A number of research sponsoring organizations have switched from a discipline orientation to a structure where the research is vertically integrated with technology, analogous to the vertically integrated research-technology performer environment described above. This includes both industrial organizations, where on the whole central research laboratories have declined and research has been shifting to the business units, and some government agencies.

The general conclusion that the author draws in the study is that for most effective and efficient conversion of science to technology, the researcher primarily and the sponsor secondarily need to be immersed in environments where the HDA principle is most operative, and where motivations and incentives are geared toward rapid transitioning. This type of physical environment is realized most efficiently when the researchers and developers are physically contiguous. If this type of physical environment structure is not readily possible, as may be the case with some fundamental university research, then attempts should be made to simulate this optimal transitioning environment through innovative management structures. This should not be interpreted as a recommendation to substitute applied research for basic research. Far too much of this substitution has occurred in the recent past. Rather, the recommendation is that basic research be conducted in an environment where there is greater awareness of the progress and potential of the research by potential transitionees and users, and opportunities to understand the needs of the developers are made available to the researchers.

The author further concludes that, for mission-oriented agencies, to enhance the simulation of optimal transitioning physical structures, joint university-federal or national or corporate laboratory projects should be expanded. In parallel, as the author's personal observations have also shown, the potential user needs to become involved in the research project as early, broadly, and intensely as

possible. This early involvement provides the user a sense of 'ownership', and produces a more seamless transition process. In the author's experience, incorporating the potential user from the research proposal evaluation phase is not too soon for successful downstream transitions of the research products to technology.

In the study (and above), it was asserted that the HDA principle is applicable to the research sponsor environment as well as the research performer environment. Since the publication of the study in late 1997, the author has been examining and developing metrics which could determine how well research has been vertically integrated with technology and mission capability requirements in a science and technology sponsor environment. Some preliminary conclusions from these collaborative metrics studies will be presented. First, as necessary background, different types of integrated programs will be discussed, in the context of Federal agency programs.

The target of global optimization for achieving aggregated agency long range goals leads to two top-level requirements which must be considered in formulating research evaluation recommendations. Is the research of high intrinsic quality and horizontally (cross-agency) and laterally (cross-discipline) integrated among the funding agencies and balanced across the different disciplines to ensure an optimal national pool of high quality knowledge, and is the research vertically (cross-category) integrated within the agencies to ensure that long range agency objectives will have a maximal chance of being impacted? Horizontal and lateral integration tend to be associated with QUALITY (is the job being done right?) and vertical integration with RELEVANCE (is the right job being done?), with the ultimate assessment issue being QUALITY-RELEVANCE (is the right job being done right?).

HORIZONTAL COUPLING/ INTEGRATION

Under the present national structure of public research sponsorship, responsibility for funding any research discipline is divided up among different Federal agencies. Each agency focuses on sponsoring the research necessary to impact the agency's unique long range objectives. Because of the unified nature of research, the different components of a research discipline funded by the different agencies are related, and there are multiple relationships among different disciplines.

From a national perspective, the aggregated research components in any research discipline should be complementary. There should be minimal duplication, and there should be minimal gaps in the research requirements and opportunities addressed for the funding available. Thus, there should be some measure of horizontal coupling among the agencies to ensure the research discipline components are complementary on a national scale.

The degree of horizontal coupling can be divided into three categories: horizontal awareness, horizontal coordination, and horizontal integration. In horizontal awareness, an agency's research managers are aware of other agencies' efforts in the discipline and plan their programs accordingly, but there is no joint planning, execution, or evaluation within the discipline. In horizontal coordination, there may be some combination of joint planning, execution, and evaluation at different intensity levels. In horizontal integration, joint efforts are strengthened while allowing each agency to retain autonomy for managing the research necessary to optimize its overall objectives.

LATERAL COUPLING/ INTEGRATION

From the national program perspective, different research disciplines which have intrinsic relationships should be conducted and managed in a complementary manner. Thus, there should be some measure of lateral (cross-discipline) intra- and inter-agency coupling to ensure that intrinsically related disciplines are complementary on a national scale.

The degree of lateral coupling can be divided into three categories: lateral awareness, lateral coordination, and lateral integration. In lateral awareness, research discipline managers are aware of other intra- and inter-agency efforts in related disciplines and plan their programs accordingly, but there is no joint planning, execution, or evaluation among the related disciplines. In lateral coordination, there may be some combination of joint planning, execution, and evaluation of related disciplines at different intensity levels. In lateral integration, joint efforts among related intra- and inter-agency disciplines are strengthened while allowing each agency to retain autonomy for managing the research to optimize its overall objectives.

VERTICAL COUPLING/ INTEGRATION

Analogous to the horizontal and lateral coupling categories are vertical coupling categories. While the main focus of vertical coupling is within a given agency, vertical coupling can transcend agencies. Because of the unified nature of research, products of research from one agency can transition to other agencies' programs. Thus, planners of vertically coupled R&D programs in one agency must be continually aware of existing and planned R&D programs of other agencies. The key point to be made is that vertical coupling is not independent of horizontal or lateral coupling. Vertical integration is linked with horizontal and lateral integration. One major focus of agency research assessment from the national perspective should be the degree to which **DIAGONAL INTEGRATION** (**horizontal**, **lateral**, and **vertical** integration) is being achieved.

The vertical coupling categorization is vertical awareness, vertical coordination, and vertical integration. In vertical awareness, the research and development managers are aware of each other's efforts in the vertical structure and plan their programs accordingly, but there is no joint planning, execution, or evaluation within the structure. In vertical coordination, there is some combination of different degrees of joint planning, execution, and evaluation within the vertical structure.

Vertical integration (VI) in an S&T program is a linkage among related programs in different phases of development. Research and development programs which have a common goal are run as a unit. There could be time differences and lags between the various programs, or they could be run with different degrees of concurrence. A research component of a vertically integrated program may be undergoing execution. Its development component may be in the early planning stage, with execution well into the future. Some of the higher category components may thus exist as planning wedges while the lower category components are being executed. The development process is not linear because of the inherent feed-forward and feed-back loops within and among categories. As Attachment 1 in Kostoff [1997n] shows, to achieve total VI, the program has to be planned and executed in a vertically integrated manner, and has to be assessed using the same taxonomy as was used for planning and execution. Because a vertically integrated program in one agency could draw upon programs managed by other agencies, the vertical linkages operate under the constraint that

each agency must have management autonomy to ensure that its overall objectives are met in the most expeditious manner.

Management Integration Metrics

With this background, the integration metrics can now be discussed. While horizontal, lateral, and vertical integration are all important for contributing to the efficient conversion of science to technology, the focus in this section is on indicators for vertical integration. In particular, for consistency with commonly used practice, the vertical integration metrics are assumed to apply to one organization only. The diverse vertical integration metrics examined can be arbitrarily divided into three generic types. The first type can be categorized as management integration metrics. This grouping contains the most primitive and least complex of the metrics. It includes measures which indicate how well different levels of development funds are mixed by managers at different levels in the organizational hierarchy. It is a limited metric, since it focuses on intraorganizational funds mixing only, and does not account for 'virtual' related programs from other organizations which contribute to the vertical structures and improve the effective mixing ratios. It could potentially indicate fragmentation where none exists, and therefore it should not be used as a stand-alone metric without substantial interpretation. In addition, insufficient understanding exists of the theoretically ultimate or desireable values for this type of metric (or for any of the collaboration metrics), and the operational value of these metrics becomes severely limited for application as management targets. These metrics become indicators in practice rather than controls.

As an example of this type of metric's application, suppose a sponsor organization manages basic research, applied research, early technology, and advanced technology programs. The funds mixing metric would indicate the combination of basic research/ applied research/ early technology/ advanced technology funds overseen/ managed at the program officer/ section manager/ division manager/ department manager/ office manager levels in the organization. This type of metric provides no indication of actual program integration or output product integration, but does provide an indication that the first step toward vertical integration is being seriously pursued. Appendix 13 describes how the thermodynamic concept of entropy, which is used sometimes as a measure of chemical mixing, can be extrapolated to indicate the mixing of funds.

Other management integration metrics can be defined, such as numbers of joint (multi-discipline, multi-development category, multi-organization, multi-sponsor) papers, patents, reports, projects, programs, conferences, meetings, and committees. Some of these metrics could begin to address horizontal and lateral integration as well. One has to be careful here. Joint ventures of any type require substantial amounts of effort in the coordination process, and overemphasis on this type of metric as an organizational target can lead to large inefficiencies and costs in time expenditures devoted to joint arrangements. At some point in the jointness process, diminishing returns become evident. The degree of jointness employed to manage or conduct any science and technology program needs to be carefully impedence matched to the intrinsic technical requirements of the program. Bureaucratic jointness requirements dictated independently of the particular needs of a given program are a recipe for inefficiency and failure.

Technical Integration Metrics

The second type of vertical integration metric can be defined as technical integration metrics. This category provides some indication of how well the basic research through advanced development programs have become aligned to each other and to the mission capability requirements in a technical sense. These metrics are typically more complex than those of the first category, since more than simple counting of elements is usually required. Again, there is the perceptual fragmentation danger when these alignment metrics are restricted to intraorganizational programs only. As before, there are no theoretical studies of desireable values, and the metrics serve as indicators rather than controls.

The simplest type of technical integration metric borders on being classified as a management integration metric. This metric indicates recognition by one of the development levels of work being performed of other development levels. For example, this could involve citations by i) early technology papers or patents or reports (in a given vertically integrated structure) of ii) papers or reports or patents emanating from basic research programs in the same vertically integrated structure.
A more complex technical integration metric involves measuring the conceptual alignment of the technical thrusts with semantic tools, such as computational linguistics approaches. Narratives describing the programs at different development levels in the vertical structure would be examined. Word or phrase similarities among the narratives would be quantified using a technique such as co-word analysis. The major limitations with this approach, objective though it may be, are that the language describing projects or programs at different levels of development changes substantially with development level. The language describing a basic research project in a vertical structure will probably be far different from the language describing an advanced technology project in the same structure. Thus, using one of the existing objective computational linguistic techniques will probably give artificially low indications of alignment among different levels in the structure.

A more valid, although more subjective, metric requires the use of subject experts to quantify the degree of relatedness among programs in the different levels of development in each vertical structure. While this approach can be relatively labor intensive, especially for vertical structures which contain large numbers of projects or programs, it probably is the most credible and provides enormous insight in generating the input data for the metrics. One method for quantifying this type of metric starts with generating a network of the kind shown in Appendix 9-A for the programs in a given vertical structure. For a network of program-level resolution, all the programs at the different levels of development in a vertical structure would be portrayed as nodes in a network. All nearest-neighbor nodes would be connected by links (While there is no intrinsic limitation to only connection of nearest-neighbor nodes, most of the obvious strong relationships will be among nearest-neighbors). Experts would then quantify the values of the links according to the strength of the relationships among the nodes connected by the links. A metric figure-of-merit for the network, such as the sum of the link value products along all the possible pathways in the network, would be computed (See Kostoff [1994i] describing use of a similarly-computed metric for a different application). It could be used to compare the relatedness of the programs in one vertical structure with the relatedness of programs in another vertical structure. Or, it could compare the relatedness of programs in one vertical structure against some maximum value of relatedness of programs in that structure, to provide a relatedness efficiency for the structure. Obviously, the method could be applied to vertical sub-structures, or to combinations of vertical structures, and other useful metrics could be obtained.

Product Integration Metric

The third type of vertical integration metric can be defined as product integration metric. Whereas the previous two classes of metrics addressed essentially the program management and program goals/ approaches, this class of metrics focuses on the science and technology product delivered by the sponsoring organization. It quantifies the intrinsic technical quality of the product transitioned to the next level of development (or to end use, depending on the charter of the organization being studied), the relevance (and the magnitude of the importance) to the organizational mission of the final product transitioned, the numbers of products transitioned, and the time elapsed in transitioning a product from one development level to the next. The same cautions to perceptual fragmentation resulting from concentration on intraorganization products only apply here as well.

Quality metrics, depending on the level of development being examined, could include patent citations or R&D Magazine 100 awards, or a myriad of other similar measures. While many of these quality metrics are the same as would be used to quantify quality of transitions in non-vertically integrated structures, the goal is to identify increases in quality due to the management and technical integration process. The transition metrics in this class require the ability to identify the different types of transitions that occur, and to place bounds on the different transition parameters such that they can be quantified accordingly. Again, because the equivalent of Carnot efficiencies for these types of metrics have not been identified, they are limited to serve as indicators rather than controls.

Relating Cause to Effect

This discussion on vertical integration metrics began with the desire to determine how well research has been vertically integrated with technology and mission capability requirements in a science and technology sponsor environment. Assume that the metrics proposed above have been employed to assess this degree of vertical integration, and assume further that changes have been observed relative to the non-vertically integrated mode of operation. How can the cause for the changes in the metrics' values be related to the effect of the change in organizational structure? This is not a simple question, especially in today's world, since many variables (e.g., geopolitical, funding, domestic political, etc.) could be changing in parallel with vertical integration changes.

For example, if transitions are improved after vertical integration has been instituted, they could be due to improved jointness at the sponsor and performer level. However, they could also be due to the research becoming more applied compared to its previous incarnation (an omnipresent possibility when vertical integration exists), and therefore intrinsically closer to the technology level to which it would transition. Each of these potential causes would have to be investigated in detail before definitive conclusions could be drawn.

On the other hand, suppose transitions are not improved, but are worse. One obvious potential cause is more inefficiencies due to the vertical integration. Another could be the changes in the geopolitical environment. Some technical areas may have blossomed, others may have decreased in importance. Good research in an area whose potential application significance has declined because of geopolitical considerations would be less likely to transition. If funding has decreased in the higher developmental categories, there will be fewer developmental programs to which research

could transition, and transitions will be reduced proportionately. The key conclusion here is that there can be many reasons for transitions to increase or decrease. Intrinsic program quality or program vertical integration are only a few of the many factors which determine transitions. Major determinants of transition success may have little to do with the underlying quality of the work, but more to do with environmental factors beyond the control of the organization's management. This is why even these types of vertical integration metrics are relatively limited as stand-alone measures of success, but need to be considered along with many other factors for a more thorough understanding of the science and technology evolution mechanisms.

IV-B-6-ii-g. Other Indicators

This section contains S&T metrics that do not fit precisely into the other more focused sections.

Pragmatic construction professionals, accustomed to intense price competition and focused on the bottom line, have difficulty justifying investments in advanced technology. Researchers and industry professionals need improved tools to analyze how technology affects the performance of the firm. This study [Hampson, 1997] reports the results of research to begin answering the question, "does technology matter?" The researchers developed a set of five dimensions for technology strategy, collected information regarding these dimensions along with four measures of competitive performance in five bridge construction firms, and analyzed the information to identify relationships between technology strategy and competitive performance. Three technology strategy dimensions-competitive positioning, depth of technology strategy and organizational fit-showed particularly strong correlations with the competitive performance indicators of absolute growth in contract awards and contract award value per technical employee. These findings indicate that technology does matter. The research also provides ways to analyze options for approaching technology and ways to relate technology to competitive performance for use by managers. It also provides a valuable set of research measures for technology strategy.

This cross-sectional study [Kahn, 1997] investigated predictors of research productivity and science-related career goals in a sample of 267 doctoral students (representing a response rate of 55%)from 15 randomly selected APA-accredited counseling psychology doctoral programs. A structural equation modeling procedure revealed that career goals and research productivity could be predicted by Holland personality type, perceptions of the research training environment, interest in research, and research self-efficacy. Student's gender and year in the doctoral program also contributed to this causal model as additional predictor variables, providing a very good fit to the data. The present findings contribute to theories of research training by presenting a comprehensive examination of the major factor previously investigated in the literature as predictors of research productivity and science-related career goals within the context or a structural equation model.

Although there are several methods for determining the quality of scientific research, there is no satisfactory method known that can measure the utilization of it. Earlier proposed methods measure a particular kind of utilization, but are - in practice - a poor indication for the utilization on the whole, a concept for which a definition is hard to make. These methods do not comply with the construct validity. The main problem in this case is the great diversity of what is meant by use of results of scientific research, resulting in a lack of consensus on the criteria for assessing the utilization. In the present study [Vancaulil, 1996], the authors propose and discuss two methods. To evaluate

utilization in a broad sense, the four-dimensional model describes the degree of utilization with three, mostly independent, aspects: the involvement of the user, the availability of a transferable research product, and the commercial benefits resulting from the research results. In the other method, the utilization of the research results is described first, and subsequently the utilization is quantified by a jury, who group the different projects in five classes, based on a Guttman scale.

Managing new product development (NPD) is, to a great extent, a process of separating the winners from the losers. At the project level, tough go/no-go decisions must be made throughout each development effort to ensure that resources are being allocated appropriately. At the company level, benchmarking is helpful for identifying the critical success factors that set the most successful firms apart from their competitors. This company- or macro-level analysis also has the potential for uncovering success factors that are not readily apparent through examination of specific projects.

To improve understanding of the company-level drivers of NPD success, Robert Cooper and Elko Kleinschmidt describe the results of a multi-firn benchmarking study [Cooper, 1995]. They propose that a compnny's overall new product performance depends on the following elements: the NPD process and the specific activities within this process; the organization of the NPD program; the firm's NPD strategy; the firm's culture and climate for innovation; and senior management commitment to NPD. Given the multidimensional nature of NPD performance, the studyinvolves 10 performance measures of a company's new product program: success rate, percent of sales, profitability relative to spending, technical success rating, sales impact, profit impact, success in meeting sales objectives, success in meeting profit objectives, profitability relative to competitors, and overall success.

The 10 performance metrics are reduced to two underlying dimensions: program profitability and program impact. These performance factors become the X- and Y-axes of a performance map, a visual summary of the relative performance of the 135 companies responding to the survey. The performance map further breaks down the respondents into four groups: solid performers, high-impact technical winners, low-impact performers, and dogs. Again, the objective of this analysis is to determine what separates the solid performers from the companies in the other groups. The analysis identifies nine constructs that drive performance. In rank order of their impact on performance, the main performance drivers that separate the solid performers from the dogs are: a high-quality new product process; a clear, well-communicated new product strategy for the company; adequate resources for new products; senior management commitment to new products; an entrepreneurial climate for product innovation; senior management accountability; strategic focus and synergy (i.e., new products close to the firm's existing markets and leveraging existing technologies); high-quality development teams; and cross-functional teams.

The final study in this section [Soderqvist, 1994] examines participation in scientific meetings. To handle the enormous amount of sources in modern and contemporary science, the historian can use different quantitative methods, particularly varieties of citation analysis. So far, all these methods have been based on publication data. Taking as its point of departure the fact that meetings constitute a pervasive, yet neglected, aspect of science, this study introduces analysis of participation in scientific meetings. The strength of this new prosopographical method is illustrated by an analysis of international immunological meetings in the period 1951-72. Frequency of participation in meetings seems to be correlated to professional standing in immunology. By means of cluster analysis of

participation data, the subdisciplinary structure and dynamics of immunology can be reconstructed.

IV-B-6-ii-h. Multiple Indicators

There is a growing consensus in the research evaluation community that single metrics provide too limited a perspective on research impact, and that an eclectic approach of suites of indicators used in concert is more appropriate for the evaluation of research. This section provides a small sampling of studies incorporating multiple indicators.

The first study reported in this section [Martin, 1996] argues that evaluations of basic research are best carried out using a range of indicators. After setting out the reasons why assessments of government-funded basic research are increasingly needed, the study author examines the multi-dimensional nature of basic research. This is followed by a conceptual analysis of what the different indicators of basic research actually measure. Having discussed the limitations of various indicators, the author describes the method of converging partial indicators used in several SPRU evaluations. Yet although most of those who now use science indicators would agree that a combination of indicators is desirable, analysis of a sample of Scientometrics articles suggests that in practice many continue to use just one or two indicators. The study also reports the results of a survey of academic researchers. They, too, are strongly in favour of research evaluations being based on multiple indicators combined with peer review. The study ends with a discussion as to why multiple indicators are not used more frequently.

In the next study, an approach for evaluation of research is described [Geisler, 1996] that integrates output indicators of four stages downstream the innovation process: immediate, intermediate, pre-ultimate and ultimate outputs. Indexes of leading output indicators are constructed. The indexes are integrated cumulatively to form an overall index of key output indicators, which is the integrated figure of merit (IFM). Data for the indicators are obtained from records and key informants, and the indicators are grouped by normalized weights. The study also discusses the limitations and the methodological, conceptual and political/organizational issues of such an approach to research evaluation.

The third study in this section [Hauser, 1997] relates choice of metric suites to program goals. Metrics affect research decisions, research efforts, and the researchers themselves. From a review of the literature, interviews at ten research-intensive organizations, and formal mathematical analyses, the authors conclude that the best metrics depend upon the goals of the R,D&E activity as they vary from applied projects to competency-building programs to basic research explorations. For applied projects, market outcome metrics (sales, customer satisfaction, margins profit) are relevant if they are adjusted via corporate subsidies to account for short-termism, risk aversion, scope, and options thinking. The magnitude of the subsidy should vary by project according to a well-defined formula.

For R,D&E programs that match or create core technological competence, outcome metrics must be moderated with "effort" metrics. Too large a weight on market outcomes leads to false rejection of promising programs. The large weight encourages the selection of lesser-value programs that provide short-trm, certain results concentrated in a few business units. This, in turn, leads a firm to use up its "research stock." Instead, to align R,D&E with the goals of the firm, the metric system should balance market outcome metrics wit h metrics system should balance market outcome metrics

with metrics that attempt to measure research effort more directly. Such metrics include many traditional indicators.

For long-term research explorations, the right metrics encourage a breadth of ideas. For example, many firms seek to identify their "best people" by rewarding them for successful completion of research exploration. However, metrics implied by this practice lead directly to "not-invented-here" attitudes and result in research empires that are larger than necessary but lead to fewer total ideas. Alternatively, by using metrics that encourage "research tourism," the firm can take advantage of the potential for research spillovers and be more profitable.

This study [Werner, 1997a] examines German and American philosophies and practices for R&D performance measure select ion. Comparative interviews wit h German and U.S. executives who used the R&D performance measures reported in a previous article (1) reveal differences in both right philosophy of measurement and the perception of its usefulness. Among U.S. managers, the most popular methods are patent counts, financial measures like rate-of-return, total quality, management, audits, and cost/time performance assessments. The emphasis is on measuring outputs per input (e.g., patents per dollar spent). Most U.S. managers were distrustful of simple metrics, preferring an integrated combination of quantitative and qualitative methods. In contrast, the German managers distrusted most R&D metrics, particularly output measures, although they commonly used input measures like annual expense per R&D employee. These differences are related to a fundamental difference in the philosophy of science between the U.S. and Germany. However, the survey results show that a measurement philosophy somewhere between the U.S. and German extremes may be appropriate for both countries, and that they are actually moving in that direction.

A related study [Werner, 1997b] reviews the state-of-the-art in measuring R&D perf ormance. Many R&D performance measurement techniques have been developed in response to the unique needs of various organizations. An extensive search of the literature from 1956 to 1995 identified over 90 articles, 12 books and two research reports describing various techniques. Integrated metrics that combine several types of quantitative and qualitative measures were found to be the most effective, but also the most complex and costly to develop and use. The choice of an appropriate R&D measurement metric depends on the user's needs for comprehensiveness of measurement, the type of R&D being measured the available data, and the amount of effort the user can afford to all ocate to it. Guidelines are provided for selecting an appropriate measurement method within these parameters.

The following study [Brown, 1997] describes the results of an evaluation of the Energy-Related Inventions Program (ERIP), one of the longest- running commercialization assistance programs in the USA. The program has been subjected to a series of evaluations since 1984. The performance metrics produced over this decade of data collection, when compared with metrics from other technology innovation efforts, suggest that the Energy-Related Inventions Program has been highly successful. The process of generating these metrics has underscored some of the difficult issues that must be addressed to fairly appraise public investments in technology commercialization. These include: (1) the need to track the progress of program participants for extended periods; (2) complexities associated with accounting for spin-off technologies; (3) determining the external and internal validity of program evaluations; and (4) dealing with performance data that are dominated by a small number of highly successful technologies.

In the next study in this section [Sylvain, 1993], analysis of the Canadian publications in the field of aquaculture reveals that Canada is one of the world's major contributors in this area. This confirms that Canada's expertise in science and technology often finds its stimulus in its resource-based industries. Several bibliometric indicators were used to enlighten the peculiar features of the Canadian research system. These include the channels of communication used by scientists, the authorship pattern, the level of collaboration, the identification of the institutions in which the research is performed and the uneven research effort distribution inside the country. The relevance of such quantitative measures for science policy-making is emphasized. The present study shows how bibliometric analysis, by describing the actual strengths and weaknesses of Canadian research and identifying the agents of this research activity, might foster a better understanding of the Canadian research enterprise as a whole.

The next study examines the utility and limitations of formal evaluation methods [Lepair, 1995]. After some comments on evaluation as an integral part of science, the emphasis in this study is on evaluation for policy purposes. Early attempts to validate the use of bibliometric indicators are outlined. Three lessons emerge: 1. Best results with a variety of methods 2. Reliable results if publication is the major means of communication 3. Useless in technology (applicable science) Next, the measurement of a Citation Gap in applicable science is described. Examples are given of the use of bibliometrics in actual policy decisions about the selection of advisors, personnel and budgets. Bibliometrics for policy purposes should never be used on its own. In a final chapter a description is given of the evaluation method to select research projects for financial support, as applied by STW, the technology branch of the Netherlands' research council, NWO.

This study [Hodges, 1996] examines the use of an algorithmic approach for the assessment of research quality. Recent years have seen a growing interest in the use of quantitative parameters for assessing the quality of research carried out at universities. In the UK, university departments are now subject to regular investigations of their research standing. As part of these investigations, a considerable amount of quantitative (as well as qualitative) information is collected from each department. This is made available to the panels appointed to assess research quality in each subject area. One question that has been raised is whether the data can be combined in some way to provide an index which can help guide the panels' deliberations. This question is looked at in this study via a detailed examination of the returns from four universities for the most recent (1992) research assessment exercise. The results suggest that attempts to derive an algorithm are only likely to be helpful for a limited range of subjects.

Another study [Johnes, 1996] focuses on performance assessment in higher-education in Britain. All public sector organisations in the UK have witnessed changes in funding arrangements during the 1980s as part of the Government's drive to make them more accountable to the tax-payer. The development of performance indicators is seen as an essential step to ensure that such organisations provide value for money. This study examines the possibility of constructing measures of the performance of UK universities. A methodology is developed in the framework of production theory and uses multiple regression techniques to estimate the relationship between the outputs and inputs of universities. Around 80% of the inter-university variation in four output measures can be explained by corresponding variations in several input measures. This highlights the need to take into account the inputs available to a university when comparing its output performance with that achieved by other institutions. The problems of interpreting an array of performance indicators are

also clearly demonstrated.

This study [Yang, 1997] examines the performance indicators for science and technology projects in Taiwan. To help the Taiwanese private sector to compete globally, the Ministry of Economic Affairs (MOEA) in Taiwan initiated a programme called the 'Science and Technology Project (STP)' in 1982. Through this programme, the government offers over 10 billion NT dollars per year to support technological research and development. Furthermore, the STP is executed by statutory bodies (non-profit research institutes) funded by the MOEA.

For the purposes of budget allocation and control, an annual performance evaluation of STP is needed, though it is a difficult task. Although the MOEA has established a system of performance evaluation and has practised it for years, there is no consensus on the fairness of this system among research institutes and other interested parties competing for funds. A more elegant evaluation system is needed. The purpose of this research is to establish a reliable system of performance indicators for the STP. The study reviewed the whole performance indicators system of R&D projects and proposed a feasible revision. The system of performance indicators can be further divided into three subsystems: (1) indicators for research results, (2) indicators for industrial co-operation, (3) indicators for technology diffusion.

The next study in this section addressed faculty usage of higher education journals [Koong, 1989]. A taxonomy and framework for evaluating the quality of journals in higher education are proposed in this study. The significance of acquiring and disseminating professional information to faculty and administrators in higher education is discussed, and it is noted that the journals in which a faculty member publishes are sometimes used as critical factors in promotion and tenure decisions. Following a review of the literature about hierarchies in higher education publishing, a model is presented which offers five constructs that affect journal quality: (1) perception, which gauges the opinions of selected peers about a journal's quality; (2) citations, which measure the number of times a work is cited in subsequent research in the area; (3) usage (publishing), a measure that shows the number of times fellow educators publish in that journal; (4) usage (readership), identifying how often the source is referred to by peers; and (5) factual information, which can be obtained from reference publications about journals. A mathematical model encompassing flexibility for faculty and academic departments with diverse needs is also introduced to help evaluate journals using the proposed constructs. The combination of the constructs and method are based on the fact that the strength of one can compensate for the limitations of the other. A figure illustrates the concept.

The final study in this section [Spann, 1995] surveys measures of technology transfer effectiveness. Federally funded R&D has been viewed as a key source of advanced technologies that, if successfully transferred to the private sector, could help rebuild America's global competitiveness. The growing perception that the nation is not getting an adequate return from its federal R&D budget is accompanied by a growing demand for more measurable technology transfer results. Yet measures of technology transfer effectiveness are neither well defined nor universally accepted. This exploratory study focused on defining and describing the measures or metrics used in the process of transferring government-funded technologies to private sector firms. The paper presents an initial conceptual framework and an exploratory, empirically based taxonomy of metrics used in technology transfer. This taxonomy and specific measures were used to help determine which technology transfer metrics were used by various players across the federal technology transfer

process. Individuals who played roles as either sponsors, developers or adopters of federally funded technologies were surveyed on their roles and the measures of transfer effectiveness used in their work units. The data showed statistically significant differences in frequency of use of the transfer measures by the three roles. Secondly, a broad set of measures were used in varying degrees by all roles. Most importantly, all three roles used most measures rather infrequently. Recommendations to guide future research are included. Recommendations are also made for technology transfer practitioners.

IV-B-6-ii-i. Indicators Integrated with other Techniques

The first study in this section [Johnston, 1995] examines the broad implications of research impact quantification. The development of methods for the quantification of research impact has taken a variety of forms: the impact of research outputs on other research, through various forms of citation analysis; the impact of research and technology, through patent-derived data; the economic impact of research projects and programs, through a variety of cost-benefit analyses; the impact of research on company performance, where there is no relationship with profit, but a strong positive correlation with sales growth has been established; and calculations of the rates of social return on the investment in research.

However, each of these approaches, which have had varying degrees of success, are being challenged by substantial revision in the understanding of the ways in which research interacts, and contributes to, other human activities. First, advances in the sociology of scientific knowledge have revealed the complex negotiation processes involved in the establishment of research outcomes and their meanings. In this process, citation is little more than a peripheral formalisation. Second, the demonstration of the limitations of neo-classical economics in explaining the role of knowledge in the generation of wealth, and the importance of learning processes, and interaction, in innovation within organisations, has finally overturned the linear model on which so many research impact assessments have been based. A wider examination of the political economy of research evaluation itself reveals the growth of a strong movement towards managerialism, with the application of a variety of mechanisms - foresight, priority setting, research evaluation, research planning - to improve the efficiency of this component of economic activity. However, there are grounds for questioning whether the resulting improved efficiencies have, indeed, improved overall performances. A variety of mechanisms are currently being experimented with in a number of countries which provide both the desired accountability and direction for research, but which rely less on the precision of measures and more on promoting a research environment that is conducive to interaction, invention, and connection.

The next study [Vanraan, 1996] gives an overview of the potentials and limitations of bibliometric methods for the assessment of strengths and weaknesses in research performance, and for monitoring scientific developments. The study author distinguishes two different methods. In the first application, research performance assessment, the bibliometric method is based on advanced analysis of publication and citation data. The author shows that the resulting indicators are very useful, and in fact an indispensable element next to peer review in research evaluation procedures. Indicators based on advanced bibliometric methods offer much more than 'only numbers'. They provide insight into the position of actors at the research front in terms of influence and specializations, as well as into patterns of scientific communication and processes of knowledge

dissemination. After a discussion of technical and methodological problems, the author presents practical examples of the use of research performance indicators. In the second application, monitoring scientific developments, bibliometric methods based on advanced mapping techniques are essential. The author discusses these techniques briefly and indicate their most important potentials, particularly their role in foresight exercises. Finally, he gives a first outline of how both bibliometric approaches can be combined to a broader and powerful methodology to observe scientific advancement and the role of actors.

The final study in this section [Nagpaul, 1995] argues that research performance is essentially a multidimensional concept which cannot be encapsulated into a single universal criterion. Various indicators used in quantitative studies on research performance at micro or meso-levels can be classified into two broad categories: (i) objective or quantitative indicators (e.g. counts of publications, patents, algorithms or other artifacts of research output) and (ii) subjective or qualitative indicators which represent evaluative judgement of peers, usually measured on Likert or semantic different ial scales. Because of their weak measurement properti es, subjective indicators can also be designated as quasi-quantitative measures. This study is concerned with the factorial structure and construct validity of quasi-quantitative measures of research performance used in a large-scale empirical study carried out in India. In this study, a reflective measurement model incorporating four latent variables (R and D effectiveness, Recognition, User-oriented effectiveness and Administrative effectiveness) is assumed. The latent variables are operationalized through thirteen indicators measured on 5-point semantic differential scales. Convergent validity, discriminant validity and reliability of the measurement model are tested through LISREL procedure.

IV-C. COST-BENEFIT/ ECONOMIC ANALYSES

IV-C-1. Background

A comprehensive survey examined the application of economic measures to the return on research and development as an investment in individual industries and at the national level [OTA, 1986]. This document concluded that while econometric methods have been useful for tracking private R&D investment within industries, the methods failed to produce consistent and useful results when applied to Federal R&D support.

An intermediate study published by the Commission of the European Communities [Capron, 1992] concluded that "the economic quantitative methods, particularly econometric models, should be viewed as an ex post quantitative evaluation tool of the economic impacts of science and technology policy. They have their shortcomings and limits. They are an instrument in the toolbox of policy evaluation which can be used for structured quantitative analyses of the economic impact of R&D policy.....The economic impact of government financed R&D might be evaluated by using simultaneously existing pinpoint methods and extended macroeconometric models. While existing pinpoint methods are numerous, the most commonly used ones are the productivity and the investment approaches. Extended macroeconometric models might be conceived by adapting present macromodels or developing adequate models."

A later analysis focused on economic/ cost-benefit approaches used for research evaluati on [Averch,

1994]. The methods involve computing impacts using market information, monetizing the impacts, then comparing the value of the impacts with the cost of research. Principal measures described include surplus measures and productivity measures. With known benefit and cost time streams, internal rates of return to R&D investments are then computed. The paper notes both the standard technical difficulties with these approaches and the political and organizational difficulties in implementing them.

IV-C-2. Classical Microlevel Application

Cost-benefit analysis has limited accuracy when applied to basic research because **of the quality of both the cost and benefit data due to the large uncertainties characteristic of the research process, as well as selection of a credible origin of time for the discounting computations**. As an illustrative example, a cost-benefit analysis performed on a fusion reactor variant (the fusion-fission hybrid, essentially a fission reactor driven by fusion neutrons which can produce both fissile fuel and power) will be described in some detail.

Rutherford's experiments in 1934 involving interaction of a deuteron beam with solid deuterium can be viewed as the genesis of fusion fuel cycle research [Kostoff, 1983a]. Almost since the formation of the AEC in the mid-1940s, the Federal government has invested significant sums of money for the potential promise of controlled fusion as an essentially limitless source of energy. In 1979, an economic analysis based on capital costs was performed on the fusion hybrid and a comparison was made with two major contenders for the same type of product, fast breeders and accelerator breeders [Kostoff, 1979]. The results showed projected cost savings (for different parameter variations) for developed fusion hybrid systems but did not address the time distribution or magnitude of development costs. Subsequent technical studies showed ranges of favorable operating conditions based on fusion reactor cycling times [Kostoff, 1981, 1982, 1983b, 1985].

To evaluate the economic potential of the fusion-fission hybrid, an incremental cost-benefit analysis was performed [Kostoff, 1983a]. While fusion-related expenditures could be traced back to Rutherford's experiments in 1934, this study ignored fusion hybrid research expenditures before 1980 (sunk costs from the perspective of 1980). For the parameter ranges chosen, it was shown that there was a broad region over which hybrid development could prove cost-effective. However, **had this same analysis been done in 1934 (around the beginning of identifiable basic research for fusion), using the same cost and benefit streams as in the 1983 study plus adding costs incurred between 1934 and 1980 and discounting back to 1934, then the result would have been much different from the 1983 study.**

In the 1983 study, the problem was treated deterministically; uncertainties or probabilities of success of the different parameter values being achieved were not taken into account. **The real problem, which pervades and limits any attempt to perform a cost-benefit analysis on a concept in the basic research stage, was the inherent uncertainty of controlling the fusion process. This translated to the inability to predict the probabilities of success and time and cost schedules for overcoming fundamental plasma research problems (e.g., plasma stabilities and confinement times); no credible methods were available**. Thus, the main value of the cost-benefit approach was to show that the potential existed for positive payoff from the hybrid reactor development, that there was a credible region in parameter space in which controlled fusion development could prove

cost effective; **what was missing was the likelihood of achieving that payoff**.

IV-C-3. Macrolevel Analyses

Much of the major economic work relating economic growth/ productivity increases to R&D spending has been performed by three economists [Mansfield, 1980, 1991; Terleckyj, 1977, 1985; Griliches, 1979]. Probably the mo st widely publicized work over the past decade to examine rates of return from basic research has been that of Mansfield [e.g., Mansfield, 1980, 1991]. His results indicated that substantial social rates of return can be attributed to basic research. While use of his methods by government officials has not been reported in the literature, the methods have received widespread attention among research policy-makers. Because of the potential impact of these methods if adopted, both his production function and recent marginal cost-benefit approaches will be discussed.

IV-C-4. Production Function Approach

The earlier study [Mansfield, 1980] attempted to determine whether an industry's or firm's rate of productivity change was related to the amount of basic research it performed. Mansfield developed a production function which disaggregated basic and applied research, then regressed rate of productivity increase with many different variables. The regressions showed a strong relationship between the amount of basic research carried out by an industry and the industry's rate of productivity increase during 1948-1966.

However, many assumptions were necessary to solve the equations: constancy of ratios of variables over time; neglect in the actual regression equations solved of the (long) lag time between when the research is performed and when the productivity change is measured (though this point is recognized and discussed by Mansfield); and the inherent uncertainties in the data used in the equations. The results have to be treated as highly uncertain. In fact, Mansfield's results are somewhat inconsistent with the findings of the second part of his study, which showed, for 119 major firms surveyed, that the proportion of R&D expenditures devoted to basic research and to relatively risky projects declined between 1967 and 1977 in most industries. **Would firms reduce their own basic research expenditures if they felt that their own basic research expenditures would result in increased productivity?**

Finally, there is the problem inherent in multiple regression analyses: that of determining cause and effect from what is essentially correlation. As Mansfield points out, "It is possible that industries and firms with high rates of productivity growth tend to spend relatively large amounts on basic research, but that their high rates of productivity growth are not due to these expenditures" [Mansfield, 1980]. Nor does Mansfield's model specify the path(s) by which R&D investment supposedly leads to productivity improvements.

IV-C-5. Macrolevel Marginal Cost-Benefit Application

A 1991 study weighed the costs of academic research against the benefits realized from the earlier introduction of innovative products and processes due to the academic research [Mansfield, 1991]. A survey of corporate R&D executives showed that an average of seven years elapsed between a

research finding and commercialization, and that commercialization would have been delayed an average of eight years without academic research. A cost-benefit analysis using this survey data showed a very high social rate of return resulting from academic research.

However, the data were not validated independently by a document-based type of analysis (such as TRACES or Hindsight, retrospective studies of innovations) of a sample number of the products and processes. The time between the research findings and commercialization is very short compared to the results of Hindsight or the TRACES studies, and is more in line with the lag time between the **end** of basic research and commercialization shown by Hindsight/TRACES. Use of a shorter lag time in the discounting process increases the benefit/cost ratio and the social rate of return. While the method is innovative, a more objective data source would provide higher confidence in the computed rates of return.

IV-C-6. Specific Cost-Benefit Studies with Different Approaches

The initial studies in this section address conceptual issues and problems associated with the application of cost-benefit approaches to science and technology evaluation. The later studies focus more on specific applications of cost-benefit analysis to determining S&T impact.

Macroeconomic Aspects

The first paper in this section [Kyriakou, 1995] examines the broader macroeconomic aspects of S/T program evaluation. Understanding the macroeconomic aspects of S/T programme evaluation exercises must be anchored in exploring S/T and its impact in the context of the modern competitive economy, starting at the level of the firm and moving up to the country and EU regional level. Whereas monitoring focuses on the continuous managerial review of project operations, evaluation is concerned with what is being achieved, with maximizing the programme's impact, and with providing guidelines for new ones. The economic context and the placement of S/T in it, in crucial in both ex-ante evaluation, setting goals and projecting evolution corridors, as well as ex-post evaluation of proximity to targets, and/or assessment/updating of projected technological and economic paths followed.

The study briefly draws this connection and then proceeds to explore the multi-level interface between S/T and the economic context, whose characteristics should inform ex-ante and ex-post evaluation efforts. Particular emphasis is placed on the role of S/T - and hence in evaluating S/T programmes - visa-vis the effects of S/T on market structure, sustainability and European Union (EU) cohesion. S/T is viewed in terms of its projected effects on the viability of monopolistic/oligopolistic arrangements, and on the incontestability of markets, namely the ability of incumbents to deter entry by new challengers. It is also argued that S/T is, and should be, the bridge linking growth and sustainability, the two towering preoccupations that are often deemed to be at odds. Finally, and most immediately critical for the EU, the vicissitudes of cohesion in the EU are explored, and the role of S/T in alleviating them is underscored. Successful and properly evaluated S/T programmes can help steer the EU away from the tensions generated by asymmetric shocks to liberalizing, integrating economies, specializing on the basis of comparative advantage.

The second study in this section [Martin, 1997] examines the role of producer surplus in evaluating

R&D investments. Comparison of producer surplus with definitive measures based on the profit function reveals potential problems with using changes in producer surplus to measure the benefits of some common types of technical change. Some illustrative applications indicate that the conventional producer surplus measures may seriously under-estimate the change in profit induced by new technology, depending on the characteristics of the underlying technology which define the nature of the supply function, and the nature of the technical change. The study authors provide guidelines for identifying cases where producer surplus will under-estimate producer research benefits, and suggest alternative measures.

The next study [BREMEN, 1992] focuses on assessing energy projects from the viewpoint of individual economic branches and total economy. It addresses the role of economic efficiency analysis, cost-benefit analysis and mult icriteria methods. Energy is an extremely important good and means of production not only for the individual branches of economy but, due to its essential meaning to the development of a region or a national economy and its external effects connected with production and consumption, also of great interest to all economic branches. This article deals with the relation of analyses in individual economical branches and those in total economy and with the question of what the importance of cost-benefit analyses and other methods is in the analysis in total economy. The author also mentions the planning as in the special literature the planning and evaluation phases are not analytically separated which is seen especially in the discussion about the multi-criteria methods.

The final macroeconomic study presented [PRICE, 1995] contains an assessment of the costs and benefits of regulatory decision making. This study outlines the framework within which cost-benefit analyses of regulation may be undertaken. The general framework is consistent for any cost-benefit analysis. The particular needs or individual structure of the industry to which the regulation is targeted and the particular nature of the regulation will affect the methodologies chosen to execute specific steps within that framework. The discussion also includes insight into the approach to cost-benefit analysis used in other jurisdictions, specifically the U.S. Nuclear Regulatory Commission, the Health and Safety Executive, Nuclear Safety Division in the United Kingdom, Transport Canada and Environment Canada. Vario us methodologies, and their relative strengths and weaknesses in the context of regulation in the nuclear industry, are outlined in the discussions of each phase of the cost-benefit framework. Those individual methodologies and approaches in other jurisdictions that are best suited to the assessment of regulat ions administered by the Atomic Energy Control Board are incorporated into a proposed framework.

Intergenerational Equity

The first study in this group [Lind, 1995] examines intergenerational equity, discounting, and the role of cost-benefit analysis in evaluating global climate policy. When public policies with impacts far into the future are being debated, the question inevitably is raised whether cost-benefit analysis which discounts future costs and benefits is not biased against future generations and whether, if such discounting is appropriate at all, a lower rate should be used to avoid such bias, The debate on global climate change is no exception. This study sketches and analyses the welfare foundations of cost-benefit analysis and from this perspective analyses the role of cost-benefit analysis in the climate policy debate, particularly with reference to intergenerational effects. The study concludes that the cost-benefit criterion cannot provide a definitive basis for deciding whether society should

commit to a longer-term programme to moderate climate change; the issues of intergenerational equity are not that global climate change will significantly lower the GNP of future generations, but relate to the possibility of science fiction-like changes in the planet that will produce catastrophic effects in the future; and the typical way in which the cost-benefit problem is posed obscures the basic choices that we should be evaluating.

The next study [Spash, 1994] also examines economic implications of potential climate modifications. Economic decisions over what action, if any, to take concerning the greenhouse effect tend to revolve around the social discount rate. Implicitly the debate concerns how to attribute intertemporal weights to welfare and implies a moral stance that is rarely given explicit recognition. Refocusing on the outcomes of current actions emphasises the role of "compensation". A conflict is apparent between the view that the current generation need be unconcerned over the loss or injury caused to future generations because they will benefit from advances in technology, investments in both man-made and natural capital, and direct bequests; and the requirement to avoid harming the innocent. Changes in units of welfare cannot be viewed as equivalent regardless of their direction. In general, doing harm is not cancelled out by doing good. The result is a rejection of the potential compensation principle which underlies the current economic stance, and a reconsideration of the acceptability of "compensation" altogether. The concept of human rights and a non-utilitarian perspective are used to show how cost-benefit analysis denies the existence of inalienable rights, and economics limits the moral considerability of harm.

Another study in this group on climate effects [Hasselmann, 1996] examines optimization of CO(sub 2) emissions using coupled integral climate response and simplified cost models. A cost-benefit analysis for greenhouse warming based on a structurally simplified globally integrated coupled climate-economic costs model SIAM (Structural Integrated Assesment Model) is used to compute optimal paths of global CO(sub 2) emissions which minimize the net sum of climate damage and mitigation costs. The climate model is represented by a linearized inpulse-response model calibrated against a coupled ocean-atmosphere general circulation climate model and a three-dimensional global carbon-cycle model. The cost terms are represented by strongly simplified expressions designed for the study of the sensitivity of the computed optimal emission paths with respect to critical input assumptions. These include the discount rates assumed for mitigation and damage costs, the inertia of the socio-economic system, and the dependence of climate damages on the change in temperature and the rate of change of temperature. Different assumptions regarding these parameters are believed to be the origin of the marked divergences of existing cost-benefit analyses based on more sophisticated economic models. The long memory of the climate system implies that very long time horizons of several hundred years are needed to optimize CO(sub 2) emissions on time scales relevant for a policy of sustainable development. Cost-benefit analyses over shorter time scales of a century or two can lead to dangerous underestimates of the long term climatic impact of increasing greenhouse-gas emissions.

This final study in this climate-focused group [Backlund, 1995], an economic analysis of forest carbon sequestration, examines global warming and dynamic cost-benefit analysis under uncertainty. This paper provides an economic analysis that integrates dynamic and stochastic features into the global warming problem. The aim is to provide a framework for analyzing alternative policy measures. We show in what sense a free-market solution is different from the first best command optimum, and we discuss an appropriate policy instrument to implement the first best solution. We

also introduce a numerical model, and simulate the optimal path for consumption, GHG emissions, etc under different assumptions. It turns out that an endogenous discount rate, minimizing the probability of a doomsday scenario, leads to a more even consumption path, than the corresponding path under a lower and constant discount rate.

Quantification of Distributive Justice

Another study on environmental and risk-related public policy [Ellis, 1993] examines the quantification of distributive justice. The most fundamental philosophical objection to cost-benefit analysis is that it fails to account for the distinction between more-necessary and less-necessary benefits. For example, it provides no way to avoid trading off a few cancer deaths in exchange for a more cost-effective but also more hazardous technology which provides cheaper paper or plastic products for the many. Since unjust distribution of benefits and burdens results primarily from the failure to prefer more-necessary goods (such as health and safety) over less-necessary ones (such as cheaper plastic razors), the authors then show that a correct calculation of the rate at which marginal utilities diminish in value (as they become less necessary to their users) can determine 'degrees of necessity' and thus the most just possible distribution of benefits and burdens. One way to measure the rate of diminishing marginal utility is provided by the 'wealth effect' in occupational risk studies. Wealthier workers will not assume the same risk in exchange for a given salary increment (which to them is not very necessary) as poorer workers would assume for that same salary increment (which to them is more necessary). It is therefore possible to construct a mathematical model for the effect of necessity/non-necessity on quantitative decision principles for environmental and risk-related public policy, thus making such decisions more distributively just than traditional cost-benefit analysis would allow.

A related study [Ganiats, 1997] examines the issue of determining the value of future health. Cost-effectiveness is an integral part of health care policy, influencing both medical and administrative decisions. However, current research methodology for evaluating cost-effectiveness produces several paradoxes, perhaps because it incorrectly represents the general population's view of future health states. Recent work introduces clinical and, demographic factors to the traditional cost-benefit model for discounting health outcomes. It suggests a revised model that provides a more accurate basis for health policy decision-making. This revised model will likely improve the apparent cost-effectiveness of prevention programs, which are at a distinct disadvantage in present models. This study presents examples of current paradoxes resulting from the standard discounting methodology, findings on the variability of health outcomes discount rates in patients, and preliminary thoughts on developing a revised model for discounting future health outcomes. This revised model should present the value of health promotion programs more accurately.

Use of Uncertain Data

The next two studies address a central problem in the prospective application of cost-benefit analysis to S&T: namely, decision-making using very uncertain data. The first study [Dompere, 1997] presents a theory of efficient prices for cost-benefit analysis in a fuzzy space. The approach proceeds by taking consumers' income, and producers' outputs and costs as given. The price preferences of consumers and producers are elicited and then embedded in a fuzzy space through fuzzy mappings to obtain a fuzzy compact price space where fuzzy price decisions are constructed. Solutions to the

fuzzy price decision problems are abstracted through fuzzy mathematical programming to obtain fuzzy equilibrium prices. From the fuzzy price space measures of price disagreement, fuzzy consumer surplus and fuzzy producer surplus are advanced. Theorems of existence and uniqueness are stated. The total result is a theory of fuzzy prices for cost-benefit analysis for decision problem, in general including cases where market imputations of prices may not be available to us as well as those cases where market failure may yield price distortions. The theory is not only compatible with either contigent variation method (direct information elicitation) and revealed preference method (market-based evaluation) but provides a direction for cases where problems may exist in both. A computational example is provided to illustrate the working mechanism of the theory.

The second of these studies [Hogarth, 1995] concerns decision- making under ignorance. The metaphor of gambling has had great influence on the topic of choice under uncertainty. However, in many real-world situations, people must make choices when they lack information about the relevant economic features of gambles, i.e., probabilities and outcomes. The authors refer to this as choice under ignorance as opposed to choice under risk or uncertainty. They propose that people handle these decisions by generating rationales or arguments that allow them to resolve the choice conflict. Moreover, these rationales often do not correspond to principles derived from the cost-benefit framework of economic models. These ideas are explored in two experiments in which subjects simulated the purchase of warranties for consumer durables. The principal findings of this study are, first, that observable behaviors differ between situations where subjects do and do not have information on probabilities and outcomes. Second, economic cost-benefit models did not yield good descriptions of the experimental subjects' decisions. Third, the nature of arguments used, and thus the processes invoked, differed as a function of the information available to subjects. And fourth, subjects' arguments indicated two types of strategies for reaching decisions. In one, they processed the particular characteristics of each choice option; in the other, they invoked a "meta-rule" or principle that resolved the choice conflict and was insensitive to the particular features of different options. Finally, the authors discuss the implications of their results. This includes questioning the appropriateness of using the gamble as a metaphor for choice in future research.

Economies of Scale

The first of two studies examining economy of scale effects [Henderson, 1996] focuses on the determinants of research productivity in drug discovery. The authors examine the relationship between firm size and research productivity in the pharmaceutical industry. Using detailed internal firm data, the authors find that larger research efforts are more productive, not only because they enjoy economies of scale, but also because they realize economies of scope by sustaining diverse portfolios of research projects that capture internal and external knowledge spillovers, In pharmaceuticals, economies of scope in research are important in shaping the boundaries of the firm, and it may be worth tolerating the static efficiency loss attributable to the market power of large firms in exchange for their superior innovative performance.

The second study, also of pharmaceutical research-and-development [Omta, 1994], compares management control and innovative effectiveness in European and Anglo-American companies. Drug regulation and pricing have put strong pressure on the cost-benefit ratio of the innovative pharmaceutical industry. Therefore, a study has been conducted in fourteen large and medium sized companies to determine some important organisational and managerial factors influencing success in

pharmaceutical innovation. The study consists of structured interviews with Research Directors and questionnaires submitted to the heads of the different research departments. The following conclusions are tentatively drawn. Firstly, the data suggest that a threshold investment of approximately $150-200 million is needed to maintain the innovative potential. Above approximately $750 million, 'economies of scale' seem to appear in pharmaceutical innovation. Secondly, an incremental strategy aimed at reducing the duration of the development process seems to be more successful than a radical strategy which lays more emphasis on discovery. Thirdly, pure play pharmaceuticals seem to be more successful than the pharmaceutical divisions of conglomerates. Management control, especially the way in which reorganisations are performed, is assessed more positively in pure play pharmaceuticals. Fourthly, the greater emphasis on human resources management in Anglo-American companies, in comparison to continental European companies, seems to be an important explanatory factor for their greater success on the pharmaceutical market.

A health industry-related study [Jonsson, 1994] focuses on economic evaluation of new medical technology. Safety and efficacy are not the only parameters of interest for choice of medical technology - costs play an increasingly important role. There is a growing interest in 'value for money', which can be assessed by economic evaluation comparing the costs and consequences of alternative courses of action. A number of different economic evaluation methods may be used: cost-minimization (looking only at costs with no consideration of consequences); cost-effectiveness (in which a unidimensional clinical outcome is assessed, for example, life-years gained); cost-utility (measuring multidimensional outcomes, for example quantity and quality of life); and cost-benefit (where outcome is considered in monetary terms). A Swedish cost-of-illness study showed that the direct health care costs increased and the indirect cost (in terms of production loss) associated with treatment of peptic ulcer fell following the introduction of H-2-receptor antagonists. In a study of reflux oesophagitis, omeprazole was shown to be more cost-effective than ranitidine. With omeprazole, the costs were lower and the effectiveness better than with the H-2-receptor antagonist.

Applications

The final group of studies focuses more on the applications of cost-benefit analysis to the measurement of science and technology impacts. The first study in this large applications group [Williams, 1984] contains a methodology for economic evaluation of process technologies in the early research and development stages. A systematic methodology has been developed by the author for building, combining, and exercising a set of specially devised performance, design, and cost models in a form suitable for process economic assessments in the presence of major technological uncertainties. This document describes the development and utilization of the new methodology. Via simulation, a cohesive spectrum or distribution of the resulting performance and cost figure-of-merit values, along with their associated probabilities, is calculated. The appropriate format for development of the user's modeling system, which includes the capability to re optimize the proposed process for each set of process inputs considered is presented, along with the required stepwise approach for selection of values and ranges of the major uncertain process variables or inputs. The basic principles of this combined methodology can be applied to many new processes or technologies - particularly those in their early R and D stages.

Interpretation of the probabilistic output data is also discussed. Such data can be useful to the

experimentalists as well as to those decision makers who must recommend or decide whether a particular process should be further developed, or which of several competing technologies should be selected for continued support. Recent experiences with this methodology in the assessment of advanced uranium isotope separation processes and in assessment of a photochemical syngas cleanup system allow two major conclusions to be drawn; that disappointments in process-performance related areas rather than hardware cost issues tend to have the most deleterious effects on unit cost, and that the process proponent's earliest single-point best guess unit cost estimates are usually found to fall in the most optimistic fringes of the computed uncertainty ranges.

A follow-on related study [Williams, 1986] develops a methodology for economic evaluation of technologies in the early research and development stages. A systematic methodology has been developed for building, combining, and exercising a set of specially devised performance, design, and cost models in a form suitable for economic assessments in the presence of major technological uncertainties. This document describes the development and utilization of the methodology that incorporates model development and multivariable uncertainty analysis for the projection of potentially competitive, full-scale performance and costs of a first-of-a-kind process or systems technology still in the early research and development stages. By Monte Carlo simulation, a spectrum or distribution of the resulting performance or life-cycle cost figure-of-merit value, along with its associated probabilities, is calculated. The appropriate format for development of the user's modeling system, which includes the capability to reoptimize the proposed systems for each set of process inputs considered, along with the required stepwise approach for selection of values and ranges of the major system variables (inputs), is presented. The basic principles of this methodology can be applied to many new technologies - including those relevant to the Strategic Defense Initiative (SDI). Interpretation of the probabilistic output data is also discussed. Such data can be useful to the experimentalists, as well as to those decision makers who must recommend or decide (1) whether a particular process should be further developed or (2) which of several competing technologies should be selected for continued support. Recent experiences with this methodology in the assessment of advanced energy technologies for the US Department of Energy are discussed. Potential applications to the SDI are also suggested.

Another applications study [Chapman, 1996] examines benefits and costs of research, using two case studies in building technology. The report is the outgrowth of a series of microstudies prepared by NIST's Building and Fire Research Laboratory (BFRL). This report has four major purposes. First, it examines five standardized methods for evaluating existing and past research projects. Second, it establishes a framework for identifying, classifying, quantifying, and analyzing the benefits and costs of a research project, of a research program, or of a new technology. Third, it presents a generic format and a set of guidelines for summarizing the economic impacts of alternative research investments. Fourth, it illustrates--by way of two case studies--how the framework and standardized methods would be applied in practice.

The next applications study [NASA, 1985] focuses on research- concept evaluation; concepts are ranked according to their potential benefit/cost ratios. The citation summarizes a one-page announcement of technology available for utilization. The ARINC Research Concept Evaluation Methodology (ARCEM) program was developed to assist in the rank ordering of research concepts in terms of their potential benefit-to-cost ratios. In particular, ARCEM resulted from the development of a planning methodology that provides NASA with a framework for generating and

analyzing control- and guidance-system concepts and for selecting concepts that maximize the benefits to the aviation community. The ARCEM program and the methodology it supports can provide a powerful tool for the organization and planning of research activities. It can indicate which concepts should provide the greatest benefit for the investment, and it can determine the number of concepts that must be implemented to justify expenditures for development of generic technologies. The ARCEM is written in BASIC for the TRS80 Model III microcomputer with a minimum configuration requirement of 48K of memory and one disk drive. Program use also requires a light-pen input device such as the 3-G Company unit.

The next series of applications reports focuses on energy-related applications. The first report in this group [CHICAGO, 1981] examines benefit and cost analysis of research and development projects. A major aspect of this project was the joint effort of researchers at the University of Chicago and researchers at Argonne National Laboratories. The main cooperation and complementarity was on the R&D Evaluation System and analysis applied explicitly to the case for electric vehicles. With respect to the former, the economic conceptualization, market penetration modeling and data collection were carried out mainly by researchers at the University of Chicago. Persons at the University of Chicago also contributed to the writing of the software package. This final report is contained in seven volumes. Volume 1 contains the technical explanation of the RD&D evaluation system, including the user's guide and the documentation manual. The second part of Volume 1 contains the software manual. Volume 2 contains a theoretical explanation of the R&D portfolio model, and extends the work presented by Tolley, Fishelson, and Tiwari. In Volume 3, the advanced benefit-cost model is adapted to the market penetration potential for electric vehicles. Volume 4 addresses the issue of industrial energy storage technology. Volume 5 discusses the relationship between market penetration rates and the potential costs savings associated with an innovative technology. Volume 6 is a threefold analysis of the firm's reaction to innovative technologies. In Volume 7, the household decision to adopt alternative air conditioning systems is modeled conceptually and demonstrated empirically using discrete choice econometric tools.

The second energy-related study [Spanner, 1992] computes expected benefits of federally-funded thermal energy storage research. Pacific Northwest Laboratory (PNL) conducted this study for the Office of Advanced Utility Concepts of the US Department of Energy (DOE). The objective of this study was to develop a series of graphs that depict the long-term benefits of continuing DOE's thermal energy storage (TES) research program in four sectors: building heating, building cooling, utility power production, and transportation. The study was conducted in three steps- The first step was to assess the maximum possible benefits technically achievable in each sector. In some sectors, the maximum benefit was determined by a "supply side" limitation, and in other sectors, the maximum benefit is determined by a "demand side" limitation. The second step was to apply economic cost and diffusion models to estimate the benefits that are likely to be achieved by TES under two scenarios: (1) with continuing DOE funding of TES research, and (2) without continued funding. The models all cover the 20-year period from 1990 to 2010. The third step was to prepare graphs that show the maximum technical benefits achievable, the estimated benefits with TES research funding, and the estimated benefits in the absence of TES research funding. The benefits of federally-funded TES research are largely in four areas: displacement of primary energy, displacement of oil and natural gas, reduction in peak electric loads, and emissions reductions.

The third energy-related report [Grey, 1983] summarizes an energy efficient engine program

technology benefit/cost study. Turbofan engine technologies required for the years 2000 to 2010 were studied, to assess the benefits of those technologies, and to formulate programs for developing the technologies required for that time period. Preliminary technology concepts that might be amenable to future development were ranked. Cycle studies, flowpath definition studies, and mechanical configuration studies were used to identify and establish the feasibility of the technologies that would be required in the 2000 to 2010 time frame. It is shown that a turbofan engine with advancements in aerodynamics, mechanical arrangements, and materials offer significant performance improvements over 1988 technology. The benefits of technologies are assessed using fuel burn and direct operating cost plus interest (DOC+I). The concepts could yield thrust specific fuel consumption benefits of almost 16%, fuel burn benefits of up to 24% and DOC+I benefits up to 14% in a long-range airplane relative to energy efficient engine technology levels. Technology development programs are formulated and recommended to realize those benefits

The next two energy-related studies [Pine, 1987] quantified ratepayer economic benefits of completed research at GRI. In the first study, the economic benefits for ratepayers are estimated for 44 technologies developed through GRI research that are in use in specific products, processes or techniques. Because the benefits of some technologies are difficult to quantify, approximate benefits were quantified only for a subset of 34 commercialized technologies in which the extent of use and associated cost savings could be estimated. The net value of these benefits was calculated at $3.5-7.0 billion (1986 dollars), about four to eight times the cumulative cost of the entire GRI R&D program from its inception through 1986. The analysis indicates that the GRI R&D program is beneficial and cost effective for gas industry and gas customers.

This later study [Pine, 1990] updated economic benefits to gas customers from completed research and development at GRI. Conducted in cooperation with gas industry partners, GRI's R and D program brought 93 gas products, processes and techniques, and 53 information items to the marketplace during 1987-1990. Quantitative estimates of economic benefits to the gas industry and its customers are provided for 60 of the technologies. The net present value is approximately $7.4 billion. While not accounting for R and D efforts in progress, the figure is 4.3 times the cumulative net present value of the cost of the entire GRI R and D program from its inception and represents a rate of return to ratepayers of almost 20%. When compared with the cost of completed R and D, the benefit-to-cost ratio is 8.1 to 1.

This report [Griffis, 1995] presents an analysis of benefits attributable to the Dredging Research Program (DRP). Each product developed by the DRP was catalogued. Each operation and maintenance dredging project was analyzed to determine whether a DRP product has been used or could be used on that project. The benefits were categorized as direct, cost avoidance, environmental enhancement, mission enhancement, and other indirect benefits. These benefits were arranged into a database. Due to uncertainty associated with each benefit estimate, each benefit estimate was assumed to follow a specific probability distribution. The sum of all benefits was then subjected to a Monte Carlo analysis and the relative frequency histogram of the final sum of all benefits was calculated.

This study [Fan, 1997] examines research, productivity, and output growth in Chinese agriculture. Recent attempts to quantify the sources of growth in Chinese agriculture have attributed an exceptionally large share of this growth to the contemporary institutional and market reforms within

China. To analyze this important issue, the authors use a newly constructed panel data set that includes an agricultural research or stock-of-knowledge variable. Their results suggest that while still a significant source of growth, the direct growth promoting consequence of institutional change and market reforms have been overstated by these earlier studies. Research-induced technical change accounts for a significant share (20%) of the growth in agricultural output since 1965.

The next study [McKeen, 1994] is a comparative-analysis of management information science (MIS) project selection mechanisms. MIS projects are selected by any of four different groups within organizations: top management, steering committees, user departments, and MIS departments. Because of their inherent differences, each of these groups is likely to favor different types of MIS projects. That is, they exhibit different selection biasing. An investigation of the nature and extent of this biasing is examined in this research. Data were collected from 176 MIS projects selected from 60 organizations. Projects were categorized as being selected by top management, steering committees, user departments, or MIS departments, and specific characteristics (e.g., size, risk, and organizational commitment) were measured for each project. As hypothesized, the research showed that projects selected by different groups did indeed differ significantly with respect to these characteristics.

Projects selected by top management do not tend to be more strategic, profitable, resource consuming, larger risk, or related to organizational well-being than other project selection groups. These projects, however, did tend to experience the longest start delay and elapsed development time. Projects selected by steering committees tended to be larger and riskier, and required more organizational change. Formal cost-benefit analysis is more predominant, but surprisingly, projects selected are not more cross-functional in scope. User department-selected projects, comparatively, are smaller, more quickly developed, and involve the fewest users, layers of management, and business functions. MIS-selected projects have more of an integration focus and follow more logical sequences in development. Their projects experience fewer delays in deliberation and duration, and less concern is given to cost-benefit analysis. The individual biasing attributable to each of the four selection mechanisms is described. The study concludes by presenting the implications of having each of these groups select MIS projects. Using this information, organizations can establish or assess the effect of using different mechanisms for selecting MIS projects.

This study [Bach, 1995] deals with an evaluation performed by BETA group about the economic effects of EU R & D programmes (Brite, Euram and Brite-Euram I) on the European industry. The approach used is based on an original methodology designed by BETA, which aims at evaluating those effects at a micro level (i.e. the participants to the programmes) by means of direct interviews of 176 partners involved in 50 projects. The definition of these economic effects is firstly described, as well as the different steps of the evaluation work. Then the overall results of the study are presented, showing the importance of both "direct" and "indirect" observed effects in monetary terms. Finally, some more detailed results highlight the positive impact of some aspects of the organization structure set up for the analyzed R & D projects on the amount of observed effects: i) the participation of a university lab; ii) the participation of at least one partner involved in a fundamental research work; iii) the diversity of research tasks over a scale ranging from fundamental research to industrialization work; iv) the combination of "user-type" and "producer-type" of activity in one given organisation (integration effect) or in one given project (consortia effect), etc.

The next three studies address cost benefit analysis in military manpower and training research and development. The goal of the first study in this group [McMichael, 1985] was to determine what current theory and practice in cost-benefit analysis (CBA) may have to offer toward improving the application of CBA tools in the Department of Defense, specifically their application to decision making in the human resources areas of manpower, personnel and training (MPT). A survey was made of the cost-benefit analysis literature to develop a taxonomy of generally accepted and widely used techniques and analytic precepts. The survey identified fourteen economic precepts and principles applicable to CBA; they were associated with two major foundations of CBA, financial analysis and welfare economics. Associated with financial analysis were the following seven elements; formulating the objective; specifying alternatives; determining the accounting stance; establishing decision criteria; discounting; conducting sensitivity analyses; formulating production functions. Associated with welfare economics were the following six elements; shadow pricing; establishing commensurability of costs and benefits; evaluating risk bearing; accounting for externalities; evaluating intangibles; measuring distributional effects. An additional element, conducting retrospective evaluations, was also included.

The goal of the second study in this group [Fast, 1992] is to measure benefits of manpower, personnel, and training (MPT) Research and Development. The Air Force is constantly trying to develop new or improve existing tools to increase the efficiency in the way personnel life cycle resources are managed. One metric commonly used is based on utility. This research produced a utility assessment technology to aid decision makers. This technology involves the process of identifying, measuring, and combining attributes to create an explicit value structure to form a basis for evaluating MPT research projects and selecting the most beneficial and cost effective portfolio of MPT research efforts. Four different techniques were evaluated and compared, those being utility analysis, cost benefit analysis, production functions, and decision theory. The research identified cost benefit analysis and decision analysis as being most applicable to MPT research projects.

The final study in this group [Belcher, 1997] describes a methodology for analyzing the costs and benefits of video teletraining (VTT). New technology is changing the way people are being trained. The Director of Naval Training (N7) has stated that the Navy needs to incorporate more of this new technology into its training environments. To achieve this goal, the training community must meet several challenges. N7 asked CNA for help in structuring a cost-benefit analysis of training technology. It wanted CNA to develop a methodology for analyzing and evaluating the potential benefits that new technologies can bring to Navy training. N7 stated that the methodology should define quantitative measures for assessing the benefits, specify mathematical relationships and procedures for computing these measures, and identify the data to be collected.

This report [Rey, 1996] addresses development of Green Box sensor module technologies for rail applications. Results of a joint Sandia National Laboratories, University of New Mexico, and New Mexico Engineering Research Institute project to investigate an architecture implementing real-time monitoring and tracking technologies in the railroad industry are presented. The work, supported by the New Mexico State Transportation Authority, examines a family of smart sensor products that can be tailored to the specific needs of the user. The concept uses a strap-on sensor package, designed as a value-added component, integrated into existing industry systems and standards. Advances in sensor microelectronics and digital signal processing permit us to produce a class of smart sensors

that interpret raw data and transmit inferred information. As applied to freight trains, the sensors' primary purpose is to minimize operating costs by decreasing losses due to theft, and by reducing the number, severity, and consequence of hazardous materials incidents. The system would be capable of numerous activities including: monitoring cargo integrity, controlling system braking and vehicle acceleration, recognizing component failure conditions, and logging sensor data. A cost-benefit analysis examines the loss of revenue resulting from theft, hazardous materials incidents, and accidents. Customer survey data are combined with the cost benefit analysis and used to guide the product requirements definition for a series of specific applications. A common electrical architecture is developed to support the product line and permit rapid product realization. Results of a concept validation, which used commercial hardware and was conducted on a revenue-generating train, are also reported.

This study [Nordham, 1993] describes an automated ship auxiliary systems design process/ benefit analysis program. Current design procedures often do not optimize the system characteristics (e.g., weight, volume, and cost) of auxiliary systems aboard U.S. Navy combatants. As a result, an automated design process was developed to examine the effect of design changes made to a surface ship auxiliary system on these characteristics. This process will allow comparison of different auxiliary system concepts for the selection of the best system configuration in a given combatant based on weight, volume, and cost impact on the ship. In addition, the design process will uniquely allow the examination of how design changes to an auxiliary system will impact different sized combatants. The automated design process is composed of two main programs -- a Ship Parametric Modeling Program in which the ship and auxiliary system model is developed in a parametric computer program for the NAVSEA CAD-2 system, and a Benefit Analysis Program in which the auxiliary system's characteristics are calculated for comparison to alternative components and system concepts. This report highlights the work done on the automated design process in FY 1993, specifically the work done on the Benefit Analysis Program. A description for use of the automated design process is also given.

The final study in the applications section [Boardman, 1994] addresses the lessons to be learned from ex-ante ex-post cost-benefit comparisons. According to the authors, the purpose of cost-benefit analysis (CBA) is to help public sector decision-making. The "help" varies according to when it is performed. CBA can be performed ex ante (EA), ex post (EP), or in the interim-in medias res (IMR) of a project. The authors propose a fourth class of CBA-one that compares EA with EP or with IMR CBA on the same project. In fact, this type of comparison has not been conducted in the literature. The authors suggest that without such research it is impossible to evaluate the practical value of CBA as a decision-making tool.

This study demonstrates the value of such comparisons, and contrasts them with other classes of CBA. Specifically: (1) it compares the advantages of comparison studies with other classes of CBA; (2) it categorizes four major types of error in CBA studies-omission errors, forecasting errors, measurement errors, and valuation errors-and models the impact of these errors on actual and estimated net benefits over time; (3) it examines the causes of the four different types of error; and (4) it compares three different classes of CBA on the same highway project: one clearly EA, one 18 months later (an IMR study) and one 7 years later (which we treat as an EP study). There are major differences in the estimates of net benefits. Contrary to what might have been expected, the largest source of difference was not due to errors in forecasts, nor differences in evaluation of intangible

benefits, but from major differences in declared and actual construction costs of the project. That is, the largest errors arose from what most analysts would have thought were the most reliable figures entered into the CBA. The authors conclude that comparison studies are potentially the most useful for learning about the accuracy and efficacy of cost-benefit analysis to decision-makers and evaluators.

Bibliography

This cost benefit analysis methods bibliography [NERAC, 1996] contains citations concerning innovations, improvements, approaches, and application methods for cost-benefit analyses. Analysis of costs and benefits for power plant productivity improvement is discussed. Use of cost-benefit analysis in establishing protection standards, and techniques for assessing benefits and cost effectiveness are examined for various systems including power production, air pollution, and waste remediation. (Contains many citations and includes a subject term index and title list.)

IV-D. COST-EFFICIENCY

A late 1980s production function approach to cost-efficiency of basic research essentially used a regression analysis between outputs and inputs [Averch, 1987, 1989]. In its latest incarnation, performed on NSF Chemistry proposals when Averch was at NSF, the method involved regressing output variables (citations per dollar, graduate students per dollar) against input variables (e.g., quality of the investigator's department, quality of the investigator, etc.). The results gave some idea of the importance of the input variables, alone or in combination, on the output variables. One obvious potential application would be prediction of proposals likely to have high productivity based on prior (input) knowledge. **Much, however, remains to be done in identifying the appropriate output measures, the appropriate input measures, and the nature of the interactions among these measures for different disciplines**.

IV-E. CO-OCCURRENCE PHENOMENA

IV-E-1. Background

Modern quantitative techniques utilize computer technology extensively, usually supplemented by network analytic approaches, and attempt to integrate disparate fields of research. One class of techniques which tends to focus more on macroscale impacts of research exploits the use of co-occurrence phenomena. In co-occurrence analysis, **phenomena that occur together frequently in some domain are assumed to be related, and the strength of that relationship is assumed to be related to the co-occurrence frequency**. Networks of these co-occurring phenomena are constructed, and then maps of evolving scientific fields are generated using the link-node values of the networks. Using these maps of science structure and evolution, the research policy analyst can develop a deeper understanding of the interrelationships among the different research fields and the impacts of external intervention, and can recommend new directions for more desirable research portfolios.

Little evidence of Federal use of these techniques (co-citation, co-word, co-nomination, and co-classification analysis) has been reported in the open literature. However, as computerized databases

get larger, and more powerful computer software and hardware become readily available, their utilization in assessing research impact should increase substantially. These techniques are discussed in more detail in Kostoff [1992a- Appendix III, 1993b, 1994j]; Tijssen [1994]. The Tijssen paper contains an excellent exposition on mapping techniques for displaying the structure of related science and technology fields.

IV-E-2. Overview Summary

Co-citation analysis has been applied to scientific fields, and co-citation clusters have been mapped to represent research-front specialties [Tijssen, 1994]. Co-word has been utilized to map the evolution of science under European (mainly French) government support, and has the potential to supplement other research impact evaluation approaches. Co-nomination, in its different incarnations, has been used to construct social networks of researchers and has the potential, if expanded to include research and technology impacts in the network link values, for evaluating direct and indirect impacts of research. Co-classification is based on co-occurrences of classification codes in patents, and is used to construct maps of technology clusters [Engelsman, 1991].

IV-E-3. Co-citation Analysis

Three of the more applicable co-occurrence techniques to the science evolution problem, listed in order of level of development and frequency of utilization, are co-citation, co-word, and co-nomination. In co-citation analysis, the frequencies with which references in published documents are cited together are obtained, and are eventually used to generate maps of clusters of cohesive research themes. Co-citation analysis was developed about two decades ago, when the Science Citation Index became more readily available for computer analysis, and it has spawned a number of studies and reviews, a few of which are listed here [Small, 1973, 1977, 1978; Garfield, 1978; Small, 1980, 1985a, 1985b, 1986; Franklin, 1988; Oberski, 1988; Braam, 1991a, 1991b].

It should be noted that co-citation is a rather indirect approach to obtaining connectivity among research areas, and it involves a number of abstract steps. Querying the author(s) of a research paper about what other research areas are related to their work would be the most direct method of obtaining the desired data [Kostoff, 1991c, 1992a-Appendix I, 1994j]. Obtaining this information by analyzing the words in the paper and related papers would be the next most direct method. Obtaining this information by examining citations and co-citations restricts the types of documents which can be analyzed (essentially published papers) and requires the additional assumption that the themes of two articles co-cited many times by authors must be strongly related. While the co-citation proponents claim that "many potentially useful applications have been demonstrated" [Franklin, 1988], others conclude that "results of co-citation cluster analyses cannot be taken seriously as evidence relevant to the formulation of research policy" [Oberski, 1988].

IV-E-4. Co-nomination and Co-classification Analyses

Co-nomination is a particular example of the more general social network analysis used to study communication among workers in the fields of science and technology. Generally, in co-nomination, experts in a given field are asked to identify other experts, and then a network is generated which shows the different linkages (and the strengths of these linkages) among all the

experts (and possibly their organizations and technical disciplines) identified. A 1988 survey [Shrum, 1988] of the development of social network analysis traces studies in this area back at least three decades. Two of these studies are particularly relevant to the specific co-nomination approach which will be described, and these two studies are outlined briefly.

In a study of theoretical high energy physicists [Libbey, 1967], respondents were asked to name two persons outside their institution with whom they exchanged research information most frequently and no more than three who they believed to be doing the most important work in their area. A network analysis was done to identify communication linkages. In a later study of theoretical high energy physicists [Blau, 1978], respondents were asked to name two persons outside their institution with whom they exchanged information most frequently about their research. Again, communication networks were generated.

Co-nomination was developed to circumvent co-citation's dependence upon databases consisting of refereed scientific publications. **It is a more direct approach of obtaining links among researchers and, if combined with other network approaches which include both links between technical fields and the link strengths [Kostoff, 1991c, 1992a-Appendix I, 1994i, Appendix 9-A-A in the present monograph], could potentially incorporate links among researchers and technical fields**. Since co-nomination is known less well than co-citation, its latest embodiment will be described briefly.

Researchers are sent a questionnaire inviting them to nominate other researchers whose work is most similar or relevant to their own. Based on the responses, networks are then constructed by assuming that links exist between co-nominated researchers and that the strength of each link is proportional to the frequency of co-nomination [Georghiou, 1988]. However, as is the case with co-citation, frequency of co-occurrence may not be a unique indicator of strength. One could postulate two cases: 1) researchers co-nominated were doing essentially identical work, and their linkages were very strong; and 2) researchers were doing vaguely similar work, and their linkages were very weak. In both cases, the frequency of co-occurrence would be the same, and the links on the network would have the same strength.

Co-classification analysis operates on the co-occurrence of terms (or codes) which are used to classify publications for ease of access in bibliographic databases. These indexer-given information items are derived from a thesaurus and may represent scientific (or technological) topics, specialties, or fields. Compared to key-words, subject classification terms have a well-defined and consistent meaning over the entire knowledge domain, which makes them particularly attractive for studying and depicting the main cognitive structure access large scientific and technological areas. The main practical restrictions are imposed by the fixed classification scheme. Moreover, classification codes are assigned primarily for information retrieval purposes and do not necessarily reflect intellectual concepts.

Key examples include Van Raan and Peters [1989], who use the co-occurrence of classification codes to construct MDS maps depicting the dynamics in the structure of chemical engineering. Tijssen [1992b] uses an MDS mapping of co-classification structures together with network analysis methods for identifying temporal changes in the cognitive links between fields of energy research. Engelsman and Van Raan [1992] present a co-classification map depicting the structure of relations

among all technological fields, according to the International Patent Classification scheme, and compare its configuration to a map of technology derived by means of co-word analysis.

IV-E-5. Co-word Analysis

The origins of co-word analysis in linguistics, lexicography, and especially computational linguistics can be found in Hornby [1942], De Saussure [1949], Firth [1957], Chomsky [1965], Halliday [1966], Harris [1968], Sparck Jones [1971], McKinnon [1977], Van Rijsbergen [1979], Melcuk [1981], Bahl [1983], Choueka [1983], Salton [1983], Sparck Jones [1984]; Benson [1986], Kittredge [1986], Choueka [1988], McCardell [1988], Nirenberg [1988], Smadja [1988], Amsler [1989], Church [1989], Maarek [1989], Salton [1989]; Smadja [1989], Church [1990], Iordanskaja [1990], Mays [1990], McDonald [1990], Smadja [1991]. These origins of co-word analysis are summarized in Kostoff [1991c, 1992a, 1993b, 1994j], along with a detailed description of modern day development and applications of co-word analysis to research policy and issues.

In summary, co-word has been utilized to map the evolution of science under European (mainly French and Dutch) government support [Callon, 1979, 1983; Rip, 1984; Bauin, 1986; Callon, 1986; Courtial, 1986; Healey, 1986; Leydesdorff, 1987a, 1987b; Bauin, 1988; Rip, 1988; Turner, 1988; Courtial, 1989; Leydesdorff, 1989; Whittaker, 1989; Courtial, 1990a, 1990b; Call on, 1991a; Braam, 1991a, 1991b; Callon, 1991b; Peters, 1991; Van Raan, 1991; Tijssen, 1994]. Until recently, the database used was essentially limited to journal papers. The frequency of co-occurrence of index or key words for these papers was the starting point for the maps which followed. Use of index words led to a biasing termed the 'indexer effect' [Healey, 1986] and effectively restricted the acceptability of co-word analysis for many years.

IV-E-5-i. Database Tomography

A new co-word approach that deals directly with full text and requires no indexing or key words was developed [Kostoff, 1991c, 1992a, 1993b, 1994j]. The methodology can be applied to any text database, consisting of published papers, reports, memos, etc., which can be placed on computer storage media. This revolutionary approach has been used to identify pervasive thrust areas of science and technology, the connectivity among these areas, and sub-thrust areas closely related to and supportive of the pervasive thrust areas.

The approach utilizes a computer-based algorithm to extract and order data from a large body of textual material which, for example, may describe a broad spectrum of science. The algorithm extracts words and word phrases which are repeated throughout this large database, and allows the user to create a taxonomy of pervasive research thrusts from this extracted data. The algorithm then extracts words and phrases which occur **physically close** to the pervasive research thrusts throughout the text, and allows the user to determine interconnectivity among the research thrusts, as well as determine research sub-thrusts strongly related to the pervasive thrusts. While the focus of applications has been to identify **technical** thrusts and their interrelationships, the raw data obtained by the extraction algorithms allows the user to relate **technical thrusts to institutions, journals, people, geographical locations, and other categories**.

Examples of the Database Tomography concept and diverse studies that have been performed since

its inception are presented in Appendix 7. Of particular interest to the present monograph, the recent studies covered by the examples include Database Tomography along with bibliometrics and expert analyses.

IV-E-6. Specific Co-occurrence Studies with Different Indicators

Co-occurrence indicators have some relation to collaborative indicators in that they provide some measure of relationships among disciplines, themes, institutions, performers, etc. The first five studies reported focus on co-citation studies, the next two studies reported focus on co-word analysis, and the final study presented focuses on combined approaches.

Co-citation Analysis

Co-citation analysis, already applied to the natural sciences' literature, was applied to the social and behavioral sciences' literature, as represented in that of the Social Sciences Citation Index [Griffith, 1983]. The major finding was that the analysis could cluster documents so that related works appeared together and could display relationships among documents and among clusters of documents which reflect scientific content. In contrast to the natural sciences, the social and behavioral sciences utilized older documents and placed greater emphasis on scholarly monographs. This proved true even in those areas most closely related to biological sciences, such as parts of experimental psychology. Generally published work in the social and behavioral sciences seems especially influenced by exceedingly small groups of researchers, who are represented often by quite old documents and who are not readily displaced by new research.

An author co-citation analysis (ACA) on the research into scholarly communication in sociology of science and in information science within a 20-year period is presented [Karki, 1996]. The question at issue is: to what extent and in what ways the research on scholarly communication brings together the sociology of science and information science, i.e. if the research on scholarly communication acts as a bridge between these two disciplines. It is natural to think of the research on scholarly communication as a common field for these two disciplines, but, by analysing the co-citations accorded to the researchers within both disciplines, one can define the intensity of the relationship or whether it really exists. The ACA suggests that the research of scholarly communication is not enough to be their common denominator: sociologists and information scientists mostly stay in their own respective territories. Finally, as the feasibility of ACA is evaluated in the light of the results, the weaknesses of the method become evident.

The third study in this section [Small, 1993] addresses macrolevel changes in the structure of co-citation clusters from 1983 to 1989. At ISI, a consistent method for clustering the combined Science Citation Index and Social Sciences Citation Index for the last seven years (1983 to 1989) has been used, according to the author. This method involves clustering highly cited documents by single-link clustering and then clustering the resultant clusters, a total of four times. This gives a hierarchical or nested structure of clusters four levels deep. Relationships among clusters at a given level can be depicted by multidimensional scaling, and by comparing successive year maps the analyst can then see how the relationships of major disciplines have changed from year to year. The analysts focus mainly on the two highest levels of aggregation, C4 and C5, to make observations about structural changes in science involving the major disciplines. Distinction is made between changes which

appear to be cyclic or oscillatory in nature, and those which appear to be more permanent or unidirectional.

The author of the previous study, Dr. Small, has been a leader in developing and advancing many aspects of co-citation analysis and mapping, and those interested in researching this area are well-advised to examine the full scope of his works. A brief summary of (mainly) his efforts in co-citation mapping follows.

In 1973, Small and Marshakova independently proposed using highly cited papers and their frequency of co-citation as the building blocks for a mapping of science [Small, 1973; Marshakova, 1973]. In 1974, Small and Griffith extended this approach to large Institute of Scientific Information citation data file [Small & Griffith 1974; Griffith et al., 1974]. Maps were constructed for both the microstructure of individual specialties, and macrostructure of broad fields, showing several scientific specialties in a common configuration. The technique of multidimensional scaling was used to display structure.

Eventually full annual files of Institute of Scientific Information (ISI) data were used, and up to four nested levels of clustering were performed, each level using the clusters obtained in the previous level as objects to cluster again [Small, Sweeney, & Greenlee, 1985]. After about four iterations it was possible to create global maps which showed relationships between disciplines in physical and biological science [Small & Garfield, 1985]. The advantages of this approach to mapping were, first, that co-citation provided a coefficient of similarity between documents,and a metric that could differentiate distances between objects. Second, clustering provided a chunking of the citation network, so that the complexity of document citation patterns could be hidden with a hierarchy of objects [Small, 1997].

Unlike the historiograph approach, co-citation maps use two dimensions to depict subject relationships. Change over time is analyzed by comparing maps from successive time periods. The time variable is usually taken as the year of the citing papers. The patterns of co-citation in that year define the collective perceptions of citing authors and give rise to clusters of highly cited and co-cited works. Shifts in highly cited papers are then used to study the rate of intellectual change. A sudden shift in the cited papers is then used to study the rate of intellectual change. A sudden shift in the cited papers set of a specialty can signal a revolution in the field. Rapidly growing fields such as AIDS can be tracked from their birth, as they spawn multiple lines of research, and eventually emerge as major fields in their own right [Small & Greenlee, 1990].

The co-citation methodology was also extended to authors, using the primary author rather than the document as the unit of analysis. Here the analysis focuses on individuals whose collective citation patterns can be mapped with multidimensional scaling [White & Griffith, 1981]. A recent interesting example of co-citation combined with word analysis is Braam et al. [1991a,b] focusing on the relatedness of different co-citation clusters through keyword similarity analysis.

As the final co-citation study shows, although co-citation techniques are very powerful structuring tools, the use of science policy indicators based on co-citation has often been criticized, especially on ISI research fronts. A major issue is the small fraction of literature retrieved, i.e. the "recall rate" problem. This recent investigation [Zitt, 1996] indicates that at the level of micro/meso studies high

recall rates can be achieved by (a) the use of appropriate clustering techniques limiting singletons and (b) the enrichment of cocited cores by medium-cited items. This combination of appropriate clustering and extension of recall proves to be efficient, provided that careful trade-offs are sought between the extension and relevance of recall. It leads to a reassessment of the performance of the co-citation approach for structuring scientific fields and providing related indicators not limited to the 'leading edge'. It also opens new opportunities for comparison/combination with other relational methods such as co-word analysis.

Co-word Analysis

This co-word analysis study [Coulter, 1996] applies various tools, techniques, and methods that the Software Engineering Institute is evaluating for analyzing information being produced at a very rapid rate in the discipline-both in practice and in research. The focus here is on mapping the evolution of the research literature as a means to characterize software engineering and distinguish it from other disciplines. Software engineering is a term often used to describe Programming in the large activities. Yet, any precise empirical characterization of its conceptual contours and their evolution is lacking. In this study, a large number of publications from 1982-1994 are analyzed to determine themes and trends in software engineering. The method used to analyze the publications was co-word analysis. This methodology identifies associations among publication descriptors (indexing terms) from the Computing Classification System and produces networks of terms that reveal patterns of associations. The results suggest that certain research themes in software engineering remain constant, but with changing thrusts. Other themes mature and then diminish as major research topics, while still others seem transient or immature. Certain themes are emerging as predominate for the most recent time period covered (1991-1994): object-oriented methods and user interlaces are identifiable as central themes.

The next study in this section [Courtial, 1993] focuses on the use of patent titles for identifying the topics of invention and forecasting trends. Co-word analysis applied to patents through WPIL normalized title words appears to give a useful picture of a given field: we obtain both qualitative (themes) and quantitative information (weight of themes). It also gives information about the strategic aspects of the themes. Furthermore, in some cases, it is an indication of the future of certain themes that may help forecasting and management studies. Finally, it provides information about what could be a real technology growth process, in relation to the so-called translation model used in co-word analysis.

Co-occurrence Maps

The final combined approach study [Tijssen, 1994] addresses mapping changes in science and technology; bibliometric co-occurrence analysis of the R-and-D literature. This study presents basic principles and examples of spatial representations derived from the analysis of co-occurrence frequency data pertaining to bibliographic information elements, such as key words and citations in research publications and patents. These bibliometric maps provide a means for communicating information on relational features of the science and technology (S&T) system-either for analytical or representational purposes. Characteristics of the main types of bibliometric maps are outlined and their potential for practical applications in S&T policy and research and development management are discussed. An emphasis is placed on more recent developments, in particular bibliometric maps

produced by the Centre for Science and Technology Studies (CWTS) for depicting temporal changes in the S&T system. Three empirical examples of such maps are presented with a focus on their application for impact assessment in both scientific as well as technological fields: (1) the emergence of new research topics in worldwide research on manufacturing technology, (2) changes in patterns of (inter)national collaboration within Dutch research on coal and coal products, and (3) the role of instruments in materials science.

IV-F. NETWORK MODELING FOR DIRECT/INDIRECT IMPACTS

IV-F-1. Background

In a mission-oriented research-sponsoring organization, the selection and continuation of research programs must be made on the basis of outstanding science and potential contribution to the organization's mission. There have been increasing pressures to link science and technology programs and goals more closely and clearly to organizational as well as broader societal goals [Carnegie, 1992]. The process of estimating **potential impact** of research, especially basic research, on organizational and societal goals is complex due to the myriad of pathways by which the research product can effect its impact. In fact, as Appendix 2 states, the process of accounting for total realized impact of research is very incomplete, again because of the nonlinear influence and impacts of research through a diverse multitude of pathways.

IV-F-2. Summary of Methodology

As a first step in addressing this multiple pathway impact issue in a more tangible way than has been done previously, a method was developed to quantify the impacts of research. The method is able to identify indirect impacts of research, and the pathways through which they are disseminated. A fully connected network is constructed whose nodes represent research, technology, and mission areas. The total impact of a given research node on any other node is the sum of the impacts (link value products) along every path in the network, and includes research-research, research-technology, and technology-research impacts. A pilot study was performed using a taxonomy of research and development nodes, with the raw input data (the link values) obtained from a survey of experts. An algorithm processed the data to provide total impact results. See Appendix 9-A for a more detailed description of the pilot study and results. See Appendix 9-B for the description of a computer algorithm which, as one of its capabilities, can display the structure and numerics of the multipath network architecture.

IV-G. EXPERT NETWORKS

Research Impact Assessment is, at its essence, a diagnostic process with many diagnostic tools. In other fields of endeavor, such as Medicine and Machinery Repair, expert systems are increasingly being used as diagnostic tools or as support to diagnostic processes. There have been some innovative efforts to develop expert system approaches combined with artificial neural networks (expert networks) for use in R&D management, including Research Impact Assessment [Odeyale, 1993; Odeyale and Kostoff, 1994a, 1994b]. The foundation of these approaches is the use of S&T metrics (and other associated metrics as well) in a computerized semi-autonomous decision aid. These efforts are summarized in Appendix 10. Much of the appendix was contributed by Dr.

Charles Odeyale, a true visionary in the application of Expert Networks to the broad area of R&D management.

IV-H. THE METRICS OF SCIENCE AND TECHNOLOGY

Since the initial Web version of the present report was published in 1998, a classic text on science and technology metrics has been published (Geisler, 2000). Anyone interested in S&T metrics should read this book. The present section presents the author's assessment of Professor Geisler's book, and emphasizes issues to be considered when implementing S&T metrics.

The book begins with a historical overview of technology's evolution as a major social force, then provides the theoretical background of the concepts and approaches for evaluating science and technology (S&T), and finishes with applications related to the evaluation of technology. The focus is on quantitative metrics (economic and financial, bibliometrics, co-analysis and mapping, and patents), but there is a section on qualitative metrics (peer review) as well. The innovation continuum addressed spans the range from fundamental science/ research to advanced technology development, and the subsequent transformation of technology into products.

The book starts from the fundamentals of measurement and metrics, addresses specific metrics from multiple perspectives, shows the benefits of aggregation of metrics into integrative indices, describes how these indices fit into the strategic management of S&T, and finally shows how S&T should be evaluated and treated as part of the overall organization's business strategy.

After an excellent discussion of inputs, outputs, and outcomes from S&T, the book presents an exhaustive evaluation of the strengths and weaknesses of each metric. Many of these different types of metrics are integrated spatially and temporally in a process-outcomes model. This multi-temporal stage dynamic model links the S&T process with the social and economic systems, and allows tracking of the innovation process from inputs/ activity to outputs, impacts, and outcomes.

The book is very eclectic; it draws from a variety of global references and experiences. While much of the analysis relates to United States experiences, both European and Asian experiences are highlighted as well. The three relatively standardized frameworks of scientific indicators for multi-country multi-parameter evaluation (OECD, U. S. National Science Board, Japanese Science Indicators System) discussed in the book reflect this national diversity.

In the last section of the book, a variety of applications to the academic, industrial, and public sectors are reviewed. The differences in the metrics used for each application, and particularly the context and larger processes in which they are used, are emphasized. Because the book's scope includes both science and technology, and because the scientists and technologists in these respective segments of the innovation continuum have different objectives and responsibilities, the differences in metrics applied to these two groups are also emphasized.

For academic institutions, Geisler distinguishes between teaching institutions (universities and colleges) and research institutions. Further, Geisler also includes academic institution spin-offs, such as research parks and cooperative programs with industry, in this metrics applications

section. For industrial institutions, Geisler describes metrics used in the evaluation of S&T projects, followed by industries and sectors. The purpose here is to provide a framework for metrics classification as implemented operationally. For public-sector institutions, Geisler discusses the relation of evaluation processes and their component metrics with the objectives of the multiple stakeholders that oversee and control the institutions. The relationship of The Government Results and Performance Act of 1993 (GPRA) to stakeholder interests is discussed with an excellent illustrative example.

Throughout the book, multiple perspectives are examined for each metric, each dynamic process, and each application. In this respect, the book is not only of the highest levels of academic scholarship, but is eminently practical for use as an operational handbook. However, the reader should not expect to be spoon-fed with fixed protocols for employing metrics. Much thought and judgement will be required to decide among the cornucopia of metrics presented, and the dynamic models in which they should be imbedded, given the breadth of strengths and weaknesses presented for each measure/ indicator/ metric.

The reader should pay particular emphasis to the following issues when reading the book, and when considering the implementation of metrics.

1) GLOBAL VS LOCAL OPTIMA

There are two fundamental incompatibilities of metrics with S&T, especially science. First, the main product of science/ research is understanding of fundamental phenomena. This understanding is not amenable to metrics. Only the expressions of understanding on the physical plane, such as science/ research documents, hardware, software, etc., are amenable to metrics. Thus, metrics will intrinsically be incomplete in describing the performance and progress of science/ research.

For this reason, metrics have not been used extensively in the evaluation of science/ research. Only recently, when laws such as GPRA were passed in the U. S., has there been more intense interest in metrics for science/ research evaluation. There is concomitantly a major concern that metrics could be mis-applied to science/ research as a result of these external pressures for accountability.

The second incompatibility applies to the economics of science/ research, and derives from the difference between global and local optimization. For the most part, fundamental science/ research is not cost-effective for industrial sponsors, because of their short-term time horizons for financial returns, and the type of <u>locally-optimized</u> economic analyses they use to compute these returns. There are three intrinsic reasons for this statement.

a) True fundamental science/ research is very risky, with many failures and few payoffs. This effect is masked today, because much science and technology as well has been classified as fundamental science/ research, and consequently the large failure rate is not observed with this much less risky applied science/ research and technology.
b) For the few science/ research projects that do succeed, the benefits may not necessarily accrue to the sponsor of the science/ research. In many cases, it is difficult to identify a

single sponsor for a successful science/ research product, or even to allocate benefits to particular sponsors.

c) Even if the benefits accrue to the sponsor, there historically has been a long time lapse between the expenditures of funds for science/ research, and the revenues from the commercial applications. This severely degrades benefit-cost ratios that are based on the time value of money. With some of the more recent information technology disciplines that have characteristically shorter development times, the time lapse may not be as large as the more imbedded physical and engineering science disciplines.

Because of these reasons, true fundamental science/ research has not been supported extensively by industry. While some so-called industrial research centers were created to provide short- and mid-term results to offer the company a competitive advantage, many existed for public relations purposes. When economic downturns occurred (e.g., the aerospace industry in the early 1970s), these research centers were the first organizational components to be eliminated. Some pockets of industrial research may exist today in a few selected disciplines (e.g., biotech, information science), but for the most part, it is government that supports basic science/ research. In this case, the metrics are quite different. The government metrics tend to be derived using global optimization over space (many beneficiaries) and time (longer horizons are acceptable). Other measures than standard benefit-cost analyses tend to be used. In plain language, *what is good for society may not be good for a firm, and vice versa.*

2) PURPOSE AND MOTIVE OF METRICS EVALUATIONS

While the specific metrics and dynamic models used, and their operational mechanics, are important in S&T evaluation and monitoring, much more important are the purpose behind the evaluation and the manager of the full evaluation. It is critical that the organization that selects the metrics and evaluation processes, and performs the analyses, be as independent and objective as possible.

In the recent Departmental reviews for which the author has been responsible, he has contracted with an arm of the U. S. National Research Council, the administrative unit of the National Academies of Science and Engineering, and the Institute of Medicine, to conduct the evaluations. The author considers having this independent unit, the Naval Studies Board (NSB), as the most important component of the evaluations, more important than any specific metrics chosen, or any agenda structure. The benefits of the NSB go beyond the strictly measurable. The panel has the flexibility to make subjective judgements, and arrive at unpopular conclusions and recommendations. Dr. Geisler addresses different types of evaluation organizations in this book, but should have emphasized the potential for strong deficiencies and inherent biases of self-evaluation (for purposes other than operational monitoring) more emphatically.

3) INTEGRATION INTO STRATEGIC MANAGEMENT

Most organizations use metrics today in isolation from dynamic models, from other management decision aids, and from effective decision-making. As such, metrics contribute more to public relations than public policy. Under such conditions of isolation, operational data derived from normal business practices is all that is available to quantify the metrics. This restricted data in

turn limits the universe of goals and objectives whose progress can be gauged by the metrics chosen. When metrics and the other complementary management decision aids are fully integrated into the strategic management process, the organizationally-appropriate objectives and goals can be selected first, the best metrics to gauge progress toward these objectives can then be chosen, and the data to quantify these metrics can be generated finally. Thus, data gathered for monitoring tactical and strategic business operations will correctly derive from objectives, and not the converse situation that exists in practice today. If metrics are to play an effective role in evaluation and monitoring, they need to be integrated into the strategic management of the organization.

Geisler correctly points out the need for fully integrated organizational behavior models, where key variables can be identified, and selected as the metrics for effective monitoring. *It is imperative that every S&T metric, and its associated data, presented in a study or briefing have a decision focus. It should contribute to the answer of a question that in turn would be the basis of a recommendation for future action.* Metrics and associated data that do not perform this function become an end in themselves, offer no insight to the central focus of the study or briefing, and provide no contribution to decision-making. They dilute the theme of the study, and, over time, tend to devalue the worth of metrics in credible S&T evaluations. Because of the present political popularity and subsequent proliferation of S&T metrics, the widespread availability of data, and the ease with which this data can be electronically gathered/ aggregated/ displayed, most S&T metrics briefings and studies are immersed in isolated data geared to impress rather than inform

4) INTEGRATION INTO STRATEGIC GOAL SELECTION

In some cases, the process of metrics development can be of equal importance to the final metrics developed. The following strategic goal selection example illustrates this point. In 1998, the author placed a document on the Web entitled Science and Technology Metrics (www.dtic.mil/dtic/kostoff/index.html). Immediately, the author was deluged with requests from S&T sponsor and laboratory managers to discuss the selection of metrics for strategic goal progress measurements. These requests derived from the burgeoning interests of the technical community in metrics as a result of the impending requirements from the newly-instituted GPRA legislation.

The author found that the process of relating metrics to strategic goals offered substantial insight into the objectives formation process, and in most cases drastically revised the number and structure of the goals themselves. A very different perspective of an organization's response to its mission can result when quantifiable goals are the target. It was instructive for the author to see how many organizational goals, across many government agencies, were more public relations statements than targets amenable to quantified evaluation. The main value that eventually results from GPRA may very well be the restructuring of organizational goals to a form where they can be evaluated with some degree of quantification, and identifying the metrics that will help perform this function.

5) PUBLIC SECTOR S&T SPONSOR RESPONSIBILITIES

In Geisler's chapter on public sector S&T evaluation, there is an illustrative example on metrics that the National Institute for Occupational Safety and Health (NIOSH) could use to evaluate progress towards its strategic goals. This example and its accompanying discussion impinge upon the mission and goals of an S&T sponsor, and the types of metrics needed to evaluate progress made toward these goals. However, the goals and accompanying metrics in the illustrative example address only part of the broader goals and metrics applicable to all S&T sponsors.

Public-sector S&T sponsors have two major responsibilities: a) to sponsor high quality S&T that has high potential for eventually being used to improve systems and operations of the sponsor's stakeholders/ customers for national benefit, and b) to make the downstream developers/ acquisitioners of these final products aware of global S&T being performed that could impact their downstream development and acquisition. These S&T sponsors have little control over the fate of their sponsored S&T after the S&T is completed, and especially after the S&T transitions to other organizations for further downstream development and acquisition. Some of the many external factors that determine the eventual fate of S&T other than technical quality include geopolitical, local political, economic, financial, legal, environmental, cultural, etc. The only control the S&T sponsors can actually exert over potential applications is to produce a high quality product that has positive transitionability characteristics (e.g., affordable, maintainable, reliable, addresses stakeholder and customer need, high technical quality, etc). Succinctly, ***S&T sponsors control outputs, not outcomes.***

Yet, present metrics systems for evaluating public sector S&T sponsors do not address the reality of the two responsibilities described above. Public sector S&T sponsors are held accountable for both outputs and outcomes. Many public sector S&T sponsor evaluations contain metrics that address downstream outcomes. Public sector S&T sponsors are held accountable, to some degree, for S&T products that do not transition for further development, or that do not eventually result in envisioned outcomes. This is an example where ***the appropriateness of the metric is perhaps more important than its measurement capability.***

Conversely, public sector S&T sponsors, for the most part, are **not** held accountable for providing their acquisition partners/ stakeholders with information about global S&T that could impact final operational systems. This is particularly egregious for two reasons: a) any public sector agency is financially limited to funding only a small fraction of global S&T, while many agencies' stakeholders have eclectic S&T needs that span many technologies being developed globally; b) of all public sector organizations, the S&T sponsors (and their associated performers) have the technical personnel who are most qualified to interpret global S&T developments, and identify those that offer the most potential. Yet, ***metrics to evaluate S&T sponsors for their performance on the crucial awareness responsibility have not even been conceived.*** Geisler's book (nor anyone else's) does not address this latter metrics group.

6) BIBLIOMETRICS DEFICIENCIES

While Geisler identified many strengths and weaknesses related to bibliometrics, there were a few issues that were understated, or not stated at all. Bibliometrics are document-based; they make sense only when adequate documentation exists. However, as pointed out in a recent paper

(2), much of S&T performed globally is not documented, and of the portion that is documented, much of the information does not reach the analyst in usable form. While there are many reasons for lack of documentation, basically there are far more *disincentives* to publishing than *incentives*. Thus, in areas that: a) relate to national security; b) involve proprietary material; or c) have a strong base external to academia, bibliometrics could provide a false impression of the discipline.

Along the same lines, bibliometrics tend to be employed in a passive operational mode. Lotka's Law, the distribution function that relates the number of authors to the number of papers they publish, shows that most researchers publish very little. Why haven't these results been used to increase the population of the lower tail of the distribution function? While there will always be differences between the prolific producers and the remainder of the researchers, why does it have to be so large? Much of the difference may be due to the lethargy of the bulk of the research community for documentation, and the absence of mandates and requirements for documentation of sponsored research. This is an example of how metrics could be used in an active feedback mode to influence what is being measured. **The passive bibliometrics operational mode is a direct result of the non-integration of metrics into the strategic management process!**

Finally, much bibliometrics is used in a comparative mode. One group's outputs, or citations, are compared to those of another group. But what happens if neither group is particularly efficient or productive? Specifically, what if an entire sub-discipline is not overly productive, or impactful? Bibliometrics does not address these cases. Bibliometrics needs to be supplemented with a capability to address absolute impacts, or outputs. A recent study (3) suggested one possible approach for citations, based on an analog to Carnot efficiency in thermodynamics. This approach related citations actually achieved to citations that could have been achieved, and went well beyond the relatively ineffectual comparison-only mode that has been the bibliometrics standard for generations. **More absolute output metrics need to be developed for science/ research and technology, as exist for many other human endeavors.**

7) INTEGRATIVE METRICS MONITORING

Geisler has an excellent chapter describing process outcomes, based in large extent on his outstanding work in this area. He generates integrated metric indices that cover many different metrics (weighted) over different time segments in a dynamic model. Such an approach lends itself to semi-automated organizational S&T-activity based monitoring. The index values would serve as warning flags for large-scale organizational performance problems. These indices could then be easily de-convoluted to the specific metrics that identify the key problem areas. This allows for monitoring at many different hierarchical levels in the metrics aggregation structure, and in a parallel sense in the organizational hierarchy as well.

In summary, Professor Geisler has produced a seminal work in science and technology metrics, and anyone directly or peripherally involved in science and technology would be well-advised to read this volume.

IV-I. S&T METRICS - SUMMARY AND CONCLUSIONS

To summarize this S&T metrics monograph, the implementation of GPRA has resulted in exponentially increased interest by the Federal agencies in the use of quantitative methods for science and technology evaluation. However, few Federal agencies report use of bibliometrics to evaluate programs and influence research planning in the published literature. Cost-benefit and other economic approaches have been reported in the published literature over the years. The foundation on which these approaches rest needs to be strengthened to improve their credibility. As Averch [1991] states, after describing the huge social rates of return to investments in hybrid corn reported by Griliches [1958]: "In general, economists compute high social rates-of-return to most kinds of research. The rates, in fact, are usually much higher than those computed for other kinds of public investment. So there is a puzzle as to why research investments do not increase until their marginal return just equals returns from other public investments."

However, for the global reasons stated in the introductory section of this paper about the increased need for accountability, and especially due to the impending implementation of GPRA to institutionalize this accounting requirement, S&T metrics will see (and are already seeing) greatly expanded use in the future (see Appendix 1-A for further description of S&T metrics issues related to GPRA. See Appendix 1-B for examples of metrics that support peer review of basic research, and Appendix 1-c for an example of metrics that support peer review of advanced technology development). Unfortunately, this expanded use of metrics derives from a reactive reflex to imposed requirements from oversight organizations, rather than an intrinsic desire to employ metrics for improving organizational performance. In fact, the GPRA-imposed requirements present an extraordinary opportunity. They provide an impetus to incorporate S&T metrics into an expanded corporate strategic vision for organizational management in the 21st century.

Present and forthcoming Information Technology capabilities allow the mechanical system principle of Condition-Based Management (CBM) to be applied to the management of organizations. CBM requires that maintenance be performed on a system when indicators signal that it is required, unlike scheduled periodic maintenance (SPM) which requires maintenance at pre-determined intervals. CBM is not only more cost-effective, since un-needed maintenance is eliminated, but it has the capability to prevent serious damage from problems which occur unexpectedly before the scheduled maintenance. Under the scenario of organizational CBM, all aspects of an organization's operation would be quantified and tracked in an integrated manner. Thus, financial transactions, resource flows, S&T inputs and outputs, strategic and tactical financial/ economic/ production/ research/ development targets and goals, etc., would be quantified and tracked. Figures of merit that integrate many of these diverse metrics would be generated.

Analogous to a physical system, these figures of merit would serve as indicators of the health or sickness of the organization. Parallel to a CBM for physical systems, when these organizational figures of merit exceeded pre-specified bounds, warning signals would sound. These messages would focus management attention on potential problem areas, and allow corrective action to be taken with sufficient lead time to avoid disaster. This is the correct use of metrics in science and technology: a component in a sophisticated management system that allows the sponsoring organizations to take corrective action when problems are about to occur, and which rewards those responsible for science and technology outputs which positively influenced the social order.

V. APPENDICES

TABLE OF CONTENTS FOR SECTION V

APPENDIX 1: METRICS IN SUPPORT OF PEER REVIEW
1-A: PEER REVIEW: THE APPROPRIATE GPRA METRIC FOR RESEARCH
1-B: METRICS FOR PEER REVIEW OF BASIC AND APPLIED RESEARCH
1-C: METRICS FOR PEER REVIEW OF ADVANCED TECHNOLOGY DEVELOPMENT
APPENDIX 2: THE UNDER-REPORTING OF RESEARCH IMPACT
APPENDIX 3: UTILITY OF CITATION ANALYSES
3-A: CHARACTERISTICS OF HIGHLY-CITED AND POORLY-CITED PAPERS
3-B: CITATION ANALYSIS OF RESEARCH PERFORMER QUALITY
3-C: CITATION DIFFERENTIALS
APPENDIX 4: DISPLAY OF BIBLIOMETRICS RESULTS
APPENDIX 5-A: CITATION NORMALIZATION APPROACHES
APPENDIX 5-B: CITATION ANALYSIS CROSS-FIELD NORMALIZATION: A NEW PARADIGM
APPENDIX 5-C: IS CITATION NORMALIZATION REALISTIC?
APPENDIX 5-D: CAB – CITATION-ASSISTED BACKGROUND.
APPENDIX 6: THE PIED PIPER EFFECT: A SPECIFIC EXAMPLE
APPENDIX 7: EXAMPLES OF S&T BIBLIOMETRICS STUDIES
7A: FULLERENES RESEARCH
7B: AIRCRAFT
7C: ANALYTICAL CHEMISTRY
7D: ELECTRIC POWER SOURCES
7E: ELECTROCHEMICAL POWER SOURCES
7F: NONLINEAR DYNAMICS
7G: FRACTALS
7H: CITATION MINING-GRANULAR SYSTEM DYNAMICS
7I: CITATION MINING-MACROMOLECULAR MASS SPECTROMETRY
APPENDIX 8: SCIENCE AND TECHNOLOGY TRANSITIONS
APPENDIX 9-A: NETWORK MODELING FOR DIRECT/INDIRECT IMPACTS
APPENDIX 9-B: NETWORK MODELING FOR ROADMAPS
APPENDIX 10: EXPERT NETWORKS
APPENDIX 11: POTENTIAL USE OF ENTROPY IN RESEARCH EVALUATION
APPENDIX 12 INFRASTRUCTURE OF S&T METRICS LITERATURE

APPENDIX 1

METRICS IN SUPPORT OF PEER REVIEW

1-A. Peer Review: The Appropriate GPRA Metric For Research [Kostoff, 1997a]

The federal government is the largest single sponsor of fundamental science research today. Increased scrutiny of federal programs in the drive toward deficit reduction requires increased public accountability for the stewards of the government's research funds. The Government Performance and Results Act (GPRA) of 1993 [GPRA, 1993] was passed to improve the accountability of government funded programs by measurements of performance against planned targets. Federal agencies are required to initiate implementation of GPRA in FY1997; pilot projects [Brown, 1996] will help identify performance measures for different types of programs. However, it is extremely important that the tools used to enforce research accountability do not destroy basic research.

There are three major components to GPRA: Strategic plans, annual performance plans, and metrics to show how well the annual plans are being met [GPRA, 1993]. Classical strategic planning derives from the military and commercial world, focuses on the application of knowledge toward a pre-defined goal rather than the search for knowledge, and assumes that the links between plans and targets are understood.

Annual performance plans are derived from production and service industries, where efficiency in the use of known resources to achieve well defined targets over the performance period is the main goal. Revolutionary basic research, which has yielded some of the largest downstream payoffs historically, has an inherently large uncertainty and failure rate, and may take many years before results are forthcoming. This intrinsic long-time scale characteristic of basic research conflicts with the short-term emphasis of much of the corporate world, where annual reports and requirements for quarterly financial performance shorten the production period for research results. This near-term focus on financial performance has essentially eliminated long-range high-risk fundamental research financed from corporate funds in most industries.

Metrics that gauge adherence to annual performance plans derive, in modern times, from the time and motion study component of industrial engineering. Again, these tools measure efficiency of the use of known resources to achieve specific goals over a set time period. At present, such output metrics are applied informally to research for purposes of academic analysis [Kostoff, 1995c], and these analytical results may provide useful insights to research activity. Annual application of these quantitative indicators is more appropriate for measuring the short-term observable outputs that characterize activity and productivity (cars produced, papers published) than the long-term outcomes that characterize mission and societal impact (improving health, enhancing safety).

A major concern of researchers is that the short-term services and production orientation of the GPRA planning and metrics components could re-focus the research away from long-range high-risk revolutionary science challenges to shorter-term low-risk evolutionary product-oriented goals. Annual application of these metrics to basic research in the formal bureaucratic sense of GPRA could convert the nature of the research being conducted from a quest for knowledge and understanding to a drive for output metrics. Uncertainties inherent in basic research bring into

question the validity and credibility of any long range plans to achieve specific goals, since long-term research effectiveness and impact will depend on economic, environmental, and geopolitical factors not evident during the research phase [Kostoff, 1997n].

A more subtle concern is that application of the present GPRA approach to basic research may effectively yield the same results as government imposed censorship. The requirements of federal agencies to display compliance with the GPRA metrics may reorient their selection of research proposals to maximize these arbitrary measures. Concepts that could improve understanding and the unification of science, but would not optimally satisfy the GPRA metrics, might no longer be proposed for federal funding because of lower funding probability. (I am reminded of Solzhenetzyn's views that the worst part of documents being censored was not that sections were rejected; the worst part was the loss of those ideas which were not even expressed and eventually no longer considered because of the knowledge that they would be censored). Safe, short-term, low-risk evolutionary research would become the accepted practice. Basic research needs to be decoupled from 'strategic' targets and GPRA metrics, and the scientific roadblocks and challenges alone should be the stimuli for research activity.

A more appropriate accountability approach for basic research is: i) articulation of a rational investment strategy; ii) long and short-term retrospective studies that show the diverse benefits from past research and potential future benefits; iii) quality control of expert peer review. An organization's research investment strategy is a rationale for the prioritization and allocation of resources to address knowledge deficiencies which impede attainment of the organization's goals. Short-term retrospective studies show how recent research has affected fields of science, and may contain projections of future impacts of research on technologies, systems, and operations. Long-term retrospective studies of major innovations and outcomes in systems and technology show the origins of critical research and development advances in a broad spectrum of fundamental research performed many decades earlier [IITRI, 1968; BATTELLE, 1973; IDA, 1991]. Expert peer review on a periodic basis will validate the soundness of the investment strategy and the importance of the research accomplishments and subsequent technology impacts.

Peer review properly designed to support GPRA would provide credible indication to the research sponsors of intrinsic program quality, program relevance, management quality, and appropriateness of direction, and has the potential to improve the quality of the research program as well [Kostoff, 2004q]. Before such a review process is implemented, a number of considerations have to be addressed.

The primary requirements of excellent peer review are the dedication of an organization's senior management to the highest quality objective review, and the motivation of the review manager to conduct a technically credible review. In particular, the review manager selects the review process, criteria, and reviewers, guides the panel questions and discussion, summarizes reviewers' comments, and recommends follow-up actions. The selection of panelists by the review manager can substantially influence the review outcome.

Excellent peer review that provides an accurate picture of the intrinsic quality of the research being reviewed requires highly competent reviewers, and no injection of additional distortions in the reviewers' evaluations as a result of biases, conflict, fraud, or insufficient work. Not only should

each reviewer be technically competent for his or her subject area, but the competence of the review group should cover the multiple facets of research issues (specific research area reviewed, allied research areas, technology, systems, missions). In addition, panel expertise should not be limited to subdisciplines of the program under review (which addresses the quest ion of whether the job is being done right), but should be broadened to the area covered by the overall program's highest level objectives (which addresses the question of whether the right job is being done). Broadening the panel in this manner will ease introduction of new paradigms.

If GPRA reports are used to support the budgetary process, the results of different panels evaluating different technical disciplines must be normalized so that parametric comparison becomes meaningful. Biases, interpretation differences, scoring differences, different review processes, and the myriad of other causes for panel differences over and above intrinsic technical quality differences must be identified and mitigated. Differences in repeatability, reliability, and precision should also be identified and minimized.

Finally, peer review costs, which include more than direct, out-of-pocket costs, should not be neglected in establishing a specific review process. With high quality performers and reviewers, time/ opportunity costs are high, and represent the major contribution to total costs. The total review costs can be a non-negligible fraction of total program costs, depending on the review frequency, the level of technical detail desired, and whether the programs are labor or hardware intensive.

In summary, peer review is the appropriate central evaluation mechanism for basic research under GPRA, but careful thought and planning will be required to implement a viable and credible peer review process.

1-B. Metrics for Peer Review of Basic and Applied Research [Kostoff, 1997n]

1-B-i. CRITERIA FOR AGENCY REVIEWS
(ONR, circa early 1990s)

The following are generic guidelines that the author used when conducting research program reviews in the mid-1980s to late 1990s. They provided a framework for the more detailed questioning and analyses that followed. Attributes like 'creativity' and 'innovation' were subsumed under topics like approach, revolutionary research, etc, and were certainly focal points of ensuing discussions.

1. Scientific quality and uniqueness of ongoing and proposed efforts

2. Scientific opportunities in areas of likely user importance

3. Balance between revolutionary and evolutionary research

4. Position of research relative to forefront of other scientific efforts

5. Responsiveness to present and future user requirements

6. Possibilities of follow-on programs in higher R&D categories

7. Appropriateness of research for agency vice other Federal agencies.

1-B-ii. QUESTIONS FOR AGENCY PROGRAMS
(ONR circa early 1990s)

These questions supplemented the previous ones listed, and offered other perspectives on attributes and characteristics of high quality research programs.

1. What is the investment strategy of the larger management unit. This would include the relative program priorities, the actual investment allocation to the different programs, and the rationale for the investment allocation. For each program being reviewed, what is the investment strategy for its thrust areas.

2. Can specific advantage to customer be identified if program is successful?

3. Would efforts be supported if they were not already underway?

4. What is the technological context of the program and how does it fit with other ongoing research in academia, industry, and other Federal agencies?

5. Is the program appropriately coordinated with programs at other research organizations?

6. What are the research objectives of the program? What are the "mid term" and "final assessment criteria?" How much will the program cost?

7. What is the program trying to do?

8. How is the program (effort) done today? What are the limitations of the current practice?

9. What is new in the approach? Why will approach be successful?

10. What are the major risks of the program?

11. Assuming program is successful, what difference will the result make to customer capabilities?

1-B-iii. INSTRUCTIONS FOR COMPLETING PROJECT RATING FORMS – BASIC AND APPLIED RESEARCH

(DOE, circa mid-1980s)

The following form contains criteria the author used when conducting research project reviews in the early 1980s. This form is fundamentally no different from the previous forms shown, although the specific criteria listed may have slight differences. Innovation is spelled out in the approach criteria. A key feature in all the forms shown is the inclusion of an overall project quality rating. This is extremely important, since it allows the inclusion of any criteria that the reviewers believe are important in determining overall project quality, but were not called out specifically in the specific criteria on the form.

Peer Review Questionnaire (Form 1)

Reviewers individually rate the project in each of six areas and choose an overall rating: scientific (technical) merit, importance of project, quality of project team, scientific (technical) approach, productivity, and probability of success. Ratings in these categories use a scale composed of integer values from zero to ten, with the ends of the scale representing seriously deficient and outstanding attributes, respectively.

For Item Q1, "Scientific (Technical) Merit," reviewers assess the importance of the scientific (technical) question or problem addressed, including the potential importance or value to science (technology) of meeting the project objectives. This judgment is based primarily on the reviewer's knowledge of the scientific (technical) field.

In Item Q2, "Importance of Project," the reviewer is to assess the importance of the project's objectives in terms of contributing to the program's mission.

For Item Q3, "Quality of Project Team," reviewers consider the composition and quality of the team through examination of contributions by individual and associated team members relevant to the objectives of this project, honors and awards, experience relevant to the project area, and the balance of appropriate skills (including collaborators), for accomplishing the project objectives.

For Item Q4, "Scientific (Technical) Approach," reviewers consider the appropriateness of the experimental and analytical methods used and the level of insight and innovation demonstrated in relation to the requirements of the project's objectives.

For Item Q5, "Productivity," the reviewers consider the impact, volume, quality, and usefulness of work produced by the project team as a whole and relate this output to the resources available and costs incurred.

For Item Q6, "Probability of Success," reviewers assess the likelihood that the project will accomplish its stated objectives.

<u>Overall Project Evaluation</u>

The overall project evaluation score is a weighted judgment by the individual reviewer based on his/her experience and on the ratings given for Items Q1 to Q6. It is not mathematically derived from the factor scores. Criteria for choosing an overall project evaluation are also on Form 1.

PROJECT RATING FORMS

FORM 1 Reviewer # _____

Panel/Project: _____ Date of Review: _____

PEER REVIEW QUESTIONNAIRE

Q1. Scientific or Technical Merit of the Project Objectives

0 1 2 3 4 5 6 7 8 9 10

Project objectives of central importance to advancing the science, technology, discipline, or research area rate 9-10, project objectives that address significant issues rate 7-8, project objectives providing information of general usefulness and interest rate 5-6, Routine project objectives rate 3-4, and project objectives of doubtful or peripheral interest would rate 0-2. Circle the appropriate number for your rating.

Supporting Comments:

_____

Q2. Importance of Project Objectives to Mission

State your estimate of the importance of this project's stated objectives in terms of contributing to the program's stated mission. Circle the appropriate number for your rating.

Not Important Very Important

0 1 2 3 4 5 6 7 8 9 10

Supporting Comments:

Q3. Quality of Project Team

0 1 2 3 4 5 6 7 8 9 10

An outstanding team rates 9-10, a strong, balanced team of experienced investigators rates 7-8, a good team that would benefit from additional skills rates 5-6, a team that requires strengthening rates 3-4, and a team with serious shortcomings rates 0-2.

Supporting Comments:

Q4. Scientific or Technical Approach

0 1 2 3 4 5 6 7 8 9 10

An expert and innovative approach rates 9-10, a skillful and logical approach rates 7-8, a reasonable approach with potential for improvement rates 5-6, an approach with key shortcomings or an approach that is out-of-date rates 3-4, and an inappropriate or illogical approach rates 0-2. Circle the appropriate number for your rating.

Supporting Comments:

Q5. Productivity

0 1 2 3 4 5 6 7 8 9 10

With respect to the resources available: 9-10 indicates high impact, exceptional output, 7-8 indicates significant results at an extensive rate, 5-6 indicates interesting results at a reasonable rate, 3-4 indicates marginal output, and 0-2 denotes little evidence of progress. Circle the appropriate number for your rating. If the project has not been under way long enough to be rated for productivity, so state.

Supporting Comments:

Q6. Probability of Success

State your estimate of the probability of success of this project accomplishing its stated

objectives. Circle the appropriate number for your rating.

Low High

0 1 2 3 4 5 6 7 8 9 10

Supporting Comments:

OVERALL PROJECT EVALUATION

0 1 2 3 4 5 6 7 8 9 10

An outstanding project rates 9-10. A strong project deserving of priority continuation rates 7-8, while a good project, deserving of continuation, that may have some shortcomings which can be addressed by the Principal Investigator rates 5-6. A weak project, or one with some deficiencies requiring program management attention rates 3-4, and a poor project with serious deficiencies which warrants close reevaluation by program management rates 0-2. Circle the appropriate number for your rating.

Supporting Comments:

FORM 2 Reviewer # _____

Panel/Project: _____ Date of Review:_____

REVIEWER SELF-RATING

1. Please rate your knowledge in the scientific/technical research area or discipline covered in this project.

Novice Understand Knowledgeable Expert

0 1 2 3 4 5 6 7 8 9 10

1-B-iv. EVALUATION FORMS FOR EXISTING PROGRAMS - LONG FORM (ONR, circa mid-1990s)

PROGRAM EVALUATION FORM

TITLE OF PROGRAM.....

REVIEWER NAME.....

.....

1A. RESEARCH MERIT (CIRCLE ONE NUMBER OR -)

1-----2-----3-----4-----5-----6-----7-----8-----9-----10

*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****

.....

1B. RESEARCH APPROACH/ PLAN/ FOCUS/ COORDINATION

1-----2-----3-----4-----5-----6-----7-----8-----9-----10

*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****

.....

1C. MATCH BETWEEN RESOURCES AND OBJECTIVES

1-----2-----3-----4-----5-----6-----7-----8-----9-----10

*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****

.....

1D. QUALITY OF RESEARCH PERFORMERS

1-----2-----3-----4-----5-----6-----7-----8-----9-----10

*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****

.....

1E. PROBABILITY OF ACHIEVING RESEARCH OBJECTIVES

1-----2-----3-----4-----5-----6-----7-----8-----9-----10

*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****

.....

1F. PROGRAM PRODUCTIVITY

1-----2-----3-----4-----5-----6-----7-----8-----9-----10

*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****

.....

2A. POTENTIAL IMPACT ON MISSION NEEDS (RESEARCH/ TECHNOLOGY/ OPERATIONS)

1-----2-----3-----4-----5-----6-----7-----8-----9-----10

*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****

.....

2B. PROBABILITY OF ACHIEVING POTENTIAL IMPACT ON MISSION NEEDS

1-----2-----3-----4-----5-----6-----7-----8-----9-----10

*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****

.....

2C. POTENTIAL FOR TRANSITION OR UTILITY

1-----2-----3-----4-----5-----6-----7-----8-----9-----10

*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****

.....

2D. PHASE OF R&D (DOD TERMINOLOGY)

6.1----------6.2----------6.3

BASIC RES *APPLIED RES** **EXPLORATORY DEV.* *ADV DEV***

.....

3. REVIEWER'S EXPERTISE IN THE RESEARCH AREA OF THIS PROGRAM

1-----2-----3-----4-----5-----6-----7-----8-----9-----10

*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****
.....
4. O V E R A L L P R O G R A M E V A L U A T I O N
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****
.....

EVALUATION CRITERIA FOR EXISTING PROGRAMS

SCORING CRITERIA

The evaluation form contains factors generally related to research and naval relevance issues. The scoring bands for all criteria except 2D are identical, and are: 1-2 (LOW); 2.5-4 (FAIR); 4.5-6.5 (AVERAGE); 7-8.5 (GOOD); 9-10 (HIGH). Criterion 2D has its own scoring range defined.

DEFINITIONS OF CRITERIA ON PROGRAM EVALUATION FORM

1A. RESEARCH MERIT - Importance to the advancement of science of thequestion or problem addressed by the program. Consider the technical objectives, potential advancement of state-of-art, and uniqueness ofcontribution.
1B. RESEARCH APPROACH/ PLAN/ FOCUS/ COORDINATION - Quality of process employed to solve the research problem, including the quality and focus of the research plan, definition of research milestones, degree ofinnovation, understanding of field, balance between experiment and theory, and coordination with (or cognizance of) other related programs to minimize duplication or gaps.

1C. MATCH BETWEEN RESOURCES AND OBJECTIVES - Relationship between scientific objectives proposed and total resources requested. Also, adequacy of resources at performer level to ensure 'critical mass' for each performing unit.

1D. QUALITY OF RESEARCH PERFORMERS - Consider publications, honors, and awards, relevant experience, and other less tangible factors which contribute to team quality.

1E. PROBABILITY OF ACHIEVING RESEARCH OBJECTIVES - Probability that the program's research objectives will be achieved.

1F. PROGRAM PRODUCTIVITY - Volume and quality of work produced and relationship of this output to the resources available, costs incurred, and time elapsed since program initiation.

2A. POTENTIAL IMPACT ON MISSION NEEDS - Potential impact of this program on mission research/ technology/ operational needs if successful.

2B. PROBABILITY OF ACHIEVING POTENTIAL IMPACT ON MISSION NEEDS - Probability that the program will achieve its potential mission impact assuming that its research objectives have been met.

2C. POTENTIAL FOR TRANSITION OR UTILITY - Probability that results from this program will be transitioned to or utilized by technical community assuming that its research objectives have been met.

2D. PHASE OF R&D - Level of program development. Scale ranges from basic research (6.1) through exploratory development (6.2) to advanced development (6.3).

4. OVERALL PROGRAM EVALUATION - Single number description of overall program

quality based on all relevant criteria. Provide detailed narrative of pros and cons and any recommendations under COMMENTS.

1-B-v. EVALUATION FORMS FOR PROPOSED PROGRAMS - LONG FORM (ONR, circa mid-1990s)

PROPOSED PROGRAM EVALUATION FORM

TITLE OF PROPOSED PROGRAM.....
REVIEWER NAME.....

.....
1A. RESEARCH MERIT (CIRCLE ONE NUMBER OR -)
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****
.....
1B. RESEARCH APPROACH/ PLAN/ FOCUS/ COORDINATION
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****
.....
1C. MATCH BETWEEN RESOURCES AND OBJECTIVES
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****
.....
1D. BALANCE BETWEEN EXPERIMENT AND THEORY
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****
.....
1E. PROBABILITY OF ACHIEVING RESEARCH OBJECTIVES
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****
.....
2A. MISSION NEED (PROBLEM OR NEED WHICH THIS RESEARCH ADDRESSES)

.....
2B. POTENTIAL IMPACT ON MISSION NEEDS (RESEARCH/ TECHNOLOGY/OPERATIONS)
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****
.....
2C. PROBABILITY OF ACHIEVING POTENTIAL IMPACT ON MISSION NEEDS
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****
.....

2D. POTENTIAL FOR TRANSITION OR UTILITY
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****
.....
2E. PHASE OF R&D (DOD TERMINOLOGY)
6.1----------6.2----------6.3
BASIC RES *APPLIED RES** **EXPLORATORY DEV.* *ADV DEV***
.....
3. REVIEWER'S EXPERTISE IN THE RESEARCH AREA OF THIS PROGRAM
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****
.....
4. O V E R A L L P R O G R A M E V A L U A T I O N
1-----2-----3-----4-----5-----6-----7-----8-----9-----10
*****LOW** ***FAIR*** ***AVERAGE**** ****GOOD**** **HIGH****
.....

EVALUATION CRITERIA FOR PROPOSED PROGRAMS

SCORING CRITERIA

The evaluation form contains factors generally related to research and mission relevance issues. The scoring bands for all criteria except 2A and 2D are identical, and are: 1-2 (LOW); 2.5-4 (FAIR); 4.5-6.5 (AVERAGE); 7-8.5 (GOOD); 9-10 (HIGH). Criterion 2A has no scoring range, and criterion 2E has its own scoring range defined.

DEFINITIONS OF CRITERIA ON PROPOSED PROGRAM EVALUATION FORM

1A. RESEARCH MERIT - Importance to the advancement of science of the question or problem addressed by the program. Consider the technical objectives, potential advancement of state-of-art, and uniqueness of contribution.

1B. RESEARCH APPROACH/ PLAN/ FOCUS/ COORDINATION - Quality of process employed to solve the research problem, including the quality and focus of the research plan, definition of research milestones, degree of innovation, understanding of field, and coordination with (or cognizance of) other related programs to minimize duplication or gaps.

1C. MATCH BETWEEN RESOURCES AND OBJECTIVES - Relationship between scientific objectives proposed and total resources requested. 1D. BALANCE BETWEEN EXPERIMENT AND THEORY - Balance between experiment and theory proposed relative to optimum required to achieve performance targets.

1E. PROBABILITY OF ACHIEVING RESEARCH OBJECTIVES - Probability that the program's research objectives will be achieved.

2A. MISSION NEED - Identify the mission need or problem (operational, technological, research) to which this research relates.

2B. POTENTIAL IMPACT ON MISSION NEEDS - Potential impact of this program on mission research/ technology/ operational needs if successful.

2C. PROBABILITY OF ACHIEVING POTENTIAL IMPACT ON MISSION NEEDS - Probability that the program will achieve its potential mission impact assuming that its research objectives have been met.

2D. POTENTIAL FOR TRANSITION OR UTILITY - Probability that results from this program will be transitioned to or utilized by technical community assuming that its research objectives have been met.

2E. PHASE OF R&D - Level of program development. Scale ranges from basic research (6.1) through exploratory development (6.2) to advanced development (6.3).

4. OVERALL PROGRAM EVALUATION - Single number description of overall program quality based on all relevant criteria. Provide detailednarrative of pros and cons and any recommendations under COMMENTS.

1-B-vi. IDENTIFYING KEY REVIEWER CRITERIA

Background

During the 1980s, a competitive process among all of ONR's claimants was used to select new Accelerated Research Initiatives (ARIs). In the mid to late 1980s, panels of experts external to ONR were used to evaluate these proposed ARIs (Research Options - ROs). From 1986-1990, 105 ROs were evaluated, and the factors which the reviewers evaluated and scored for each RO remained essentially the same. In 1990, the following analysis was made of the reviewers' scores.

Purpose

1. It was decided to analyze the patterns of the scores of these 105 ROs. This analysis woul d have the following benefits:

2. Future ROs could be improved through the feedback of observed trends and patterns to the proposers

3. The evaluation questionnaire could be simplified if some of the factors proved to be unimportant in determining the final score

4. The review process could be altered if different factors were important for different claimants or for different technical areas

5. The development categories (early 6.1 [6.1 is DOD terminology for basic research], late 6.1, etc.) of different claimants' ROs could be checked against the claimants' charters to determine whether these charters were being followed

Overview of Contents

The present document contains an analysis of the panel reviewers' scores. Categorizations of the data base are made to allow parametric studies. The first section of this report contains regressions and correlations of the scoring factors as a function of claimant, winners/losers, technical discipline,

single/multi, size, and Phase of R&D (development category). The purpose of this first section is to identify which factors were important to the reviewers in determining their final score for each RO, and whether these key factors change for different parametric values. The second section of this report contains plots of dollars vs Phase of R&D, as a function of claimant, POM year, technical discipline, RO size, number of claimants proposing the RO, and winners/ losers. The third section of this report contains plots of dollars vs Overall Program Score (OPE - the reviewers' bottom line score), as a function of the same parameters as above.

1. REGRESSION ANALYSIS RESULTS

The factors from the reviewers' questionnaires which are used in the regression analyses are: Research Merit (RM); Research Approach (RA); Match Between Resources and Objectives (MBRO); Balance Between Experiment and Theory (BBET); Potential Impact on Naval Needs (PINN); Potential for Transition or Utility (PTU); Overall Program Evaluation score (OPE); and Phase of R&D (in DOD terminology, research and development category). For the main regression analysis, fifteen different parametric variations were made with the seven factors RM, RA, MBRO, BBET, PINN, PTU, OPE, and one run was made to show intercorrelations among these seven evaluation factors for the total data base. The same type of analysis was performed in each of the fifteen runs.

First, a six factor model was obtained from the multiple regression analysis to predict OPE: (OPE=b0+b1*RM+b2*RA+b3*MBRO +b4*BBET+b5*PINN+b6*PTU). The three independent variables (x1, x2, x3) with the highest regression coefficients (b1, b2, b3) were then used in a three factor model (OPE=b0+b1*x1+b2*x2+b3*x3), and the resultant R-Squared values (R-Squared represents the fraction of the total variability removed by the regression) were compared to determine the effectiveness of a three factor model relative to a six factor model. After the highest R-Squared three factor model was run, the independent variables (x1, x2) with the two highest regression coefficients (b1, b2) were used in a two factor model (OPE=b0+b1*x1+b2*x2). The process was repeated again going to a one factor model (OPE=b0+b1*x1).

In addition to the fifteen cases mentioned above, seven other regressions were run. OPE score was regressed against RO size (where size is the amount of funds requested for the RO's first year) for all ONR, CRP (an ONR unit at the time), and non-CRP; and OPE score was regressed against Phase of R&D for all ONR, CRP, and non-CRP. CRP Physical Sciences ROs were analyzed similarly to the fifteen cases above.

The results of the first fifteen cases are summarized in Table 1 below. Starting from the left-hand side, the first column describes the subdivision of the total RO data base to which the regression applies. The second column contains the value of R-Squared for the six factor model. The third, fourth, and fifth columns contain the three evaluation factors which produce the highest value of R-Squared of any three factor model. These three factors always had the highest regression coefficients in the six factor model, and these factors are shown from left to right in order of descending magnitude of their regression coefficients. The sixth column contains the value of R-Squared for the model which consists of the factors contained in the previous three columns. The seventh and eighth columns contain the two evaluation factors which produce the highest value of R-Squared of any two factor model. These two factors are shown from left to right in order of descending magnitude of their regression coefficients. The ninth column contains the value of R-Squared for the model which consists of the factors contained in the previous two columns. The

tenth column contains the evaluation factor which produced the highest value of R-Squared of any one factor model. The eleventh column contains the value of R-Squared for this one factor model.

TABLE 1

SUMMARY OF REGRESSION RESULTS

1.....2.....3.....4.....5....6....7....8....9....10....11

.....6.....3.....2.....1

.....FAC.....FAC.....FAC.....FAC

.....MOD.....MOD.....MOD.....MOD

CASE.....R^2.....FACTORS...R^2..FACTORS...R^2...FACT...R^2

ALL ONR.....903...RM....PTU...RA....901..RM...PTU..871..RM...783

ALL

WINNING.....866...RM....RA....PTU...863..RM...PTU..824..RM...703

ALL

LOSING.....775...PTU...RM....RA....768..RM...PTU..741..RM...561

PHYS SCI.....899...RM....BBET..RA....888..RM...RA...869..RM...779

ENV SCI.....914...RM....MBRO..PTU...904..RM...MBRO.897..RM...840

ENG

SCI.....971...PTU...RM....RA....960..PTU..RM...953..RM...729

LIFE

SCI.....962...RM....PTU...RA....936..RM...PTU..919..RM...824

CRP.....892...RM....RA....PTU...889..RM...RA...865..RM...777

NRL.....885...BBET..RM....RA....874..BBET.RM...860..BBET.774

NON-CRP.....915...RM....PTU...BBET..904..RM...PTU..891..RM...782

SINGLE

CLAIM.....899...RM....PTU...RA....897..RM...PTU..870..RM...766

MULTI

CLAIM.....975...RM....MBRO..PTU...955..RM...MBRO.954..RM...920

CRP SING

CL.....874...RM....RA....PTU...873..RM...RA...829..RM...709

NRL SING

CL.....885...RM....BBET..RA....873..BBET.RM...859..BBET.770

NON-CRP SING

CL.....910...RM....PTU...BBET..898..RM...PTU..885..RM...776

a. General Results

In all cases examined, with the exception of losing ROs, the values of R-Squared range from about 0.85 to 0.95 for a six factor model. Since an R-Squared value of 1.0 means the regression model precisely explains the data set, the above results mean that the factors selected in the ONR evaluation capture the main considerations used by the reviewers to determine their OPE scores. In all cases examined, the values of R-Squared for a three factor model are within 3% of the values of R-Squared for a six factor model, and usually within 1%. These three factor models consist of RM, RA or one of its surrogates (MBRO, BBET, which used to be included under RA), and except in the Physical Sciences RO case, PTU.

In all cases examined, the values of R-Squared for a two factor model are within 4% of the values of R-Squared for a three factor model, and usually within 2%. These two factor models consist of RM, and either PTU, or RA or one of its surrogates. In all cases, the drop in the value of R-Squared in going from a two factor model to a one factor model ranges from 0.04 to about 0.2, usually averaging about 0.1. The one factor models consist of RM, with the exception of BBET for NRL.

The relatively small gradients in the magnitude of the value of R-Squared in going from a six factor model to a two factor model implies that the reviewers used two, and sometimes three, main factors in deciding the worth of a proposal. The choice of factors differed for claimants, technical areas, etc., but the number of key factors always remained small.

b. Key Specific Results

For the CRP, research considerations (RM, RA) predominate in determining OPE, while for the non-CRP, mission relevance considerations (PTU) play a secondary but non-negligible role relative to RM in determining OPE. This implies that, to some extent, the reviewers are applying weightings to different factors which go beyond the technical discipline under consideration and depend on the proposing organization

For NRL, BBET plays the primary role in determining OPE, and RM plays a secondary but non-negligible role in determining OPE

In the regressions of OPE against RO size, no correlations were observed. Thus, OPE score is independent of RO size.

In the regressions of OPE score against Phase of R&D, no correlations were observed (R-Squared approximately zero). The conclusion is that OPE score is independent of Phase of R&D.

2. PHASE OF R&D ANALYSIS RESULTS

The Phase of R&D factor reflects the reviewers' judgement as to where an RO lies along the 6.1 - 6.2 - 6.3 spectrum. A picture of how all ONR ROs, or subdivisions thereof, are distributed across this spectrum is valuable for understanding whether ONR claimants are following their charters relative to basic/ applied research, and for gaining general insight into the program. Forty nine separate cases were analyzed, and the results are presented as histograms (distributions by discrete bands) of ROs' first year dollars across the different phases of R&D.

The results for the first level ONR categorizations are summarized in Figures 2-A to G. These figures contain distributions (by discrete bands) of Research Options' first year dollars across the different phases of R&D for different parameter combinations. On all of these figures, the top band represents the first year dollar value of Research Options whose panel-averaged Phase of R&D scores placed them in the earliest stages of basic research. The next to the top band contains ROs judged to be in the intermediate stages of basic research. Within the band which bounds basic and applied research (labeled basic/appl), the specific programs above the midpoint of the band are counted as basic research and those below are counted as applied research. As the bands proceed further downward, the research becomes more applied.

.....

.....ALL ONR ANALYSIS-FIGURE 2-A

VERY BASIC.....:xxxxxxxxx

BASIC.....:xxxxxxxxxxxxxxxxxxxxx

BASIC/APPL.....:xxxxxxxxxxx

APPLIED.....:xxxx

VERY APPL.....:x

.....0.....20

.....$M

For ALL ONR, the distribution is reflective of a mission-oriented basic research program, with the highest dollar amplitude in the middle of the basic research region, and a modest dollar amplitude at the upper and lower bounds of the basic research region. About 84% of the total RO funds are in basic research, and the remainder are in applied research. Since the ONR annual guidance to the claimants suggests a basic/ applied research split of about 80% basic and 20% applied, it can be inferred that the claimants are indeed following the guidance for the present case.

.....

.....CLAIMANT ANALYSIS-FIGURE 2-B

.....CRP.....NRL

VERY BASIC...:xxxxxxxxxxx.....:

BASIC.....:xxxxxxxxxxxxxxxxxxxx.....:xxxxxxxxxxxxxxxxx

BASIC/APPL...:xxxxxxxxx.....:xxxxxxxxxxxxxxxx

APPLIED.....:x.....:xxxxxxxxxxxxxx

VERY APPL....:.....:xxx

.....0.....50.....0.....6

.....ARP.....SMALL.CLAIMANTS

VERY.BASIC...:xxxxx.....:xxxxxxxxx

BASIC.....:xxxxxxxx.....:xxxxxxxxx

BASIC/APPL...:xxxxxxxxxxxxxx.....:xxxxxxxxxxxxxxxxxxxx

APPLIED.....:xxxxxxxxxxxxxxxxxxxx.....:xxxxxxxxxxxx

VERY APPL....:xxxxxxxxx.....:xxx

.....0.....4.....0.....3

.....$M.....$M

The CRP's distribution is centered in the basic research region, while NRL's distribution is centered on the basic/ applied research boundary. Since NRL is a full spectrum R&D laboratory, the researchers would probably be intermixed with, or may also be working in, the higher category levels of development. The more applied flavor of the proposed NRL research relative to that of the CRP may be a reflection of the closer ties of the NRL researchers to the ongoing NRL development work, and would also be reflective of more definable transition paths for the research.

Compared to the CRP and NRL, the ARP's (an applied research unit within ONR) distribution is distinctly different, peaking near the center of the applied research region. In particular, the CRP and ARP distributions appear to form a complementary set, overlapping at the basic/applied research boundary. This is a heartening result, for it reflects the separate but tandem missions established for these two organizations. It shows further that the ARP has been able to sustain the precarious position of remaining centered within the applied research region without drifting into exploratory development.

.....

.....<u>TIME TREND ANALYSIS</u>-FIGURE 2-C

.....<u>POM.87</u>.....<u>POM.88</u>

VERY.BASIC...:xxxxxxxxxx.....:xxxxx

BASIC.....:xxxxxxxxxxxxxxxxxxxx.....:xxxxxxxxxxxxxxx

BASIC/APPL...:xxxxxxxxxx.....:xxxxxx

APPLIED.....:xxxx.....:

VERY.APPL....:.....:x

.....0.....15.....0.....22

.....$M.....$M

.....<u>POM.89</u>.....<u>POM.90</u>

VERY.BASIC...:xxxxxx.....:xxxxxxxxxxxxxxxxxx

BASIC.....:xxxxxxxxxxxxxxxxxxxx.....:xxxxxxxxxxxxx

BASIC/APPL...:xxxxxxxxxxxxxxxxxx.....:xxxxxxxxxxxxx

APPLIED.....:xxxxxxxxxxx.....:xxxx

VERY.APPL....:.....:xx

.....0.....13.....0.....13

.....$M.....$M

When POM year is varied, there do not appear to be any time monotonic trends discernible

.....

.....<u>TECHNICAL.DISCIPLINE.ANALYSIS-FIGURE.ES2-D</u>

.....<u>PHYSICAL.SCIENCE</u>.....<u>ENVIRONMENTAL.SCIENCE</u>

VERY.BASIC...:xxxxxxxxxxxxxxx.....:xxxxxxxxxxxxxx

BASIC.....:xxxxxxxxxxxxxxx.....:xxxxxxx

BASIC/APPL...:xxxxxxxxxx.....:xxxxxxxxxxxxxxxxxx

APPLIED.....:xxxx.....:xxx

VERY.APPL....:xx.....:

.....0.....18.....0.....17

.....$M.....$M

.....<u>ENGINEERING.SCIENCE</u>.....<u>LIFE.SCIENCE</u>

VERY.BASIC...:.....:xx

BASIC.....:xxxxxxxxxxxxxxxxxxx.....:xxxxxxxxxxxxxxxxxx

BASIC/APPL...:xxxxxxxxx.....:x

APPLIED.....:xxxxx.....:xxx

VERY.APPL....:x.....:

.....0.....17.....0.....18

.....$M.....$M

The ONR Physical Science ROs are concentrated mainly in the basic research region, with a very modest amount tapering off into the applied research region. The Environmental Sciences ROs appear to have a deficiency in the center of the basic research region. One partial explanation results from the following observations over the past five POMs. The Ocean Sciences/Atmospheric Sciences components of Environmental Sciences tend to be fairly fundamental in nature, and many of them would fit in the top band. However, many Acoustics ROs have been quite sizable, and tend to be more in the direction of applied research. These would probably populate the band on the boundary of basic/applied research.

The ONR Engineering Sciences ROs have an absence of dollars in the most fundamental research band, which also correlates with observations over the past five POMs. The remainder of the Engineering Sciences distribution parallels that of the Physical Sciences ROs very closely. The Life Sciences RO distribution appears almost totally concentrated in the middle of the basic research region.

.....

.....SIZE ANALYSIS-FIGURE 2-E

.....LARGE.ROs.....SMALL.ROs

VERY.BASIC...:xxxxxxxx.....:xxxxxxxxxxx

BASIC.....:xxxxxxxxxxxxxxxxxxxx.....:xxxxxxxxxxxxx

BASIC/APPL...:xxxxxxx.....:xxxxxxxxxxxxxxxxxx

APPLIED.....:xx.....:xxxxxxxxxx

VERY.APPL....:.....:xxx

.....0.....48.....0.....16

.....$M.....$M

By arbitrary definition, large ROs have first year funding greater than $1 million, and small ROs have first year funding less than or equal to $1 million. While the distribution for small ROs is broader than the distribution of large ROs, there appears to be little difference in Phase of R&D, for the distribution means, between the large and small ROs for all ONR, for the CRP, and for the non-CRP.

.....SINGLE VS MULTI-CLAIMANT ANALYSIS-FIGURE 2-F

.....SINGLE.CLAIMANT.....MULTICLAIMANT

VERY.BASIC...:xxxxxxxxxx.....:xxxxxxx

BASIC.....:xxxxxxxxxxxxxxxxxxxx.....:xxxxxxxxxxxxxxxxxx

BASIC/APPL...:xxxxxxxxxxx.....:xxxxxxxxxxxxxx

APPLIED.....:xxxxx.....:xxx

VERY.APPL....:x.....:

.....0.....44.....0.....15

.....$M.....$M

The ONR single and multi-claimant distributions appear to have about the same means. The bands on both extremes of the single claimant distribution are either reduced or eliminated on the multi claimant distribution. Personal observations over the past five POMs lead to the conclusion that the addition of claimants to an RO proposal tends to have the effect of adding 'filters', with extremes being eliminated. Further, because of the diversities in Phase of R&D contributed by each of the claimants, and the requirement that each RO be given only one score for this factor, there tends to be an averaging by the reviewers, a diffusive process which has the effect of 'trimming the wings' of the factor distribution.

.....WINNERS VS LOSERS ANALYSIS-FIGURE 2-G

.....WINNING.ROs.....LOSING.ROs

VERY.BASIC...:xxxxxxxxxx.....:xxxxxxxxxxx

BASIC.....:xxxxxxxxxxxxxxxxxxxx.....:xxxxxxxxxxxxxxxxxx

BASIC/APPL...:xxxxxxxxxxx.....:xxxxxxxxxxxxxx

APPLIED.....:xxxx.....:xxxxxx

VERY.APPL....:.....:xx

.....0.....44.....0.....15

.....$M.....$M

***Phase of R&D score appears to have no discernable impact on whether an RO will win or lose, for ONR as a whole, or for the CRP. Phase of R&D may have a slight influence on whether a non-CRP RO will win or lose, but this may be due to some other factor which is highly correlated with Phase of R&D.**

.....

3. Overall Program Evaluation Score Analysis

OPE is the factor which has the strongest influence on the final RO score. Study of the distribution of dollars among the OPE scoring bands for all ONR ROs, or subdivisions thereof, can identify strengths or weaknesses in various components of the program. Forty nine separate cases were analyzed, and the results are presented as histograms (distributions by discrete bands) of ROs' first year dollars across the different OPE scoring bands.

The results for first level ONR categorizations are summarized in Figures 3-A to G. These figures contain distributions (by discrete bands) of Research Options' first year dollars as a function of Overall Program Score for different parameter combinations. On all of these figures, the top band represents the first year dollar value of Research Options whose panel consensus Overall Program Evaluation Scores placed these ROs in the Fair-Average category. The next band to the top can be viewed as Average-Good; the next band below can be viewed as Good-Very Good; and the bottom band can be viewed as High or Outstanding.

.....

.....ALL ONR ANALYSIS-FIGURE 3-A

FAIR/AVER.....:xx

AVER/GOOD.....:xxxxxxx

GOOD/VERYGOOD...:xxxxxxxxxxxxxxxxxxxxx

HIGH.....:xxxxx

.....0.....82

.....$M

For all ONR proposed ROs, the bulk are in the Good - Very Good range, which corroborates personal observation over the past five POMs. The proposed ROs which come from the claimants for the overall competition typically have not been reviewed formally by expert external panels. It is conjectured that a rigorous pre-review by external expert panels convened by the claimants would filter out the Fair-rated and most of the Average-rated ROs.

.....

CLAIMANT ANALYSIS-FIGURE 3-B

.....CRP.....NRL

FAIR/AVER.....:x.....:x

AVER/GOOD.....:xxxxx.....:xxxxxxxxx

GOOD/VERYGOOD..:xxxxxxxxxxxxxxxxxxxx.....:xxxxxxxxxxxxxxxxxx

HIGH.....:xxxxxx.....:xx

.....0.....64.....0.....11

.....$M.....$M

.....ARP.....SMALL.CLAIMANTS

FAIR/AVER.....:xxxxx.....:xxxxxxxxxxxxxxxx

AVER/GOOD.....:xxxxxxxxxxxxxxxx.....:xxxxxxxxxxxxxxxxxx

GOOD/VERYGOOD..:xxxxxxxxxxxxxxxxxxxx.....:xxxxxxxxxxxxxxxxxxx

HIGH.....:xxxxxxx.....:.....

.....0.....5.....0.....3

.....$M.....$M

The CRP distribution is very similar to that of the total ONR, with the exception that there are slightly less dollar fractions in the two lower score bands. The major differences between the CRP and NRL distributions seem to be that the CRP has a higher dollar fraction

in the Outstanding band and the NRL has a somewhat higher dollar fraction in the Average-Good band.

.....

.....TIME TREND ANALYSIS-FIGURE 3-C

.....POM.87.....POM.88

FAIR/AVER.....:xx.....:

AVER/GOOD.....:xxxxx.....:xxxxxxx

GOOD/VERYGOOD....:xxxxxxxxxxxxxxxxxxxxx.....:xxxxxxxxxxxxxxxxxx

HIGH.....:xxxxxxxxxx.....:xx

.....0.....18.....0.....25

.....$M.....$M

.....POM.89.....POM.90

FAIR/AVER.....:x.....:xxxx

AVER/GOOD.....:xxxxxxxxx.....:xxxxx

GOOD/VERYGOOD.....:xxxxxxxxxxxxxxxxxxxxx.....:xxxxxxxxxxxxxxxxxxx

HIGH.....:xxxxx.....:xxxx

.....0.....20.....0.....20

.....$M.....$M

***There do not seem to be any major observable trends with time, and the main common feature among the different POM year results is that the highest proportion of ROs are scored in the Good-Very Good band. Unfortunately, no method appears to have been discovered for eliminating proposals in the Fair-Aver band or improving the overall average quality of a POM year's proposals.**

.....

.....TECHNICAL DISCIPLINE ANALYSIS-FIGURE 3-D

.....PHYSICAL.SCIENCE.....ENVIRONMENTAL.SCIENCE

FAIR/AVER.....:x.....:xxxx

AVER/GOOD.....:xxxx.....:xxxxxxxxxxxxx

GOOD/VERYGOOD.....:xxxxxxxxxxxxxxxxxxxxx.....:xxxxxxxxxxxxxxxxxxx

HIGH.....:xxxx.....:xxxxx

.....0.....32.....0.....19

.....$M.....$M

.....ENGINEERING.SCIENCE.....LIFE.SCIENCE

FAIR/AVER.....:x.....:xxxx

AVER/GOOD.....:xxxxx.....:xxxxxxx

GOOD/VERYGOOD.....:xxxxxxxxxxxxxxxxxxxxx.....:xxxxxxxxxxxxxxxxxx

HIGH.....:xx.....:xxxxxxxxxxxx

.....0.....21.....0.....11

.....$M.....$M

ONR Physical Sciences and Life Sciences distributions are quite similar. Relative to these two distributions, the Environmental Sciences distribution has a greater dollar fraction in the Average-Good band (the other three bands having about the same dollar fraction) and the Life Sciences distribution has a greater dollar fraction in the Outstanding band.

The OPE scores presented here are actual non-normalized panel consensus scores. Each of the technical areas discussed here was nominally evaluated by one or more expert panels. Thus, differences in distributions and mean scores among panels could be due to differences in quality of the proposals, or could be due to differences in how reviewers interpret the definitions of the scoring bands. There has been a normalization done on panel scores for the past three POM years. In the normalization, it is assumed that half the difference between any two panels' mean scores is due to a quality difference in the proposals, and the other half of the difference is due to the relative severity of the panelists in assigning scores. It is the normalized scores which determine the final scores and prioritizations of the proposals. However, personal observations and informal 'shadow' reviews over the past five

POMs confirm the findings of the distributions in this section. Most notably, the Life Science ROs tend to have a few more Outstanding contributors than those of the other disciplines, and the Environmental Science ROs tend to have more of a contribution of Average members.

.....

.....SIZE.ANALYSIS-FIGURE.3-E

.....LARGE.ROs.....SMALL.ROs

FAIR/AVER.....:x.....:xxx.....

AVER/GOOD.....:xxxx.....:xxxxxxxxxxxx

GOOD/VERYGOOD.....:xxxxxxxxxxxxxxxxxxxx.....:xxxxxxxxxxxxxxxxxxx

HIGH.....:xxxxxx.....:xxx

.....0.....58.....0.....24

.....$M.....$M

***The large ROs seem to score slightly higher than the small ROs. However, this may be due to the arbitrary choice of a dividing line between large and small. In the regression section of this report, OPE was correlated with RO size, with no arbitrary dividing lines present, and OPE score was shown to be independent of RO size.**

.....

.....SINGLE VS MULTI-CLAIMANT ANALYSIS-FIGURE 3-F

.....SINGLE.CLAIMANT.....MULTICLAIMANT

FAIR/AVER.....:*.....:***

AVER/GOOD.....:****.....:************

GOOD/VERYGOOD.....:******************.....:********************

HIGH.....:**.....:**********

.....0.....67.....0.....16

.....$M.....$M

The distributions of ONR single and multi claimancy are quite similar, and the means appear about the same. The CRP single and multi claimancy distributions are very similar. While the non-CRP multiclaimant ROs have a higher fraction of Outstanding/Very Good dollars, they also have a higher fraction of Average/Very Good dollars. There appears to be no major difference between the two distributions. The CRP single claimant distribution has a smaller dollar fraction in the lower bands, and a larger dollar fraction in the higher bands, than the non-CRP single claimant distribution. The same holds true for the CRP multiclaimant distribution relative to the non-CRP multiclaimant distribution. Since the CRP is essentially a partner to all multiclaimant ROs (with a few exceptions), if it had the same share of all multiclaimant ROs, the CRP and non-CRP multiclaimant distributions would be identical. The fact that the CRP distribution reflects higher scores than the non-CRP distribution means that the multiclaimant ROs with higher CRP contribution score higher than those with lower contribution.

.....

.....<u>WINNERS VS LOSERS ANALYSIS</u>-FIGURE 3-G

.....<u>WINNING.ROs</u>.....<u>LOSING.ROs</u>

FAIR/AVER.....:.....:xxxxxxxx

AVER/GOOD.....:xxx.....:xxxxxxxxxxxxxxxxxxx

GOOD/VERYGOOD.....:xxxxxxxxxxxxxxxxxxxxx.....:xxxxxxxxxxxxxxxxx

HIGH.....:xxxxxx.....:

.....0.....67.....0.....18

.....$M.....$M

***The bulk of the winning ONR ROs are in the Good range or higher; the bulk of the losing ROs are below the Good range, and there is some overlap. It should be noted that the next to the bottom band contains ROs whose OPE scores range from 7.0 to 8.5. Personal observations over the past five POMs lead to the conclusion that there is a substantial difference between ROs at the upper end of this range and at the lower end. Most of the losing ROs in this range scored at the lower end. There is a small fraction of winners in the Average-Good band. These are un-normalized scores; some of the final scores were increased due to the normalization procedure. Also, in different POM years, the threshhold values for funding ROs differed.**

1-B-vii. TECHNICAL/PROGRAMMATIC ISSUES FOR APPLIED RESEARCH

PROGRAM REVIEW
(ONR, circa mid-1990s)

A) TECHNICAL ISSUES

1. FOR EACH COMPONENT OF THE APPLIED RESEARCH PROGRAM, ADDRESS THE FOLLOWING:

a. WHAT ARE THE TECHNICAL OBJECTIVES?

b. WHAT ARE THE KEY TECHNICAL ROADBLOCKS TO BE OVERCOME

c. WHY WAS THE PARTICULAR TECHNICAL APPROACH CHOSEN?

d. WHAT IS THE FEASIBILITY OF THE TECHNICAL APPROACH FOR ACHIEVING THE TECHNICAL OBJECTIVES?

e. IDENTIFY THE PROGRESS AND ACCOMPLISHMENTS MADE TOWARD ACHIEVING THE OBJECTIVES.

f. IDENTIFY THE RISK IN ACHIEVING THE OBJECTIVES.

g. WHAT ARE THE PROJECTED CAPABILITIES THE COMPONENT WILL PROVIDE AND HOW WILL THEY CONTRIBUTE TO THE TOTAL PROGRAM; HOW DO THESE CAPABILITIES COMPARE WITH THE STATE-OF-THE-ART AND WITH POTENTIAL CAPABILITIES OF OTHER TECHNICAL APPROACHES?

h. WHAT MORE FUNDAMENTAL RESEARCH RESULTS ARE UTILIZED TO INSURE SUCCESS OF THE PROGRAM? IF NEEDED FUNDAMENTAL RESEARCH INFORMATION IS NOT AVAILABLE, WHAT FALLBACK POSITIONS EXIST?

2. IF THE PROGRAM OBJECTIVES ARE ACHIEVED, WHAT IS THE PROBABILITY THAT THE INDIVIDUAL COMPONENTS AND/OR THE TOTAL PROGRAM ARE TRANSITIONABLE. WHAT IS THE EVIDENCE TO SUPPORT YOUR RESPONSE.

3. WHAT IS THE LOGICAL STRUCTURE AND PROGRESSION OF THE TEST PROGRAM? WHAT VALIDATIONS WILL BE ACHIEVED FROM EACH STEP OF THE TEST PROGRAM, INCLUDING LAB TESTS AND FIELD TESTS?

4. WHAT IS THE TECHNICAL FOCUS OF THE TOTAL PROGRAM? HOW ARE DISCRETE COMPONENTS BEING INTEGRATED INTO A UNIFIED PROGRAM?

5. WHAT IS THE BALANCE BETWEEN RESOURCES AND TECHNICAL OBJECTIVES? IS THE TOTAL PROGRAM SUFFICIENTLY FOCUSED FOR THE RESOURCES, OR IS IT TOO DILUTED AMONG THE DIFFERENT COMPONENTS?

B) PROGRAMMATIC ISSUES

1. WHAT IS THE MANAGEMENT AND WORK BREAKDOWN STRUCTURE OF THE PROGRAM?

2. WHAT ARE THE MILESTONES TO ACHIEVE THE PROGRAM OBJECTIVES; WHAT WILL BE DEMONSTRATED, AND WHEN?

3. WHAT ARE THE CRITICAL PATHS, AND HOW COULD THEY IMPACT THE SCHEDULE?

4. FUNDING DISTRIBUTION BY TASK AND PERFORMER FOR EACH YEAR.

5. CHANGES IN SCOPE FROM ORIGINAL PLANS, AND RATIONALE SUPPORTING THESE CHANGES.

6. PROGRAM SHORTFALLS TO DATE, IMPACT ON OVERALL GOALS, AND PLANS FOR MITIGATION

7. PROGRAM COORDINATION WITH OTHER AGENCIES AND WITH INDUSTRY, BOTH DOMESTIC AND FOREIGN.

8. HOW WOULD THE PROGRAM BE AFFECTED IF THE MONEY WERE SPREAD OVER FOUR YEARS INSTEAD OF THREE YEARS; TWO YEARS INSTEAD OF THREE YEARS; HOW WOULD THIS AFFECT RISK?

EVALUATION CRITERIA FOR APPLIED RESEARCH PROGRAM REVIEW

I) TECHNICAL CRITERIA

PROVIDE COMMENTS ON THE TECHNICAL ISSUES IDENTIFIED ABOVE AND ANY OTHER TECHNICAL ISSUES WHICH YOU FEEL ARE RELEVANT TO THIS PROGRAM. ADDRESS STRENGTHS AND WEAKNESSES, AND INCLUDE RECOMMENDATIONS FOR IMPROVING THE PROGRAM.

II) PROGRAMMATIC CRITERIA

PROVIDE COMMENTS ON THE PROGRAMMATIC ISSUES IDENTIFIED ABOVE AND ANY OTHER PROGRAMMATIC ISSUES WHICH YOU FEEL ARE RELEVANT TO THIS PROGRAM. ADDRESS STRENGTHS AND WEAKNESSES, AND INCLUDE RECOMMENDATIONS FOR IMPROVING THE PROGRAM.

ALTERNATIVE APPLIED RESEARCH PROGRAM EVALUATION FORM

REVIEWER'S NAME_____

1. IS THE INVESTMENT STRATEGY APPROPRIATE FOR AN APPLIED XXXXXXXXXX

RESEARCH PROGRAM? WAS THE PRIORITIZATION AND ALLOCATION OF RESOURCES AMONG RESEARCH COMPONENTS SUPPORTED BY A LOGICAL RATIONALE? IS THERE AN APPROPRIATE BALANCE BETWEEN REQUIREMENTS-DRIVEN (TOP-DOWN) AND OPPORTUNITIES-DRIVEN (BOTTOM-UP) APPLIED RESEARCH IN THE PROGRAM? HOW CAN VERTICAL INTEGRATION WITHIN THE PROGRAM BE IMPROVED?

2. FOR EACH RESEARCH COMPONENT OF THE XXXXXXXXXXXX RESEARCH PROGRAM, ADDRESS THE FOLLOWING:

2a. ARE THE TECHNICAL OBJECTIVES CLEAR AND RELATED TO THOSE OF THE TOTAL PROGRAM?

2b. ARE THE KEY TECHNICAL ROADBLOCKS TO BE OVERCOME IDENTIFIED?

2c. IS THE PARTICULAR TECHNICAL APPROACH CHOSEN APPROPRIATE?

2d. IS THE TECHNICAL APPROACH FOR ACHIEVING THE TECHNICAL OBJECTIVES FEASIBLE?

2e. ARE THE PROGRESS AND ACCOMPLISHMENTS MADE TOWARD ACHIEVING THE OBJECTIVES ACCEPTABLE?

2f. ARE THE RESEARCH TECHNICAL QUALITY AND PRODUCTIVITY SUFFICIENT?

2g. IS THE RISK APPROPRIATE IN ACHIEVING THE OBJECTIVES.

2h. ARE THE PROJECTED CAPABILITIES THE COMPONENT WILL PROVIDE AND CONTRIBUTE TO THE TOTAL PROGRAM ADEQUATE; HOW DO THESE CAPABILITIES COMPARE WITH THE STATE-OF-THE-ART AND WITH POTENTIAL CAPABILITIES OF OTHER TECHNICAL APPROACHES?

3. IF THE PROGRAM OBJECTIVES ARE ACHIEVED, WHAT IS THE PROBABILITY THAT THE INDIVIDUAL COMPONENTS AND/OR THE TOTAL PROGRAM ARE TRANSITIONABLE? WHAT IS THE EVIDENCE TO SUPPORT YOUR RESPONSE?

4. WHAT IS THE TECHNICAL FOCUS OF THE TOTAL PROGRAM? HOW ARE DISCRETE COMPONENTS BEING INTEGRATED INTO A UNIFIED PROGRAM?

5. WHAT IS THE BALANCE BETWEEN RESOURCES AND TECHNICAL OBJECTIVES? IS THE TOTAL PROGRAM SUFFICIENTLY FOCUSED FOR THE RESOURCES, OR IS IT TOO DILUTED AMONG THE DIFFERENT COMPONENTS? IS THERE AN APPROPRIATE BALANCE AMONG ANALYSIS, THEORY, COMPUTER MODELING, LAB TESTING, FIELD TESTING, AND HARDWARE DEVELOPMENT?

6. IS THE PROGRAM COORDINATION WITH OTHER FEDERAL AND STATE AGENCIES AND INDUSTRY (AND FOREIGN, IF APPLICABLE) ADEQUATE? IS THERE SUFFICIENT LEVERAGING OF THESE LARGER EXTERNAL PROGRAMS?

PROVIDE COMMENTS ON THE TECHNICAL ISSUES IDENTIFIED ABOVE AND ANY OTHER TECHNICAL ISSUES WHICH YOU FEEL ARE RELEVANT TO THIS PROGRAM.

ADDRESS STRENGTHS AND WEAKNESSES, AND INCLUDE RECOMMENDATIONS FOR IMPROVING THE PROGRAM.

1-C. Metrics for Peer Review of Advanced Technology Development

1. EXECUTIVE SUMMARY

The science and technology (S&T) programs sponsored by the United States Department of the Navy (DoN) are divided into three major budget categories:

1) Basic Research (6.1)
2) Applied Research (6.2)
3) Advanced Technology Development (6.3)

In 1999, DoN commissioned an internal review of the 6.3 program. A thirty-one member review panel met for one week to rate and comment on six evaluation criteria (Military Goal, Military Impact, Technical Approach/ Payoff, Program Executability, Transitionability (to more advanced development/ engineering budget categories or acquisition), Overall Item Evaluation) for each of the fifty-five presentation topics into which the mid-$500 million per year 6.3 program was categorized. This appendix describes the review process, documents insights gained from the review, summarizes key principles for a high-quality S&T evaluation process, and presents a network-centric protocol for future large-scale S&T reviews.

Overall 6.3 Program Results

For the evaluation criteria Military Impact, Technical Approach, Program Execution, Transitionability, and Overall Item Evaluation, distribution functions of numbers of programs vs. rating bands (Low, Medium, High) were presented. No systemic overall 6.3 problems were uncovered.

Programs Related to Future Naval Capabilities (FNC)

In 1999, the naval services had identified twelve FNCs that were deemed as high priority targets for development. For the evaluation criterion Military Goal, the number of programs related to each FNC with strengths of relationships above parametrically-varied thresholds was obtained. In addition, the number of programs related to multiple FNCs was calculated. All 6.3 programs were related to at least one FNC with a strength of relationship of Medium or higher, and 95% of the 6.3 programs were related to at least one FNC with a strength of relationship of High. Some 6.3 programs were related to as many as eight FNCs with a strength of relationship of Medium or higher, and a few 6.3 programs were related to as many as four FNCs with a strength of relationship of High. Having this understanding of inter-relationships will be invaluable in helping the Execution Managers coordinate the program management and output among the IPTs.

Individual Program Results

The panel-averaged ratings for each 6.3 item for the six criteria were generated. These data were used to determine the aggregate relationships noted above. A regression analysis of the five component criteria against the Overall Item Evaluation criterion was performed, to determine

which criteria had the most influence on bottom-line score (Overall Item Evaluation). Two criteria, Military Impact and Technical Approach, provided the bulk of the influence on the determination of bottom-line score. A model consisting of these two criteria predicted the bottom-line score to within two per cent. This is consistent with other large-scale reviews (DOE, 1982; Kostoff, 1997n).

Recommendations for Action

Numerical results were used to place the fifty-five 6.3 items in broad quality categories. Specific actions recommended for each item depended heavily on the comments from the reviewers, with special attention paid to the comments from the user/ customer representatives. In general, no corrective action was recommended for items that had good performance and execution, good transition potential, and strong relation to at least one FNC. Various levels of correction, including termination, were recommended for items that had the following characteristics:

- Insufficient commitment to transition
- ”Core-Program” structure
  -Insufficient FNC focus
  -Insufficient demonstration focus
- Potential for high cost over-run

Insights gained from both the planning and conduct of the review should be of considerable value when conducting future large-scale 6.3-type reviews, and include the following:

1) Provision of detailed programmatic descriptive material to the panelists and audience before the review is very useful; its value could be enhanced by e-mail interchange between the presenter or facilitator and the panelists before the presentations to clarify outstanding issues and allow for more effective use of actual meeting time.

2) Appropriate use of Group-Ware could allow:

- Streamlining the review process with real-time data analysis and aggregation
- Remote reviewer participation, thereby minimizing travel and logistics problems
- More reviewers to participate in the process, producing a more representative sample of the technical community
- Reviewers to be selected for expertise in specific evaluation criteria only, thereby enhancing the credibility of each rating
- Sufficient expertise on the panel such that the Jury function (fully independent decision-making) can be separated from the Expert Witness function (potentially conflicted technical judgment and testimony)

3) When assessing and comparing quality of programs representing multiple disciplines, it is necessary to normalize. Evaluating all programs in one setting is an excellent way to accomplish this objective. Because of the realistic time constraints associated with a single-setting review, depth must be traded off for breadth. This trade-off is acceptable, as long as depth is evaluated by some means during the S&T operational management cycle.

2. OBJECTIVES AND GOALS OF REVIEW

2.1. Background

The science and technology (S&T) programs sponsored by the United States Department of the Navy (DoN) are divided into three major budget categories:

1) Basic Research (6.1)
2) Applied Research (6.2)
3) Advanced Technology Development (6.3)

These categories are reviewed periodically to insure that a high level of technical quality is maintained, and that their constituent programs are relevant and responsive to intermediate and long-term naval services' goals. Typically, the programs within these categories are reviewed either individually or in aggregate about some central technical or mission theme.

2.2. Major Review Objectives

In 1999, DoN commissioned an internal review of the total 6.3 budget category. The objectives of the review were twofold: technical quality control and military relevance quality control for the total budget category.

2.2.1. Technical Quality Control

For the total 6.3 program review, assessing technical quality meant addressing issues such as technical approach and potential payoff relative to alternate technologies, demonstrating achievement of technical targets on schedule and cost, and ability to transition to more advanced development/ engineering budget categories (or acquisition) if demonstration succeeds.

2.2.2. Military Relevance Quality Control

In 1999, the naval services had identified twelve Future Naval Capabilities (FNC) that were deemed as high priority targets for development. It was desired specifically to ascertain the relation between the existing 6.3 program and the FNCs, in order to determine the level of management attention required to insure that the program would evolve seamlessly toward better alignment with the FNCs.

2.3. Review Sub-Objectives

Supporting these two major objectives were four important sub-objectives that drove the timing and structure of the review:

- Identifying systemic problems;
- Identifying FNCs requiring additional management attention;
- Increasing awareness of all DoN S&T stakeholders of technology development criteria important to DoN S&T management; and

- Optimizing the S&T portfolio for total FNC satisfaction.

2.3.1. Identifying Systemic Problems

One sub-objective was to ascertain whether there were any systemic strengths or weaknesses that transcended individual program characteristics, and required higher-level management attention than would be necessary for individual program problems. Attainment of this sub-objective required that the individual programs be evaluated on as common and standardized a basis as possible. This normalization procedure necessitated that the total 6.3 budget category be evaluated in one setting, using common evaluation criteria, with the same panel.

2.3.2. Identifying FNCs Requiring Additional Management Attention

A second sub-objective derived from the management structure instituted to insure S&T program responsiveness to the twelve FNCs. An Integrated Product Team (IPT) was established for each of the twelve FNCs. Each IPT had broad representation from the S&T, requirements, and acquisition communities. Each IPT had the charter of developing S&T programs that would respond to its particular FNC. This second review sub-objective was to ascertain the magnitude and quality of the existing 6.3 program relative to each of the IPTs S&T responsibility areas, as a starting point for relating the total existing 6.3 program to the totality of programs required, and therefore to what new programs had to be established by each IPT. Simply put, this sub-objective was to determine the supply-demand imbalance (if any) of the present 6.3 program for each of the FNCs.

2.3.3 Increasing Awareness of All DoN S&T Stakeholders of Technology Development Criteria Important to DoN S&T Management

A third sub-objective related to the composition of the IPTs, since the membership was drawn from very diverse communities. It was desired to increase the IPTs' awareness of the S&T criteria that are important to DoN S&T management in the development of technology. Toward that end, the IPT Chairpersons were invited to participate directly in the review, and the other IPT members were invited to attend the review as audience.

2.3.4. Optimizing S&T Portfolio for Total FNC Satisfaction

A fourth sub-objective was to insure that technology portfolio development for the total 6.3 program was aimed at optimizing total FNC satisfaction. Achievement of this sub-objective required that the goals of each IPT be presented in one setting in a standardized manner, and the multiple application characteristics of each program be understood and appreciated. These complex interactions between technologies and capabilities also required a single setting for enhanced understanding.

3. STRUCTURE AND CONDUCT OF 6.3 REVIEW

3.1. Ground-rules of Review

A number of ground-rules were established for the 6.3 review at the outset. These rules are summarized below in Table 1.

Table 1. Summary of 6.3 Program Review Ground Rules.

No.	Ground Rule
1	All programs within the 6.3 budget category that received funding in Fiscal Year 2000 (FY00) would be included in the review
2	The taxonomy used for structuring the review presentations would be the most recent one also used for program selection and management
3	For logistics purposes, the review presentations would be limited to one week duration
4	Information Technology Group-Ware would be used where feasible
5	The principles of high quality program review would be followed wherever feasible. These principles have been summarized in the main document.

The main elements of the 6.3 review were:

- presentations of the 6.3 program by the DoN S&T Execution Managers to an evaluation panel,
- ratings and comments by the panelists,
- analysis, interpretation, and recommendations by the review's operational managers, and
- final decisions by DoN S&T senior management.

Within this scenario, the three major foundational blocks were selection of the evaluation criteria, selection of the evaluation panel, and selection of a taxonomy for categorizing presentations.

3.2. Selection of Evaluation Criteria

The prime objectives, as stated above, were to evaluate technical quality and military relevance of the 6.3 budget category, especially relevance to the FNCs. In addition, since the 6.3 budget category has an underlying demonstration and product motivation, it was desired to see how well the individual programs met these hard deliverable targets. Five component criteria were defined to address both the potential technical and military payoffs, and the probability that this potential would be realized. These criteria are:

- Military Goal (relevance of program to military target),
- Military Impact (probability of producing military product),
- Technical Approach (potential technical payoff using specific approach),
- Program Executability (probability that technical targets can be demonstrated on time and budget), and

- Transitionability (likelihood that development would go to higher budget category or to acquisition after successful demonstration).

These were the component evaluation criteria selected. The specific definitions used, and sample evaluation forms, are shown in Appendix 1 (the generic term 'item' used in Appendix 1 refers to the funded technology development represented by each of the fifty-five presentations). In addition to the five component criteria, a sixth 'bottom-line' evaluation criterion (Overall Item Evaluation) was used, as shown on the sample form. The purpose of this overall criterion was to account for any factors that the reviewers thought might be important in evaluating a particular program, but that were not included in the component criteria. As will be shown later, the five component criteria captured all the major factors that were used by the reviewers in arriving at their 'bottom-line' scores.

3.3. Selection of the Evaluation Panel

Evaluation panels for S&T programs are usually of two limiting forms. One type consists of personnel completely external to the program(s) being evaluated, and if such personnel are also experts in the program's technical area, this review is termed a peer review (NRC, 1998; USNRC, 1988). Typically (not always), when peer reviews are used, they tend to focus primarily on detailed technical issues, and secondarily on mission-relevance and management-related issues. The second type consists of personnel associated with the organization that manages the program(s); this review is termed an internal review. Typically (not always), when internal reviews are used, they tend to concentrate primarily on higher level mission-relevance management-oriented issues, and secondarily on detailed technical issues.

It was decided to perform an internal review using naval personnel entirely with some ONR management representation, for the following reason. The second sub-objective described above (Identify FNCs Requiring Additional Management Attention) reflected a transition of the 6.3 program from having a major 'core-like' structure to being much more strongly aligned and focused toward the critical FNCs. This new structure enhances the role of the technology customer/ user in the S&T decision-making process. The panel composition, with its relatively high representation from the requirements community, reflected this shift in emphasis. Also, as will be discussed later, recommendations resulting from the review were strongly influenced by the views of the user community representation on the panel.

In addition, because depth was traded for breadth in the 6.3 review, it was believed to be more important to have personnel represented on the panel that had a breadth focus rather than a depth focus. The panel members were also required to represent a diverse group of naval organizations, since the evaluation criteria spanned areas of authority of different naval organizations.

Four types of reviewers were included in the panel. These were:

- The Executive Steering Committee, the senior managers of the Office of Naval Research (ONR)
- Representatives from the Marine Corps

- Representatives from the DoN S&T resource sponsor (OPNAV 911)
- Advisors
  - Representatives from the Operational Navy organizations responsible for setting requirements.
  - Department Heads from ONR

A total of thirty-one reviewers were on the evaluation panel. Their civilian and military ranks were high-level, mainly civilians drawn from the Senior Executive Service and active military drawn from the Flag (Admiral) level.

3.4. Selection of a Presentation Taxonomy

The FY00 6.3 program was estimated (from the vantage point of FY99) to eventually be between $500 and $600 million. To complete the presentations within one week (a necessary ground-rule due to logistics considerations), about ten presentations per day seemed to be a reasonable limit. There were a couple of options for dividing the 6.3 budget category into separate presentations that would allow sufficient material to be shown for credible criteria evaluation. For the review, it was decided to use the taxonomy by which recent programs were selected and managed. This resulted in fifty-five separate presentations.

3.5. Conduct of the Review

With these foundational review blocks in place, the review proceeded as follows. A letter from the Chief of Naval Research was sent to all the major participants (presenters, reviewers, audience) initiating the review process. The letter included guidelines to the presenters (6.3 program Execution Managers) for generating canonical vugraphs that would address each of the evaluation criteria. The presenters generated the vugraphs (and backup material), and posted password-protected copies on the Internet a few weeks before the review. This allowed the reviewers and audience to become familiar with the fifty-five 6.3 programs before the actual presentations.

In parallel with the dissemination of background material, and logistics to prepare for the actual presentations, a Group-Ware software package was developed to help streamline the review process. This package would document the information flow from data entry of the reviewers' ratings and comments to final display of the results at the Executive Session at the end of the review. Time constraints did not allow a fully tested Group-Ware package to be implemented at the review, and only a portion of the capability was actually utilized. The package that was completed eventually, and processes in which it could be imbedded, offer the capability of a much enhanced peer or internal review approach. The software package is described in Appendix 2. A network-centric review process that would utilize this package, the experience of the 6.3 review and previous reviews, as well as reasonable extrapolations from these experiences, is described in Appendix 3.

The presentation sessions were classified at the SECRET level, and therefore no technical details will be presented in this report. The first segment of the presentation sessions consisted of the Chairpersons of the IPTs describing the scope and objectives of their FNCs. Because of the

synergistic and symbiotic nature of many of the FNCs (e.g., Information Distribution contributes to Missile Defense, Autonomous Operations contributes to Warfighter Protection), exposition of the FNC details in one setting before one audience and one panel allowed each participant to understand 1) the sub-capability inter-relations within each FNC and among the FNCs, and 2) how to best leverage and exploit these inter-relations for maximum aggregate FNC benefit.

For the remainder of the presentation week, the fifty-five Execution Managers presented their programs. The nominal presentation period was twenty minutes for actual presentation, ten minutes for questions and answers, and an additional five minutes for the reviewers to complete the evaluation forms. Some larger and more complex programs required more than twenty minutes, and smaller programs required less than twenty minutes.

Shortly after the review, the panel-averaged numerical results and integrative statistics were e-mailed to all the reviewers. The review managers then performed analyses and interpretations of the numerical results, and summarized the reviewers' comments in preparation for an Executive Session. These comment summaries were sent to the Executive Session audience shortly before the meeting; a summary of all the results was presented at the Executive Session. The final results and recommendations were used by senior DoN S&T management in the planning and budget allocation projections for the future DoN S&T program.

4. RESULTS OF REVIEW/ RECOMMENDATIONS

Because of the classified nature of the review, detailed results will not be presented. Instead, the types of results obtained, and the recommendations for action based on these results, will be outlined. Results were categorized into three types:

1) Overall 6.3 program results
2) Programs related to FNCs
3) Individual program results

4.1. Overall 6.3 Program Results

For the evaluation criteria Military Impact, Technical Approach, Program Execution, Transitionability, and Overall Item Evaluation, distribution functions of numbers of programs vs. rating bands (Low, Medium, High) were presented. No systemic overall 6.3 problems were uncovered.

4.2. Programs Related to FNCs

For the evaluation criterion Military Goal, the number of programs related to each FNC with strengths of relationships above parametrically-varied thresholds was obtained. In addition, the number of programs related to multiple FNCs was calculated. All 6.3 programs were related to at least one FNC with a strength of relationship of Medium or higher, and 95% of the 6.3 programs were related to at least one FNC with a strength of relationship of High. Some 6.3 programs were related to as many as eight FNCs with a strength of relationship of Medium or higher, and a few 6.3 programs were related to as many as four FNCs with a strength of relationship of High. Having this understanding of inter-relationships will be invaluable in helping the Execution Managers coordinate the program management and output among the IPTs.

The 6.3 programs were ranked by strength of relationship to each FNC. At the Executive Session, the principal S&T representative to each IPT discussed the potential role of the strongly related programs to addressing the FNC's goals.

4.3. Individual Program Results

The panel-averaged ratings for each 6.3 item for the six criteria were generated. These data were used to determine the aggregate relationships noted above. A regression analysis of the five component criteria against the Overall Item Evaluation criterion was performed, to determine which criteria had the most influence on bottom-line score (Overall Item Evaluation). Two criteria, Military Impact and Technical Approach, provided the bulk of the influence on the determination of bottom-line score. A model consisting of these two criteria predicted the bottom-line score to within two per cent. This is consistent with other large-scale reviews (DOE, 1982; Kostoff, 1997d).

This result should not be interpreted that the other three component evaluation criteria were unimportant. Rather, construction of a correlation matrix showed that the component criteria were strongly correlated, and the other three component criteria were subsumed under the two dominant criteria (Military Impact, Technical Approach).

For each of the fifty-five 6.3 items reviewed, a short description of the item's objectives and a summarization and integration of comments made by the Review Panel (categorized by the six review criteria) were generated. To arrive at these summary comments, the unabridged comments generated by the reviewers were read, and the main themes and messages were extracted. Where significant differences occurred between reviewers, minority and majority viewpoints were included.

4.4. Recommendations for Action

Numerical results were used to place the fifty-five 6.3 items in broad quality categories. Specific actions recommended for each item depended heavily on the comments from the reviewers, with special attention paid to the comments from the user/ customer representatives. In general, no corrective action was recommended for items that had good performance and execution, good transition potential, and strong relation to at least one FNC. Various levels of correction, including termination, were recommended for items that had the following characteristics:

- Insufficient commitment to transition
- ”Core-Program” structure
  - Insufficient FNC focus
  - Insufficient demonstration focus
- Potential for high cost over-run

5. LESSONS LEARNED FROM REVIEW

There were many lessons learned from all phases of the 6.3 review, including the planning and consideration of alternative approaches, the conduct of the actual 6.3 review, and the post mortem analysis of the review's results and processes. Five of the major lessons will be described in this section. These lessons include:

1) value of performing a total S&T budget category review in one setting;
2) differences between 6.3 review and 6.1/ 6.2 reviews;
3) understanding effective use of information technology in program reviews;
4) value of adequate background material and review preparation, and
5) improving match between reviewer expertise and specific evaluation criteria requirements.

5.1. Value of Performing a Total S&T Budget Category Review in One Setting

There are two limiting cases by which an assemblage of programs can be reviewed. One method is to review the assemblage as a group, the other is to review the programs individually. Group reviews allow comparisons to be made across programs, but two compromises are necessary in real-world logistics-limited environments. Breadth is covered at the expense of depth, and the reviewer expertise per program will be smaller. Countering these compromises is the excellent normalization obtained with a single panel in a single setting. Individual reviews allow more in-depth assessment, and more specialty-focused reviewers. In addition, for a vertically-structured organization such as DoN S&T, individual program reviews (e.g., one 6.3 program) allow the other members of the vertical structure (e.g., related 6.1 and 6.2 programs) to be reviewed as well.

The typical DoN S&T review examines sub-groups of programs, usually spanning budget categories. The total 6.3 review showed that there was equal value in examining the total budget category at one setting, because of the comparative value. Selection of individual vs. group review of programs should depend on the overall review's objectives. An interspersing of both types of reviews over an organization's operational cycle is probably optimal. Neither approach is intrinsically superior.

5.2. Differences between 6.3 Review and 6.1/ 6.2 Reviews

Fundamentally, the objectives of reviewing 6.3 are not very different from those of reviewing 6.1 and 6.2. In both cases, military relevance and technical quality are the main drivers. However, while the 6.1 programs aim at achieving enhanced understanding of fundamental processes, the 6.3 programs aim at demonstrating products with desired affordability and performance characteristics. These differences tend to be reflected in the selection of specific criteria for each review type, in how the presentations address those criteria, and in the balance of types of reviewers selected for panel evaluations.

The 6.1 reviews focus on evaluating the advances in knowledge and the research questions answered, using criteria such as research merit, research approach, balance between experiment

and theory, degree of innovation, and potential applications, while the 6.3 reviews use the criteria mentioned previously. The metrics have a different time scale involved. The 6.1 programs have a long-range focus; the 6.1 output metrics (papers, patents, etc) may have a short-term focus, but the 6.1 outcome metrics (benefit-cost ratio, rate of return, dollars saved, quality of life improvements) have a long-term focus. Many times, the 6.1 outcome metrics results can no longer be related to the research managers or performers or programs that they were designed to measure, and their operational utility can be called into question. For 6.3, the outcome metrics are much more closely related in time to the programs, managers, and performers these metrics were designed to measure, and a greater degree of accountability can be obtained from using the 6.3 outcome metrics.

While 6.1, 6.2, and 6.3 review panels all have S&T and customer/ user representation, the differences among panels tend to be in the relative emphasis of representation from the different communities. Across agencies, the 6.1 panels typically consist mainly of scientists and technologists, with some user/ customer representation, while the 6.3 panels typically have a much larger user/ customer fraction.

In those cases where 6.1 programs are reviewed with their 6.2 and 6.3 counterparts, as part of a larger vertical structure review (e.g., ONR's Department reviews), the panels tend to be relatively balanced with respect to community participation. These types of vertically-integrated structure reviews tend to be very informative, with substantial exchange of cross-category information. Any 'impedance mis-matches' across categories are easily detected, and corrections can be readily recommended that will maximize vertical structure quality, as opposed to maximizing single category quality.

To repeat, single category and vertically-integrated structure reviews each have a unique role to play in an organization's overall strategic management process, and these roles depend on the review's specific objectives.

5.3. Understanding Effective Use of Information Technology in Program Reviews

One point became crystal clear in selecting appropriate information technology to support the review process. The following sequence should be obeyed religiously: Review objectives determine the metrics to be used; metrics determine the data to be gathered; metrics and data determine the types of reviewers selected; and metrics and data and reviewers jointly determine the process and supporting tools to be used. In particular, the Group-Ware selected should support the process and objectives, not drive them as is the all too familiar case in practice today. Furthermore, the Group-Ware needs to be specifically tailored to the process and objectives selected. The Group-Ware needs to be an integral component of the operational process, just as a particular scalpel serves as an integral component of a surgeon's repertoire. Efficient use of Group-Ware in the context of a network-centric review process (see Appendix 3) is discussed in Appendix 2.

5.4. Value of Adequate Background Material and Review Preparation

A major purpose of providing background material to all review participants before the presentations, especially to the review panel, is to insure that each participant will have a threshold level of understanding about each aspect of each program. A balance needs to be reached between the amount of material provided, and the amount that will be read by the reviewers. This balance will affect the structure of the material.

The 6.3 reviewers and audience were provided draft copies of the vugraphs to be presented at the actual review, about a week before the presentations. The vugraphs were posted on a password-protected Web site, and any other supportive material the presenters believed was important was added to the Web site as well. This background material proved adequate for the intended purpose. In other program reviews, the first author has tended to provide two or three page narrative summaries for each program component to be presented. For example, if a $40 million Aircraft program review consists of presenting eight $5 million Aircraft component briefings (e.g., propulsion, aerodynamics, avionics), then the background material might consist of two or three page narrative summaries for each of the eight component areas, plus perhaps a three page summary of the total Aircraft program. This amount of background material is probably near the limit of what reviewers can be expected to read in traditional presentation-centered reviews, especially when their participation is pro bono, or near pro bono.

However, except for reviewers' time constraints, there appears to be no fundamental reason that much of the evaluation groundwork could not be done prior to the presentations. The Dutch STW (a government S&T sponsoring organization), for example, conducts one type of review entirely by mail (Van Den Beemt, 1991, 1997). If presentations are desired, and if sufficient programmatic material could be sent to the reviewers before the presentations, then much of the evaluation could be completed in advance of the presentations. Use of the new information technology, embedded in a facilitated process that encourages extensive interactions among reviewers and presenters, could enable this groundwork to be performed very efficiently, and not be overly burdensome on reviewers' time. One method for achieving this pre-presentation evaluation, based on experience gained with an innovation workshop [Kostoff, 1999ba and some experiences with other program reviews, is included in the description of a proposed network-centric review process (Appendix 3).

5.5. Improving Match between Reviewer Expertise and Specific Evaluation Criteria Requirements

In the 6.3 review, all the reviewers rated all the evaluation criteria. Yet some of the reviewers had substantial experience in technology development and less in military operations, whereas with other reviewers the converse was true. As a body, the reviewers covered all the evaluation criteria quite well with their aggregate expertise.

While the review results would probably be unchanged, it might be more efficient to have each reviewer's expertise matched more closely with each evaluation criterion. This can be accomplished in at least two ways. First, a weighting could be applied to each reviewer's rating for each evaluation criterion, based on the reviewer's expertise relative to that criterion. Second, reviewers could be selected to rate specific criteria only.

The latter approach would probably be most desirable. Because of the large number of individuals that would be required as reviewers, implementation of such an approach has presented logistical difficulties in the past. Use of the new information technology, imbedded in a process that includes extensive interactions before the actual presentations (outlined above), would allow a much closer match between reviewers' expertise and specific evaluation criteria. It would allow the large number of reviewers required to achieve statistical significance for each criterion's ratings to be utilized efficiently.

One method of achieving this desirable match-up is included in the network-centric review process proposed in Appendix 3.

All the above lessons learned from the 6.3 review, lessons learned from other S&T reviews, and reasonable extrapolations therefrom, have been integrated into the proposed network-centric program review process described in Appendix 3. The key features of this network-centric S&T evaluation process are:

- Use of Group-Ware for real-time data entry and summary statistical displays
- Larger representation from technical communities due to logistics management with Group-Ware support
  - a) Use of many reviewers allows separation of Jury function (management decision-making) from Expert Witness function (technical judgment and testimony)
  - b) Use of many reviewers allows selection of reviewers with expertise in specific evaluation criterion for specific technical areas
- Expanded distribution of background material using Internet/ e-mail transmission
- Extensive e-mail interactions and preliminary evaluations before actual presentations
- Potential for completely remote reviews

6. SUMMARY AND CONCLUSIONS

A review of the total DoN S&T FY00 6.3 program was conducted by a senior DoN review panel. The review's purpose was to assess the 6.3 program from the perspectives of military relevance, technical quality, transitionability, and demonstration executability.

6.1. Evaluation Criteria

Five specific component criteria were used by the evaluation panel:

- Military Goal;
- Military Impact;
- Technical Approach/ Payoff;
- Program Executability; and
- Transitionability.

A sixth bottom-line criterion, Overall Item Evaluation, was also used

6.2. Evaluation Panel

The evaluation panel consisted of:

- ONR Executive Steering Committee;
- DoN S&T resource sponsor representatives;
- Marine Corps representatives;
- Advisors
  - 4a) FNC IPT Chairpersons
  - 4b) ONR Department Heads

6.3. Review Components

The major review components were:

1) Situation report presentations to the evaluation panel by the Chairpersons of the twelve FNC IPTs;
2) Technical presentations to the evaluation panel by the Execution Managers of the fifty-five 6.3 items;
3) Ratings and comments by the reviewers for each of the evaluation criteria for each 6.3 item
4) Processing of individual numerical entries to generate panel-averaged ratings, FNC distributions, and overall 6.3 program distributions; and
5) An Executive Session in which the numerical results were presented and placed in the larger FNC context.

6.4 Lessons Learned

Insights gained from both the planning and conduct of the review should be of considerable

value when conducting future large-scale 6.3-type reviews, and include the following:

1) Provision of detailed programmatic descriptive material to the panelists and audience before the review is very useful; its value could be enhanced by e-mail interchange between the presenter or facilitator and the panelists before the presentations to clarify outstanding issues and allow for more effective use of actual meeting time.

2) Appropriate use of Group-Ware could allow
   -Streamlining the review process with real-time data analysis and aggregation
   -Remote reviewer participation, thereby minimizing travel and logistics problems
   -More reviewers to participate in the process, producing a more representative sample of the technical community
   -Reviewers to be selected for expertise in specific evaluation criteria only, thereby enhancing the credibility of each rating
   -Sufficient expertise on the panel such that the Jury function (fully independent decision-making) can be separated from the Expert Witness (potentially conflicted technical judgment and testimony) function

3) When assessing quality of programs representing multiple disciplines, it is necessary to normalize. Evaluating all programs in one setting is an excellent way to accomplish this objective. Because of the realistic time constraints associated with a single-setting review, depth must be traded off for breadth. This trade-off is acceptable, as long as depth is evaluated by some means during the S&T operational management cycle.

7. REFERENCES AND BIBLIOGRAPHY TO APPENDIX 1-C

Altura-BT, "Is Anonymous Peer-Review the Best Way to Review and Accept Manuscripts", MAGNESIUM AND TRACE ELEMENTS, 1990, Vol 9, Iss 3, pp 117-118.

Armstrong, J.S., "Why Conduct Journal Peer Review: Quality Control, Fairness, or Innovation", SCIENCE AND ENGINEERING ETHICS, 3:1, 1997.

ASTEC, "Funding the Fabric - Should Commonwealth Government Competitive Research Granting Schemes Contribute More to Research Infrastructure Costs?", Canberra: Australian Government Publishing Service, 1991.

Brown, E. A., "Conforming the Government R&D Function with the Requirements of the Government Performance and Results Act: Planning the Unplannable? Measuring the Unmeasurable?", SCIENTOMETRICS, 36:3, 1996.

Buechner, Q., "Proposal Costs," JOURNAL OF THE SOCIETY OF RESEARCH ADMINISTRATORS, 5, 47-50, 1974.

Ceci, S. J. & Peters, D., "How Blind Is Blind Review?," AMERICAN PSYCHOLOGIST, 39 (2), 1491-1494, 1984.

Chubin, D. E. and Hackett, E. J., "Peerless Science: Peer Review and U. S. Science Policy", State University of New York Press, Albany, NY, 1990.

Clayson-DB, "Anonymity in Peer-Review - Time for a Change - Comment", REGULATORY TOXICOLOGY AND PHARMACOLOGY, 1995, Vol 22, Iss 1, pp 101-101

Cox-D Gleser-L Perlman-M Reid-N Roeder-K, "Report of the AD-Hoc-Committee-on-Double-Blind-Refereeing", STATISTICAL SCIENCE, 1993, Vol 8, Iss 3, pp 310-317

Delcomyn-F, "Peer-Review - Explicit Criteria and Training Can Help", BEHAVIORAL AND BRAIN SCIENCES, 1991, Vol 14, Iss 1, pp 144-144

DOE, "An Assessment of the Basic Energy Sciences Program", Office of Energy Research, Office of Program Analysis, Report No. DOE/ER-0123, March 1982.

DOE, “Evaluation of Alternate Magnetic Fusion Concepts 1977", DOE/ET-0047, U.S. Department of Energy, Assistant Secretary for Energy Technology, Office of Fusion Energy, May 1978.

Fielder-JH, "Disposable Doctors - Incentives to Abuse Physician Peer-Review", JOURNAL OF

CLINICAL ETHICS, 1995, Vol 6, Iss 4, pp 327-332

Goodstein-D, "Ethics and Peer-Review - Commentary", STEM CELLS, 1995, Vol 13, Iss 5, pp 574-574

GPRA, Government Performance and Results Act of 1993 (PL 103-62), 1993

Greengrass, E, “Information Retrieval: an Overview”, National Security Agency, Report Number TR-R52-02-96, 28 February 1997.

Gresty-MA, "Peer-Review and Anonymity", NEURO-OPHTHALMOLOGY, 1995, Vol 15, Iss 6, pp 281-282

Gupta-VK, "Should Intellectual Property Be Disseminated by Forwarding Rejected Letters Without Permission", JOURNAL OF MEDICAL ETHICS, 1996, Vol 22, Iss 4, pp 243-244

Hall, D., and Nauda, A., "An Interactive Approach for Selecting IR&D Projects," IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT, Vol. 37, No. 2, May 1990.

Hensley, O., Gulley, B., and Eddleman, J., "Evaluating Development Costs for a Proposal to a Federal Agency," JOURNAL OF THE SOCIETY OF RESEARCH ADMINISTRATORS, 12, 35-39, 1980.

Keown-D, "The Journal of Buddhist Ethics - An Online Journal", LEARNED PUBLISHING, 1996, Vol 9, Iss 3, pp 141-145

Kostoff, R. N., "GPRA Science and Technology Peer Review", SciCentral, www.scicentral.com, 1998.

Kostoff, R. N., "Peer Review: The Appropriate GPRA Metric for Research", SCIENCE, Volume 277, 1 August 1997a.

Kostoff, R. N., "Research Program Peer Review: Principles, Practices, Protocols", http://www.dtic.mil/dtic/kostoff/index.html, 1997b.

Kostoff, R. N., "The Principles and Practices of Peer Review", in: Stamps, A. E., (ed.), SCIENCE AND ENGINEERING ETHICS, Special Issue on Peer Review, 3:1, 1997c.

Kostoff, R. N., “The Handbook of Research Impact Assessment,” http://www.dtic.mil/dtic/kostoff/index.html, 1997d. Also, DTIC Technical Report Number ADA296021.

Kostoff, R. N., Science and Technology Innovation, TECHNOVATION. 19:10, 1999a. Also, http://www.dtic.mil/dtic/kostoff/index.html.

Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. "Hypersonic and Supersonic Flow Roadmaps Using Bibliometrics and Database Tomography". JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE. 15 April, 1999b.

Laband, D. N., "A Citation Analysis of the Impact of Blinded Peer-Review", JAMA, 272:2, 1994.

Moran-G, "Ethical Questions About Peer-Review", JOURNAL OF MEDICAL ETHICS, 1992, Vol 18, Iss 3, pp 160-160

Neetens-A, "Should Peer Reviewers Shed the Mask of Anonymity", NEURO-OPHTHALMOLOGY, 1995, Vol 15, Iss 3, pp 109-109.

NRC. “Peer Review in Environmental Technology Development Programs”. National Academy Press. Washington, DC. 1998.

NRC. Evaluating Federal Research Programs: Research and the Government Performance and Results Act. National Academy Press. Washington, DC. 1999.

NRC. Peer Review in Environmental Technology Development Programs. National Academy Press. Washington, DC. 1998.

Nylenna-M Riis-P Karlsson-Y, "Multiple Blinded Reviews of the 2 Manuscripts - Effects of Referee Characteristics and Publication Language", JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION, 1994, Vol 272, Iss 2, pp 149-151

Sutherland-HJ Meslin-EM Dacunha-R Till-JE, "Judging Clinical Research Questions - What Criteria Are Used", SOCIAL SCIENCE & MEDICINE, 1993, Vol 37, Iss 12, pp 1427-1430

USNRC, “Peer Review for High-Level Nuclear Waste Repositories: Generic Technical Position”. NUREG-1297. Washington, DC. U.S. Nuclear Regulatory Commission, 1988.

Van den Beemt, F.C.H.D, "The Right Mix: Review by Peers as Well as by Highly Qualified Persons (Non-Peers)", Australian Research Council Commissioned Report: "Peer Review Process" No.54 (1997), 153-164.

Van den Beemt, F.C.H.D. and Le Pair, C., "Grading the Grain: Consistent Evaluation of Research Proposals", RESEARCH EVALUATION, 1:1, 1991.

Weinberg, A. M., "Criteria for Evaluation, a Generation Later", In The Evaluation of Scientific Research (ed. Ciba Foundation), pp. 3-12.Chichester: John Wiley & Sons, 1989.

8. APPENDICES

APPENDIX 1 TO APPENDIX 1-C - EVALUATION CRITERIA USED IN 6.3 REVIEW

Evaluator Name: **Date: Monday – 2 August**
Evaluator Organization: **Time: 1345**

S&T 6.3 Thrust/ATD/MDD Program Title: Advanced Multi-Function RF System

1) MILITARY GOAL (Enter ONE INTEGER between 1 and 10 for each FNC)

HI			MED			**LO**			
10	9	8	7	6	5	4	3	2	1

FNC		FNC	
Information Distribution		Missile Defense	
Time Critical Strike		Platform Protection	
Decision Support Systems		Expeditionary Logistics	
Autonomous Operations		Warfighter Protection	
Littoral ASW		Capable Manpower	
Total Ownership Cost Reduction		Organic MCM	

(Circle ONLY ONE number for each criterion)

	HI			MED			LO			
1. MILITARY IMPACT	10	9	8	7	6	5	4	3	2	1
2. TECHNICAL APPROACH	10	9	8	7	6	5	4	3	2	1
3. PROGRAM EXECUTABILITY	10	9	8	7	6	5	4	3	2	1
4. TRANSITIONABILITY	10	9	8	7	6	5	4	3	2	1
5. OVERALL ITEM EVALUATION	10	9	8	7	6	5	4	3	2	1

Comments:

6.3 Review Scoring Definitions and Values

1) MILITARY GOAL

How important is the Thrust's 6.3 component or the ATD/Maritime Defense Demonstration to the designated Future Naval Capabilities?

HI - Critical to one or more of the 12 designated Future Naval Capabilities
MED - Addresses one or more of the 12 designated Future Naval Capabilities
LO - Does not address one of the 12 designated Future Naval Capabilities

2) MILITARY IMPACT

What is the Thrust's 6.3 component or ATD/Maritime Defense Demonstration's potential for military capability improvement? What are the products?

HI - Revolutionary
MED - Substantial
LO - Incremental

3) TECHNICAL APPROACH

Why was this approach taken?

HI - Better technical payoff than alternate approaches
MED - Equivalent technical payoff to alternate approaches
LO - Worse technical payoff than alternate approaches

4) PROGRAM EXECUTABILITY

What is the probability that the Thrust's 6.3 component or ATD/Maritime Defense Demonstration's technical targets can be demonstrated at the stated costs and schedule?

HI – Near certainty
MED – Probably
LO – Unlikely

5) TRANSITIONABILITY

What is the probability that the Thrust's 6.3 component or ATD/Maritime Defense Demonstration will result in transition to higher category development or acquisition if successful?

HI – Solid financial commitment by transitionee
MED – Solid support without financial commitment by transitionee
LO – No support (including negative support) by transitionee

6) OVERALL ITEM EVALUATION

What is the bottom-line Thrust's 6.3 component or ATD/Maritime Defense Demonstration's quality score, based on evaluation criteria above and any other criteria deemed important by reviewers?

HI - Revolutionary improvements in military and technology capabilities
MED - Substantial improvements in military and technology capabilities
LO - Incremental improvements in military and technology capabilities

APPENDIX 2 TO APPENDIX 1-C - INTEGRATED GROUP-WARE FOR PROGRAM PEER REVIEW

A2-1) Group-Ware Software System

The main intention in using groupware was to allow electronic collection of data, ratings and comments, that could be used for immediate analysis, documentation, and display. Two groupware systems were considered in preparation for the 6.3 Program Review – the first option was commercially available (Ventana System's Group Systems), whereas the second was developed in-house. Time constraints lead to the use of a hybrid of the two systems.

The commercial groupware system used at the 6.3 Program Review is a proven software, typically used in a voting / rating scenario. The software was networked to several computers, that allowed data entry personnel to input data simultaneously. It also allowed for real-time compilation of data, including basic analysis such as calculated mean values, distribution functions of the ratings, standard deviations, and histogram plots of the voting results. Drawbacks in this groupware system included the limited types of output, and incompatibility with other commercial softwares such as Microsoft (MS) Excel or MS Powerpoint. Output files had to be manipulated by experts to allow further analyses not performed by the groupware system.

A groupware simulating database systems was developed as an alternative. This approach was later tested, and proved to be far more powerful than the commercial system for the specific application due to its flexibility. The groupware system used readily available and internally compatible software (Microsoft ACCESS, Excel, PowerPoint). The database approach could be tailored for any review scenario requiring electronic data collection and instantaneous analysis, documentation, and display. This system could be pre-programmed with user defined requirements, such that only desired / specific outputs or analyses are performed. Outputs could be manipulated in various ways (filtering, sorting, variety of plots, etc.). Numerical ratings and text comments could be automatically documented in a presentable pre-formatted report. Outputs are fully compatible with all word processing and spreadsheet software packages.

One of the premiere features of the developed database system is the ability to develop and tailor graphical user interfaces (GUI), with simple icons to facilitate data entry, and thereby reduce the probability of error. GUIs can also be programmed such that the user can navigate through the program and retrieve and display the desired outputs. This system is now available for use by the FNC IPTs for decision-making processes, or by other users for DoN S&T reviews.

APPENDIX 3 TO APPENDIX 1-C – NETWORK-CENTRIC PEER REVIEW

I) INTRODUCTION

The objective of the proposed network-centric peer review is to evaluate a large ongoing S&T program, using a representative segment of the technical community, and employing whatever information technology is required to substantially enhance the quality of the review. Network-centric peer review uses the power of modern communication networks and information technology to expand greatly the number of people that can participate in real-time peer reviews, and expands greatly the access to data that can support all aspects of peer review. This technology allows diverse review operational modes such as the Science Court to be considered seriously, and allows the jury function of peer review to be independent from the higher conflict potential expert reviewer/ witness function. The operational architecture required for network-centric peer review may differ little from the architecture required for its parent network-centric strategic management. Since all strategic management components need to be integrated for optimal synergistic benefits, implementation of network-centric peer review should occur in parallel with implementation of the other components of network-centric strategic management.

This appendix addresses:

*information technology advances and their potential impact on peer review;
*an implementation procedure for a network-centric peer review process;
*research opportunities for network-centric peer review.

II) INFORMATION TECHNOLOGY ADVANCES

In recent years, advances in computer hardware have resulted in much higher computational speed systems with massive amounts of rapidly-accessible storage space. In parallel with the hardware advances are software improvements that allow organization and 'mining' of the transmitted data, and architecture implementations that allow large networks of disparate data sources (whether sensors, humans, structured databases, or other types) to be linked. With such network architectures readily available, one person can communicate with many individuals at once, and the input from many individuals and data sources can be collected, integrated, and analyzed in real time. The implications for peer review in particular, and for strategic management in general, are enormous. One of the major (justified) criticisms of peer review (and of road-maps, metrics, data mining, information retrieval, S&T planning, S&T evaluation, S&T transitioning, and other strategic management decision support aids) has been that only a small fraction of the relevant communities and available data are being accessed when these decision aids are being exercised. Logistics costs and time delays have limited the magnitude of information and people available to contribute to these decision aids' outputs, especially when time frames approximating real-time are required. Now, the hardware and software in combination with the network architectures, and especially supported by *individuals who understand the relation between the information technology capabilities and the decision aid*

requirements, allow these logistics-based limitations to be removed.

III - POTENTIAL IMPACT OF INFORMATION TECHNOLOGY ADVANCES ON PEER REVIEW

First, the potential impact of information technology advances on the different temporal segments of peer review will be estimated. Then, the potential impact of information technology advances on the different quality principles will be discussed. In the following section, these concepts and estimates will be crystallized and integrated into a proposed network-centric review process.

III -1) Impact on Temporal Segments

This discussion will be based on the assumption that one component of a research program peer review will be a meeting that some, not necessarily all, of the participants will attend. Conduct of a meeting-based research program peer review can be categorized into three stages: a pre-meeting phase, the actual meeting, and a post-meeting phase.

III - 1 - A) Pre-Meeting Phase

The main goal of the pre-meeting phase is to inform and prepare all the participants sufficiently that little time is wasted during the actual meeting phase. Standard peer reviews today allow the various review participants to receive summary background material, to be read by the time of the meeting. An interdisciplinary workshop conducted by the author in December 1997 [Kostoff, 1999a] went one step further. Participants exchanged ideas by e-mail, and all participants were involved in each e-mail. By the time of the meeting, many of the issues had been greatly clarified. However, what could be envisioned in this pre-meeting phase if network-centric peer review were operable, utilizing much of the power of available information technology?

First, a substantially larger amount of data could be made accessible to each review participant, since the network could be structured to allow each node (participant) ready access to every other node (data source/ participant). Second, a substantially larger number of participants could be involved in the review, limited only by the extent of the network architecture. Third, a real time iterative rating, learning, and subsequent presentation modification process could be established. New concepts could be dialogued and improved, presentations could be critiqued and rated preliminarily, and greatly modified for the meeting. Some types of reviews could be conducted entirely without physical presence, whereas those that required an actual meeting would have most of the time-delaying issues examined beforehand. In summary, this phase could accommodate substantially more data and participants than at present, could integrate and analyze this data in real-time, and could provide feedback in a continuous short-turnaround mode. It could also provide a period of reflection and gestation, as concepts became more integrated with the passage of time. How could this network-centric pre-meeting phase be envisioned to affect the next actual meeting phase?

III - 1 - B) Meeting Phase

First, the actual review panel could consist of hundreds or more of experts, some of whom are on-site and the remainder are off-site. All would be linked through the network architecture, and the off-site participants may be video-tele-conferenced to the presentation material as well. These features allow the review process to be decentralized, either partially or fully, and provide much greater flexibility in time and location scheduling. They also allow a greater diversity of reviewers to be used, in technical areas ranging from closely aligned with the focused presentation themes to very disparate disciplines that could contribute innovative insights to the target themes and offer the possibility of real breakthroughs.

All data input would be mechanized, and instantly recorded. Statistical analyses could be performed on the data, at the level of each presentation and integrated over all presentations. This integrative analysis would show how each project's ratings would influence overall rankings and overall parametric criteria, thus placing local decisions in their global context. All the background data, the reviewers' ratings and comments, and other supportive data, would be available instantly to all participants. This latter feature would allow real-time Delphi processes, or modifications of comments and ratings, to be conducted at the end of the presentation period, or in dedicated Executive Sessions. The availability of large amounts of data of all types and large numbers of experts in diverse areas might allow the addition of extra evaluation criteria to be employed usefully, and offer additional perspectives on the S&T being reviewed. What impact could a network-centric meeting process have on the final post-meeting phase?

III - 1 - C) Post-Meeting Phase

The post-meeting phase would have some analogies to the pre-meeting phase, with more focus on integration of new concepts and identification of solutions/ modifications to problem areas identified, stimulated by the intense interactions from the highly efficient meeting phase. Final rankings, comments, and decisions would be obtained iteratively with the availability of the integrated comments and statistics, and a comprehensive integrated report could be assembled from the diverse reviewers effortlessly.

III - 2) Impact on Principles of High Quality

III - 2 - A) Need for Synergy and Integration

In the preface to the high quality principles section, the main theme expounded was that peer review, and the complementary decision aids as well, needed to be an integral component of the overall strategic management process. If peer review, or any of these decision aids, are treated as add-ons or independent entities, the power of these techniques and value to the sponsoring organization are diminished substantially. These techniques are interlocking, their operation is symbiotic, and their benefits are synergistic. For network-centric peer review to achieve its full potential, it must be integrated fully into the network-centric strategic management process. Thus, the requirements for successful operation of network-centric peer review are more severe than for traditional peer review, because the operational targets and potential roadblocks are at a higher level.

For example, if data mining is not performed using all the global data sources available as well

as the human and computer analytic and interpretive capabilities, then a gap will exist in the data available for comparing programs under review with the state-of-the-art. This in turn will affect the use of metrics to gauge the comparisons, and road-maps to show project and technology linkages. The impact of data-deficient peer review on strategic planning will result in greater uncertainty in the planning process and products, and will be translated into greater uncertainty in the project selection, management, and transition processes and products.

Thus, a full-scale network-centric strategic management process must eventually be developed, of which the peer review component is one element. However, once the architecture has been established for a network that links the S&T performer/ management/ oversight/ acquisition/ operational/ vendor communities, then

- peer review can be accomplished readily in the network-centric mode,
- road-maps can be easily generated in the network-centric mode,
- planning can be performed efficiently in a network-centric mode,
- multi-discipline multi-category multi-performer multi-user programs can be coordinated and managed effectively in the network-centric mode,
- Integrated Product Teams can conduct planning and operations in a highly decentralized network-centric mode, and
- even marketing and sales can be conducted in a network-centric mode using all the resources of organizations/ nations/ and international communities.

The key point here is that it is the architectural structure, and the inherent logic that links the nodes of the network, that are central to the effective operation of all these seemingly diverse components of strategic management. Once the architecture has been constructed, and the data control established, successful operation of the strategic management tactical elements ceases to be a critical path item.

III - 2 - B) Impact on Specific Principles

The first three principles of high quality peer review listed in Appendix 1 focus on management commitment, incentives, motivation, and statement of objectives. These provide a context, or set the stage, for conducting a high quality peer review, but would not be impacted by the specific tools employed during the review.

The fourth principle, Evaluator Competency, could be impacted substantially by network-centric operation. Three of the critiques related to evaluator competency in peer reviews are:

- that not all technical areas are covered adequately by relatively small panels used in peer reviews,
- even in those covered areas, the sample of the community is too small to be representative, and
- there are many facets of related technical and non-technical areas that the panel does not

cover as a body because of the narrow technical focus.

Network-centric operation would allow many representatives from any technical speciality of interest, representatives from all technical areas involved, and representatives from areas that go beyond the purely technical (users of the technology, impactees, environmental, regulatory, etc.). Because time commitments of reviewers would be reduced due to less need for travel, and because high quality reviewers tend to be busy time-restricted people, more high quality reviewers would be available to participate in the review process, further raising the quality level of the review.

There is another potential benefit related to the Evaluator Competency criterion that deals with the evaluators' operational mode. In the vast majority of traditional S&T peer reviews, the panel has a dual role/ function. It serves as (hopefully) an impartial jury, and serves as an expert witness/ reviewer body as well. This is intrinsically different from the legal system, where the jury and the witnesses/ experts are separate bodies, with separate responsibilities and separate individual requirements. Combining the jury and witness/ expert functions has the potential for serious conflict. The combination problem arises mainly due to the finite panel size, and the logistical inability to handle large numbers of witnesses/ experts in parallel with panel operation.

There have been attempts to conduct peer reviews in which the jury function is executed by one group, and the expert/ witness function is executed by an entirely distinct group (DOE, 1978; Van den Beemt, 1997). The Science Court procedure used by the first author to evaluate competing alternate magnetic fusion concepts is one example (DOE, 1978; Kostoff, 1997d). The first author's experience with the Science Court was that it was a very valuable process, but very time consuming and unwieldy. Network-centric operation would convert the Science Court into a much more manageable and powerful process.

Thus, network-centric operation offers potential benefits in either panel mode of operation. In the case where the panel operates as both the jury and expert/ witness body, network-centric operation expands the number of participants to insure expertise coverage of all criteria. In the case where the jury and witness/ expert body are separate, network-centric operation still insures expert coverage of all criteria, but allows the panel to function as a relatively independent conflict-free jury.

The next principle that could be affected by network-centric operation is Evaluation Criteria. With the expanded access to data allowed by network-centric operation, criteria could be added for which data could be obtained straight-forwardly. For example, suppose knowledge of specific types of impact was an important criterion, but the data by which impact would be evaluated were not readily available. Under traditional peer review, that criterion might not be used, but under network-centric operation, that criterion could be employed due to ready data availability on impact.

The criterion of Reliability would be impacted substantially by network-centric operation. With a large sample from the relevant communities, degree of representativeness is no longer an issue,

and the repeatability of the results over different panels becomes a moot point. In addition, much more data becomes available for incorporation into the evaluation, and statistical representativeness effectively disappears as a data issue.

The Data Awareness criterion would obviously be affected to a large extent. Network-centric operation allows massive amounts of global data to be accessed, filtered, mined, interpreted, and evaluated. Bibliometric analysis capabilities will allow the performers, institutions, and countries that are sponsoring/ performing S&T to be identified, thereby enhancing the potential for leveraging and exploitation, and minimizing the opportunities for excessive redundancy. Along with limited numbers of reviewers, limited access to data is a major deficiency of present day peer reviews that would be overcome by network-centric operation.

The Secrecy criterion could be impacted to some degree. Network-centric operation could allow people at remote sites to participate as reviewers/ expert witnesses without their identity being revealed to other participants in the process. This enhanced anonymity would allow for greater open-ness and frank-ness, ultimately yielding a more useful product.

The Cost criterion would be impacted, due to the reduced travel requirement, and the reduced facilities requirement. Since time commitments would be reduced as well, high caliber typically busy people would be more likely to serve, and a higher quality product would also result concomitant with the lower cost.

IV - IMPLEMENTATION OF A NETWORK-CENTRIC REVIEW PROCESS

IV - 1) Background

The first author has conducted meetings/ reviews that have made some use of network capabilities. These include the review of the Department of the Navy's total Advanced Technology Development program described in the text, and an innovation workshop on Autonomous Flying Systems. The lessons learned from conducting these meetings/ reviews will be integrated with the principles of high quality peer review in Appendix 1 and the network concepts of this appendix to outline an operational implementation for a high quality network-centric S&T program peer review.

The objective of the review is to evaluate a large ongoing S&T program, using a representative segment of the technical community, and employing whatever information technology is required to substantially enhance the quality of the review. For illustrative purposes only, the parameters of the Department of the Navy Advanced Technology Development program review described in the main text will be used in the following discussion.

IV - 2) Definition of Evaluation Criteria

In the proposed network-centric review, after the objectives and goals have been specified, the first operational step would be to define the evaluation criteria. These are the metrics that would

allow quantitative determination of progress toward the goals and objectives. For mission-oriented organizations, there tend to be two over-arching evaluation criteria: mission-relevance and technical quality. For a variety of reasons, including the analysis of progress in achieving sub-goals and objectives, additional supportive criteria tend to be employed in reviews. For the proposed review, assume the same criteria are used as were employed in the Department of the Navy illustrative example: Military Goal; Military Impact; Technical Approach/ Payoff; Program Executability; and Transitionability. In combination, these criteria will help answer the question: Will this program result in a high impact high-quality militarily relevant product with high probability of meeting cost, schedule, and performance targets?

IV - 3) Selection of Review Taxonomy

The second operational step is selection of a taxonomy for the review. *A cardinal rule in assessment is that a program should be reviewed using the same taxonomy by which it was selected and managed.* Otherwise, the program integration (linkages among the program's sub-components) will appear fragmented, even though the sub-components may appear of high quality individually.

A taxonomy is analogous to a mathematical coordinate system, and the requirements for a high quality S&T taxonomy parallel those of a high quality coordinate system. These requirements/ characteristics are:

IV - 3 - A) Orthogonality - a good coordinate system has orthogonal axes, where the inner product between any two axes is zero. This avoids multiple counting and axis redundancy. Similarly, a good taxonomy should have categories as independent as possible.

IV - 3 - B) Completeness - a good coordinate system has sufficient degrees of freedom to cover the full range of dimensionality of the physical problem. A 2-D coordinate system would be insufficient for representing a 3-D problem. Similarly, a good program taxonomy will have a sufficient range of categories to include the different technical disciplines that could occur.

IV - 3 - C) Unit basis vectors - a good coordinate system has the unit vector for each dimension the same size. This avoids resolution mis-matches. In addition, the computational grid size should have adequate resolution to allow computational results to be compared to experimental results. Similarly, a good program taxonomy should include technical disciplines of relatively equal importance with relatively equal amounts of funding, with sufficient category resolution to allow equal levels of coherence about a central theme.

IV - 3 - D) Alignment - a good coordinate system is aligned with the structure of the physical problem. This simplifies the solution by reducing the conversion/ translation between the grid and the structure. A spherical coordinate system is more appropriate to representing a spherical body than a cartesian rectangular system. Similarly, a good program taxonomy should be impedance-matched to data availability.

Assume that these guidelines are followed in taxonomy selection for the proposed review, and a taxonomy of forty categories is defined to represent the total program.

IV - 4) Review Panel Selection

The third operational step is review panel selection. The availability of information technology capabilities will allow the following substantial panel enhancements relative to traditional peer review procedures.

IV - 4 - A) Use of Group-Ware for entering data and computing summary rating statistics in real-time will allow a much larger and more representative segment of the technical community to actively participate in the process;

IV - 4 - B) Having a larger panel will allow the expert witness function and the jury function to be de-coupled, similar to the procedure of the Science Court (DOE, 1978);

IV - 4 - C) Having a larger panel will also allow reviewers to be selected with expertise in a particular evaluation criterion for a specific technical area;

IV - 4 - D) Use of data mining techniques in different literatures will allow a larger pool of experts to be identified as potential process participants.

For the proposed review, assume there is a central panel of perhaps fifteen individuals, and there are one hundred expert reviewers. The fifteen central panelists would not necessarily be expert in any of the areas reviewed, but would be high caliber individuals as free as possible of potential conflict with the programs under review. In the legal analogy, they would serve as the jury. The hundred expert reviewers would be divided equally among the five criteria, or twenty per evaluation criterion. In the legal analogy, they would serve as the expert witnesses. While complete independence from the programs reviewed would be preferable for the expert reviewers, it would not be the absolute requirement used for the fifteen central panelists.

The fifteen central panelists would be selected based on national reputation and absence of conflict. Their function would be to provide final ratings and comments on all the evaluation criteria for all forty programs under review. Their inputs would consist of background material provided by the program presenters, actual program presentations, and preliminary comments and ratings by the one hundred expert reviewers.

Expert reviewer selection would proceed as follows, using the Technical Approach/ Payoff criterion as an example. In parallel with recommendations for experts in the forty technical areas under review, the literature would be ‘mined’ using key phrases that describe the forty technical areas. A large number of reviewer candidates would be obtained. Bibliometrics would be employed to winnow this list through identification of those candidates with extensive publishing and citation records. Other reviewer selection criteria would be employed, to insure that bright younger people, who have not yet established a publication track record, would be included in

the review process. All four of these selection approaches were used to nominate participants for the innovation workshop referred to previously, and have been used in part by the first author for other types of reviews as well.

The twenty candidates selected as expert reviewers for the Technical Approach/ Payoff criterion would have two required output products. They would provide comments and preliminary ratings only on the single evaluation criterion for each of the forty programs. In order not to overwhelm the fifteen central panelists with comments and preliminary ratings from each of the twenty expert reviewers for each of the five criteria for each of the forty programs, one of the expert reviewers for each criterion for each program would be assigned the task of aggregating and summarizing the comments and preliminary ratings for the given criterion and program. To insure a balanced summary is presented from the expert reviewers to the central panelists, another of the expert reviewers for the criterion would have to approve the summary generated by the expert with primary authority. This expert with secondary authority would be selected based on maximum divergence with the viewpoints of the expert with primary authority, to the extent known beforehand. In the illustrative example, each expert reviewer would serve as the primary authority for Technical Approach/ Payoff for two programs, and would serve as the secondary authority for Technical Approach/ Payoff for two other programs.

IV - 5) Operational Review Process

Selection of the goals and objectives, evaluation criteria, review taxonomy, and reviewers, and definition of assignments and responsibilities, establish the structure of the review. The structure, in turn, provides the foundation for the operational review procedure that follows. The complete review process proposed here will consist of three phases: pre-presentation, presentation, post-presentation. The steps emphasized are those in which the use of information technology, especially in the network-centric mode, will enhance the efficiency and quality of the peer review process. Most of the procedures proposed have either been used or tested to some degree by the first author, and their feasibility has been demonstrated.

IV - 5 - A) Pre-Presentation Phase

The objectives of this phase are to provide as much information to all the review participants as is possible before the meeting occurs, and to clarify any outstanding questions and issues. This will allow the participants in the presentation phase to start on a much higher plane, and use the presentation period much more efficiently.

This pre-presentation phase has three distinct sub-phases. First is the distribution of background material. This sub-phase objective is to provide maximal information about the programs to be reviewed and about global efforts in the programs' technical areas and allied disciplines. Since all reviewers are required to provide a preliminary rating on one criterion for every one of the forty programs, this sub-phase will provide the threshold level of understanding about each program necessary for casting an intelligent vote.

The second sub-phase consists of e-mail interaction among reviewers, where comments are exchanged about the program material and issues are clarified. At the end of this sub-phase, each reviewer has transmitted his/ her comments on the assigned evaluation criterion for each of the forty programs to the individuals assigned primary and secondary responsibility for the specific criterion for each program.

The third sub-phase consists of the primary and secondary principals responsible for each criterion for each program writing a brief summary based on the inputs of the other reviewers assigned to each criterion for each program. At the end of this sub-phase, these brief summaries will have been transmitted to the fifteen member central panel, along with the preliminary summary rating statistics for each criterion for each program.

IV - 5 - A - i) Distribution of Background Material

This phase begins with the distribution of background material for the reviewers (and audience, if an audience is desired). In order for the background process to be most effective, the material should be distributed at least three months prior to the actual presentations. Two types of material are proposed.

First are narratives and vugraphs describing in detail the material to be reviewed. The first author distributes this type of background information routinely for S&T peer reviews. Requirements for this material have been detailed elsewhere [Kostoff, 1998]. To maximize distribution efficiency, the material should be made available on the Internet, and the reviewers/ audience informed of its location. If distribution of some of the material has to be restricted for proprietary or other reasons, then the Web site should be password-protected.

The second type of material is information related to the programs to be presented. This material is 'data-mined' from appropriate source S&T databases (e.g., Science Citation Index (basic research), Engineering Compendex (applied research and technology), NTIS Technical Reports (government-sponsored S&T reports), Medline (medical S&T), RADIUS (narratives of on-going government R&D programs). The first author has distributed 'data-mined' information to support reviews of technical areas of modest breadth. This information can be very valuable in identifying the scope of S&T performed globally in the specific technical area under review, in allied areas, and in disparate fields that have some thread of commonality with the specific area under review.

However, even for fields of moderate breadth, substantial effort is required to provide useful background information of this type. The query used has to be refined to satisfy two conditions: the coverage (records retrieved) should be comprehensive (large signal), and have minimal extraneous material (large signal-to-noise). Then, for most recipients, the records retrieved need to be summarized. The first author has used the Database Tomography approach (Kostoff, 1999b) to develop queries with these properties, and to summarize the main pervasive technical themes in such retrieved record databases, and the relationships among these themes. While these computational linguistics and bibliometrics tools help substantially, they do not obviate the

need for technical experts to spend substantial time and effort in developing this background material.

For the illustrative example used in this report, a forty sub-program Advanced Technology Development naval S&T program, the effort required for global data mining of the technical disciplines to be reviewed would be enormous. Nevertheless, if each reviewer's rating is to be meaningful, then the reviewer needs to have some threshold level of understanding about each program reviewed. A substantial effort is necessary to provide such information, especially in summary form.

IV - 5 - A - ii) Individual Reviewer's Comments

The discussion in this sub-section follows the experience of the innovation workshop in Autonomous Flying Systems mentioned previously. Even though the objectives of a workshop are different from those of a peer review, nevertheless, the principles learned from the workshop's pre-presentation phase can be readily extrapolated to peer review application.

In the innovation workshop, each participant sent new concepts relating to the workshop theme to all the other participants by e-mail. An e-mail-based interactive discussion ensued among the participants to 'flesh-out' the concepts, and either clarify and/ or embellish them in preparation for the actual presentations. In order to stimulate this e-mail discussion, a facilitator was required to raise numerous questions. The discussion proved extremely successful in clarifying the concepts, but the need, and effort required, for facilitation of the discussion was appreciated only after the pre-presentation phase had begun.

In this phase of the peer review process, after the reviewers have received the background material, they would be expected to spend the next few weeks digesting the material and clarifying any outstanding or problematic issues. The primary and secondary principals for each criterion for each program would be expected to act as facilitators, to stimulate discussion on these issues. The total review group would not be involved in each e-mail discussion group; this would overwhelm the communication channels. Each e-mail discussion group, in the present example, would consist of the twenty experts for a given evaluation criterion for a given program, plus the individual who will be presenting the information. At the end of this phase, approximately two months before the presentations, each of the twenty experts would provide his/ her comments and preliminary ratings on the given evaluation criterion for the given program to the appropriate primary and secondary principals.

IV - 5 - A - iii) Summary Comments to Central Panel

After receiving the individual comments and preliminary ratings from each reviewer, the primary and secondary principals for each criterion for each program will generate a brief summary for each criterion for each program. If the two principals cannot agree on a specific summary, the secondary principal will contribute a dissenting addendum to the summary transmitted by the primary principal to the central panel. In any case, both the comment summary and a summary

of the preliminary rating statistics are transmitted to each member of the central panel. In order for the central panel members to have time to absorb all the summary material, they would need to receive it no later than one month before the presentations.

In summary, the total pre-presentation time-line is as follows:

*Distribution of background material to expert reviewers - three months before presentations
*Transmission of comments and preliminary ratings to primary and secondary principals - two months before presentations
*Transmission of summary comments and preliminary rating statistics to central panel members - one month before presentations.

IV - 5 - B) Presentation Phase

In network-centric peer review, this phase is optional. There is no fundamental requirement for presentations. All of the review could be conducted through the network by e-mail, Internet, etc. However, there is a cultural aspect to peer review that rivals the information technology aspects in shaping the conduct of the review. Many cultures are not yet at the required comfort level with purely remote operation. In addition, there is value in real-time discourse with the presenters. Therefore, this presentation phase will be included in the present paper.

For the scenario proposed in this paper, presentations will be made to an on-site audience consisting of the fifteen member central panel and the one hundred member reviewer group. Presentations can also be made to a remote audience by video tele-conferencing. Under the present scenario, the role of the remote audience is observation.

All the members of the on-site audience will be linked by Group-Ware. During the presentations, the reviewers will enter final ratings and any additional comments they believe are important based on last-minute observations or insights. At the end of each presentation day, the remote transmission link will be closed, and the reviewers and central panel will meet in Executive Session. The Group-Ware algorithms will have computed each program's statistics (panel averages for each evaluation criterion rating, etc) and any desired integrative statistics over multiple program groups as well. All these numerical results will be displayed graphically to all the on-site audience. The Group-Ware will have also aggregated the additional comments, and these comments will be displayed to all the participants. Both the ratings and the comments will be discussed for each evaluation criterion for each program presented. The central panel will then rate each evaluation criterion for each program presented, and these final program and integrative statistics will be displayed in real-time.

A note about Group-Ware. In the naval Advanced Technology Development review described in the text, Group-Ware was used in part. It had two components: computing summary and integrative statistics, and aggregating comments. Both these features operated in real-time. The immediate summary and integrative statistics feedback provides for high efficiency discussions, and its value increases as the number of programs reviewed and the number of experts used

increase. The comment aggregation is valuable for documentation purposes. For an on-site panel, comment aggregation has little value, can serve to bias reviewers' initial comments, and can be a distraction to some reviewers. For reviewers from remote locations, comment aggregation should prove to be of substantial value.

IV - 5 - C) Post-Presentation Phase

This phase consists of writing the final review report. Depending on the contractual structure of the review, either the staff of the organization sponsoring the review will write the report, or the central panel will write the report. Because of the extensive pre-presentation preparation, the involvement of a large segment of the community, and the extensive interactions that occurred during all prior phases of the review, much of the available information will be ready for direct insertion into the report.

V - RESEARCH OPPORTUNITIES IN NETWORK-CENTRIC PEER REVIEW

Opportunities for research into network-centric peer review abound. Issues to be addressed include the following:

*How is peer review quality defined, especially in a network-centric mode? What are the metrics of quality; how can they be measured? What data is required to quantify these metrics, and how is this data obtained?

*What incentives and rewards have been employed to produce higher quality reviews, and what incentives and rewards should be tested for efficiency?

*What types of network architectures should be developed for optimal review operation? How extensive should the networks be for successful operation? What are the implications of reviewer anonymity protection on the network architectures? What other types of security and verification procedures are required to minimize review disruption and corruption problems? What levels of fault-tolerance need to be incorporated into the network? What are the hardware and software requirements for optimal large-scale operation?

*What are optimal reviewer selection processes, and what are the trade-offs among these processes?

*What are the cost-benefit considerations related to panel sizes, for different types of review objectives? What are the trade-offs of adding more experts in a given technical area for statistical reliability and validity purposes verses broadening the expertise representation across many different fields? How far should the expertise diverge from the target S&T being evaluated, in order to access insights from other disciplines that could benefit the target discipline?

*What are the trade-offs involved in Science Court operation verses dual function jury-witness

panel? What other panel operational modes are possible with network-centric operation? What has been the experience of these other operational modes; what is the potential of other operational modes, whether or not there has been some past history of operation?

*What credible processes exist, or could be devised, to normalize across panels and disciplines? How does network-centric operation complicate or simplify these diverse processes?

*How does the expanded capability of network-centric operation impact the selection of diverse evaluation criteria, and how does it impact the development of, and accession to, the data required to address these criteria?

*How are reliability and repeatability impacted by network-centric operation?

*How should the different types and sources of global data be accessed and integrated with the peer review process? What are the implications on the process operation and results on the availability of these different types of data? What data sources need to be developed and constructed to provide required information for peer reviews, and how does network-centric operation influence the composition and structure of these sources?

*What are the true costs and benefits of network-centric peer review, and what are the main parameters that affect cost-sensitivities? What steps could be instituted now to reduce potential high cost components of the network-centric peer review process?

*How should the larger network-centric strategic management process be constructed in order to maximize benefits from network-centric peer review, as well as optimize benefits organizationally and nationally from the strategic management process? What constraints do the other elements of the network-centric strategic management process place on efficient operation of the network-centric peer review component, and what enhanced capabilities for the peer review component do these other components offer? What are the common elements of all the components of the strategic management process, and what are the unique elements required for network-centric peer review? Are there benefits to constructing architectures that will encompass all the network-centric strategic management components, such that specific requirements for the peer review component will require a minimal additional requirement for resources?

VI - SUMMARY AND CONCLUSIONS

Network-centric peer review uses the power of modern communication networks and information technology to expand greatly the number of people that can participate in real-time peer reviews, and expands greatly the access to data that can support all aspects of peer review. This technology allows diverse review operational modes such as the Science Court to be considered seriously, and allows the jury function of peer review to be independent from the higher conflict potential expert reviewer/ witness function. The operational architecture required for network-centric peer review may differ little from the architecture required for its parent

network-centric strategic management, and since all strategic management components need to be integrated for optimal synergistic benefits, implementation of network-centric peer review should occur in parallel with implementation of the other components of network-centric strategic management.

APPENDIX 2

<u>THE UNDER-REPORTING OF RESEARCH</u> IMPACT [Kostoff, 1998b]

As the federal debt has increased dramatically, competition for federal funds has become more severe. However, the combination of a strong economy and weak inflation in the mid-1990s has kept interest rates low, and has shielded federal funds recipients from the full consequences of the large debt. In the research arena, NSF and NIH research budgets have increased, DOE and DOD budgets have decreased. However, even a one percent rise in interest rates would have a $50 billion dollar yearly impact on the federal budget, and would place all federal funds recipients in much greater jeapordy. A doubling of interest rates or worse, as occurred in the late 1970s/ early 1980s could have disasterous consequences for all federal recipients, especially those with long-horizon benefits such as research.

For research to compete strongly for federal funds, the benefits from research need to receive full accounting and be articulated clearly. The implementation of the Government Performance and Results Act of 1993 (GPRA) [Public Law 103-62] has begun to place even more emphasis on this research accounting requirement. Unfortunately, the present informal 'system' for tracking and disseminating research products and downstream impacts has many deficiencies, resulting in a gross under-reporting of the broad range of research products, benefits and outcomes. Historically, there has been no central mechanism for documenting impacts, and no collective will among the federal agencies and their industrial counterparts to expend the resources necessary for a full accounting of benefits. This problem is compounded by the lack of universal agreement on: the definitions and scopes of research impacts, outcomes, and benefits; the types of studies necessary to ascertain and document these benefits; the total data which would be required to perform these studies and the interpretation of the results of such studies.

Long-term benefits of research are presently tabulated from retrospective studies (e.g., see Kostoff [1997q, Section IV-B], for diverse retrospective study examples and more discussion on the lack of indirect impact accounting]), econometric studies (e.g., cost-benefit), and anecdotal studies (e.g., accomplishments books). Most of the benefits addressed by these studies are direct: evolution of research through development along disciplinary lines. The common thread to the success of almost all the long-term benefit government and corporate studies examined by the author is reliance upon corporate memory. How many research products have "fallen through the cracks" because of corporate amnesia, or with present-day downsizing, corporate lobotomies? While technology to account for these benefits may not have existed in the past, in this day and age of high speed computers with large storage capabilities and intelligent algorithms, the technology now exists to track and identify these research benefits.

Additionally, research intrinsically has multiple impacts on other research and technology through myriad pathways. However, these indirect long and short-term impacts and benefits of research are often overlooked. The indirect impacts tend to cross diverse disciplines, which complicates their tracking; the impact sequence is not necessarily linear from basic research to final product, which further complicates the tracking; and the more sophisticated information technology and databases

required to systematically track these impacts have not existed in the past.

Matrix approaches (e.g., Dean [1972]) can account mainly for forward impacts: the impact of a research program on a variety of technologies, and subsequent impact of these technologies on a variety of systems. While these forward impacts represent only the tip of the iceberg of total research impacts, even these limited matrix approaches are rarely used. Network approaches (e.g., Kostoff [1994i]) can account for forward, lateral, and backward impacts: the impact of a research program on other research programs and other technologies, and subsequent impact of these technologies on other research programs and technologies and systems. Network studies have shown the potential orders of magnitude impact enhancement due to inclusion of these types of indirect impacts [Ibid.]; the massive increase is due to the summation of an extremely large number of modest size indirect impacts. The under-reporting of indirect impacts stems from the lack of data needed for the matrices and networks, from lack of a coordinated research tracking system integral to the research execution and transition process.

This lack of coordination among all the principals in the national research enterprise contributes to poor product and impact accounting procedures throughout the research evolution process, and results in an under-reporting of the full research benefits. This could result (and may have already resulted) in research receiving less funding than is warranted by the full scope of its socially useful benefits and impacts. Research product tracking and monitoring need to be made an integral part of the research planning/ selection/ outlay/ execution/ transition/ evaluation process, and not be treated as an afterthought, as is presently the case.

SCIENCE CITATION INDEX

What type of research product tracking system should be developed? The system should have the capability of tracking long-term research impacts as well as near-term. It should be able to follow indirect impacts of research, as well as direct impacts. The system should be simple to operate, not require substantial resources from the data providers or the system maintainers, and cover as broad a spectrum of development categories and sponsors and users as is possible. For ease of introduction, the system should have some basis in an existing process, where there is a substantial body of operational experience.

One very limited prototype of such a system is the Science Citation Index (SCI). Through its manipulation and tracking of references in papers, it is able to follow the flow of information over time, and the evolution and impacts of research. However, for the research product tracking purposes suggested in this paper, the present structure of the SCI has severe limitations. It is focused on basic and applied research only, and does not span the gamut of research to technology product. It does not contain sponsor information, does not contain funding information, and does not contain unique representations for performers and organizations. Would the credit card companies give identical cards to all the John Smiths in the world; why should the SCI? This latter problem is more than one of appearances. Much sponsor credit can be under-reported because of the errors and ambiguity of performer and organization information (see e.g., Kostoff, [1997e]).

Equally important, even in the case of examining impacts on basic and applied research, there are severe problems with the SCI. These problems stem from the structure of the basic SCI unit, the published peer-reviewed research paper. The typical paper focuses, in priority order, on research approach, research product, and intellectual heritage (references). This focus derives from performer priorities, not sponsor tracking priorities. The completeness of the references, the adequacy of the references, and the relative importance of each reference, are governed by the performer's subjectivity and the limited space available for the paper. In particular, under the present highly competitive climate for research funds, how motivated are researchers to give more credit than absolutely necessary (in print) to the origins of new concepts or paradigms? Thus, the present structure and design of the research paper is not the optimal structure required for tracking.

PROPOSED EXPANSION OF CITATION INDEXES

The SCI can be viewed as a beta test prototype for an expanded system to address the needs of tracking broader research impacts. The proposed system would cover the range from basic research to product development and testing. It would consist of a science tracking component, and a development, engineering, and testing component. It should be viewed as a first step in the improved tracking and documentation of research benefits, not as a final solution. In particular, it is limited to tracking the evolution and technology transfer of that segment of research that has been documented in the open literature, and will therefore not include the tracking of proprietary, classified, and other types of non-published research.

1) Science Component

The science component would be an expanded version of the SCI. It would contain additional journals, sponsor information, funding information (resource expenditures covered by the paper), and would uniquely and unambiguously identify the performers and their institutions. Some idea of relative importance of the references would be provided. There may be other useful information which could be supplied as well. Modification of the SCI in the manner suggested would require the cooperation of the journals as well, since they would have the responsibility of requesting this additional information from the authors. The journals would also be requested to have their peer reviewers assign more importance to the completeness and prioritization of the references, and would transmit this requirement to the authors as well.

2) Development, Engineering, and Testing Component

This component would consist of one or more databases which would have citations and citation tracking similar to the modified SCI proposed above. The documents in these databases would not be limited to refereed published papers; they could include patents, non-refereed reports and published papers, book chapters, and other documents which contain references. Each category could have its own database, or there could be combinations of categories is specific databases.

3) Potential Studies

Construction of such an expanded system is possible now because of the advances made in computer speed, storage, and information manipulation algorithms. Implementation of this expanded citation tracking system would allow long and short-term impacts of research to be followed. These studies would not be a substitute for expert involvement in retrospective studies, but rather would serve as directional maps or guides which allow the experts to identify and probe the different impact pathways. The capabilities inherent in this process would allow the indirect impacts of research to be documented over many pathways, and the full benefits of basic research to be collected and articulated more thoroughly.

APPENDIX 3

UTILITY OF CITATION ANALYSES [Kostoff, 1998c]

Leydesdorff [1998] addresses the history of citations and citation analysis, and the transformation of a reference mechanism into a purportedly quantitative measure of research impact/ quality. Following his lead, the present appendix examines different facets of citations and citation analysis, and discusses the validity of citation analysis as a useful measure of research impact/ quality.

I. CITATIONS

I-a. Citations as Bookmarks

The starting point for this appendix centers around the need for citations. Why are citations used in a paper? There are obviously many reasons for citations, ranging from contributions to the advancement of science and knowledge to less noble purposes for inclusion in text. Some of these reasons will be enumerated in the following paragraphs.

Start with the bookmark function of citations. The average reader of a technical paper typically does not have the luxury to expend large amounts of time on extracting useful information from the paper. The shorter the paper, the greater is the likelihood that it will be read in its entirety. Citations, like acronyms or mathematical symbols or 'laws', provide a condensed reference to a much larger body of data. The relatively few readers who would be interested in such details can examine them at a later date.

One could write a paper including Lotka's law without providing a reference to Lotka's law, or without even mentioning the name 'Lotka's law'. Whenever the need to include Lotka's law arose, one would write out the definition. This unabridged approach to writing would lead to an unnecessarily lengthy document, and would lose the average reader quite rapidly. Using the abridged description 'Lotka's law' allows for an efficiency of presentation. Including such a citation allows the reader to access more details, shows evidence of the author's awareness of other related works, and probably provides more credibility to the paper in the reader's eyes.

I-b. Citations as Intellectual Heritage Linkages

Other than the shorthand function, citations provide links to the intellectual heritage foundation for the citing paper, and help provide the historical context for displaying the unique contributions of the citing paper. While the intellectual heritage linkage role of citations is probably the dominant consideration when viewing citations as a measure of research impact, one needs to be careful on this point of important contributors to intellectual heritage. In the best of all worlds, only a small fraction of all potential intellectual sources will be and can be acknowledged. Especially in any technical field, there are thousands of papers and other sources which have contributed to the intellectual foundation, as there are thousands of bricks which contribute to the support structure of a building's roof. In particular, there may be sources which are not obvious, at least consciously, to

the paper's author. Perhaps a major foundational concept for a paper came from attendance at a seminar or a lunchtime discussion, either of which have escaped the author's memory. Intrinsically, the intellectual attribution process is very incomplete.

Given the finite space allowed in the journals, only a small sampling of the total true intellectual foundation for a paper can be cited, even if all these sources were tangible and identifiable by the author. The selection process used by an author to include a relatively few citations in the bibliography for identifying the intellectual heritage is poorly understood. While some sort of Lotka's law approach is assumed to be at work in selecting only the seminal contributions to the foundation, serious questions exist: what are the selection criteria; what are the cutoff criteria? This uncertainty therefore translates into an undefined role for citations as a measure of intellectual heritage. Some studies [MacRoberts, 1996] have attempted to measure the fraction of intellectual heritage that selected papers included in their bibliographies. While these studies are insightful and useful, the benchmark used (the analyst's perception of what the main intellectual heritage is) is also selective and arbitrary, and limits the utility of such analyses. A more useful approach might be a few case studies where all the references in a sample of published papers are discussed with the authors, and the reasons for inclusion of each reference (and exclusion of other potential references) in the papers are enumerated.

I-c. Citations for Tracking Research Impacts

One critical element of the research management process is identifying and articulating the impacts and benefits of research. This helps convince the research sponsors that there has been (or will be) payoff from their research investment, and provides the rationale for continuing the research investment. However, tracking the impacts of research is notoriously difficult. In the process of having impact, research undergoes a transformation to development and engineering, and is effectively camoflouged. Also, basic research typically has a multiplicity of impacts in diverse fields. Many of these fields are unfamiliar to the researcher and the sponsor, and therefore any impacts far afield from the researcher's discipline go unrecognized.

For basic research, these latter indirect impacts are an important component of the research's total impact [Kostoff, 1994i]. The magnitude of these indirect impacts may be small in many (not all) cases. However, because of the large number of indirect impact pathways, the cumulative effect of all the small indirect impacts resulting from a body of research may be quite large. In fact, in some cases this cumulative effect of indirect impacts could dominate the direct impacts of research [Kostoff, 1994i].

One largely unutilized role of citations is to serve as a 'radioactive tracer' of research impacts. Citations allow the analyst to track the documented flow and evolution of research over time until the linkages to far downstream products can be identified. Citations allow the different types of impacts to be identified as well. For example, the sponsors of mission-oriented research may want to ascertain whether: 1) certain types of technical disciplines are accessing the research products; 2) certain types of organizations, or specified countries, are utilizing the research products; 3) the research is having its initial direct impact on other basic research or applied research or

development. Citations are a documented approach to generating this important diagnostic information.

However, using citations for this diagnostic purpose is much more difficult, complex, and time-consuming than the mainstream application of counting citations for relative impact. The mainstream use of citation counts is algorithm based, and large volumes of data can be processed rapidly to provide copious relative impact results. The tracking application is intrinsically slow and laborious, requiring judgement of the appropriateness and quality of the impact as well as impact quantity. Because of the potential information available from the tracking application, this is a very fruitful area for future citation research and analysis.
Other positive (and negative) uses of citations can be found in MacRoberts [1996] and Kostoff [1997b, 1998b].

I-d. Citations for Self-Serving Purposes

Citations also play other roles, of a less positive (to the advancement of science, anyway) nature. One role is self-aggrandizement, or the ego satisfaction of self-citation for purposes not justified technically. Another role for citations is political. Including citations to journal editors or potential reviewers or 'politically correct' papers will help a paper's chances of being accepted for publication in a specific journal.

Because citations can impact rewards such as promotion/ tenure/ grant consideration, there is a financial self-interest role based on increasing citation volume. This is where 'citation clubs' are formed, and each member cites the other members regularly. Each member has increased citation volume, which eventually translates to more money for each member due to promotions or contracts or other benefits. In addition, there is a potential exclusivity role for citations, whereby they are used mutually among closed groups of researchers to exclude (by sheer volume of citations) competitive concepts which threaten existing mainline infrastructures (see the 'Pied Piper Effect' in section II).

II. CITATION ANALYSIS

II-a. Conclusions from Section I

Section I described some of the many possible uses of citations, including bookmark, intellectual heritage, impact tracker, and self-serving purposes. Since the main published uses of citation analyses tend to focus on absolute and relative measures of impact (and inferred measures of quality), the discussion in this section will concentrate on the applicability of citation analyses as an impact or quality measure. The main message to be derived from section I is that there are many reasons for an individual to select particular references for inclusion in a paper, only one of which is the dominant contribution of citations to research impact, significant intellectual heritage. Trying to draw conclusions about the quality or impact of a specific reference based on one particular paper's list of references is akin to solving the inverse problem in science: there may be many solutions; they are not unique; the correct solution cannot be determined without other information. What meaning, then, can be ascribed to the field of citation analysis and the metric of citation counts if the basic unit

has such associated uncertainty? More importantly, what is the purpose of using such a metric, and why is its use so widespread?

II-b. Expanded Utilization of Quantitative Measures

While there may be many reasons for the growth and utilization of citation analysis, its expanded use stems (from the author's perspective) from the evolution of research sponsorship. Technical research has evolved from a rich man's pastime [Science, 1998] to industrial support to almost exclusive government support. The approaches used by industry to assess the value of basic research were primarily based on economics. Existing economic tools show that basic research, with its short term costs and long-term high risk payoff horizons, could not be justified as economically cost-effective by most industries. Therefore, since research is viewed by society as a necessity, the support for basic research has by default almost exclusively shifted to government.

As the U.S. national debt has increased drastically in the last two decades, competition for scarce funds in the Federal arena has increased substantially as well. Basic research, with its long-term payoff horizon, now has to compete strongly with medicare, welfare, and other service provision and development programs. In Europe and Asia, basic research has undergone a similar transformation, with more of a strategic focus to the research.

In this environment of scarce government funds, accountability of all government programs has increased substantially. There are two major characteristics of this increased accountability: more detailed programmatic information is requested by the program assessors, and more quantified information is requested. The upsurge in computer availability over the past decade has enabled large quantities of detailed information to be stored, tracked, and interpreted, and has driven the request for the large volumes of detailed program information. The request for increased quantitative information also derives from the increased computer capabilities for handling and analyzing large amounts of this type of data. In addition, there is substantial motivation from the assessors to have simple quantitative indicators which could drive the resource allocation process, and substantiate and justify the resource allocation decisions that are generated, rather than use the more complex and expensive and subjective qualitative peer review evaluation processes.

This desire for increased accountability, focused on quantitative measures of research output and impact, counterbalanced by the intrinsic long-term uncertain payoff from research, has produced a dilemma. The simple research outputs, such as published papers and patents, can be easily quantified in the short term. However, they are intermediate measures, not long-tern benefit measures. The quantifiable impacts from research such as societal outcomes or economic payoffs are long-term phenomena and cannot be generated in the short term. Because the research oversight organizations want valid performance metrics applicable to existing research, the question arises whether credible short term proxies for long-term research impacts and outcomes can be defined.

Citation analyses generate relatively short-term quantifiable items, they have the appearance of short-term research impacts, and are therefore attractive candidates as short-term proxies for research impact and perhaps quality. The real question becomes: what, if anything, do they measure?

II-c. Enhanced Value of Aggregating Citations

The previous section showed that any citation, or group of citations, in a particular paper's bibliography does not provide a unique indicator of positive impact of the cited source on the citing paper. Is there any combination of citations possible which could translate into research impact or quality?

Possibly. Consider the following analogy to gas dynamics. Assume there is a flowing gas with gross velocity V and constant temperature T and pressure P. If one examines a group of molecules in the gas, each member of the group will have a different direction and magnitude to its velocity vector. Thus, the aggregate characteristics of the gas cannot be related to the velocity and 'kinetic temperature' of any one molecule. However, by summing over the velocity distribution functions of large groups of molecules (i.e., taking 'moments' of the velocity distribution function), gross gas properties such as V and P and T can be obtained.

In gas kinetics, one way of viewing each component molecule in its relation to the aggregate is to conceptualize the molecule's velocity vector as consisting of a component with mean velocity V (the aggregate velocity) and a component with random velocity. In the summation process used to derive aggregated gas properties, the random component is integrated out, leaving only the mean component V. Can an analogous model be applied to citation analysis?

Possibly. Assume that some, if not most, citations reflect intellectual heritage. For any single paper, the citations which reflect intellectual heritage may not be obvious, and of those citations which do reflect intellectual heritage, the dominant or highest priority ones may not be obvious. However, from the nature of the positive and negative reasons for citing shown above, it appears that the main positive reason (intellectual heritage) for citation impact or quality purposes is tied to or reflective of intrinsic technical considerations, and the negative reasons are related to non-technical self-serving individual characteristics. Thus, if a paper's bibliography is viewed as consisting of a directed (research impact or quality) component related to intellectual heritage and random components related to specific self-interest topics, then for large numbers of citations from many different citing papers, the most significant intellectual heritage (research impact or quality) citations will aggregate and the random author-specific self-serving citations will be scattered and not accumulate.

II-d. Limitations of Citations as Stand-alone Measures of Impact

While corroborations of large numbers of citations with other indicators of substantial research impact and quality have shown general agreement, especially with use of large citing and cited universes, there are at least two limitations to this model of citation analysis for stand-alone use as a measure of research impact or quality. First, the reference to intellectual heritage can be positive or negative. A paper could be highly cited because it contributed to the growth of a field, or it could be highly cited because its flaws were obvious to many people, and they wanted to correct the record. Second, there could be systemic biases which affect the aggregate results, one of which has been termed the "Pied Piper Effect" [Kostoff, 1997q], (See section IV-B-5-v for a brief description of the

Pied Piper Effect; also see Appendix 6 for a more detailed description).

II-e. Early Case Study of Comparative Citations

The present sub-section summarizes a short citation study which eventually led to a citing comparison of some Russian/ American papers in different technical fields. The questions raised in interpreting the data highlight a few of the difficulties in attempting to interpret citation results without supplementary information.

In a 1999 Text Mining study [Kostoff, 1999] of hypersonic/ supersonic flow over aerodynamic bodies, publication and citation distribution functions for different parameters (authors/ journals/ organizations/ countries) were generated. Large numbers of authors/ papers/ journals with relatively few citations each were observed, and a few authors/ papers/ journals with large numbers of citations were seen. Small focused studies were then performed to determine the characteristic differences between highly cited and lowly cited papers in hypersonic flow.

Appendix 3-A-1 (extracted from a larger paper on the study [Kostoff, 1999]) summarizes the results from these focused studies. A key point is that Russian publications tended to populate the poorly cited papers sample, and NASA (U.S.A.) publications tended to populate the highly cited papers sample. To study this Russian/ American difference further, all the papers in the Science Citation Index (SCI) written by the three most prolific Russian authors and the three most prolific American authors in hypersonic/ supersonic flow (names were obtained from the larger Data Mining study) were examined. The results were equally striking. Essentially, the Russian papers in this field are not being cited by the larger technical community, or even the Russian technical community.

Because of these findings, another small focused study on the field of near-earth space was performed. This field was chosen since it had been examined for a previous Text Mining study [Kostoff, 1998]. All English language papers published in 1993 in the SCI (with Russian-Acad-Sci authors only) which contained the word SATELLITE* were selected. Russian-Acad-Sci authors were chosen because they were the most prolific according to the larger space Data Mining study.

There were 29 such papers, of which 16 were both relevant to satellites in space and were written by Russian authors only. For each of the 16 papers, an attempt was made to identify a paper published by American authors only in 1993 which had at least one reference in common with the Russian paper, and had an approximately similar theme. Because of the Related Records field in the SCI, which identifies all records (papers) in the total SCI database which have at least one reference in common with the target paper, pairing (where pairs exist) can be done rapidly. Seven of these pairs were found; unfortunately, there were not always American papers which met the arbitrary criteria used (published in 1993; approximately similar theme; at least one common reference) for pairing with the Russian papers.

Of the 16 relevant Russian papers, 14 had zero cites, one had four cites (two self cites), and one had six cites (two self cites). For the seven pairs of Russian/ American papers, the Russian citation average was 1.4 cites per paper, and the American citation average was about 34 cites per paper (of

which about 6.5 were self cites, or about 20%). Also, for these seven pairs, the Russian median was zero cites per paper, and the American median was 37 cites per paper. This is not a large sample, but the differences are so great that the suspicion exists a large sample would give about the same message.

Finally, a small focused study on fullerenes was performed. All English language papers in the SCI published in 1993/ 1994 which contained the phrase CARBON NANOTUBE* were selected. This is one of the 'hottest' areas of fullerene research. There were 131 such papers, all were relevant to the desired topic. Citation patterns of papers written by Russian authors only and American authors only were examined.

There were 44 papers published by American authors only, and three papers by Russian authors only. The American papers averaged 27.3 cites per paper, while the Russian papers averaged 6 cites per paper. The American median was 20 cites per paper, while the Russian median was 4 cites per paper. (As an aside, the Japanese papers appeared to very numerous and well cited, followed by the Western European papers).

The author may examine other fields and may use larger samples, but there seems to be a loud and clear message coming through. Whether or not the Russians are prolific in a field in terms of paper production, their works are not getting cited by the larger technical community. Possible explanations are:

1) They could be doing good (citeable) work, and not reporting it;

2) The work reported may be good, but very applied, and not amenable to citing in the literature; i.e., citation is not the appropriate measure of quality or utility or impact in this case;

3) The work reported could be good, but might not be published in the forefront literature, and the technical community therefore might not be very aware of this work.

4) The work could be poor, and the citations pinpoint this.

The author has asked perhaps a dozen experts for explanations of these findings, and the number of reasons given approaches the number of experts. This potential diversity of explanations for citation analysis results pinpoints the major operational problem with using these indicators in stand-alone mode.

In the mid-1970s, the author led two delegations on Controlled Fusion to the Soviet Union. He visited the Kurchatov Institute in Moscow, and Academgorod near Novosibirsk. Both times, he was impressed by the technical quality of the Russian work in Fusion (both fast-pulsed systems and near-steady state), although there were obvious gaps. At the time, the author had the impression that this high technical quality extended to other fields, with obvious exceptions in computers, microelectronics, etc. The present citation results seem to reflect a different level of technical performance than what the author thought he had seen in the mid-1970s.

Did the author have a misperception then? Had the author examined citation performance 20 years ago, would he have arrived at the same conclusions as today? Or, has the dissolution of the Soviet Union resulted in a real degradation of their technical performance? Or, are the author's study approach and groundrules overly limited and not applicable? Or do all of the above explanations and questions have some validity, and point out graphically the deficiencies of trying to use simple quantitative indicators in a stand-alone mode (such as citat ion counts) to measure extremely complex and sophisticated issues.

II-g. Citation Analysis as a Warning signal

Perhaps this particular example has shown the value, if any exists, of using quant itative metrics such as citation counts for research quality or impact studies. The quantitative results serve as the 'red flags' or warning lights that a problem may exist; they are the modern day equivalents of the 'canary in the mine' approach to volatile gas detection. However, it was uncertain exactly what killed the canary decades ago, and it is uncertain today what specific citation counts mean. This is precisely how the author uses citation studies today; they serve as indicators that further investigation into specific areas is warranted, and they are always accompanied by, and subordinate to, expert analysis/ peer review.

APPENDIX 3-A

CHARACTERISTICS OF HIGHLY-CITED AND POORLY-CITED PAPERS

3-A-1. Hypersonic/ Supersonic Flow Study [Kostoff, 1999a]

To ascertain whether any relationship between highly cited and lowly cited papers and their associated journals and performing organizat ions could be observed, the characteristics of samples of highly and lowly cited papers were analyzed. The database used to extract the samples was the expanded web version of the SCI. In contrast to the CD-ROM version of the SCI used to obtain the bulk data for this paper, the web version has 60% more journals (~5200), and is more convenient for performing citation analyses (however, the web version in its present incarnation is less convenient than the CD-ROM version for most bulk data analysis, since not all records can be downloaded at once). All records in the web version which contained the term HYPERSONIC (a small subset of the supersonic/ hypersonic field) and were published in 1993 were examined.

There were 155 raw 'hits', or records obtained by the query, of which 15 (10%) were not applicable to the topic of hypersonic flow over aerodynamic bodies. Of the remainder, 64 records (46%) had zero citations by other papers; 55 records (39%) received between one and four citations; and 21 records (15%) were cited five or more times by other documents in the expanded SCI, and were viewed as highly cited papers.

Seven of those highly cited papers (33%) were published in the AIAA JOURNAL (231-number of papers from database published in journal); three papers in the JOURNAL OF SPACECRAFT AND ROCKETS (109); three papers in the JOURNAL OF FLUID MECHANICS (48); and one paper each in a variety of journals which contained fewer papers from the total database. The median journal in the sample contained 48 of the total database papers, as contrasted to the m edian journal in the total database containing one paper. Since the number of journals which contain n published papers follows approximately a hyperbolic distribution, the journals in the highly cited sample are, on average, the very top echelon of the total database journals in terms of numbers of papers published.

In the highly cited paper sample, twelve were from foreign institutions; twelve were from universities; and six were from NASA laboratories. The five most highly cited papers were from universities. The median organization in this sample contributed thirteen papers to the total database, as contrasted to the median organization in the total database contributing one paper. Since the number of papers n contributed by an organization to the total database also follows a 1/n^2 distribution, the organizations in the highly cited sample are, on average, the very top echelon of the total database organizations in terms of numbers of papers contributed.

The 64 records with zero citations were also examined, albeit from a different perspective. Because the range of citations in the total 140 record sample was between zero and ten, it was felt that there probably was a quality stratification within the sample group with zero citations, and thus the very poor performers could not be isolated as precisely as the good performers. The following

observations were made of the zero cited papers sample.

AIAA JOURNAL contributed 3% of the zero cited papers, as contrasted to 33% of the papers in the highly cited sample; JOURNAL OF SPACECRAFT AND ROCKETS - 13% zero cited/ 14% highly cited; JOURNAL OF FLUID MECHANICS - 0% zero cited/ 14% highly cited; HIGH TEMPERATURE - 9% zero cited/ 0% highly cited; JOURNAL OF AIRCRAFT -8% zero cited/ 0% highly cited; PMM JOURNAL OF APPLIED MATHEMATICS AND MECHANICS -6% zero cited; 0% highly cited; ZEITSCHRIFT FUR FLUGWISSENSCHAFTEN UND WELTRAUMFORSCHUNG -6% zero cited/ not listed in CD-ROM database. The journals with a high ratio of highly cited papers to zero cited papers tend to emphasize the more fundamental research. The journals with a low ratio of highly cited papers to zero cited papers tend to emphasize the more applied research. The fact that the applied papers are being cited less than the more fundamental papers does not mean they are less useful or of lower quality; they may be of substantial use to developers, who publish much less than researchers, and this more practical use would not be reflected in the present type of bibliometrics study.

Industrial organizations contributed 27% of the zero cited papers, as contrasted to 10% (2 papers) of the highly cited papers (these two highly cited papers were actually one paper split into two sections and published sequentially in the same journal issue); university organizations -33% zero cited; 57% highly cited; NASA -9% zero cited/ 29% highly cited; American organizations -36% zero cited/ 43% highly cited; European organizations -25% zero cited/ 38% highly cited; Asian organizations -9% zero cited/ 14% highly cited; Middle Eastern organizations -5% zero cited; 0% highly cited; Russian organizations -23% zero cited; 5% highly cited. This last observation is quite surprising, since two of the top four paper contributing organizations in the total CD-ROM database were Russian.

In summary, this small sample analysis led to the following conclusions for hypersonic flow. Fundamental research papers are more likely to be cited than applied research papers; university papers are more likely to be cited than industry papers; the journals which contain concentrations of highly cited papers are also the core journals in terms of papers published; NASA produced many papers (147 in the total CD-ROM database), and had a substantial fraction of the highly cited papers; Russia produced slightly more papers than NASA (169 in the total CD-ROM database), and had almost no highly cited papers.

The NASA/ Russia citation differential led to another short study which examined American/ Russian differentials in supersonic/ hypersonic flow citations. Two groups of papers were generated. The first group consisted of all papers (from the web version of the SCI) published in 1993/ 1994 by the three most prolific supersonic/ hypersonic flow Russian authors identified in Kostoff [1997o]; the second group included all papers by the three most prolific supersonic/ hypersonic flow American authors from Kostoff [1997o]. There were 12 papers in the first (Russian) group, and 36 papers in the second (American) group. All papers related to supersonic/ hypersonic flow. The citations received by all these papers were examined.

Of the twelve Russian papers, nine received zero cites, two received one cite each, and one received three cites. The average cites per paper is 0.4. All of the five total cites were self-cites (There is

nothing intrinsically wrong with self cites; in those cases where the author has done the pioneering work in the field, self-cites are most appropriate. However, when all cites are self-cites, then the true impact of the paper on the larger scientific community must be called into question).

Of the 36 American papers, seven received zero cites. The total number of citations received was 106, of which 56 were self cites. The average cites per paper is three. While all these citation numbers reported are quite small, reflecting the low level of effort in this technical field, there is obviously a systemic difference between the citat ions received by the Russian and American papers. Whether these differences extend beyond supersonic/ hypersonic flow to other topical areas is an interesting question.

There are two crucial pieces of data missing from these two short studies (and from most bibliometrics analyses) which prevent harder conclusions about quality and value to be drawn. The amount of research effort represented by each paper is unknown to the analyst, and the eventual use of the results from each paper is unknown to the analyst. Thus, the number of highly cited papers per dollar of research investment (or some similar research efficiency metric), probably a better measure of value than pure numbers of papers or highly cited papers, cannot be stated. Also, the quality of the eventual hypersonic vehicles which resulted from the papers' research, probably a better measure than numbers of cited papers, was not tracked and cannot be stated. In addition, the papers in these two short studies were not read in detail independently by hypersonic flow experts, and thus their quality could not be gauged independently from another perspective and correlated to the citation results.

3-A-2. Cortex Study [Kostoff, 2005i]

Citation Comparison among *Cortex*, *Neuropsychologia*, and *Brain*

To compare citations among papers published in *Cortex*, *Neuropsychologia*, and *Brain*, three leading neuropsychology journals, the following experiment was run. All articles published in *Cortex*, *Neuropsychologia*, and *Brain* in the years 1998-1999 were retrieved from SCI. There were 110 *Cortex* articles, 278 *Neuropsychologia* articles, and 341 *Brain* articles. Then, the ten most cited articles from each retrieval (the citations from each paper used for the tabulation of most and least cited are those listed in the SCI Times Cited field, and are the total citations received by each paper from all other papers in the SCI) were extracted, as well as the ten least cited articles, and various characteristics compared. The results are shown in Table 7

TABLE 7

	CORTEX		***NEUROPSYCHOLOGIA***		***BRAIN***	
	MOST CITED	LEAST CITED	MOST CITED	LEAST CITED	MOST CITED	LEAST CITED
# AUTH						
AVER	3.9	2.8	5.2	2.6	7.1	4.6
MEDIAN	4	3	5	1	7.5	4.5
# REFS						

AVER	46.3	28	52.5	26.8	68.3	42.4
MEDIAN	49	29.5	49	26	62.5	35
# CITES						
AVER	21	0.8	71.3	0	166.8	2.8
MEDIAN	18.5	1	67.5	0	157	3
ORG						
INST	5	4	2	4	8	2
UNIV	5	6	8	6	2	8
COUNTRY	4 ITALY	2 ITALY	4 UK	5 USA	5 UK	3 JAPAN
	3 FRANCE	2 USA	4 USA	2 ITALY	2 USA	1 USA
	1 AUSTRIA	2 GERMANY	1 ITALY	1 NZ	2 CANADA	1 UK
	1 BELGIUM	2 JAPAN	1 CANADA	1 NETH	1 GERMANY	1 FRANCE
	1 GERMANY	1 NETH		1 AUSTRALIA		1 ITALY
		1 AUSTRALIA				1 CANADA
						1 GERMANY
						1 NETH
TYPE						
BEHAV	8		4			
SURGERY			1		2	
DIAG-NI	2		5		7	
DIAG-INV					1	

CODE:TYPE
BEHAV=CLINICAL BEHAVIOR STUDIES
SURGERY=SURGICAL INTERVENTIONS
DIAG-NI=NON-INVASIVE DIAGNOSTIC TESTS
DIAG-INV=INVASIVE DIAGNOSTIC TESTS

A number of interesting observations may be made from Table 7. First, the most cited articles in *Neuropsychologia* are cited, on average, more than three times as often as the most cited articles in *Cortex*, and the most cited articles in *Brain* are cited, on average, more than twice as often as the most cited articles in *Neuropsychologia*.

Second, the most cited papers have more authors than the least cited, in all three journals, and the effect is most pronounced in *Neuropsychologia*. Additionally, the average number of authors increases with the average number of citations, ranging from about four authors of the most cited *Cortex* papers to about seven authors of the most cited *Brain* papers.

Third, the most cited papers have substantially more references than the least cited, in both journals, and the effect is most pronounced in *Neuropsychologia*. Additionally, the average number of citations increases with the average number of references (an effect observed by the first author in recent unpublished text mining studies), ranging from about 46 references in the most cited *Cortex* papers to about 68 references in the most cited *Brain* papers.

Fourth, there is no clear overall trend in citations as a function of institutional representation. The institution/ (institution + university) ratio (where institution in the table cells should be interpreted as any non-university organization; e.g., research laboratory, clinic, hospital, company) for most cited papers starts at 0.5 for Cortex, drops to 0.2 for *Neuropsychologia*, and increases sharply to 0.8 for *Brain*. This ratio for least cited papers starts at 0.4 for both *Cortex* and *Neuropsychologia*, and decreases to 0.2 for *Brain*. Its most dramatic change is from 0.8 for the most cited *Brain* papers to 0.2 for the least cited *Brain* papers.

Fifth, the most cited papers in *Cortex* are all from continental Western Europe, with heavy representation from Italy and France, while the least cited papers in *Cortex* represent four different continents. The most cited papers in *Neuropsychologia* are, with the exception of Italy, from the UK and North America (with heavy representation from the UK and USA), while the least cited papers have more representation from Western Europe but none from the UK. The most cited papers in *Brain* are from the major English-speaking countries, whereas the least cited are scattered around Western Europe, Asia, and North America.

Sixth, there is a distinct shift in type of study (the bottom of Table 7) in proceeding from *Cortex* to *Neuropsychologia* to *Brain*. Clinical behavioral studies, many of them essentially case studies, predominate the most cited Cortex papers. There are only two papers characterized as Diagnostic-Non-Invasive (e.g., PET, MRI, etc). *Neuropsychologia* has more of a balance between Behavioral and Diagnostic-Non-Invasive in its ten most cited papers. *Brain* shows a heavy emphasis on Diagnostic-Non-Invasive (7/10), two papers on surgical procedures, and one on Diagnostic-Invasive.

Based on reading Abstracts from each of these journals, the types as represented in the top ten most cited articles roughly approximate the types of papers published overall. Thus, as citations increase in absolute amounts, the study type transitions from the clinically oriented behavioral focus to the correlates with more objective measurements. Also, as the results from the most cited papers section showed, as the study type transitions from the clinically oriented behavioral focus ('soft' technology) to the more objective measurements ('hard' technology), the most cited papers tend to become more recent.

APPENDIX 3-B.

<u>CITATION ANALYSIS OF RESEARCH PERFORMER QUALITY</u> [Kostoff, 2002e]

INTRODUCTION

In the evaluation of science and technology (S&T), whether ongoing or proposed programs, a key criterion is the track record of the proposer or performer. Past analyses [DOE, 1982; Kostoff, 1997a] have shown that, typically, the criterion of Team Quality is the major determinant of program or project quality. Many qualitative and quantitative approaches have been used for the purpose of determining Team Quality [Kostoff, 1997a]. None are viewed as adequate in a stand-alone mode, and present practice is to use multiple approaches to determine Team Quality [Martin, 1983; Kostoff, 1997b].

One of the more widely used of these approaches, especially applicable to research, is citation analysis. For proposer quality assessment, citation analysis consists of counting citations to documents produced by the proposer's research unit, then comparing this citation count to numbers of citations received by similar documents from other research units. The assumption is then made that documents with higher relative numbers of citation counts have more impact than those with lower citation counts, and are of higher quality from the citation metric perspective.

While this approach appears rather straight-forward and deceptively simple, it is intrinsically very complex. This appendix will illuminate the complexities, and show that high quality S&T citation analysis requires technical experts performing very manually intensive comparisons with very subjective judgements. It will show further that the automated assembly-line approaches to citation analysis, widely used by the decision aid community today, are highly uncertain at low-to-mid citation levels characteristic of most research.

After a background description of the problem, the analytical techniques developed for the citation analysis will be presented. Two illustrative examples of the use of citation analysis to support proposal review will be presented. Because of the confidentiality agreements operable for proposal review, all information that identifies either the proposing organization or the potential science and technology sponsor will be removed. The results of the analysis will then be presented, followed by summary and conclusions that emphasize the lessons learned from using these techniques. Special emphasis will be placed on requirements for thematic similarity between the target documents and the external documents against which they are compared.

BACKGROUND

In the present context, citation is referencing, in a document, the work of another individual or group. The work referenced can exist in many forms, although the most common use is reference of another document. Citation analysis is the examination of the multiple dimensions and myriad facets of citations for the purpose of understanding the many impacts of the target documents of interest.

Citation counts resulting from citation analyses are usually classified as outputs, but they are neither outputs nor outcomes. While they are closer to outputs than outcomes, since they can be used in relatively short range analyses and they do not impact the larger problems characteristic of outcomes, they are not under the direct control of the performer.

Modern day interest in studying and developing the citation process accelerated after WW2 [e.g., Zachlin, 1948, Zirkle, 1954]. However, the origins of citation analysis as a widespread bibliometrics tool can be traced to the mid-1950s, with Garfield's proposal for creating a citation index [Garfield, 1955]. As the Science Citation Index (SCI) was developed, along with companion citation indices, the computer revolution and associated information technology developed in parallel. The combination of SCI, massive information storage, and rapid information retrieval laid the foundation for a multi-application S&T evaluation capability.

The foundations of modern traditional citation analysis were established by Garfield [1955, 1963, 1964, 1965, 1966, 1970] and CHI, Inc [Narin, 1975, 1976, 1984, 1994, 1996; Albert, 1991], and extended to co-citation analysis by Small [1973, 1974, 1977, 1981, 1985], Sullivan [1977, 1979, 1980], and Marshakova [1973, 1981, 1988].. The practice of citation analysis has been extended further by groups at the Hungarian Library of Sciences [Schubert, 1986, 1993, 1996; Zsindely, 1982] and the University at Leiden [Moed, 1986; Nederhof, 1987; Braam, 1988, 1991; VanRaan, 1991, 1993, 1996; Davidse, 1997]. A broad summary of the status of citation analysis is contained in a recent festschrift to Eugene Garfield [Festschrift, 2000].

Traditional citation analysis is presently used both at the micro and macro scales. It is used at the micro level, especially in academia, to evaluate components of impact of a given published document, or the documents published by a given researcher or research group. It is used at the macro level to evaluate technical discipline or national outputs. Because of the large numbers of documents and subsequent citations that exist in macro level analyses, semi-automated techniques have been developed to handle the data efficiently. As time has proceeded, these semi-automated techniques have diffused toward micro level application.

Citation analysis has two components. The first component is <u>counting</u> of citations to a document or group of documents, depending on the purpose of the analysis. The second component is placing these citation counts in a larger context through a <u>comparison</u> and normalization process, to provide meaning to the numbers of counts obtained.

Many articles have been written about problems inherent in the traditional citation analysis process [e.g., Geisler, 2000; MacRoberts, 1989, 1996; Kostoff, 1998]. There are two main categories of problems: those associated with the <u>counts</u> of citations, and those associated with the <u>comparisons</u> of counts of citations. The problems associated with counts of citations can be sub-divided further into problems associated with the <u>quantity</u> of the underlying data, and problems associated with the <u>quality</u> of the underlying data.

Problems with Citation Counts

Problems with Quantity of Underlying Data

The main resource available for performing citation analysis today is the SCI. The number of candidate articles to be used in a citation analysis is limited to the number of articles in the total SCI. This total is limited by the following sequence of steps.

a) There is approximately $500 billion-$800 billion/ year worth of S&T being performed globally today, depending on one's definition of S&T. Only a small fraction of the S&T performed is documented. While there are many reasons for this [Kostoff, 2000a], basically there are more disincentives to publishing than incentives.

b) Of the S&T performed that eventually gets documented, only a very modest fraction is accessed by the SCI (or any single database). There are tens of thousands each of internal and external technical reports, classified reports and papers, workshop and conference proceedings, journals, magazines, newspapers, and patents resulting from the S&T performed and published annually. Yet, the SCI accesses only about 5600 journals presently. While these accessed journals tend to be the highest quality peer-reviewed research journals, they represent only a fraction of S&T that is documented.

c) Of the documented S&T that is accessed by the SCI, only a fraction reaches the average analyst performing citation analysis. The main reason is the extremely poor information retrieval techniques actually used by the technical community [Kostoff, 2000b].

Thus, the citation counts derived from the records in the SCI under-represent the total referencing of prior work by the global technical community, and there is no evidence that this under-representation is homogeneous across disciplines or sub-disciplines.

Problems with Quality of Underlying Data

The problems with citation data quality translate into problems with the citation selection process (i.e., the approach used by authors to select references for inclusion in their papers). The issues related to the sociological and cultural aspects of how people cite have been raised by the references cited above, and will not be repeated here. Suffice it to say that the combination of quantity and quality problems with citations places strong limits on the degree to which citations can be used as a stand-alone metric. This is especially true for documents that receive mid and low level numbers of citations (i.e., the vast majority of documents published); the very highly cited documents (a very small fraction of all articles published) are in a class by themselves, and modest margins of error in interpreting their citation counts don't affect overall conclusions about their impact.

Problems with Citation Comparisons

Problems with citation count comparisons form the focus of this appendix. Whether applied to

micro or macro scale problems, citation count comparisons have received insufficient attention, and offer further severe constraints on the credibility of present day citation analyses. There are two main types of potential citation count comparisons: comparison of counts to an absolute standard, and comparison of counts to a relative standard. The former comparison is analogous, in the physical sciences, to comparing actual engine efficiencies to maximum engine efficiencies possible (Carnot efficiencies). The latter comparison is analogous to an athletic competition, where one group's performance is compared to another group's performance. One problem with the latter comparison is that the performance of a group is never related to its potential, only to the performance of another 'similar' group. The latter comparison is used in essentially all citation analyses today. This issue of comparison with absolute or relative standards was examined in a 1997 paper [Kostoff, 1997c], and will not be addressed further.

Citation count comparisons are necessary because of the high variability of citation counts with different parameters. Citation counts depend strongly on the specific technical discipline, or sub-discipline, being examined. The funding and number of active researchers can vary strongly by sub-discipline, and these numbers of researchers affect the numbers of citations directly. The maturity of the sub-discipline affects the numbers of citations, since the basic research community is oriented more toward publishing than the applied research or technology development communities. The breadth of the sub-discipline can affect citation counts, since more focused disciplines will concentrate citations into fewer key researchers. The classification and proprietary levels can vary sharply by sub-discipline, and can strongly affect what gets published and therefore cited in open-literature publications. The documentation and citation culture can vary strongly by sub-discipline. Since citation counts can vary sharply across sub-disciplines, absolute counts have little meaning, especially in the absence of absolute citation count performance standards.

Thus, in order to provide meaning and context to citation counts for performance evaluation in traditional citation analysis, some type of citation count normalization is required. The main normalization approaches used in traditional citation analyses are described in an excellent review article [Schubert, 1996]. They can be summarized as follows:

1) Reference standards based on prior sub-field classification

Journals are classified into a number of science sub-fields. Since some journals are single discipline, and some multi-discipline, percentage weights are assigned to each journal indicating their connection with the different sub-fields. According to Schubert [1996], the method works only at a higher (macro) statistical level; i.e., if the sample under study is large and mixed enough to support the validity of such a statistical approach. Further according to Schubert [1996], for micro level analyses, it is sometimes unavoidable to use a classification scheme concerning not only the journals but every single paper. Schubert proceeds to point out that such classification schemes are enclosed in some specialized databases, such as in the *Physics Briefs*, to classify each paper into one or more of ten first-level and many lower-level sub-fields of physics.

2) Journals as reference standards

Primary journals in science are generally agreed to contain coherent sets of papers both in topics and professional standards. According to Schubert [1996], it seems justified to regard the set of regular authors of a journal as reference standard for any single author (or team of authors), the set of institutions regularly publishing in the journals as reference standard of any single institution, the citation rate of the set of papers published in the journal (or of a properly selected subset) as reference standard of any single paper. Also according to Schubert [1996], one may thus expect that any difference in productivity, citation rate or other scientometric indicators reflects differences in inherent qualities.

3) Related records as reference standards

Subject matter similarity between two documents is measured by the number of shared references. According to Schubert [1996], bibliographic coupling appears to be one of the most selective and flexible techniques of reference standard selection, but "because of its high requirements in time and effort, its use can be suggested only in micro or meso-level".

It is the present author's contention that none of the above normalization methods are adequate for precise normalization, since they do not provide sufficient resolution for distinguishing among the lower level sub-fields. Inability to distinguish precisely among sub-fields translates, in some cases, to substitution of far different magnitude numbers for the normalization base. The next section will show some of the effort required for more precise normalization comparisons.

ANALYSIS TECHNIQUES AND ISSUES

First proposal

The author was recently asked, by a potential sponsor, to evaluate an S&T proposal generated by organization XXXX. While there were a number of criteria that had to be evaluated relative to technical quality and relevance of the proposal to the potential sponsor's mission, one key criterion was the quality of the proposer's research team. It was decided to evaluate team quality through evaluation of the research team's various outputs and outcomes, using citation analysis and other metrics. This section focuses on the citation analysis component used.

The proposal and accompanying material presented many different types of outputs from XXXX researchers. Assessing the quality and impact of those outputs was complex, especially since they covered more than one research area. The following procedure was used as a first-order estimate of quality/ near-term impact of XXXX's output, and thereby of the research team.

The citations of selected XXXX publications were compared against those of thematically similar non-XXXX publications (a control group of publications), using a pair-wise comparison approach. Specifically, all XXXX publications for 1996 (38 documents), as identified in the

Web version of the Science Citation Index (SCI), were compared with thematically similar non-XXXX publications from the SCI.

[1996 was selected as a compromise year. The author wanted to examine recent documents that reflected current management and staff of XXXX, but also wanted to insure that sufficient time had passed since publication such that citations had a reasonable chance to accumulate. Figures 1 and 2, titled Citing Papers Time Distribution, show the yearly and cumulative numbers of citing papers as a function of time, for 1996 and 1993, respectively. For 1996, the citing papers (for all the XXXX papers published in 1996) show a linearly increasing cumulative trend up to and including 2000. For 1993, the citing papers (for all the XXXX papers published in 1993) show more of an S-curve trend. While 1993 shows a leveling off of the citations, and would therefore have been a better year to select from that perspective, it was judged to be too far in the past to be relevant for assessing the quality of present XXXX staff and management. Citations from 1996 should almost be ready to level off, if the 1993 distributions can be extrapolated to 1996, and therefore 1996 was selected.]

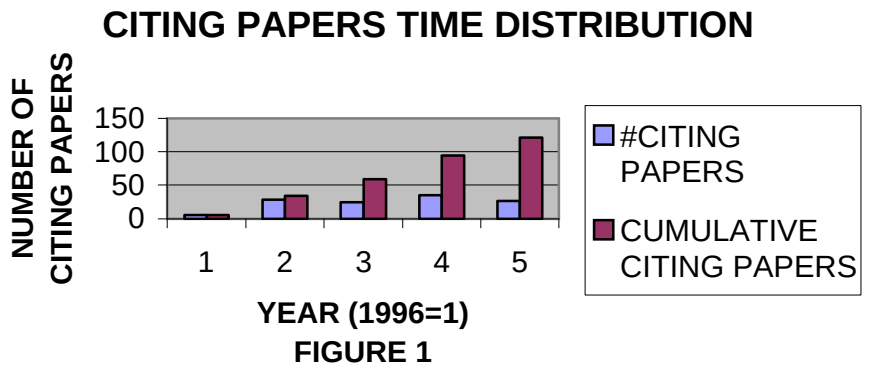


FIGURE 1

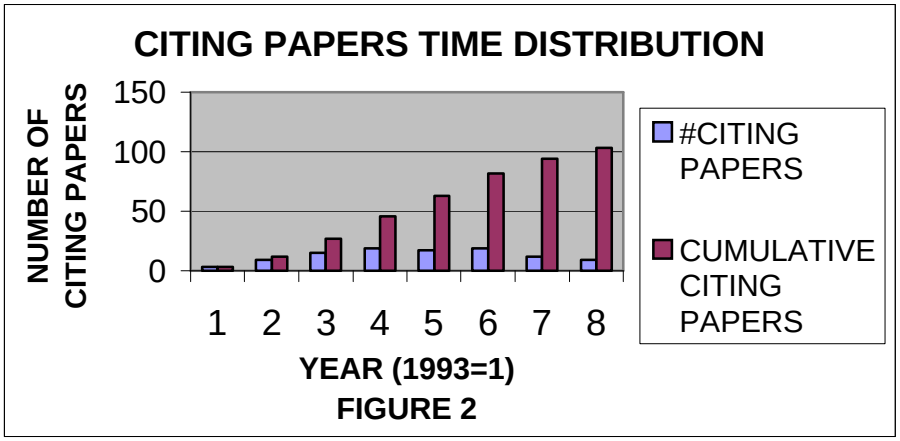


FIGURE 2

Ideally, the size of the control group for each paper should be statistically representative of the total thematically similar non-XXXX papers in the SCI, since the purpose of the citation analysis is to compare the citation performance of each proposer's paper to the aggregate of the relevant performer community.. Practically, resource and time constraints placed severe linits on the size of the control group. Specifically, for each of the 38 papers published in 1996 (hereafter referred to as the target papers), three non-XXXX papers thematically and temporally similar to the target papers were selected. If 1996 papers with the requisite thematic characteristics could be identified, they were given first priority in the selection, to insure temporal normalization. If 1996 papers could not be identified, then 1997 papers were selected. Thus, the results are conservative with respect to XXXX.

Selection of papers in the SCI thematically similar to the target paper depends strongly on the study's purpose and objectives, the mission of the performing organization, the degree of focus of the paper's theme, the size of the research paper pool from which to choose, and the level of technical description in the paper's SCI Abstract. The relation to study purpose is especially important, and is often overlooked. Specifically, is the purpose of the study to evaluate the 'job right' quality of the performer (i.e., is the specific task selected being performed with the latest tools and techniques to achieve the specific objectives?), or is the purpose of the study to evaluate the 'right job' quality of the performer (i.e., have the right task and right objectives been selected?). If the focus is on 'job right' quality, then the thematically similar papers will be limited to a very narrow area of inquiry. If the focus is on 'right job' quality, then the focus of thematically related papers can be expanded greatly.

For example, suppose that a researcher being evaluated was performing acoustic studies in the 100 KHZ small object detection regime. If the performing organization's mission in acoustics was limited to performing studies only in this regime, and if the quality determination was

phrased as how well the researcher was performing relative to other researchers studying the 100 KHZ regime, then the thematically similar papers would all be focused narrowly around frequencies of 100 KHZ. The study reduces to determining the most cited papers at 100 KHZ. If, however, the organization's mission in acoustics provided flexibility in selecting the frequency regime to study, and the organization *chose* to focus on the 100 KHZ regime, then thematically related papers could include those in a broader range of frequency regimes. The study reduces to determining the most cited paper in mid-high frequency acoustics. The choice of journal as reference standard, described previously and referenced in Schubert [1996], relates strongly to the latter definition of organization mission, where essentially any paper in an acoustics specialty journal could serve as a reference standard. The practical implications of 'job right' vs 'right job' comparisons are that papers with substantially higher citation counts could be included in the normalization pool as the allowed definition of thematic similarity becomes broadened.

Selection of papers thematically similar to the target paper was very difficult, time-consuming, and subjective. This was especially true for the broad-based analyses. The selection was more straightforward for the much more limited specific technology papers, since these more focused areas seemed to have many researchers working related problems. The author believes that the subjectivity involved in selecting thematically similar papers is a major source of uncertainty of the results. A rigorous study, in addition to having the rigorous information retrieval and statistical sampling processes mentioned in the next two paragraphs, requires the use of multiple evaluators for the same target papers to average out evaluator subjective bias.

Many of the applied research papers combined analytical technique advancement with novel application advancement. It was not always possible to have thematic similarity for both technique and application, especially in those research areas with relatively few performers, and typically a choice had to be made between technique and application for determining thematic similarity.

Two important issues were i) determining the number of thematically similar candidate papers in the pool from which to choose, and then ii) determining the number of papers to select from the pool. First, in a rigorous study, candidate thematically similar papers would be identified by the most rigorous processes available. In the author's information retrieval studies [Kostoff, 1997d, 2000b], a manually intensive iterative approach using computational linguistics and bibliometrics is used to identify the full scope of relevant literature papers for each specific topic studied. For the present study, this would have required 38 such literature searches. In the time available, even one such rigorous literature search was not feasible. A very approximate approach was used.

Second, the number of papers to select from the candidate pool should have the greatest thematic similarity, and be representative statistically. Again, this would have required poring over hundreds, or thousands, of similar papers, and selecting a substantial number of the most representative thematically. Again, a small sampling approach was used because of time exigencies.

The first selection step was to examine the Related Records field of the SCI for a given target paper. This field contains papers that have at least one reference in common with the target paper, as stated previously [Schubert, 1996]. Papers that share references tend to be similar thematically, but this is not always true, and the relation between thematic similarity and number of shared references is not always monotonic.

Because of time constraints, a limited number (three) of thematically related papers was examined for each target paper. If three records thematically similar to the target paper could be identified from the Related Records papers, the selection was completed for that target paper. If three records could not be identified, then key words from the target paper's Abstract/ Title/ Keyword fields were used to search the SCI for related records. This approach was substantially more time consuming than the already time-consuming Related Records approach.

FIGURE 3 - CITATION AND FIGURE OF MERIT DATA

A	B	C	D	E	F	G	H	I	J	K	L
REC #	PAP CIT	SELF CIT	PAP1 CIT	PAP2 CIT	PAP3 CIT	AVER CIT	FOM1	MED CITES	FOM2	STD DEV CIT	FOM3
1	4	1	3	3	23	9.667	0.293	3	0.571	11.55	-0.49
2	2	1	9	7	21	12.33	0.14	9	0.182	7.572	-1.36
3	0										
4	0		5	1	2	2.667	0	2	0	2.082	-1.28
5	0		5	6	9	6.667	0	6	0	2.082	-3.2
6	3	2	3	4	4	3.667	0.45	4	0.429	0.577	-1.15
7	0		11	14	4	9.667	0	11	0	5.132	-1.88
8	1	1	1	3	2	2	0.333	2	0.333	1	-1
9	6	3	3	7	5	5	0.545	5	0.545	2	0.5
10	5	0	2	5	16	7.667	0.395	5	0.5	7.371	-0.36
11	5	3	5	2	14	7	0.417	5	0.5	6.245	-0.32
12	2	2	3	3	2	2.667	0.429	3	0.4	0.577	-1.15
13	1	0	4	4	5	4.333	0.188	4	0.2	0.577	-5.77
14	5	2	6	4	9	6.333	0.441	6	0.455	2.517	-0.53
15	7	4	15	5	12	10.67	0.396	12	0.368	5.132	-0.71
16	5	5	3	7	1	3.667	0.577	3	0.625	3.055	0.436
17	4	4	8	4	6	6	0.4	6	0.4	2	-1
18	9	4	38	2	13	17.67	0.338	13	0.409	18.45	-0.47
19	4	2	3	7	7	5.667	0.414	7	0.364	2.309	-0.72
20	2	1	2	6	8	5.333	0.273	6	0.25	3.055	-1.09
21	0	0	2	5	16	7.667	0	5	0	7.371	-1.04
22	1	1	13	8	9	10	0.091	9	0.1	2.646	-3.4
23	24	20	5	2	7	4.667	0.837	5	0.828	2.517	7.682

24	4	0	4	22	8	11.33	0.261	8	0.333	9.452	-0.78
25	0										
26	0										
27	3	0	11	14	2	9	0.25	11	0.214	6.245	-0.96
28	2	2	3	3	4	3.333	0.375	3	0.4	0.577	-2.31
29	4	4	8	10	6	8	0.333	8	0.333	2	-2
30	2	2	3	3	13	6.333	0.24	3	0.4	5.774	-0.75
31	1	1	2	4	5	3.667	0.214	4	0.2	1.528	-1.75
32	0										
33	6	6	13	26	3	14	0.3	13	0.316	11.53	-0.69
34	0	2	2	4		3	0	3	0	1.414	-2.12
35	3	1	2	5	16	7.667	0.281	5	0.375	7.371	-0.63
36	0		2	7	1	3.333	0	2	0	3.215	-1.04
37	2	1	5	22	4	10.33	0.162	5	0.286	10.12	-0.82
38	4	1	5	3	14	7.333	0.353	5	0.444	5.859	-0.57
SUM	115	74	197	200	252	AVER	0.297		0.324		-0.98

Once thematically similar records were identified, the citations for each of the four records were tabulated. Figures of merit were generated, and the citation performance of each target paper was compared with that of the three thematically related papers. The results are shown in Figure 3. Starting from the left, column A is the number of the record, column B is the citations of the target paper, column C is the self-citations of the target paper, columns D, E, F are the citations of the thematically similar papers (the Abstracts of papers 3, 25, 26, 32 did not contain sufficient information for similar papers to be identified), column G is the average citations of the thematically similar papers, column I is the median citations of the thematically similar papers, and column K is the standard deviation of the citations of the thematically similar papers. Columns H, J, L are figures of merit FOM1, FOM2, FOM3, respectively, defined as follows:

FOM1=citations of target paper/ (citations of target paper plus average citations of related papers)

FOM2=citations of target paper/ (citations of target paper plus median citations of related papers)

FOM3=(citations of target paper minus average citations of related papers)/ standard deviations of related papers.

FOM1 and FOM2 have the desirable properties of ranging between zero and unity, as well as equaling 0.5 when the target paper citations equal those of the average or median citations of the related papers. FOM3 removes the limitations of using absolute number values, and places the citation differences in the context of standard deviations.

This section ends with a note about the four papers that could not be evaluated due to insufficient

information contained within the Abstract. Ideally, with unlimited time and resources, the full text target and control group papers would be read in their entirety. Practically, time is available for reading Abstracts only. Unfortunately, in the non-medical technical literature, and some of the medical literature, there are no requirements on the technical content of Abstracts. Consequently, many Abstracts contain very little technical detail, and they cannot be used in the citation process. This issue is addressed summarily in a letter to Science [Kostoff, 2001a], and in more detail in a letter to selected technical journal editors proposing the use of Structured Abstracts in all technical journals [Kostoff, 2001b].

Second Proposal

In early 1998, the author was asked to evaluate an S&T proposal for a different potential sponsor, generated by an organization (ZZZZ) different from the proposing organization (XXXX) of the first proposal. One critical component again was evaluation of team quality. This was a complex procedure for the second proposal, since most of the organization's publication outputs were co-authored with people from other organizations, and the author wanted to identify the quality of the contributions of researchers from organization ZZZZ only. Again, citation analysis was one of several methods used to gauge team quality, and this section reports on the citation analysis component only.

1. Database Examined and Process Used

One purpose of the study was to examine the citation impact on the technical community of the ZZZZ researchers who publish. Another purpose was to assess some estimate of the ZZZZ researchers' contribution to the published product. Two studies were performed. First, all the 1997 papers in the web version of the SCI that contained a ZZZZ author address were examined. The position of the ZZZZ author in the author list for each paper was highlighted. Citations for this group of papers were not examined, because of the recent date.

Second, all the 1993 papers that contained a ZZZZ author address were examined. 1993 was selected for two reasons. A four-year lag allows many (not all) citations to accumulate, and is sufficient to show differentiation in citation counts among papers. Also, 1993 was the third year that paper abstracts were included in the SCI, allowing more than title information to be obtained about a paper if necessary. Author position was highlighted again, and then the citations received by each paper with citations received by a non-ZZZZ authored paper of similar theme were compared.

RESULTS AND DISCUSSION

First Proposal

The results for the first proposal are as follows.

Figures 4 and 5, titled Citation Distribution Function, show the numbers of papers N(X) with X cites for 1993 and 1996, respectively. 63% of the 1993 target papers had either zero or one cites, and 37% of the 1996 target papers had either zero or one cites. For 1996, the average number of

citations per target paper was three, of which 2/3 were self-cites. (No judgements are made about including or excluding self-cites. To make such judgements rationally, each full-text paper would have to be read, and the technical rationale for self-citation other than author self-gratification would have to made. Such a level of detail is beyond the scope of this study.) For 1993, the average number of citations per target paper was about 2.5. For 1996, the average number of citations per thematically related paper was about twice the number of target paper citations.

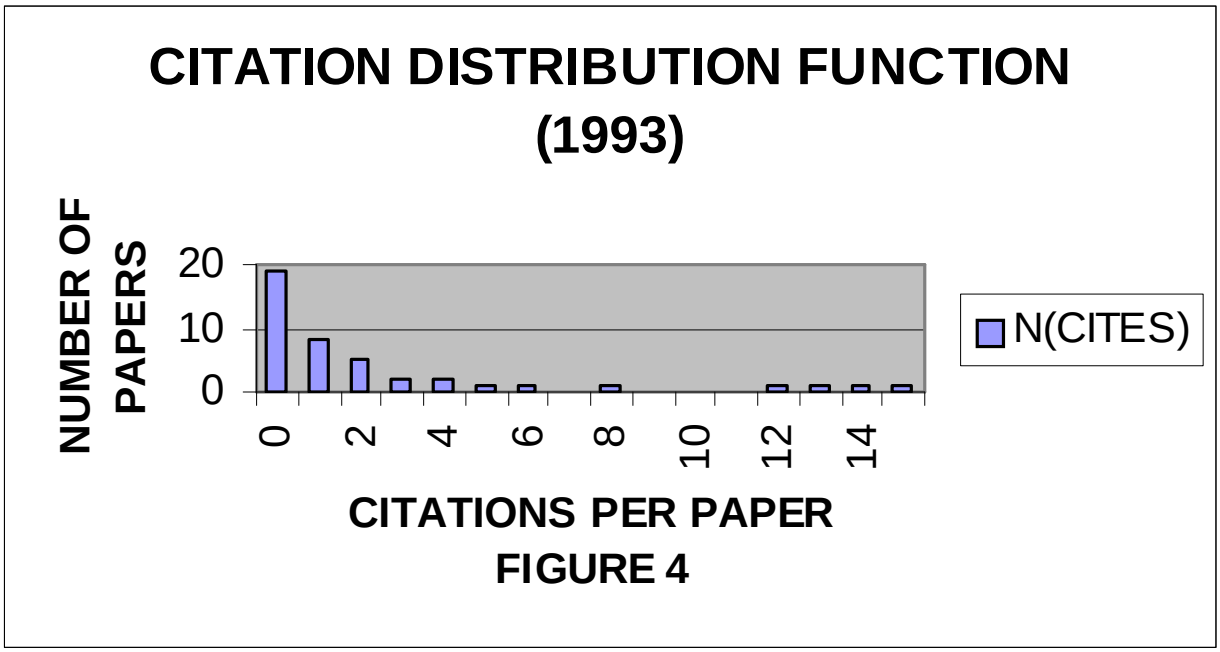


FIGURE 4

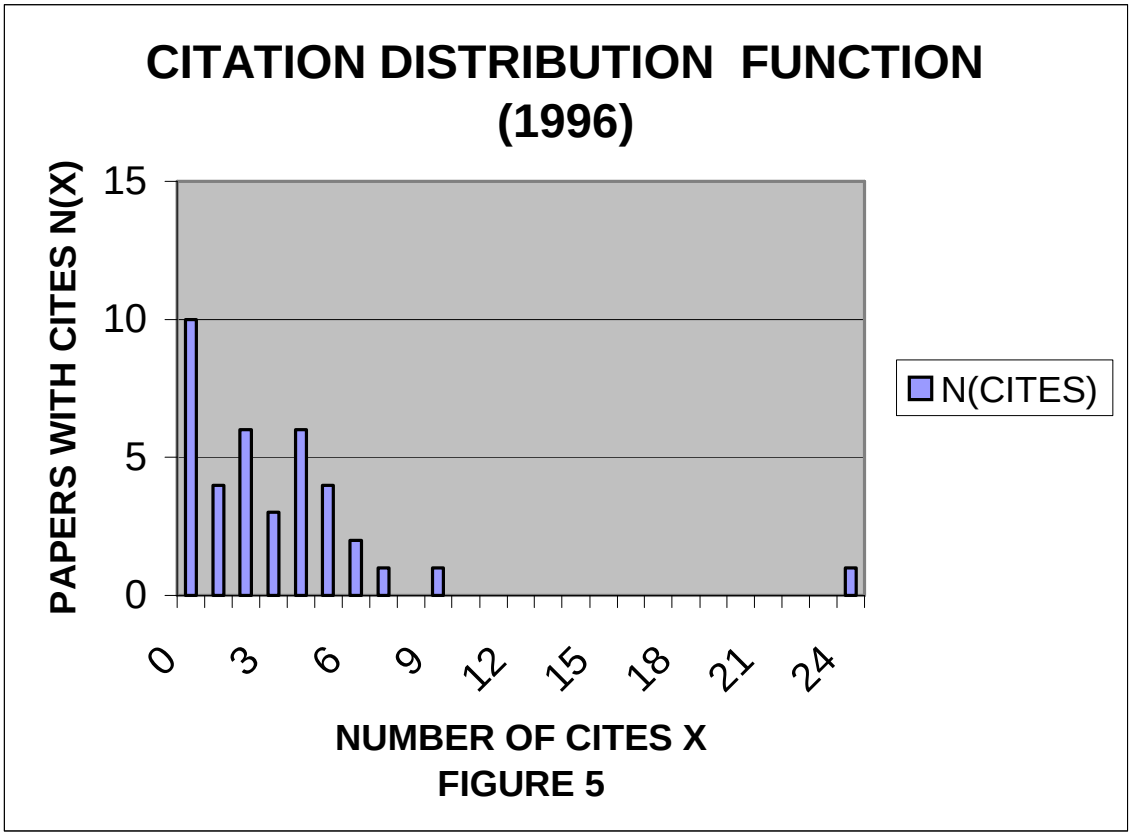


FIGURE 5

For 1996, the average value of FOM1 and FOM2 was about 0.3, and the average value of FOM3 was about minus one standard deviation. Thus, all three figures of merit gave essentially similar results. FOM1 and FOM2 were greater than 0.5 in less than ten percent of the target papers examined. In the best performing target paper, both in absolute citations and relative citations, 20 of the 24 citations were self-cites. This particular paper had many authors, and many of these authors cited the target paper in later publications.

Many of the research disciplines examined seem to have relatively few papers thematically related to the target paper. In addition, the absolute levels of citations are low, relative to other disciplines the author has examined. This suggests research into areas that have few performers, probably low funding, and therefore low citations.

Second Proposal

1. Results and Discussion

a. 1997 Database

In the 1997 database, there were 43 papers in the SCI with a ZZZZ address for the research unit. These papers had a total of 184 authors, with an average of 4.29 authors per paper, a median of 3 authors per

paper, and a mode of 3 authors per paper. A Coefficient of Author Position (CAP) was defined as a measure of the ZZZZ author's location in the total author list. The definition of CAP was:

CAP=(x-1)/(n-1)

where x was the location of the ZZZZ author in the list, and n was the total number of authors in the list. Thus, if there were three authors in the list, and the ZZZZ author was third, CAP would equal one. If the ZZZZ author was first in this case, CAP would equal zero. If the paper had only one author, CAP was set equal to zero. Thus, the higher the value of CAP, the less was the relative contribution of the ZZZZ author.

There are two assumptions here. First, the ordinal positioning of any author in the list reflects his/ her relative contribution to the paper. In the absence of large power differential relationships (e.g., advisor/ student), this is probably a very reasonable assumption. In the presence of large power differential relationships, it may or may not be reasonable, but validation of the assumption would be next to impossible.

Second, the ordinal positioning can be quantified for computational purposes. There appears to be nothing in the literature that supports or rejects this assumption. For large numbers of papers undergoing citation analyses, anomolies will disappear, and quantification for estimation purposes may be reasonable. However, because of the uncertainty of the validity of this assumption, supplementary approaches were used to estimate the contribution of organization ZZZZ's researchers to overall paper quality. In this particular case, there were no significant differences in final results among the different methods used.

The total value of CAP summed over the 43 papers was 26.27, with an average value of 0.61, a median value of .92, and a mode of 1. Most papers were multi-authored; there were only four papers with one author. To summarize these results, the preponderance of papers that include an ZZZZ research unit author address have multiple authors, and the ZZZZ author is usually at the end of this list. The typical paper in this database had about three authors, with the ZZZZ author being last.

b. 1993 Database

i. Author Position Study

In the 1993 database, there were 44 papers in the SCI with an ZZZZ address. These papers had a total of 126 authors, with an average of 2.86 authors per paper, a median of 3 authors per paper, and a mode of 3 authors per paper. The total value of CAP summed over the 44 papers was 18.97, with an average value of .43, a median double value of 0/.5 (half the papers had a CAP of zero, the other half had a CAP of .5 or greater) and a mode of 0. The typical paper in this database had about three authors, with the ZZZZ author being second.

In comparison with the 1997 database results, the total number of papers is about the same. The median and mode of authors per paper is the same, but the average has dropped by a third from 1997 papers to

1993 papers. More importantly, the average CAP value dropped by a third from 1997 to 1993, the median CAP value dropped by a half, and the mode plummeted from one to zero. Thus, in 1993, the ZZZZ authors were contributing significantly more to papers (as measured by their ordinal position in the authors list) than in 1997.

ii. Citation Comparison Study

For the 1993 database, citations of pairs of similar theme papers were compared. In particular, for a given paper with a ZZZZ author address in the list, a similar theme paper was selected from the Related Records field, and the number of citations received by each paper was transcribed and compared. The procedure used was to select the first 1993 paper from the Related Records field with a similar theme to the target paper (this procedure normalized publication date and theme), and compare each paper's citations. (In a very few cases, no 1993 papers could be found in the Related Records field, and a 1994 or 1992 paper of similar theme was used. In a very few cases, no similar theme paper could be found for 1992 or 1994.)

Then, the ratio of citations of the two papers was transcribed, and this ratio was placed in one of five bands: very high (VH), high (H), same (S), low (L), very low (VL).

'Very High', for example, meant that the ratio of citations received by the related paper to the citations received by the ZZZZ paper was very high, a subjective judgement made by observation. 'Same' meant that the numbers of citations received by the two papers were close, not necessarily identical. Typically, citations received by a few of the other related papers would be examined to ascertain the approximate range of citations, and then judgements about the significance of the differences in citation numbers would be made. Obviously, in a definitive or final study of this nature, there would need to be people involved who could judge if in fact themes were closely related, and there would need to be citation distribution studies of related papers to obtain a more quantitative basis for judging significance of differences.

The population of the five bands was as follows: 12(VH); 9(H); 14(S); 4(L); 1(VL), for a total of 40 pairs where the citations could be compared. While the mode is in the S band, the median is in the H band. Since half the papers in the database had a CAP of zero, all other things being equal one would expect six papers in the VH band to have a CAP of zero. In actuality, nine papers in the VH band had a CAP of zero. Thus, those papers with a VH figure of merit tended to have more ZZZZ lead authors than one would expect from the database overall average.

There were seven prolific ZZZZ authors, each of whom participated in three or more papers. The population of the five bands for these seven prolific authors was: 1(VH); 5(H); 9(S); 3(L); 0(VL). Compared to the overall 1993 database, where 52.5% of the ZZZZ papers were in the VH or H bands, these seven authors had 33% of papers in the VH and H bands. Also, for these seven authors, the average CAP was .6, the median CAP was 0.8, and the mode CAP was 1. For the 1993 database, the parallel numbers were .43 (av), 0/.5 (med), 0 (mode). Thus, while the more prolific authors had better relative citeability than the database average, these authors were closer to the end of the author listing than the database average.

iii. Discussion

The highlights of this author position study are:
* The preponderance of 1997 papers that include a ZZZZ author address have multiple authors, and the ZZZZ author is usually at the end of this list. The typical paper in this database had about three authors, with the ZZZZ author being last.
* In 1993, the ZZZZ authors were contributing significantly more to papers (as measured by their ordinal position in the authors list) than in 1997. The typical paper in the 1993 database had about three authors, with the ZZZZ author being second.
* Those papers with a VH figure of merit tended to have more ZZZZ lead authors than one would expect from the database overall average.
* While the more prolific ZZZZ authors in 1993 had better relative citeability than the database average, these authors were closer to the end of the author listing than the database average.
* More work needs to be done to place ordinal position quantification on a stronger scientific foundation.

In about half the cases, papers with a ZZZZ author address were cited as well as, or better than, comparable non-ZZZZ address papers. On the surface, it appears that papers with ZZZZ authors are having a reasonable impact on the technical community. However, the contribution of the ZZZZ authors to these papers, especially those where the ZZZZ author is listed last, remains unknown. It would have been useful to compare the number of authors for each paper in the pair; this might have shed some light on whether or not the ZZZZ papers are 'author heavy'. This was not done because this issue was not recognized until now. It would also be useful to ascertain why the ZZZZ authors dropped back in their ordinal position in the author list from 1993 to 1997.

SUMMARY AND CONCLUSIONS

This appendix has provided two examples of the application of citation analysis to proposal evaluation. A number of lessons were learned concerning requirements for high quality citation analysis. These lessons are summarized as follows.

A. Since citation counts can vary sharply across sub-disciplines, absolute counts have little meaning, especially in the absence of absolute citation count performance standards. In order to provide meaning and context of citation counts for performance evaluation in citation analysis, some type of citation count normalization is required.

B. Three types of reference standards are used traditionally for citation analysis: 1) Reference standards based on prior sub-field classification; 2) Journals as reference standards; 3) Related records as reference standards. None of the above normalization methods are adequate for precise normalization, since they do not provide sufficient resolution for distinguishing among the lower level sub-fields. Inability to distinguish precisely among sub-fields translates, in some cases, to substitution of far different magnitude numbers for the normalization base

C. Selection of papers in the SCI thematically similar to the target paper depends strongly on the study's purpose and objectives, the mission of the performing organization, the degree of focus of the paper's theme, the size of the research paper pool from which to choose, and the level of technical description in the paper's SCI Abstract. The relation to study purpose is especially important, and is often overlooked. If the focus is on 'job right' quality, then the thematically similar papers will be limited to a very narrow area of inquiry. If the focus is on 'right job' quality, then the focus of thematically related papers can be expanded greatly. The practical implications of 'job right' vs 'right job' comparisons are that papers with substantially higher citation counts could be included in the normalization pool as the allowed definition of thematic similarity becomes broadened.

D. Selection of papers thematically similar to the target paper was very difficult, time-consuming, and subjective. This was especially true for the broad-based analyses. The selection was more straightforward for the much more limited specific technology papers, since these more focused areas seemed to have many researchers working related problems. The subjectivity involved in selecting thematically similar papers is a major source of uncertainty of the results. A rigorous study, in addition to having the rigorous information retrieval and statistical sampling processes mentioned in the next two paragraphs, requires the use of multiple evaluators for the same target papers to average out bias.

E. Many of the applied research target papers combined analytical technique advancement with novel application advancement. It was not always possible to have thematic similarity for both technique and application, especially in those research areas with relatively few performers. Typically, a choice had to be made between technique and application for determining thematic similarity.

F. Two important issues were i) determining the number of thematically similar candidate papers in the pool from which to choose, and then ii) determining the number of papers to select from the pool. First, in a credible study, candidate thematically similar papers would be identified by the most rigorous processes available, and such processes are presently very complex and time-consuming. Second, the number of papers to select from the candidate pool should have the greatest thematic similarity, and be representative statistically. Such selection would have required poring over hundreds, or thousands, of similar papers, and selecting a substantial number of the most representative thematically.

G. Contrary to much popular thinking, the technical expertise of the citation analyst can have a major impact on the quality of the results. The type of pair-wise comparison required for credible citation studies is a highly subjective process, requiring the selection of a thematically similar normalization base. If the analyst understands the subject matter, the subjective judgements made will be reasonably accurate. If the analyst is not a technical expert in the subject area, the results will contain a high degree of uncertainty. Thus, in a rigorous citation analysis, multiple technical experts are necessary to average out individual bias and subjectivity, and much manually intensive effort is required for the normalization process.

Operationally, the above results suggest that a credible citation analysis for determining performer or team quality should have the following components:

- Multiple technical experts to average out individual bias and subjectivity
- A process for comparing performer or team output papers with a normalization base of similar papers
- A process for retrieving a substantial fraction of candidate normalization base papers
- Manual evaluation of many candidate normalization base papers to obtain high thematic similarity and statistical representation

Since the use of citation analysis as one metric for determining research performer or team quality is substantially under-utilized in government and industry at present, the addition of the above requirements to the citation analysis process would only serve to reduce its utilization further. Pragmatically, tradeoffs are required if citation analysis is to be used as an evaluative tool. The degradation in citation analysis quality as the above conditions are relaxed needs to be studied further.

REFERENCES FOR APPENDIX 3-B

Albert, M.B., Avery, D., Narin F., Mcallister P., "Direct Validation Of Citation Counts As Indicators Of Industrially Important Patents" , Research Policy 20: (3) 251-259 , Jun 1991.

Braam, R.R., Moed H.F., Vanraan A.F.J., "Mapping Of Science By Combined Co-Citation And Word Analysis .1. Structural Aspects" , Science Technology & Human Values 13: (1-2) 97-98 ,Win-Spr 1988.

Braam, R.R., Moed, H.F., Vanraan A.F.J.,"Mapping Of Science By Combined Cocitation And Word Analysis .1. Structural Aspects", Journal Of The American Society For Information Science 42: (4) 233-251, May 1991.

Davidse, R. J., and VanRaan, A. F. J., "Out of Particles: Impact of CERN, DESY, and SLAC Research to Fields other than Physics", Scientometrics, 40:2. P. 171-193, 1997.

Del Río, J. A., Kostoff, R. N., García, E. O., Ramírez, A. M., and Humenik, J. A., "Citation Mining Citing Population Profiling using Bibliometrics and Text Mining". Centro de Investigación en Energía, Universidad Nacional Autonoma de Mexico, 2001. http://www.cie.unam.mx/W_Reportes.

DOE, "An Assessment of the Basic Energy Sciences Program", Office of Energy Research, Office of Program Analysis, Report No. DOE/ER-0123, March 1982.

Festschrift, ASIST Monograph Series, Web Of Knowledge - A Festschrift In Honor Of Eugene Garfield, 2000.

Garfield, E., “Citation Indexes For Science - New Dimension In Documentation Through Association Of Ideas”, Science, 122: (3159) 108-111, 1955.

Garfield , E., Sher , I. H., “New Factors In Evaluation Of Scientific Literature Through Citation Indexing” American Documentation, 14: (3) , 1963.

Garfield , E., “Science Citation Index-New Dimension In Indexing Unique Approach Underlies Versatile Bibliographic Systems For Communicating + Evaluating Information”, Science, 144: (361), 1964.

Garfield , E., “Can Citation Indexing Be Automated”, Statistical Association Methods For Mechanized Documentation Symposium Proceedings 1964: (Nbs26) 189, 1965.

Garfield E., “Patent Citation Indexing And Notions Of Novelty Similarity And Relevance”, Journal Of Chemical Documentation 6: (2), 1966.

Garfield., E., “Citation Indexing For Studying Science”, Nature 227: (5259) , 1970.

Geisler, E., “The Metrics of Science and Technology”, Quorum Books, Westport, CT, 2000.

Kostoff, R. N., "The Handbook of Research Impact Assessment," Seventh Edition, Summer 1997, DTIC Report Number ADA296021. Also, available at http://www.dtic.mil/dtic/kostoff/index.html, 1997a.

Kostoff, R. N., "Peer Review: The Appropriate GPRA Metric for Research", Science, Volume 277, 1 August 1997b.

Kostoff, R. N., "Citation Analysis Cross-Field Normalization: A New Paradigm", Scientometrics, 39:3, 1997c.

Kostoff, R. N., Eberhart, H. J., and Toothman, D. R., "Database Tomography for Information Retrieval", Journal of Information Science, 23:4, 1997d.

Kostoff, R. N., "The Use and Misuse of Citation Analysis in Research Evaluation", Scientometrics, 43:1, September, 1998.

Kostoff, R. N., “The Underpublishing of Science and Technology Results”, The Scientist, 1 May 2000a.

Kostoff, R. N., "High Quality Information Retrieval for Improving the Conduct and Management of Research and Development", Proceedings: Twelfth International Symposium on Methodologies for Intelligent Systems, 11-14 October 2000b.

Kostoff, R. N., and Hartley, J., “Structured Abstracts For Technical Journals”, Science, 11 May 2001a.

Kostoff, R. N., and Hartley, J., “Structured Abstracts For Technical Journal Articles, Letter to Technical Journal Editors, 14 May 2000b. Letter available from author.

MacRoberts, M.H., and MacRoberts, B.R., "Problems of Cit ation Analysis: A Critical Review,"Journal of the American Society for Information Science, 40:5, 1989.

MacRoberts, M., and MacRoberts, B., "Problems of Citation Analysis", Scientometrics, 36:3, July-August, 1996.

Marshako.Iv , “System Of Document Connections Based On References”, Nauchno-Tekhnicheskaya Informatsiya Seriya 2-Informatsionnye Protsessy I Sistemy, (6) 3-8 1973

Marshakova Iv , “Citation Networks In Information-Science”, Scientometrics, 3: (1) 13-25 1981

Marshakova Iv , “On The Mapping Of Science”,
Vestnik Akademii Nauk Sssr, (5) 70-82 1988

Martin, B. and Irvine, J., "Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio-Astronomy," Research Policy, 12, 1983.

Moed H.F., Vanraan A.F.J., “Observations And Hypotheses On The Phenomenon Of Multiple Citation To A Research Groups Oeuvre.”, Scientometrics 10: (1-2) 17-33 , July 1986.

Narin. , F., Carpenter M.P., “National Publication And Citation Comparisons”, Journal Of The American Society For Information Science 26: (2) 80-93 , 1975.

Narin, F., "Evaluative Bibliometrics: The Use of Publicati on and Citation Analysis in the Evaluation of Scientific Activity" (monograph), NSF C-637, National Science Foundation, Contract NSF C-627, NTIS Accession No. PB252339/AS, March 31, 1976.

Narin, F., Carpenter M.P., Woolf P., “Technological Performance Assessments Based On Patents And Patent Citations”, IEEE Transactions On Engineering Management 31: (4) 172-183 , 1984

Narin, F., Olivastro, D., and Stevens, K. A., "Bibliometrics -Theory, Practice, and Problems", in: Kostoff, R. N., (ed.), Evaluation Review, Special Issue on Research Impact Assessment, 18:1, February 1994.

Narin F., Hamilton K.S., “Bibliometric Performance Measures”, Scientometrics 36: (3) 293-310 , Jul-Aug 1996.

Nederhof A.J., Vanraan A,F.J., “Citation Theory And The Ortega Hypothesis”, Scientometrics 12: (5-6) 325-328 , Nov 1987

Schubert A., Glanzel W. , Braun T., “Relative Indicators Of Publication Output And Citation Impact Of European Physics Research: 1978-1980”, Czechoslovak Journal Of Physics 36: (1) 126-129 , 1986

Schubert A., Braun T., ”Reference-Standards For Citation Based Assessments”, Scientometrics 26: (1) 21-35 , Jan 1993

Schubert, A., and Braun, T., "Cross-Field Normalization of Scientometric Indicators", Scientometrics, 36:3, 1996.

Small , H. G. “Relationship Between Citation Indexing And Word Indexing - Study Of Co-Occurrences Of Title Words And Cited References”, Proceedings Of The American Society For Information Science 10: 217-218 , 1973.

Small , H., “Co-Citation In Scientific Literature - New Measure Of Relationship Between 2 Documents”, Current Contents (7) 7-10, 1974.

Small , H. G., “Co-Citation Model Of A Scientific Specialty - Longitudinal-Study Of Collagen Research”, Social Studies Of Science 7: (2) 139-166, 1977.

Small. , H., “The Relationship Of Information-Science To The Social-Sciences - A Co-Citation Analysis”, Information Processing & Management 17: (1) 39-50, 1981.

Small H., Sweeney E., Greenlee E., “Clustering The Science Citation Index Using Co-Citations .2. Mapping Science”, Scientometrics , 8, 1985.

Sullivan D, White Dh, Barboni Ej, “Co-Citation Analyses Of Science – Evaluation”, Social Studies Of Science, 7: (2) 223-240 1977

Sullivan D, Koester D, White Dh, Kern R, “Understanding Rapid Theoretical Change In Particle Physics - Month-By-Month Co-Citation Analysis,” Proceedings Of The American Society For Information Science, 16: 276-285 1979

Sullivan D, Koester D, White Dh, Kern R, “Understanding Rapid Theoretical Change In Particle Physics - A Month-By-Month Co-Citation Analysis,” Scientometrics, 2: (4) 309-319 1980

Vanraan, A.F.J., “Fractal Geometry of Information Space As Represented By Co-Citation-Clustering” , Scientometrics 20: (3) 439-449 , Mar-Apr 1991.

Vanraan, A.F.J., Tijssen R.J.W., “The Neural Net Of Neural Network Research - An Exercise In Bibliometric Mapping” , Scientometrics 26: (1) 169-192 , Jan 1993.

Vanraan, A.F.J., “Advanced Bibliometric Methods As Quantitative Core Of Peer Review Based Evaluation And Foresight Exercises”, Scientometrics 36: (3) 397-420 , Jul-Aug, 1996.

Zachlin, A. C., “On Literature Citation”, Science, 107: (2777) 292-293, 1948.

Zirkle, C., “Citation of Fraudulent Data”, Science, 120: (3109) 189-190, 1954.

Zsindely, S., Schubert A., Braun T., “Citation Patterns Of Editorial Gatekeepers In International Chemistry Journals”, Scientometrics 4: (1) 69-76 , 1982.

APPENDIX 3-C.

CITATION DIFFERENTIALS IN THE SCIENCE CITATION INDEX

ABSTRACT

The Science Citation Index allows computation of citation counts for a paper by two different methods. One approach is the Times Cited field associated with the paper of interest (Pi). The other is the Cited Reference Search capability. The Times Cited field essentially counts links between the SCI record of the Pi and the other SCI records that contain references to Pi in their Cited References field. Any errors in how Pi is referenced in these other SCI records will nullify a link. The Cited Reference Search capability lists all references for Pi, and groups them by similarity. One group is those references that have been entered correctly, and have established the link to the Times Cited field.

Citation counts for ten highly cited papers were computed for each method. The first author's name, as it appeared in the SCI record of the actual paper, was the only variant used for the experiment. The Times Cited count averaged about four percent less than the Cited Reference Search. This appeared due to errors in entering the journal volume, page, or year. Any errors in entering the first author's name would exacerbate this under-representation. From observation, the greatest source of author name error appeared to be in the treatment of the middle initial (exclusion, if the middle initial appeared in the SCI record of the actual paper).

BACKGROUND

A literature citation is a reference to the work of another. In modern times, the number of literature citations received by a research unit (presented paper(s), published paper(s), patent(s), author(s), group(s), etc) has evolved into one metric for impact of the research unit. Citations are one factor in making tenure, award, and prize decisions.

Two immediate questions arise relative to citations.

1) How valid are citations as a metric of impact?
2) How reliable are the citation counts obtained?

The first question has been addressed by many authors (e.g., 1-3), and will not be discussed further. This appendix addresses some aspects of the second question.

The focus of this appendix arose during the course of text mining (4,5) studies that the author was performing. The Science Citation Index (SCI) was being used to identify the number of citations received by specific papers in the study. One of the quantities calculated during the bibliometrics portion of the study was the number of citations received by highly cited papers. The author noticed differences in the number of times that a paper was cited, depending on the method used to calculate citations. This appendix provides estimates of these differences.

Before proceeding to the analysis, a brief discussion of the meaning of citations will be presented. A complete tabulation of citations received by a paper would require identification of all documents world-wide that contain the paper as a reference. This would include all journal papers, all conference papers, and perhaps magazine and newspaper articles as well. The central problem with obtaining the complete tabulation is the lack of databases that maintain citation information. To the author's knowledge, the SCI is the only comprehensive technical database that maintains citation information. Thus, all the sources excluded by the SCI from its database represent citations that will not be included in the tabulation. Those journals included in the SCI tend to be a good representation of the major research journals in the world. Thus, not only is a substantial portion of the technical literature excluded from the tabulation, but the literature that is included is skewed toward the research end of the technical spectrum. Very applied documents that may be referenced in more trade-oriented, or heavily applications-oriented, literatures will be very under-represented in citations shown in the SCI compared to citations potentially possible from all the literatures. Thus, the starting point for the present analysis is the truncated segment of the world's technical literature as represented by the SCI database.

ANALYSIS

Assume the unit of interest for the present analysis is a published document, and it is desired to obtain the number of citations received by this document. There are two major approaches used by the SCI to compute citations.

1) Times Cited Field

One of the fields in the SCI is named Times Cited. In practice, the number displayed for this field is the number of links between the paper of interest (hereafter called cited record) and the other records in the SCI database that contain the cited record in their reference lists. If the cited record has a very similar format structure and content to a record in a reference list, a link will be established with the citing document, and registered on the Times Cited counter. If the cited record has format/ content differences with a record in a reference list, then the record in the reference list will not be registered on the Times Cited counter. The record will appear, however, as a result of the next approach.

2) Cited Reference Search

The second approach used by the SCI to compute citations is the Cited Reference Search capability. To exercise this capability, the analyst enters Cited Author, Cited Work, Cited Year, to identify citations received by a specific paper. If all the citations for a specific author are desired for a specific year, then only the first and third entries are made. If all the citations for a given author are desired over time, then only the first entry is made.

If a specific paper is entered, this capability will display all the citations to the given paper. These citations can be divided into two groups. The first group is all those references that are linked to the paper of interest because of the closeness of the format/ contents. The numbers of

links are summed up, and the resultant number of citations highlighted. The first entry in Figure 1 shows an example for Fenn's 1989 paper in Science (6). This is one of the rows that would be displayed when using the Cited Reference Search capability. In the SCI, the analyst can click on this highlighted row, and the actual SCI record of Fenn's paper will be retrieved.

FIGURE 1 – CITED REFERENCE SEARCH EXAMPLES

Hits	Cited Author	Cited Work	Volume	Page	Year
1606	**FENN JB**	**SCIENCE**	**246**	**64**	**1989**
5	FENN JB	SCIENCE	264	64	1989
8	FENN JB	SCIENCE	246	46	1989
12	FENN JB	SCIENCE	246	64	1985
FEIGENBAUM MJ	**J STAT PHYS**	**19**	**25**	**1978**	
1	FEIGENBAUM JJ	J STAT PHYS	189	25	1978
1	FEIGENBAUM MF	J STAT PHYS	19	24	1978

The second group is all those references that are not linked to the cited record because of the differences of the format/ contents. Those non-linked references that are similar to each other are also summed up, but not highlighted. The second, third, and fourth entries in Figure 1 are examples from the Cited Reference Search of Fenn's paper. In the second entry, five references have interchanged the 4 and 6 in the Volume number. In the third entry, eight references have interchanged the 4 and 6 in the page number, and in the fourth entry, twelve references have the year wrong. There were no cases where reference was made to J Fenn (middle initial excluded).

The fifth entry in Figure 1 is an example for MJ Feigenbaum's 1978 paper in Journal of Statistical Physics (7). In the SCI, the analyst can click on this highlighted row, and the actual SCI record of Feigenbaum's paper will be retrieved. The sixth and seventh entries are lines where there were errors in Feigenbaum's first and middle initials, along with errors in other fields. In addition, forty references omitted the middle initial J altogether, and were listed as a few separate entries, not linked to the actual paper or highlighted.

Thus, it appears that five quantities have to be correct for a given reference in order for it to be linked to the Times Cited counter: Cited Author, Cited Work, Volume, Page, and Year. To estimate the number of records that would not be linked to the Times Cited counter due to errors in one or more of the above five quantities would be a monumental task. The central problem is identification of all possible variants of the first author's name. In the following analysis, the first author's name was extracted verbatim from the cited record, and was the only variant used for estimating the number of records that would not be linked to the Times Cited counter due to entry errors.

Ten highly cited papers were selected for the analysis. These are papers identified from text mining studies performed by the author over the past few years. To simplify the data analysis, papers were identified that were the only publications by a given author in a given journal for a

specific year. Table 2 summarizes the results. The left column is the first author, the middle column is the number of citations shown by the Times Cited field, next column is the number of citations computed from the Cited Reference Search, and the right column is the ratio of the Cited Reference Search citations to the Times Cited citations.

TABLE 2 – CITATION DIFFERENCES IN TEN PAPERS

AUTHOR	**# CITES**	**CIT_REF**	**RATIO**
FENN (6)	1606	1657	1.031756
FEIGENBAUM (7)	1612	1651	1.024194
KARAS (8)	1336	1455	1.089072
WHITEHOUSE (9)	653	660	1.01072
HILLENKAMP (10)	985	1007	1.022335
HUNT (11)	534	557	1.043071
ROE (12)	1334	1413	1.05922
KLINE (13)	771	805	1.044099
CURZON (14)	382	389	1.018325
MANDELBROT (15)	549	577	1.051002

The differences range from about one percent to nine percent, with a weighted average difference of four percent.

CONCLUSIONS AND RECOMMENDATIONS

On average, the Times Cited field in the SCI displays about 96% of the citations that would be obtained by the more detailed Cited Reference Search. Errors in first author name entries would exacerbate this under-representation, to an unknown degree. Probably the largest source of author name entry error is the treatment of the middle initial (based on spot checks using last name stemming followed by wildcards), but this statement is not definitive.

For statistical purposes in representing numbers of citations, the Times Cited field is adequate. For a more accurate representation, the Cited Reference Search would be required. Using a stem of the author's name (followed by wildcards) to obtain estimates of the differences due to name entry errors is very time consuming, and does not fully obviate the problem, since it is not known how the error would have impacted any stem selected. For almost any conceivable application, this additional level of complexity and time would not justify the probable slight increase in citation count accuracy.

REFERENCES FOR APPENDIX 3-C

1. Kostoff, R. N. The Use and Misuse of Citation Analysis in Research Evaluation. Scientometrics. 43:1. September 1998.
2. Macroberts, MH, Macroberts, BR. Problems Of Citation Analysis - A Critical-Review. Journal Of The American Society For Information Science. 40 (5): 342-349. Sep 1989.
3. Macroberts, MH, Macroberts, BR. Problems Of Citation Analysis. Scientometrics. 36 (3): 435-444. Jul-Aug 1996.
4. Kostoff, R. N., and DeMarco, R. A. "Science and Technology Text Mining". Analytical Chemistry. 73:13. 370-378A. 1 July 2001.
5. Kostoff, R. N., Del Rio, J. A., García, E. O., Ramírez, A. M., and Humenik, J. A. "Citation Mining: Integrating Text Mining and Bibliometrics for Research User Profiling". JASIST. 52:13. 1148-1156. 52:13. November 2001.
6. Fenn, JB, Mann, M, Meng, CK, Wong, SF, Whitehouse, CM. Electrospray Ionization For Mass-Spectrometry Of Large Biomolecules. Science. 246 (4926): 64-71. Oct 6 1989.
7. Feigenbaum, MJ. Quantitative Universality For A Class Of Non-Linear Transformations. Journal Of Statistical Physics. 19 (1): 25-52 1978.
8. Karas, M, Hillenkamp, F. Laser Desorption Ionization Of Proteins With Molecular Masses Exceeding 10000 Daltons. Analytical Chemistry. 60 (20): 2299-2301. Oct 15 1988.
9. Whitehouse, CM, Dreyer, RN, Yamashita, M, Fenn, JB. Electrospray Interface For Liquid Chromatographs And Mass Spectrometers. Analytical Chemistry. 57 (3): 675-679 1985.

10. Hillenkamp, F, Karas, M, Beavis, RC, Chait, BT. Matrix-Assisted Laser Desorption Ionization Mass-Spectrometry Of Biopolymers. Analytical Chemistry. 63 (24): A1193-A1202. Dec 15 1991.
11. Hunt DF, Yates JR, Shabanowitz J, Winston S, Hauer CR. Protein Sequencing By Tandem Mass-Spectrometry. Proceedings Of The National Academy Of Sciences Of The United States Of America. 83 (17): 6233-6237. Sep 1986.

12. Roe, PL. Approximate Riemann Solvers, Parameter Vectors, And Difference-Schemes. Journal Of Computational Physics. 43 (2): 357-372 1981.

13. Kline, SJ, Reynolds, WC, Schraub, FA, Runstadl, PW. Structure Of Turbulent Boundary Layers. Journal Of Fluid Mechanics. 30: 741 1967.
14. Curzon, FL, Ahlborn, B. Efficiency Of A Carnot Engine At Maximum Power Output. American Journal Of Physics. 43 (1): 22-24 1975.

15. Mandelbrot, BB, Passoja DE, Paullay AJ. Fractal Character Of Fracture Surfaces Of Metals. Nature. 308 (5961): 721-722 1984.

APPENDIX 4

DISPLAY OF BIBLIOMETRICS RESULTS

Indicators can be arranged in one or more dimensions. Emphasis has always been laid on the necessity of multidimensional thinking while analyzing scientometric indicators. Scientific research is a multifaceted human activity, and overemphasizing any of its aspects (publication productivity, citation influence, technological applicability, etc.) may lead to serious distortions in its assessment. While each scientometric indicator represents a single component of a multidimensional manifold which itself is just one element in assessing a complex system, presentations in one or several dimensions may equally prove useful [Braun, 1993].

The most direct way of presenting scientometric indicators is in one dimensional ranked lists. While simplistic, this approach reflects the paramount competitiveness of the scientific enterprise. Linear rankings are most attractive for presentation to the larger non-specialist audience (see Braun [1993]).

Two dimensional displays can include relational charts or scatter plots for correlations. In two dimensional relational charts [Schubert, 1986; Braun, 1987], pairs of indicators (observed vs. expected citation rates or attractivity vs. activity indices)are displayed in a planar orthogonal coordinate system. Emphasis is shifted from ranking to the formation of groups or 'clusters' and other characteristic relations among various indicators.

An obvious deficiency of the relational charts is the lack of any indication of the size of the sets of publications underlying the points of the diagram. By adding the third dimension of publication size, this objection can be overcome. The basic idea of 'landscaping' national scientific performances is to represent the size by the 'mass' of a mountain-like formation. If two or more countries have similar citation characteristics, the peaks representing them may get superimposed forming chains, massifs, and other surface formations. An example is presented in Braun [1991].

There seems to be a natural limit of graphical presentation at three dimensions. There are techniques, however, to overcome this apparent restriction. A rather original method of representing multivariate data was proposed by Herman Chernoff: "Each point in k-dimensional space, k<=18, is represented by a cartoon face whose features, such as length of nose and curvature of mouth, correspond to components of the point. Thus every multivariate observation is visualized as a computer drawn face. This presentation makes it easy for the human mind to grasp many of the essential regularities and irregularities present in the data."

Braun [1993] shows a face pattern with 18 facial features applicable in representing multidimensional data. Schubert [1992] contains a four-dimensional example of applying Chernoff-faces in scientometrics: uncitedness, citation rate per cited paper, mean expected citation rate and relative citation rate are represented by the shape of face, size of eyes, length of nose and curvature and length of mouth, respectively.

APPENDIX 5-A.

CITATION NORMALIZATION APPROACHES [Schubert, 1993]

1. **The Publishing Journal as Reference Standard**

Primary journals in science are generally agreed to contain coherent sets of papers both in contents and in professional standards. This coherence stems from the fact that most journals are nowadays specialized in quite narrow subdisciplines and the "gatekeepers" (i.e., the editors and referees) controlling the journal are members of an "invisible college" sharing their views on questions like relevance, validity or quality.

It seems, therefore, justified to expect the same level of citation rate for papers published in the same journal at the same time. If two such papers receive a different number of citations, one may rightly suspect that this reflects differences in their inherent qualities. By relating the number of citations received by a paper (or the average citation rate of a subset of papers published in the same journal - the **Mean Observed Citation Rate, MOCR**) to the average citation rate of all papers in the journal (the **Mean Expected Citation Rate, MECR**) the **Relative Citation Rate (RCR)** will be obtained. This indicator shows the relative standing of the paper (or set of papers) in question among its close companions: it value is higher\lower than unity as the sample is more\less cited than the average. In general, sets of papers under investigation are published in more than one journal; in that case, the mean expected citation rate (MECR) can be defined as the average citation rate of the journals. (The weights are, of course, the publication frequencies in the respective journals.) The mean observed citation rate (**MOCR**), i.e., the average citation rate per paper can again be related to the **MECR** to result in the relative citation rate (**RCR**), indicating the relative impact of the papers in question among the average papers of the publishing journals as reference standard.

There are some weaknesses inherent in using the publishing journal as reference standard. Papers published in multidisciplinary journals are measured by common standards, which might be clearly unfair, say, for a geoscience article published in Nature together with a molecular genetics paper. Since journals form a virtually continuous spectrum from highly specialized to multidisciplinary, and different research fields or even subcommunities in the same field may typically use different segments of this spectrum, the unbiasedness of the reference standards must be thoroughly checked whenever comparative assessments are based on the **RCR** indicator.

As a rule, it can be said that in coherent research fields, where papers are usually published in specialized journals (as is the general trend in contemporary science) published journals as reference standards and **RCR** as indicator can readily be proposed for comparative assessments. It must, however, be added that even in such cases extension from one to two dimensions may multiply the effectiveness of the analysis.

2. The Set of Related Records as Reference Standard

"Bibliographic Coupling" uses the number of references a given pair of documents have in common

to measure the similarity of their subject matter. Comparing a set of papers that are "similar" in this sense to a given article of the same age will yield an ideal reference standard for citation assessments. This apparently simple and straightforward method has long been practically un accomplishable because of the technical difficulties of collecting the "coupled" papers, by using any traditional version of citation indexes.

Fortunately, the situation has radically changed with the advent of the **CD-ROM edition** of the **Science Citation Index** database. The **SCI CD Edition** uses bibliographic coupling under the name **related records**. Two records are considered "related" when they list a number of identical papers in their respective bibliographies. Related records of an article are other articles published during the same period that cite at least one of the same references that the "parent" article cited. Because they have references in common, an article and its related records are supposed to be also related by subject. In general, the more references in common, the stronger the subject similarity between two articles. The **SCI CD Edition** has a built-in possibility for searching related records: a maximum of 20 related records are available for any given record ranked by strength of relatedness.

In an exploratory study of using **SCI CD Edition** for comparative evaluation of citation impact, the publication output of the Hungarian pharmaceutical company CHINOIN in 1986 was investigated. Three conclusions from the Study are:

a. Both for CHINOIN publications and for the "related records", observed citation rates per paper fall short of expected values. Thus it seems that the research topics of CHINOIN are not the "hottest spots" of their respective subject field, which does not, however, qualify the research in any means.

b. Although the expected citation rate of CHINOIN publications is rather close to that of the standard reference set ("related records"), their actual citation rate falls far below. Earlier studies concerning longer time periods did not show such a gap between expected and observed citation rates. The relatively low rate of subsequent year citations can most probably be attributed to insufficient informal, prepublication communication of research.

c. The observed citation rate of the related records is conspicuously close to the expected citation rate of the "parent" CHINOIN publications. This finding, in a sense, validates the use of relative scientometric indicators based on the comparison of actual with expected (journal average) citation rates. At least in the case of the present sample, the much more sophisticated "customized" control group-compiled on the principle of bibliometric coupling-obtains the same ci tation level as reference standard as did the simple journal average.

In subject fields less coherent than pharmaceutical research, however, the differences might be much more substantial, and the use of the set of related records as a more reliable reference standard is certainly worth the additional effort.

3. The Set of Cited Journals as Reference Standard

The set of publications to be assessed may represent various levels of aggregation, such as research

teams, institutions, or whole research communities of a given subfield in a given country. Independently of the level of investigation, the publishing journal is a useful and reliable reference standard for citation assessments - bearing in mind the caveats earlier mentioned. In one particular case, however, this approach fails completely, namely, if journals themselves are subjected to comparative assessment. There is an ever growing interest in evaluation of journals by citation analysis and one of the crucial questions, in this case too, is the comparison of journals publishing in science subfields of inherently different citation levels.

One possible solution might be again the use of related records. It is however, practically impossible to retrieve the related records to every single article of just one volume of a medium size journal and to collect their citations.

Standardization of citation levels by subfields and comparing the standardized scores has been attempted. This approach was found to be loaded with the inherent arbitrariness in the categorization of the journals into subfields and the ambiguity of treating inter- or multidisciplinary journals.

A method which now seems to provide the most satisfactory resolution at the lowest cost in terms of computer and\or manual search is based on the **journal in the reference lists of the articles of the journal in question**. These journals were selected by the most reliable persons, the authors of the journal as references (in both senses of the word) and therefore, can justly be regarded as standards of the expected citation rate.

All but a very few journals fall far below the standard set by their references. This is perhaps because authors tend to base their statements on the most authoritative sources. In every research area, a hierarchy of journals is set-up with one or just a few journals on the top and all others tend to cite "upwards".

A detailed study has been made on 2459 journals covered continuously by **SCI** in the period 1981-1985, and publishing at least 50 papers in these five years. Only 140 of them proved to be cited above the average of their cited references. This subset may rightly be considered the "chosen few" of the community of journals.

A closer look at this subset reveals that a considerable number of these journals are **review journals**, some of them having the work "review" even in their title. This is not too surprising, since review papers are well known to be cited much above the average. It is, however, interesting to realize that analysis of cited journals provides a simple means to distinguish review journals from "ordinary" ones. The indicator is the fraction of journal self-citations in all citations. Evidently, this fraction is much lower for review journals (collecting, by their very nature, references from a much wider pool of journals) than for primary journals.

APPENDIX 5-B.

CITATION ANALYSIS CROSS-FIELD NORMALIZATION: A NEW PARADIGM

[Kostoff, 1997i]

CROSS-FIELD CITATION NORMALIZATION: THE ISSUES

Science, Nature, Physics Today, Scientometrics, and other leading science and science evaluation journals continually publish articles comparing and ranking technical disciplines, departments, institutions, countries, and people on the basis of literature citations. Because of differences in numbers of researchers in different fields and in citing cultures, normalizations of absolute citation numbers to some reference are required to assign meaning to any comparisons. As shown in a recent review of cross-field citation normalization techniques, all present methods normalize citations of a given paper to citations of similar theme papers [Schubert, 1993; Appendix 5-A of the present document]]. The two main differences among these methods are how the similar theme papers are defined (e.g., papers published in same journal issue, papers sharing a threshold number of common references, etc.), and what types of mathematical/ statistical approaches are used to normalize the position of a target paper relative to that of its competitors. This limited comparative approach allows relative comparisons among similar papers, but ignores two crucial points. Purely relative comparison with other similar papers does not allow very credible comparisons among different disciplines based on citation analysis, and does not provide an indication of citation efficiency.

To gain wider acceptance and credibility, citation analysis needs to overcome these two limitations, and offer the broader perspective of how frequently a paper was cited compared to how frequently it could have been cited. The following sections describe a citation normalization method [Kostoff, 1997i] that would overcome the above two limitations, and provide the added dimension offered by the broader perspective.

CROSS-FIELD CITATION NORMALIZATION: A NEW PARADIGM

The fundamental concept of the new paradigm was derived from the thermodynamic principle of Carnot efficiency. The thermodynamic analog will be described through an illustrative example, and the metamorphosis to citation efficiency will then be shown.

Assume that two classes of engines are being evaluated. One class of engines (hereafter called fusion engines) has been developed to convert energy being produced in very high temperature fusion reactors, and the other class (hereafter called ocean engines) has been developed to convert energy from the temperature differentials in the deep ocean. Assume that there are three different fusion engines being evaluated in the fusion class, and the demonstrated conversion efficiencies of these engines are 1, 2, and 3 percent, respectively. Assume that there are three different ocean engines being evaluated in the ocean class, and the demonstrated conversion efficiencies of these engines are also 1, 2, and 3 percent, respectively.

If it were desired to evaluate the performance quality of all six engines, with efficiency being the

metric of quality, one simplistic approach would be to rank all six engines by demonstrated efficiency. The fusion engines would, on average, have equivalent quality to the ocean engines by this approach. However, a far better indicator of performance quality would be the ratio of each engine's demonstrated efficiency to the maximum efficiency the engine could achieve in its operating environment.

From thermodynamics, this maximum theoretical efficiency that each engine could achieve is the Carnot efficiency, which is a function of the high temperature and low temperature extremes in which the engine operates. For very high maximum temperatures and near-ambient low temperatures (characteristic of fusion), the Carnot efficiency approaches unity, and for low maximum temperatures and ambient low temperatures (characteristic of ocean), the Carnot efficiency approaches zero. If the comparison figure of merit becomes the ratio of demonstrated efficiency to Carnot efficiency, then the ocean engines in this case would outperform the fusion engines by a wide margin, since the ocean engines are operating closer to their theoretical maximum than are the fusion engines. Even where the engine evaluation is limited to one field (e.g., fusion), viewing relative performance from the new efficiency ratio perspective provides an added dimension for understanding performance, while the relative engine rankings within fusion remain unchanged.

Now the crossover from thermodynamic efficiencies to citation efficiencies will be made, with use of analogs to the above example. For fusion, convert each engine into a research paper of similar theme, and convert each engine efficiency into citations received by the research paper over some unit of time. Thus, there are now three fusion research papers of similar theme being compared which have 1, 2, and 3 citations over some unit of time, respectively. Similarly, for ocean, there are now three ocean papers of similar theme being compared which have 1, 2, and 3 citations over the same unit of time, respectively.

Generically, the existing orthodox approach to cross-field citation normalization might divide the number of fusion citations by the domain average (2.0) and provide each fusion paper a normalized value and ranking in its class. Thus, the paper with 3 citations might have a normalized value of 1.5 (3/ 2), and an upper 33 percentile ranking. Using similar normalization for the ocean papers and dividing citations by 2.0 (the domain average), the paper with 3 citations might have a normalized value of 1.5 (3/ 2), and an upper 33 percentile rating. The existing orthodox approach would consider the leading paper in each class as the same quality because of identical ranking in its class (upper 33 percentile).

However, as in the Carnot cycle analogy, a better figure of merit for quality would be the ratio of <u>actual number of citations received by a paper</u> to the <u>theoretical maximum number of citations that could be received by the paper</u>, a quantity which will be termed the <u>citation efficiency</u>. Then, different papers in the same field, as well as papers in different fields, could be compared on the basis of <u>citation efficiency</u>. The citation efficiency becomes the cross-field normalizer, and indicates how well a paper performed from a citation perspective compared to how well it could have performed. It is an intrinsic measure of accomplishment.

DETERMINATION OF CITATION EFFICIENCY

There are two crucial steps involved in determining the citation efficiency, and they are not completely independent. To compare a target paper to other papers, the first step is the selection of the universe of papers to be compared and the second step is the determination of the maximum number of citing papers to be used in the computation of efficiency. For present purposes, assume that a universe of papers to be compared to the target paper has been selected using existing techniques. Again, for present purposes, assume that this universe consists of sub-universes of papers with similar themes. Thus, the universe of fusion and ocean papers consists of a fusion sub-universe with similar themes and an ocean sub-universe with similar themes.

Next comes the determination of the maximum number of potential citing papers. The following theme-centered approach is proposed for computing maximum potential citations. For the fusion papers within the similar theme sub-universe, the maximum number of times one of the fusion papers could have been cited (in the given unit of time) is assumed to be equal to the number of different citing papers in which any of the papers in the fusion sub-universe were cited. Any of these citing papers could have cited 0, 1, or all of the similar theme fusion sub-universe papers. The same procedure for determining the maximum applies to the ocean papers, but the fusion maximum will probably be quite different from the ocean maximum. Then the citation efficiency of each paper in the selected universe can be computed, and the papers compared by this figure of merit. The actual number of citations of each fusion paper would be divided by the fusion paper maximum (this maximum is the same for all the fusion sub-universe papers) to arrive at the efficiency, and the actual number of citations of each ocean paper would be divided by the ocean paper maximum (this maximum is the same for all ocean sub-universe papers) to arrive at the efficiency.

The following figures illustrate how such an efficiency computation would be performed. Figure 1 is a matrix showing how many times each citing paper (A, B, C) cites each cited paper (G, H, I) for the ocean case.

FIGURE 1 - CITING PAPER VS CITED PAPER MATRIX: OCEAN

.....CITING PAPER

.....A..B..C

.....G...x..x..x

CITED.....H...x..x

PAPER.....I...x.

The x(s) in the matrix represent a citation. Thus, citing paper A cites papers G, H, and I, while citing paper C cites only paper G. The maximum number of potential citations for papers G, H, or I is 3, because there are three citing papers. The citation efficiency of G is 1 (3/ 3); the efficiency of H is .67 (2/ 3); and the efficiency of I is .33 (1/ 3).

Figure 2 is the same type of matrix for the fusion papers. The citing pattern has been changed.

FIGURE 2 - CITING PAPER VS CITED PAPER MATRIX: FUSION

.....CITING PAPER

.....A'.B'.C'.D'.E'.F'

.....G'...x..x..x..

CITED.....H'.....x..x

PAPER.....I'.....x

Now, each citing paper (A'-->F') cites only one of the fusion papers (G'-I'). The maximum number of potential citations for papers G', H', or I' is 6, because now there are six citing papers. The citation efficiency of G' is .5 (3/ 6); the efficiency of H' is .33 (2/ 6); the efficiency of I' is .17 (1/ 6).

Under the present normalization system, paper G would have
been rated as the same quality as paper G', since each ranked first in its own thematic sub-universe, and paper I would have been rated as the same quality as paper I', since each ranked last in its own thematic sub-universe. Under the new system proposed here, paper G ranks above paper G', and paper I ranks above paper I'. This is displayed more graphically in Figure 3, where the citation efficiencies of the ocean papers are obviously higher than their fusion counterparts.

FIGURE 3 - CITATION EFFICIENCY VS NUMBER OF CITATIONS
.....OCEAN VS FUSION

.....1.*.....Gx

.....*

.....*

.....*.....Hx

CITATION.....*

.....0.5*.....yG'

EFFICIENCY.....*

.....*....Ix.....yH'

.....*.....yI'

.....*

.....0.*.....*.....*.....*.....*....

.....0.....1.....2.....3

.....NUMBER.OF.CITATIONS.

Aggregate citation efficiencies may also be defined. Assume the aggregate citation efficiency of the group of ocean papers (G, H, I from figure 1) were desired. This quantity is the ratio of the number of citations received by papers G, H, and I (the number of asterisks in figure 1) to the maximum number of times these papers could have been cited (the nuumber of matrix elements in figure 1). For the figure 1 example, this aggregate citation efficiency is .67 (6/ 9), and for figure 2 this aggregate citation efficiency is .33 (6/ 18).

This example illustrates the added dimension provided by the citation efficiency perspective; the ability to evaluate and interpret research paper utilization patterns within and across different disciplines. Is the difference in aggregate efficiencies due to a different level of awareness of ocean and fusion authors of the intellectual foundations of their respective fields, and/ or is the difference due to the different levels of quality and uniqueness of the intellectual foundation papers in the different fields, and therefore different citation desireability of these papers? What other factors are operable?

Finally, the 'quality' of different citing journals (or any other quantified parameters associated with each journal) may be incorporated in the citation efficiency by computing a quality-weighted citation efficiency, or a quality-weighted aggregate citation efficiency.

SUMMARY

A new paradigm for comparing quality of published papers across different disciplines has been proposed. This method uses a figure of merit of the ratio of actual citations received to the potential maximum number of citations that could have been received. It is analogous to approaches used to compare performance in physical systems, and appears intrinsically more useful than present approaches.

APPENDIX 5-C.

IS CITATION NORMALIZATION REALISTIC [Kostoff, 2005j]

OVERVIEW

One method for assessing quality of research outputs across different technical disciplines is comparing citations received by the research output documents. However, cross-discipline citation comparison studies require discipline normalization, in order to eliminate discipline differences in cultural citation practices and discipline differences in number of active researchers available to cite. The 'definition' of, and number of documents used to represent, a discipline become critical. This study attempted to determine whether the citation characteristics (average, median) of a discipline's domain stabilized as the domain's size was decreased. A sample of papers (classified as *research articles only*, not review articles, by the Institute for Scientific Information) published in the journal Oncogene in 1999 was clustered hierarchically, and the citation averages and medians were computed for each cluster at different cluster hierarchical levels. The citation characteristics became increasingly stratified as the clusters were reduced in size, raising serious questions about the credibility of a selected denominator for normalization studies. An interesting side result occurred when all the retrieved articles were sorted by number of citations. Thirteen of the fifty most highly cited *research articles* had 100 or more references, whereas zero of the fifty least cited *research articles* had 100 or more references.

INTRODUCTION

Citation analysis is the quantitative and qualitative analysis of references in published documents (Narin, 1976; Kostoff, 2001). It is used mainly to identify historical trends in research disciplines, identify seminal documents, identify citer characteristics, and evaluate researcher/ research organization impact. Number of citations received by a document is a function of many variables, two of the most prominent being qualit y of the document's contents and number of researchers in the discipline(s) addressed by the document. To factor out the discipline effect (researcher candidate pool), especially when comparing research units across disciplines, some type of normalization is required. Various types of normalization have been used, including discipline normalization and journal normalization (Schubert and Braun, 1996). All these methods are founded on the belief that a discipline with nominal citation characteristics can be defined, thereby allowing some type of credible normalization.

The purpose of the present article is to examine citations of published papers in a given domain, allow the domain to get smaller, and ascertain whether isocitation regions of documents become relatively size-independent (the region-average citations would remain approximately constant as the region size changes). The approach started with a collection of documents from a technical 'discipline', performed document clustering that grouped the documents by similarity, allowed the groupings to get smaller, and thereby allowed the constituent documents of each group to

become more similar in technical content. If the average group member citation value changed with size, this would raise questions as to whether any of the groups could be used as a denominator for clustering, and would raise more serious questions about whether credible normalization is possible.

Toward that end, we selected a discipline-focused journal (Oncogene), and downloaded 490 records (with Abstracts) for 1999, from the Science Citation Index (SCI). Each record was classified by the SCI as *a research article*; none were classified as review papers or otherwise. For each record, we tabulated #references, #citations, #keywords, #Abstract words, and #title words.

We examined the relationships among #Abstract words, #cites, and #refs. We first sorted based on #Abstract words, but found no significant relationship of #cites with # Abstract words. Both the top 50 and the bottom 50 records had twelve articles with 40 or more cites. However, the top 50 had zero articles with more than 100 references, whereas the bottom 50 had seven. We then sorted by #cites. Thirteen of the top fifty had 100 or more references, whereas zero of the bottom 50 had 100 or more references.

We then used our document partitional clustering algorithm (CLUTO) to generate a four level hierarchical tree (taxonomy) structure (Karypis, 2004; Zhao, 2004) from the papers' Abstracts. Most of CLUTO's clustering algorithms treat the clustering problem as an optimization process that seeks to maximize or minimize a particular clustering criterion function defined either globally or locally over the entire clustering solution space. CLUTO uses a randomized incremental optimization algorithm that is greedy in nature, and has low computational requirements.

For the first hierarchical level, the clustering algorithm split the total database into two categories. As shown in Table 1, for average cites, one of the clusters had an average document citation of 27.4 citations per document, and the other had an average citation of 27.3. For the second level, the algorithm split each first level category into two sub-categories, so that we had four second level categories. For the third level, the algorithm split each second level category into two categories, and for the fourth level, the algorithm split each third level category into two sub-categories. The lowest (fourth level) clusters averaged thirty papers each. Then, for each category in each level, we computed both the average and median number of citations.

We found that as the domains became smaller and more focused, and the Abstracts in each domain (cluster) became more similar in technical content, the average and median citations became more stratified (see Table 1). This suggests that a different method for computing citation normalization factor is required than presently used. While our demo was performed on the papers in a single journal, we wouldn't have to limit the source to a single journal in practice. We could use a query-based retrieval, and cluster the retrieved articles thematically. The key point is to arrive at thematically very similar articles in each cluster to be used as a basis for comparison.

TABLE 1

AVERAGE CITES (STANDARD DEV) TOTAL # PAPERS			
LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4
		22.84615 (17.85385) 52	20.25 (14.61734) 16
	29.45333 (47.80168) 150		24 (19.19523) 36
		32.95918 (57.50247) 98	32.2 (65.26368) 60
27.40351 (40.46126) 228			34.15789 (43.29129) 38
		19.825 (14.25030) 40	23.08696 (16.07910) 23
	23.46154 (19.51269) 78		15.41176 (10.17385) 17
		27.28947 (23.43006) 38	31.52632 (28.88746) 19
			23.05263 (16.00164) 19
		30.93902 (39.50569) 82	29.46875 (20.18300) 37
	27.98658 (34.06769) 149		31.88 (48.16537) 50
		24.37313 (25.75045) 67	23.72727 (24.57675) 33
27.27099 (33.17963) 262			25 (27.19625) 34

		22.62687 (24.02450) 67	23.41176 (30.88896) 34
	26.32743 (32.09707) 113		21.81818 (14.32317) 33
		31.71739 (40.83498) 46	25.76471(38.95434) 17
			35.2069 (42.17428) 29

MEDIAN CITES (Inner Quartile Range) TOTAL # PAPERS			
LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4
		16 (17.85) 52	21 (14.62) 16
	18 (47.80) 150		16 (19.20) 36
		18 (57.50) 98	17 (65.26) 60
18 (40.46) 228			26 (43.29) 38
		16 (14.25) 40	19 (16.08) 23
	20 (19.51) 78		12 (10.17) 17
		24 (23.43) 38	28 (28.89) 19
			22 (16.00) 19
		24 (39.51) 82	24 (20.18) 37
	19 (34.07) 149		24 (48.17) 50
		17 (25.75) 67	17 (24.58) 33
19 (33.18) 262			17 (27.20) 34
		15 (24.02) 67	14 (30.89) 34
	18 (32.10) 113		22 (14.32) 33
		21 (40.84) 46	11 (38.95) 17
			28 (42.17) 29

We then examined those articles (records) with 100 or more references, and evaluated their citation ranking in their level 4 (lowest) category. The results are shown in Table 2 below,

ordered by category number.

TABLE 2

CITATION RANK IN
TAXONOMY LEVEL
4
ARTICLES WITH 100
OR MORE REFS

CATEG#	#REFS	#CITES	RANK
3	345	471	1/60
3	111	154	2/60
4	128	232	1/38
4	137	22	20/38
5	176	50	2/23
5	101	17	13/23
7	165	133	1/19
7	187	65	2/19
7	136	31	7/19
8	141	55	1/19
9	108	19	24/32
10	213	318	1/50
10	187	56	4/50
11	157	123	1/33
11	119	56	3/33
12	106	139	1/34
12	139	39	5/34
12	127	23	8/34
15	188	162	1/17

The first row can be interpreted as follows. In the first category that had an article with over 100 references, category 3 of level 4, this article had 345 references and 471 citations, and it ranked first (out of 60 records in that category) in citations in that category. Thus, out of the 19 records in the table, 8 records were first in their respective level 4 categories, 3 were second, and 1 was third.

If we raise the threshold on cutoff to 150, or even 200 references, the results are even more striking. There are eight records with 150 or more references, of which five rank first in their respective categories, two rank second, and one ranks fourth. There are two records with 200 or more references, and both rank first in citations in their relatively large categories.

Thus, the articles that have large numbers of references tend to be highly cited, especially when compared to strongly thematically related articles.

We then examined the other end of the spectrum. Table 3 shows the metrics for articles that contained the least references. There were 15 records with 18 or less references. Three were last

in their respective categories in citation ranking, and nine were in the bottom half. However, three were in the top quarter.

TABLE 3

ARTICLES WITH 18
OR LESS REFS

CATEG#	#REFS	#CITES	RANK
1	17	6	16/16
4	16	34	13/38
4	11	13	27/38
6	15	26	3/17
6	17	18	5/17
7	16	35	5/19
9	9	6	28/32
9	14	2	32/32
12	16	9	29/34
12	16	27	8/34
14	16	52	1/33
14	17	23	15/33
14	16	11	22/33
16	18	25	16/29
16	18	4	29/29

Finally, we examined the characteristics of the 16 articles that ranked at the top of their respective categories in terms of citations, and the 16 articles that ranked at the bottom. The next two tables, 4 and 5, display the metrics.

TABLE 4

HIGHEST CITED
RECORDS IN
EACH CATEGORY
- LEVEL 4

#REFS	#ABSWD	#CITES	#TTLWD	#KEYWD	CLUST#	ORDER-
72	112	243	8	25	49	63
106	117	139	19	25	11	50
213	136	318	8	23	55	39
345	139	471	15	20	34	13
38	141	67	16	21	62	23
188	157	162	33	24	0	61
16	158	52	9	10	36	58
157	164	123	16	23	28	44
141	165	55	21	18	25	30
34	172	42	14	20	42	25
39	189	148	9	17	57	54

128	214	232	17	27	4	19
55	228	85	8	20	45	34
165	240	133	9	23	18	27
54	261	81	20	19	20	4
72	283	45	25	22	16	2
113.9375	179.75	149.75	15.4375	21.0625	<<<<<<<<<<<<	AVERAGES OF ABOVE
89	164.5	128	15.5	21.5	<<<<<<<<<<<<	MEDIANS OF ABOVE

TABLE 5

LOWEST CITED RECORDS IN EACH CATEGORY - LEVEL 4

#REFS	#ABSWD	#CITES	#TTLWD	#KEYWD	CLUST#	ORDER-
24	148	0	14	19	16	2
29	105	4	23	15	17	5
20	172	1	17	25	13	10
29	189	0	8	21	24	18
29	235	2	20	21	58	24
24	191	4	12	20	42	25
28	189	4	13	18	27	29
50	195	4	9	18	9	32
14	185	2	20	17	41	36
38	179	0	19	19	59	40
32	305	5	15	19	51	43
43	217	7	16	22	37	49
65	189	2	9	23	60	51
54	184	3	10	21	44	55
52	137	0	22	21	0	61
18	136	4	10	14	54	64
34.3125	184.75	2.625	14.8125	19.5625	<<<<<<<<<<<<	AVERAGES OF ABOVE
29	187	2.5	14.5	19.5	<<<<<<<<<<<<	MEDIANS OF ABOVE

The major difference in both the average and median values is number of references.

In summary, to compare the quality/ impact of different research papers as represented by citations, the papers should be as similar thematically and typically (research article, review article, etc) as possible. Publication dates, journals, and other factors should be normalized, where possible. For the Oncogene test case, segregation according to thematic similarity resulted in changing group citation averages. This suggests that a meaningful 'discipline' citation average may not exist, and the mainstream large-scale mass production semi-automated citation analysis comparisons may provide questionable results. It further suggests that meaningful cross-discipline citation comparisons require the manually intensive approach of identifying those few research papers most closely related to the paper of interest, and normalizing on those papers (Kostoff, 2002). Finally, it confirms what many research evaluators recognize instinctively: there are really relatively few very thematically similar technical articles in any discipline, and any metrics used to evaluate research should be based on this reality.

REFERENCES FOR APPENDIX 5-C.

Kostoff RN, Del Rio JA, García EO, Ramírez AM, Humenik JA. (2001). Citation Mining: Integrating Text Mining and Bibliometrics for Research User Profiling. Journal of the American Society for Information Science and Technology. 52:13. 1148-1156.

Narin F. (1976). Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity (monograph). NSF C-637. National Science Foundation. Contract NSF C-627. NTIS Accession No. PB252339/AS.

Kostoff, R. N. (2002). Citation Analysis for Research Performer Quality. Scientometrics. 53:1. 49-71.

Schubert, A., Braun, T. (1996). Cross-Field Normalization of Scientometric Indicators, Scientometrics. 36:3. 311-324.

Karypis G. CLUTO—A Clustering Toolkit. http://www.cs.umn.edu/~cluto

Zhao Y, Karypis G. Criterion Functions for Document Clustering: Experiments and Analysis. Machine Learning, in press.

APPENDIX 5-D

CAB - CITATION-ASSISTED BACKGROUND [Kostoff, 2005g]

ABSTRACT

A chronically weak area in research papers, reports, and reviews is the complete identification of background documents that formed the building blocks for these papers. A method for systematically determining these seminal references is presented. Citation-Assisted Background (CAB) is based on the assumption that seminal documents tend to be highly cited. CAB is being applied presently to three applications studies, and the results so far are much superior to those used by the first author for background development in any other study. An example of the application of CAB to the field of Nonlinear Dynamics is outlined. While CAB is a highly systematic approach for identifying seminal references, it is not a substitute for the judgement of the researchers, and serves as a supplement.

INTRODUCTION

Research is a method of systematically exploring the unknown to acquire knowledge and understanding. Efficient research requires awareness of all prior research and technology that could impact the research topic of interest, and builds upon these past advances to create discovery and new advances. The importance of this awareness of prior art is recognized throughout the research community. It is expressed in diverse ways, including requirements for Background sections in journal research articles, invited literature surveys in targeted research areas, and required descriptions of prior art in patent applications.

For the most part, development of Background material for any of the above applications is relatively slow and labor intensive, and limited in scope. Background material development usually involves some combination of manually sifting through outputs of massive computer searches, manually tracking references through multiple generations, and searching ones own records for personal references. The few studies that have been done on the adequacy of Background material in documents show that only a modest fraction of relevant material is included (MacRoberts and MacRoberts, 1989, 1996; Liu, 1993; Calne and Calne, 1992; Shadish et al, 1995; Moravcsik and Murugesan, 1975).

In particular, an analysis of Medline papers on the haemodynamic response to orotracheal intubation showed that recognized deficiencies in research method were not acknowledged. The authors recommended that, when submitting work for publication, investigators should provide evidence of how they searched for previous work (Smith and Goodman, 1997).

Another specific example was provided by MacRoberts and MacRoberts (1997). Replicating their earlier work in a journal on genetics which indicated that only 30% of influences evident in text are reflected in a paper's references, the text of an issue of Sida was studied by the

MacRoberts to extract influences of previous work evident therein. Influences they judged present in the text appeared in the references only 29% of the time.

Typically missing from standard Background section or review article development, as well as in the specific examples cited above, is a systematic approach for identifying the key documents and events that provided the groundwork for the research topic of interest. The present paper presents such a systematic approach for identifying the key documents, called Citation-Assisted Background (CAB). The next section describes the CAB concept, and provides an outline of its operation, with an illustrative example from the research area of Nonlinear Dynamics.

CONCEPT DESCRIPTION

The CAB concept identifies the key Background documents for a research area using citation analysis. CAB rests on the assumption that a document that is a significant building block for a specific research area will typically have been referenced positively by a substantial number of people who are active researchers in that specific area. Implementation of the CAB concept then requires the following steps:

- The research area of interest must be defined clearly
- The documents that define the area of interest must be identified and retrieved
- The references most frequently used in these documents must be identified and selected
- These critical references must be analyzed, and integrated in a cohesive narrative manner to form a comprehensive Background section or separate literature survey

These required steps are achieved in the following manner.

1. The research topic of interest is defined clearly by the researchers who are documenting their study results. For example, consider the research area of Nonlinear Dynamics. In a recent text mining study of Nonlinear Dynamics (Kostoff et al, 2004), the research area was defined as "that class of motions in deterministic physical and mathematical systems whose time evolution has a sensitive dependence on initial conditions."
2. The topical definition is sharpened further by the development of a literature retrieval query. In the text mining study mentioned above, the literature retrieval query was ((CHAO* AND (SYSTEM* OR DYNAMIC* OR PERIODIC* OR NONLINEAR OR BIFURCATION* OR MOTION* OR OSCILLAT* OR CONTROL* OR EQUATION* OR FEEDBACK* OR LYAPUNOV OR MAP* OR ORBIT* OR ALGORITHM* OR HAMILTONIAN OR LIMIT* OR QUANTUM OR REGIME* OR REGION* OR SERIES OR SIMULATION* OR THEORY OR COMMUNICATION* OR COMPLEX* OR CONVECTION OR CORRELATION* OR COUPLING OR CYCLE* OR DETERMINISTIC OR DIMENSION* OR DISTRIBUTION* OR DUFFING OR ENTROPY OR EQUILIBRIUM OR FLUCTUATION* OR FRACTAL* OR INITIAL CONDITION* OR INVARIANT* OR LASER* OR LOGISTIC OR LORENZ OR MAGNETIC FIELD* OR MECHANISM* OR MODES OR NETWORK* OR ONSET OR TIME OR FREQUENC* OR POPULATION* OR STABLE OR ADAPTIVE OR CIRCUIT* OR DISSIPAT* OR

EVOLUTION OR EXPERIMENTAL OR GROWTH OR HARMONIC* OR HOMOCLINIC OR INSTABILIT* OR OPTICAL)) OR (BIFURCATION* AND (NONLINEAR OR HOMOCLINIC OR QUASIPERIODIC OR QUASI-PERIODIC OR DOUBLING OR DYNAMICAL SYSTEM* OR EVOLUTION OR INSTABILIT* OR SADDLE-NODE* OR MOTION* OR OSCILLAT* OR TRANSCRITICAL OR BISTABILITY OR LIMIT CYCLE* OR POINCARE OR LYAPUNOV OR ORBIT*)) OR (NONLINEAR AND (PERIODIC SOLUTION* OR OSCILLAT* OR MOTION* OR HOMOCLINIC)) OR (DYNAMICAL SYSTEM* AND (NONLINEAR OR STOCHASTIC OR NON-LINEAR)) OR ATTRACTOR* OR PERIOD DOUBLING* OR CORRELATION DIMENSION* OR LYAPUNOV EXPONENT* OR PERIODIC ORBIT* OR NONLINEAR DYNAMICAL) NOT (CHAO OR CHAOBOR* OR CHAOTROP* OR CAROTID OR ARTERY OR STENOSIS OR PULMONARY OR VASCULAR OR ANEURYSM* OR ARTERIES OR VEIN* OR TUMOR* OR SURGERY)

3. The query is entered into a database search engine, and documents relevant to the topic are retrieved. In the text mining study mentioned above, 6160 documents were retrieved from the Web version of the Science Citation Index (SCI) for the year 2001. The SCI was used because it is the only major research database to contain references, in a readily extractable format.
4. These documents are combined to create a separate database, and all the references contained in these documents are extracted. Identical references are combined, the number of occurrences of each reference is tabulated, and a table of references and their occurrence frequencies is constructed. In the text mining study on Nonlinear Dynamics, 113176 separate references were extracted and tabulated. Table 1 contains the twenty highest frequency (most cited) references extracted from the Nonlinear Dynamics database.

TABLE 1 – MOST HIGHLY CITED DOCUMENTS

AUTHOR	**YEAR**	**SOURCE**	**VOL**	**PAGE**	**# CIT**
PECORA LM	1990	PHYS REV LETT	V64	P821	177
GUCKENHEIMER J	1983	NONLINEAR OSCILLATIO			149
OTT E	1990	PHYS REV LETT	V64	P1196	142
LORENZ EN	1963	J ATMOS SCI	V20	P130	115
CROSS MC	1993	REV MOD PHYS	V65	P851	105
WOLF A	1985	PHYSICA D	V16	P285	103
TAKENS F	1981	LECT NOTES MATH	V898	P366	97
OTT E	1993	CHAOS DYNAMICAL SYST			97
GRASSBERGER P	1983	PHYSICA D	V9	P189	94
GUTZWILLER MC	1990	CHAOS CLASSICAL QUAN			88
ROSENBLUM MG	1996	PHYS REV LETT	V76	P1804	77
GRASSBERGER P	1983	PHYS REV LETT	V50	P346	76
ECKMANN JP	1985	REV MOD PHYS	V57	P617	75
THEILER J	1992	PHYSICA D	V58	P77	66

NAYFEH AH	1979	NONLINEAR OSCILLATIO			62
FUJISAKA H	1983	PROG THEOR PHYS	V69	P32	61
WIGGINS S	1990	INTRO APPL NONLINEAR			61
RULKOV NF	1995	PHYS REV E	V51	P980	59
PYRAGAS K	1992	PHYS LETT A	V170	P421	59
LICHTENBERG AJ	1992	REGULAR CHAOTIC DYNA			58

Two frequencies are computed for each reference, but only the first is shown in Table 1. The frequency shown in the rightmost column is the number of times each reference was cited by the 6160 records in the retrieved database only. This number reflects the importance of a given reference to the specific discipline of Nonlinear Dynamics. The second frequency number (not shown) is the total number of citations the reference received from all sources, and reflects the importance of a given reference to all the fields of science that cited the reference. This number is obtained from the citation field or citation window in the SCI. In CAB, only the first frequency is used, since it is topic-specific. Using the first discipline-specific frequency number obviates the need to normalize citation frequencies for different disciplines (due to different levels of activity in different disciplines), as would be the case if total citation frequencies were used to determine the ordering of the references.

Before presenting a specific implementation algorithm for the Nonlinear Dynamics example, a few caveats will be discussed. First, listing and selection of the most highly cited references are dependent on the comprehensiveness and balance of the total records retrieved. Any imbalances (from skewed databases or incorrect queries) can influence the weightings of particular references, and result in some references exceeding the selection threshold where not warranted, and others falling below the threshold where not warranted.

Second, it is important that the query used for record retrieval be extensive (Khan and Khor, 2004; Harter and Hert, 1997; Kantor, 1994), as was shown for the Nonlinear Dynamics example. The query needs to be checked for precision and recall, which becomes complicated when assumptions of binary relevance and binary retrieval are relaxed (Della Mea and Mizzaro, 2004). There are a multitude of issues to be considered when evaluating queries and their impact on precision and recall. A recent systems analytic approach to analyzing the information retrieval process concludes that, for completeness, the interaction of the Environment and the information retrieval system must be considered in query development (Kagolovsky and Moehr, 2004). The first author's experiences (with the four studies done so far with CAB, including the study reported in this paper) have shown that modest query changes may substitute some papers at the citation selection threshold, but the truly seminal papers have citations of such magnitude that they are invulnerable to modest query changes. For this reason, the cutoff threshold for citations has been, and should be, set slightly lower, to compensate for query uncertainties.

Third, there may be situations where at least minimal citation representation is desired from each of the major technical thrust areas in the documents retrieved. In this case, the retrieved documents could be clustered into the major technical thrust areas, and the CAB process could

be performed additionally on the documents for each cluster. The additional references identified with the cluster-level CAB process, albeit with lower citations than from the aggregated non-clustered CAB process, would then be added to the list obtained with the aggregated CAB process. The first author has not found this cluster-level CAB process necessary for any of the four disciplines studied with CAB so far.

Fourth, there may be errors in citation counts due to references errors, and the subsequent fragmenting of a reference's occurrence frequency metric into smaller metric values. Care needs to be taken in insuring that a given reference is not fissioned into multiple large fragments, that are not subsequently combined.

How large would this fragmenting effect be? There have been a number of published studies estimating these types of data entry errors on SCI citation results (Gosling et al, 2004; Fenton et al, 2000; Putterman et al, 1991). Essentially all the articles retrieved used the same approach. They selected a sample of journal papers from a journal or journals, and compared the references against the originals. In the words of one of the retrieved papers' authors: "To evaluate the reference accuracy in the Journal of Dermatology and the Korean Journal of Dermatology, we randomly selected 100 references from each journal and checked them against the original articles." (Lee and Lee, 1999). They generated metrics for citation errors, and presented the results statistically. There was a range of results, but 'significant' errors appeared to be in the range of about ten percent.

The first author did a study in early 2003 (unpublished) examining the differences between numerical outputs in the Times Cited field in the Science Citation Index (SCI) and the Cited Reference Search capability in the SCI. This difference reflected the error in entering reference data in the SCI, and would directly lead to fragmenting of the reference occurrence frequency metrics.

The SCI allows computation of citation counts for a paper by two different methods. One approach is the Times Cited field associated with the paper of interest (Pi). The other is the Cited Reference Search capability. The Times Cited field essentially counts links between the SCI record of the Pi and the other SCI records that contain references to Pi in their Cited References field. Any errors in how Pi is referenced in these other SCI records will nullify a link. The Cited Reference Search capability lists all references for Pi, and groups them by similarity. One group is those references that have been entered correctly, and have established the link to the Times Cited field.

Citation counts for ten highly cited papers were computed for each method. The first author's name, as it appeared in the SCI record of the actual paper, was the only variant used for the experiment. The Times Cited count averaged about four percent less than the Cited Reference Search. This appeared due to errors in entering the journal volume, page, or year. Any errors in entering the first author's name would exacerbate this under-representation. From observation, the greatest source of author name error appeared to be in the treatment of the middle initial (exclusion, if the middle initial appeared in the SCI record of the actual paper). In the study

above, not all the errors made in entering data could be identified, and therefore the four percent number is a lower bound on the differential.

For statistical purposes in representing numbers of citations, the Times Cited field is adequate. For a more accurate representation, the Cited Reference Search would be required. Using a stem of the author's name (followed by wildcards) to obtain estimates of the differences due to name entry errors is very time consuming, and does not fully obviate the problem, since it is not known how the error would have impacted any stem selected. For almost any conceivable application, this additional level of complexity and time would not justify the probable slight increase in citation count accuracy.

Fifth, the CAB approach is most accurate for recent references, and its accuracy drops as the references recede into the distant past. This results from the tendency of authors to reference more recent documents and, given the restricted real estate in journals, not reference the original documents. To get better representation, and more accurate citation numbers, for early historical documents, the more recent references need to be retrieved, collected into a database, and have their references analyzed in a similar manner (essentially examining generation of citations).

As an example of what would be required for the early historical documents, assume 150 reference documents are selected for the primary Background study, and the retrieved database is for 2001. Assume there is an average of twenty references per retrieved record for a total of 3000 references. Assume half of these references are in the SCI, for a total of 1500 references. All these 1500 references could be retrieved, could constitute the new database, the critical references in this database could be identified, and the process repeated ad infinitum. Or, to make the numbers more manageable in terms of number of iterations required, an upper limit on publication date could be specified for each succeeding iteration. Thus, for an initial retrieval of 2001 as in the example, the next retrieval could be for references prior to 1980, then the following retrieval would be for references prior to 1960. However, for most literature surveys, this iterative approach would be un-necessary, since recent references tend to be of primary interest.

Sixth, high citation frequencies are not unique to seminal documents only; different types of references can have high citation frequencies. Documents that contain critical research advances, and were readily accessible in the open literature, tend to be cited highly, and represent the foundation of the CAB approach. Application of CAB to three technical research areas so far (in addition to the present Nonlinear Dynamics study) shows that this type of document is predominant in the highly cited references list. Books or review articles also appear on the highly cited references list. These documents do not usually represent new advances, but rather are summaries of the state of the art (and its Background) at the time the document was written. These types of documents are still quite useful as Background material. Finally, documents that receive large numbers of citations highly critical of the document could be included in the list of highly cited documents. In three studies so far, the first author has not identified such papers in the detailed development of the Background.

Additionally, one of the three application studies concerns high speed compressible flow, a discipline in which the first author worked decades ago. Using the CAB approach, the first author found that all the key historical documents with which he was familiar were identified, and all the historical documents identified appeared to be important. Thus, for that data point at least, the weaknesses identified above (imbalances, undervaluing early historical references, unwanted highly cited documents) did not materialize. To insure that any critical documents were not missed because of imbalance problems, the threshold was set a little bit lower to be more inclusive.

The converse problem to multiple types of highly cited references, some of which may not be the seminal documents desired, is influential references that do not have substantial citation frequencies. If the authors of these references did not publish them in widely and readily accessible forums, or if they do not contain appropriate verbiage for optimal query accessibility, then they might not have received large numbers of citations. Additionally, journal or book space tends to be limited, with limited space for references. In this zero-sum game for space, research authors tend to cite relatively recent records at the expense of the earlier historical records. Also, extremely recent but influential references have not had the time to accumulate sufficient citations to be listed above the selection threshold on the citation frequency table. Methods of including these influential records located at the wings of the temporal distribution will be described in the following implementation section. Inclusion of the references that were not widely available when published is more problemmatical, and tends to rely on the Background developers' personal knowledge of these documents, and their influence.

CONCEPT IMPLEMENTATION

To identify the total candidate references for the Background section, a table similar in structure to Table 1, but containing all the references from the retrieved records, is constructed. A threshold frequency for selection can be determined by arbitrary inspection (i.e., a Background section consisting of 150 key references is arbitrarily selected). The first author has found a dynamic selection process more useful. In this dynamic process, references are selected, analyzed, and grouped based on their order in the citation frequency table until the resulting Background is judged sufficiently complete by the Background developers.

To insure that the influential documents at the wings of the temporal distribution are included, the following total process is used. The reference frequency table is ordered by inverse frequency, as above, and a high value of the selection frequency threshold is selected initially. Then, the table is re-ordered chronologically. The early historical documents with citation frequencies substantially larger than those of their contemporaries are selected, as are the extremely recent documents with citation frequencies substantially larger than those of their contemporaries. By contemporaries, it is meant documents published in the same time frame, not limited to the same year. Then, the dynamic selection process defined above is applied to the early historical references, the intermediate time references (those falling under the high frequency threshold), and the extremely recent references.

Table 2 is an example of the final references that would have been selected for the Background section of the Nonlinear Dynamics study using CAB, had an extensive Background section been desired. The first reference listed, Einstein's 1917 paper, had many more citations than any papers published in the 1910s or 1920s. In fact, there were half a dozen papers published between 1831 and 1931 that had four citations each, and these were the closest to Einstein's paper. This is a graphic example of how we interpret a paper's having substantially more citations than its contemporaries.

TABLE 2 – SEMINAL DOCUMENTS SELECTED FOR INCLUSION IN BACKGROUND

AUTHOR	**YEAR**	**SOURCE**	**VOL**	**PAGE**	**# CIT**	**BACKGR**
EINSTEIN A	1917	VERHAND DEUT PHYS GE	V19	P82	13	Y
LAMB H	1932	HYDRODYNAMICS			14	Y
WIGNER E	1932	PHYS REV	V40	P749	11	Y
KOLMOGOROV AN	1937	B MGU A	V1	P1	10	Y
HUSIMI K	1940	P PHYS-MATH SOC JPN	V22	P264	10	Y
GABOR D	1946	J I ELEC ENG 3	V93	P429	11	Y
HODGKIN AL	1952	J PHYSIOL-LONDON	V117	P500	30	Y
TURING AM	1952	PHILOS T ROY SOC B	V237	P37	27	Y
CODDINGTON EA	1955	THEORY ORDINARY DIFF			15	Y
ANDERSON PW	1958	PHYS REV	V109	P1492	21	Y
FITZHUGH R	1961	BIOPHYS J	V1	P445	24	Y
CHANDRASEKHAR S	1961	HYDRODYNAMIC HYDROMA			23	Y
LORENZ EN	1963	J ATMOS SCI	V20	P130	115	Y
MELNIKOV VK	1963	T MOSCOW MATH SOC	V12	P1	23	Y
HENON M	1964	ASTRON J	V69	P73	18	Y
SMALE S	1967	B AM MATH SOC	V73	P747	19	Y
OSELEDEC VI	1968	T MOSCOW MATH SOC	V19	P197	25	Y
GUTZWILLER MC	1971	J MATH PHYS	V12	P343	42	Y
RUELLE D	1971	COMMUN MATH PHYS	V20	P167	23	Y
ZAKHAROV VE	1972	SOV PHYS JETP-USSR	V34	P62	21	Y
NAYFEH AH	1973	PERTURBATION METHODS			24	Y
HENON M	1976	COMMUN MATH PHYS	V50	P69	41	Y
ROSSLER OE	1976	PHYS LETT A	V57	P397	39	Y
MAY RM	1976	NATURE		P459	35	Y

			V261			
BENETTIN G	1976	PHYS REV A	V14	P2338	27	Y
MACKEY MC	1977	SCIENCE		P287	35	Y
			V197			
NICOLIS G	1977	SELF ORG NONEQUILIBR			26	Y
FEIGENBAUM MJ	1978	J STAT PHYS	V19	P25	28	Y
NAYFEH AH	1979	NONLINEAR OSCILLATIO			62	Y
CHIRIKOV BV	1979	PHYS REP	V52	P263	43	Y
PACKARD NH	1980	PHYS REV LETT	V45	P712	54	Y
LANG R	1980	IEEE J QUANTUM ELECT	V16	P347	29	Y
WINFREE AT	1980	GEOMETRY BIOL TIME			25	Y
TAKENS F	1981	LECT NOTES MATH		P366	97	Y
			V898			
BRODY TA	1981	REV MOD PHYS	V53	P385	35	Y
HOPFIELD JJ	1982	P NATL ACAD SCI-BIOL	V79	P2554	37	Y
GUCKENHEIMER J	1983	NONLINEAR OSCILLATIO			149	Y
GRASSBERGER P	1983	PHYSICA D	V9	P189	94	Y
GRASSBERGER P	1983	PHYS REV LETT	V50	P346	76	Y
FUJISAKA H	1983	PROG THEOR PHYS	V69	P32	61	Y
GREBOGI C	1983	PHYSICA D	V7	P181	26	Y
BOHIGAS O	1984	PHYS REV LETT	V52	P1	54	Y
KURAMOTO Y	1984	CHEM OSCILLATIONS WA			49	Y
HELLER EJ	1984	PHYS REV LETT	V53	P1515	44	Y
AREF H	1984	J FLUID MECH		P1	29	Y
			V143			
WOLF A	1985	PHYSICA D	V16	P285	103	Y
ECKMANN JP	1985	REV MOD PHYS	V57	P617	75	Y
BERRY MV	1985	P ROY SOC LOND A MAT		P229	35	Y
			V400			
MILNOR J	1985	COMMUN MATH PHYS	V99	P177	28	Y
FRASER AM	1986	PHYS REV A	V33	P1134	49	Y
THEILER J	1986	PHYS REV A	V34	P2427	34	Y
BROOMHEAD DS	1986	PHYSICA D	V20	P217	26	Y
FARMER JD	1987	PHYS REV LETT	V59	P845	36	Y
SKARDA CA	1987	BEHAV BRAIN SCI	V10	P161	25	Y
TEMAM R	1988	INFINITE DIMENSIONAL			31	Y
PARKER TS	1989	PRACTICAL NUMERICAL			40	Y
OTTINO JM	1989	KINEMATICS MIXING ST			35	Y
CASDAGLI M	1989	PHYSICA D	V35	P335	32	Y
OSBORNE AR	1989	PHYSICA D	V35	P357	25	Y
PECORA LM	1990	PHYS REV LETT	V64	P821	177	Y
OTT E	1990	PHYS REV LETT	V64	P1196	142	Y
GUTZWILLER MC	1990	CHAOS CLASSICAL QUAN			88	Y
WIGGINS S	1990	INTRO APPL NONLINEAR			61	Y

SUGIHARA G	1990	NATURE		P734	35	Y
			V344			
KANEKO K	1990	PHYSICA D	V41	P137	30	Y
AIHARA K	1990	PHYS LETT A		P333	30	Y
			V144			
DITTO WL	1990	PHYS REV LETT	V65	P3211	29	Y
MEHTA ML	1991	RANDOM MATRICES			51	Y
SAUER T	1991	J STAT PHYS	V65	P579	48	Y
PECORA LM	1991	PHYS REV A	V44	P2374	29	Y
HUNT ER	1991	PHYS REV LETT	V67	P1953	28	Y
THEILER J	1992	PHYSICA D	V58	P77	66	Y
PYRAGAS K	1992	PHYS LETT A		P421	59	Y
			V170			
LICHTENBERG AJ	1992	REGULAR CHAOTIC DYNA			58	Y
KENNEL MB	1992	PHYS REV A	V45	P3403	33	Y
KOCAREV L	1992	INT J BIFURCAT CHAOS	V2	P709	31	Y
PRESS WH	1992	NUMERICAL RECIPES C			29	Y
GARFINKEL A	1992	SCIENCE		P1230	27	Y
			V257			
MARCUS CM	1992	PHYS REV LETT	V69	P506	26	Y
ALEXANDER JC	1992	INT J BIFURCAT CHAOS	V2	P795	25	Y
CROSS MC	1993	REV MOD PHYS	V65	P851	105	Y
OTT E	1993	CHAOS DYNAMICAL SYST			97	Y
CUOMO KM	1993	PHYS REV LETT	V71	P65	57	Y
ABARBANEL HDI	1993	REV MOD PHYS	V65	P1331	54	Y
PLATT N	1993	PHYS REV LETT	V70	P279	38	Y
CUOMO KM	1993	IEEE T CIRCUITS-II	V40	P626	34	Y
WU CW	1993	INT J BIFURCAT CHAOS	V3	P1619	28	Y
HEAGY JF	1994	PHYS REV E	V50	P1874	40	Y
OTT E	1994	PHYS LETT A		P39	40	Y
			V188			
STROGATZ SH	1994	NONLINEAR DYNAMICS C			35	Y
ASHWIN P	1994	PHYS LETT A		P126	33	Y
			V193			
LASOTA A	1994	CHAOS FRACTALS NOISE			30	Y
HEAGY JF	1994	PHYS REV E	V49	P1140	30	Y
ROY R	1994	PHYS REV LETT	V72	P2009	28	Y
SCHIFF SJ	1994	NATURE		P615	28	Y
			V370			
RULKOV NF	1995	PHYS REV E	V51	P980	59	Y
NAYFEH AH	1995	APPL NONLINEAR DYNAM			46	Y
KOCAREV L	1995	PHYS REV LETT	V74	P5028	40	Y
KATOK A	1995	INTRO MODERN THEORY			27	Y
ROSENBLUM MG	1996	PHYS REV LETT	V76	P1804	77	Y

ABARBANEL HDI	1996	ANAL OBSERVED CHAOTI			45	Y
KOCAREV L	1996	PHYS REV LETT	V76	P1816	38	Y
LAI YC	1996	PHYS REV LETT	V77	P55	27	Y
ASHWIN P	1996	NONLINEARITY	V9	P703	27	Y
ZELEVINSKY V	1996	PHYS REP	V276	P85	26	Y
KANTZ H	1997	NONLINEAR TIME SERIE			54	Y
PIKOVSKY AS	1997	PHYSICA D	V104	P219	43	Y
PECORA LM	1997	CHAOS	V7	P520	40	Y
ROSENBLUM MG	1997	PHYS REV LETT	V78	P4193	39	Y
BEENAKKER CWJ	1997	REV MOD PHYS	V69	P731	25	Y
GAMMAITONI L	1998	REV MOD PHYS	V70	P223	52	Y
GUHR T	1998	PHYS REP	V299	P189	37	Y
VANWIGGEREN GD	1998	SCIENCE	V279	P1198	32	Y
GOEDGEBUER JP	1998	PHYS REV LETT	V80	P2249	29	Y
TASS P	1998	PHYS REV LETT	V81	P3291	29	Y
HEGGER R	1999	CHAOS	V9	P413	27	Y
FISCHER I	2000	PHYS REV A	V6201	P1801	16	Y
MATEOS JL	2000	PHYS REV LETT	V84	P258	15	Y
WANG W	2000	CHAOS	V10	P248	14	Y
VANAG VK	2000	NATURE	V406	P389	13	Y

These results were examined by the authors. They judged that all papers in the table were relevant for a Background section, or review paper. Some of the earliest papers (e.g., Wigner or Anderson) are concerned with random systems and not with chaotic systems, but the methods they employed influenced how to view and contrast with chaotic systems mathematically.

They also identified about 6% additional papers that he would have included in a Background section. These papers tended to have relatively high total citations, but relatively low citations from the Nonlinear Dynamics papers in the present database. Some of the papers omitted were straight plasma physics focused on nuclear fusion tokomak physics. The system was naturally very Nonlinear so the work involved Nonlinear Dynamics, but the purpose of the paper was fusion and not advancing the field of Nonlinear Dynamics. This could cause Nonlinear Dynamics authors not to reference these papers widely. Their references come from the plasma community. Finally, some papers are highly cited, but then get replaced by better (or more easily read) papers by the same author. The newer citations tend to cite the author's newer paper.

The analysis and discussion above have focused on the <u>contents</u> of the Background; i.e., which

documents should be included. In some cases, the Abstracts of the seminal references have been retrieved and clustered, to produce a structure for the Background. Thus, the CAB approach can be used to determine both the content and structure of the Background section. Again, CAB does not exclude content and structure determinations by the experts. CAB can be viewed as the starting point for content and structure determination, upon which the experts can build with their own insights and experience.

While the CAB approach is systematic, it is not automatic. Judgement is required to determine when an adequate number of references has been selected for the Background, and further judgement is required to analyze, group, and link the references to form a cohesive Background section. Additionally, the highly influential references that were not highly cited due to insufficient dissemination should be included by the Background developers, if they know of such documents. CAB is not meant to replace individual judgement or specification of Background material. CAB is meant to augment individual judgement and reference selection, as reflected in its name of Citation-Assisted.

CONCLUSIONS

A method for systematically determining seminal references for inclusion in literature surveys or Background sections of research documents has been described. It is based on the assumption that seminal documents tend to be highly cited. CAB is being applied presently to three applications studies, and the results so far are much superior to those used by the first author for background development in any other study.

REFERENCES FOR APPENDIX 5-D

Calne DB, Calne R. Citation of original research. Lancet**.** 340 (8813): 244-244. Jul 25 1992.
Della Mea V, Mizzaro S. Measuring retrieval effectiveness: A new proposal and a first experimental validation. Journal of the American Society for Information Science and Technology. 55 (6): 530-543. Apr 2004.
Fenton JE, Brazier H, De Souza A, Hughes JP, Mcshane DP. The accuracy of citation and quotation in otolaryngology/head and neck surgery journals. Clinical Otolaryngology. 25 (1): 40-44. Feb 2000.
Gosling CM, Cameron M, Gibbons PF. Referencing and quotation accuracy in four manual therapy journals. Manual Therapy. 9 (1): 36-40 Feb 2004.
Harter SP, Hert CA. Evaluation of information retrieval systems: Approaches, issues, and methods. Annual Review of Information Science and Technology. 32: 3-94 1997.
Kagolovsky Y, Moehr JR. A new look at information retrieval evaluation: Proposal for solutions. Journal of Medical Systems. 28 (1): 103-116. Feb 2004.
Kagolovsky Y, Moehr JR. Evaluation of information retrieval. Journal of Medical Systems. In Press.
Kantor PB. Information-retrieval techniques. Annual Review of Information Science and Technology. 29: 53-90. 1994.
Khan MS, Khor S. Enhanced Web document retrieval using automatic query expansion**.**

Journal of the American Society for Information Science And Technology. 55 (1): 29-40. Jan 1 2004.
Kostoff, R. N., Shlesinger, M., and Tshiteya, R. "Nonlinear Dynamics Roadmaps using Bibliometrics and Database Tomography". International Journal of Bifurcation and Chaos. January 2004.
Lee SY, Lee JS. A survey of reference accuracy in two Asian dermatologic journals (the Journal of Dermatology and the Korean Journal of Dermatology). International Journal of Dermatology. 38 (5): 357-360. May 1999.
Liu, M.X . Progress In Documentation - The Complexities of Citation Practice – A Review of Citation Studies. Journal of Documentation. 49 (4): 370-408 Dec 1993
Macroberts, M.H, Macroberts B.R. Problems of Citation Analysis - A Critical-Review. Journal of the American Society for Information Science. 40 (5): 342-349. 1989.
MacRoberts, M.H, MacRoberts, B.R Citation content analysis of a botany journal. Journal of the American Society for Information Science. 48 (3): 274-275 Mar 1997
MacRoberts, M.H,, MacRoberts, B.R. Problems of citation analysis. Scientometrics. 36 (3): 435-444 Jul-Aug 1996.
Moravcsik MJ, Murugesan P. Some results on function and quality of citations. Social Studies of Science. 5 (1): 86-92. 1975.
Putterman C, Lossos IS. Author, verify your references - or, the accuracy of references in Israeli medical journals. Israel Journal of Medical Sciences. 27 (2): 109-112. Feb 1991.
Shadish WR, Tolliver D, Gray M, Sengupta SK Author judgments about works they cite - 3 studies from psychology journals. Social Studies of Science. 25 (3): 477-498. Aug 1995.
Smith, A.J, Goodman, N.W. The hypertensive response to intubation. Do researchers acknowledge previous work? Canadian Journal of Anaesthesia-Journal Canadien D Anesthesie. 44 (1): 9-13 Jan 1997

APPENDIX 6

THE PIED PIPER EFFECT: A SPECIFIC EXAMPLE [Kostoff, 1997n]

An article in Science magazine purports to identify the Top 10 U.S. Universities in Clinical Medical Research from 1990-1994 [SCIENCE, 1995]. The published papers and citations per paper are ranked in decreasing frequency by medical research institution, and the institutions with the highest frequencies of publications and citations are identified as the top universities in clinical medicine research. This Science article crystallizes the problem of using metrics as a gauge of research productivity and, by inference, quality. This statement will be amplified with an illustrative example which questions the linkage between high research output and high research quality. The example focuses on cataracts, but is extrapolateable to other chronic systemic problems as well.

The author recently did a literature survey of research papers related to cataracts. The author examined four years (1991-1994) of abstracts from the Science Citation Index (SCI) and the Social Science Citation Index (SSCI). Of the many hundreds of abstracts identified, perhaps 99% dealt with different aspects of the surgical treatment of cataracts. Maybe 1% or less dealt with nutritional approaches, and these were mainly vitamin and mineral supplementation for prevention. There were no papers in these peer-reviewed journals dealing with alternative approaches to cataract treatment.

The mainstream medical community views cataracts strictly as an eye problem. The lens degenerates for unknown reasons, in their view, and when it has deteriorated sufficiently, it should be replaced surgically. This approach arises from the paradigm of viewing the eye as a separate component of the total physical system, and the lens replacement becomes equivalent conceptually to replacing a car's windshield when it has become pitted.

An alternative paradigm is that the body experiences chronic systemic problems (deficiencies of various types), and these problems manifest themselves as symptoms in specific organs. For some people, the weak organ is the eye, and the symptom is the cataract. Healing, in this paradigm, consists of identifying and eliminating the deficiencies. Surgically removing the cataract, while improving functioning (at least temporarily), does nothing to address the fundamental systemic problems which are at the foundation of the cataract's presence. It is equivalent to removing the warning light on a car's dashboard when it signifies a problem.

These alternative approaches never surface in the peer reviewed literature, as the author's survey has shown. The journal reviewers (and the funding proposal reviewers as well) are researchers trained along the orthodox paradigms, and they provide high marks to those papers (and proposals) aligned with the reviewers' backgrounds. In addition, there are institutional and commercial biases which also govern the willingness of the reviewers and editors (and sponsors) to provide positive evaluations of alternative approaches. Thus, the copious papers and citations (and grants) from this component of medical research reflect activity among a closed group whose members subscribe to essentially the same orthodox paradigm. Far from being a measure of quality, the numbers of papers and citations (and projects) from some branches of medical research could be interpreted as a measure of the extent of the problem.

The author differentiates between the two major characteristics of high quality science: doing the job right and doing the right job (in the best of all worlds, one would do the right job right) [1997n]. The Science article is an example of doing the job right. Once the research target has been selected (paradigm of using the surgical approach to eliminating cataracts), the orthodox medical research community performs an excellent and highly productive effort in finding the best ways to achieve the target. It is analogous to firing a missile very accurately at the wrong target. However, one can question seriously whether the community is doing the right job (using the right paradigm), and the present closed funding, review, and publication structure effectively precludes innovations which will address the right job.

The Science article, and the above comments, illustrate the danger of relying on metrics to infer quality from scientific activity. Metrics have their place in a comprehensive evaluation procedure of research, but as a stand-alone approach (as reflected in the Science article) metrics are subject to misinterpretation.

APPENDIX 7

EXAMPLES OF SCIENCE AND TECHNOLOGY BIBLIOMETRICS STUDIES

In the early 1990s, the author invented and patented the Database Tomography approach (Kostoff, 1995d). The initial studies using Database Tomography were focused on technical reports and organizational project databases. Starting in the mid-1990s, with the expanded availability of large journal and conference proceeding databases such as Science Citation Index, Engineering Compendex, and Medline, the author's group has performed text mining studies of a number of diverse technical disciplines as represented by their open literature publications.

These latter studies have contained two major components. One is bibliometrics, to identify the infrastructure of the technical discipline (authors, journals, institutions), as well as provide some indications of the extent and productivity of the discipline. The other is computational linguistics, to identify the categorical structure of the technical discipline. This appendix provides some selected examples of the bibliometrics component of these studies. The examples are in chronological order, so the reader can see how the analytical methodology and information displayed have evolved with time.

The computational linguistics component provides two generic types of outputs. One is qualitative, represented by taxonomies of the technical discipline, or the technical categories and sub-categories into which the discipline can be divided. The other is quantitative, and is characterized by the levels of effort or emphasis that are devoted to each of the categories/ sub-categories in the taxonomy. This compositional metric reflects the investment strategy at whatever level the discipline is being described by the database used (organizational, national, global). This metric is a measure of how well the actual investment decisions reflect the optimal investment strategy for accelerating the progress of science and technology efficiently, consistent with the mission goals of the organization(s) sponsoring the efforts in the discipline. The reader is referred to the full studies for descriptions of the computational linguistics component and metrics [Kostoff et al, 1997g, 1997h, 1998a, 1999a, 2000a, 2000d, 2001b, 2001c, 2001i, 2002a, 2002c, 2003c, 2003d, 2003j, 2003l, 2003n, 2003q, 2003u, 2004a, 2004c, 2004j, 2004k, 2004l, 2004n, 2004p, 2004r, 2005b, 2005c. 2005f, 2005i, 2005k]

APPENDIX 7-A.

FULLERENE DATA MINING USING BIBLIOMETRICS AND DATABASE TOMOGRAPHY [Kostoff et al, 2000a]

1. INTRODUCTION

The present Appendix describes use of the DT process, supplemented by literature bibliometric analyses, to derive technical intelligence from the published literature of fullerene science and technology.

Fullerene, as defined by the authors for this study, consists of theory/ experiment/ computation/ applications related to large ordered carbon atom clusters. It is defined operationally by the following query, obtained by the iterative technique referenced in the next paragraph:

"C-60" OR "C-70" OR "C60" OR "C70" OR FULLERENE* OR CARBON NANOTUBES OR BUCKMINSTERFULLERENE OR FULLERIDE* OR FULLERITE OR METALLOFULLERENES OR METHANOFULLERENE OR ENDOHEDRAL OR SOCCERBALL OR BUCKEYTUBE OR "C-78"

To execute the study reported in this Appendix, a database of relevant fullerene articles is generated using the iterative search approach of Simulated Nucleation (4,5). Then, the database is analyzed to produce the following characteristics and key features of the fullerene field: recent prolific fullerene authors; journals that contain numerous fullerene papers; institutions that produce numerous fullerene papers; keywords most frequently specified by the fullerene authors; authors whose works are cited most frequently; particular papers and journals cited most frequently; pervasive themes of fullerene; and relationships among the pervasive themes and sub-themes. Finally, the lessons learned from this study (and two parallel studies) from integrating the topical domain experts with the analytical data mining tools are summarized.

2. BACKGROUND

2.1 Overview

The information sciences background for the approach used in this Appendix is presented in Kostoff (6). This reference shows the unique features of the computer and co-word-based DT process relative to other roadmap techniques. It describes the two main roadmap categories (expert-based and computer-based), summarizes the different approaches to computer-based roadmaps (citation and co-occurrence techniques), presents the key features of classical co-word analysis, and shows the evolution of DT from its co-word roots to its present form.

The study reported in the present Appendix differs from the previous published papers in this category (6,7,8) in four respects. First, the topical domain (fullerenes) is completely different. Second, a much more comprehensive bibliometrics cross-discipline comparison is performed.

Third, the balance of effort has shifted from computer-centric (where the primary emphasis was on the computer results, and the secondary emphasis was on the expert analysis of the computer results) to expert-centric (where the primary emphasis is on expert analysis of the computer results and raw data, and the computer results serve to augment the capabilities of the expert). There are two reasons for this shift in emphasis. Expert-centric S&T data mining provides an in-depth understanding/ identification of the technical concepts and their inter-relationships, whereas the computer-centric approach focused on the more superficial level of context-free phrases. Also, as shown in later sections of this paper, one of the major products of a serious data mining study is the 'educated expert', who has had his/ her horizons broadened substantially by the data mining experience. The study experience should center around maximum enhancement of the capabilities of the expert in the topical area.

Fourth, the study describes the data mining lessons learned from focusing on the integration of the technical domain expert with the computational tools.

3. DATABASE GENERATION

The key step in the fullerene literature analysis is the generation of the database. For the present study, two databases were used.

3.1 Science Citation Index (9)

The first database consists of selected journal records (including authors, titles, journals, author addresses, author keywords, abstract narratives, and references cited for each paper) obtained by searching the web version of the Science Citation Index (SCI) for fullerene articles. At the time the present paper was written (late 1998), the version of the SCI used accessed about 5300 journals (mainly in physical, engineering, and life sciences basic research).

The SCI database selected represents a fraction of the available fullerene (mainly research) literature. It does not include the large body of classified literature, or company proprietary technology literature. It does not include technical reports or books or patents on fullerenes. It covers a finite slice of time (1991 to mid-1998). The database used represents the bulk of the peer-reviewed high quality fullerene science and technology, and is a representative sample of all fullerene science and technology in recent times.

To extract the relevant articles from the SCI, the title, keyword, and abstract fields were searched using keywords relevant to fullerenes, although different procedures were used to search the title and abstract fields (4). The resultant abstracts were culled to those relevant to fullerenes. The search was performed with the aid of two powerful DT tools (multi-word phrase frequency analysis and phrase

proximity analysis) using the process of Simulated Nucleation (4).

An initial query of FULLERENE* and related terms produced two groups of papers: one group was judged by domain experts to be relevant to the subject matter, the other was judged to be non-relevant. Gradations of relevancy or non-relevancy were not considered. An initial database of titles, keywords, and abstracts was created for each of the two groups of papers. Phrase frequency and proximity analyses were performed on this textual database for each group. The high frequency single, double, and triple word phrases characteristic of the relevant group, and their boolean combinations, were then added to the query to expand the papers retrieved. Similar phrases characteristic of the non-relevant group were effectively subtracted from the query to contract the papers retrieved. The process was repeated on the new database of titles, keywords, and abstracts obtained from the search. A few more iterations were performed until the number of records retrieved stabilized (convergence).

The final query used for the fullerene study, shown in the Introduction, contained 15 terms. In other studies, such as Aircraft S&T, the final query contained over 200 terms. There are two main reasons for the difference in query complexity. First, in the Aircraft study, the coverage is much broader than in the fullerene study. Second, but perhaps more importantly, the contents of the SCI database are more aligned with the objectives of the fullerene study than those of the Aircraft study. As will be shown later by the results, the journal literature on fullerenes describes a research field well aligned with the contents of the SCI research database. Aircraft is both a science/ technology area as well as a tool/ platform for performing research. While the SCI is well aligned with the science/ technology component of Aircraft (e.g., aircraft structures, aircraft propulsion), the SCI also includes papers relating to the use of Aircraft as a platform from which to perform research (e.g., crop spraying, buffalo tracking). If the search philosophy is to start the iterative query process with AIRCRAFT and subtract terms not applicable to the platform function of Aircraft, then a large SCI query will be required for Aircraft to remove these platform-oriented terms. This type of dual usage does not exist yet for fullerenes in the published journal literature, and is therefore reflected in the much simpler fullerene query.

The situation is analogous to selection of a mathematical coordinate system for solving a physical problem. If the coordinate system is aligned naturally with the body geometry (e.g., a spherical coordinate system used to model flow around a sphere), then a minimal number of equation terms is necessary. If the coordinate system is mis-matched to the body geometry (e.g., a spherical coordinate system used to model the flow around a parallel-piped), then a large number of equation terms will be required to effectively translate between the two geometries.

The authors believe that queries of these magnitudes and complexities are required when necessary to provide a tailored database of relevant records that encompasses the broader aspects of target disciplines. In particular, if it is desired to enhance the transfer of ideas across disparate disciplines, and thereby stimulate the potential for innovation and discovery from complementary literatures (10), then even more complex queries using Simulated Nucleation may be required.

The authors believe that the 'purity' and completeness of the database of topically relevant records

obtained using Simulated Nucleation is a key reason that the invariance of most of the normalized bibliometric distributions across different topical domains can be displayed (see the normalized bibliometric distribution functions in later sections). One beneficial value of utilizing Simulated Nucleation is that the search terms are obtained from the words of the authors in the SCI and EC databases, not by guessing on the part of the searcher.

3.2 Engineering Compendex (11)

The second database consists of selected journal and conference proceeding records (including authors, titles, journals, author addresses, author keywords, abstract narratives, and references cited for each paper) obtained by searching the CD-ROM version of the Engineering Compendex (EC) for fullerene articles. In late 1998, this version of the EC accessed about 2600 journals, mainly in physical and engineering sciences applied research and technology).

The EC database selected represents a fraction of the available fullerene (mainly applied research and technology) literature. It does not include either the large body of classified and company proprietary technology literature, or the large body of technical reports on fullerenes. It covers a finite slice of time (1991 to mid-1998). Because of the monolithic research nature of fullerenes, the same query used for searching the SCI was used to search the EC.

4. RESULTS

The results from the publications bibliometric analyses are presented in section 4.1, followed by the results from the citations bibliometrics analysis in section 4.2. Results from the DT analyses are shown in section 4.3. The SCI and EC bibliometric fields incorporated into the database included, for each paper, the author, journal, institution, and keywords. In addition, the SCI included references for each paper. Due to the fundamental research orientation of fullerenes as reflected in the published journal literature used for this study, most of the EC results were included in the SCI results. Therefore, only the SCI results will be presented in this paper.

The bibliometrics sections (4.1, 4.2) have two components. Important numerical indicators are presented that illuminate some aspect of the fullerenes research literature (e.g., average authors per paper, number of journals, papers per institution), and distribution functions of publication and citation parameters (e.g., numbers of authors f(n) who publish 'n' papers) are compared with those of other technical discipline studies that used a similar approach.

The DT sections contain three components. First, the high frequency keywords are grouped into 'natural' categories, and the picture they provide of the fullerenes literature (research, open literature, unclassified, non-proprietary) is described. Second, the high frequency phrases from the abstracts are grouped into 'natural' categories, and the picture they provide of the fullerenes literature is presented. Third, the high numerical indicator phrases from the proximity analyses of the abstracts and other portions of the database (author names, article titles, journal names, author addresses) are grouped into 'natural' categories, and the picture they provide of the fullerenes literature is shown. The meaning of the term 'natural' is that these categories were not prescribed beforehand. From

observation of the hundreds of different phrases and their frequencies, categories useful for interpreting and describing the main literature findings appeared to emerge..

The analytical approaches taken for the first three components (keyword phrase frequency, abstract phrase frequency, phrase proximity) are based on their fundamental data structures. The keyword and abstract phrase frequencies are essentially quantity measures. They lend themselves to 'binning', and addressing adequacies and deficiencies in levels of effort. They do not contain relational information, and therefore offer little insight into S&T linkages.

The phrase proximity results are essentially relational measures, although some of the proximity results imply levels of effort that support specific S&T areas. The phrase proximity results mainly offer insight into S&T linkages, and have the potential to help identify innovative concepts from disparate disciplines (10). Thus, the keyword and abstract phrase frequency analyses will be addressed to adequacy of effort, and the phrase proximity analyses will be addressed to relationships primarily and supporting levels of effort secondarily.

4.1 Publication Statistics on Authors, Journals, Organizations, Countries

The first group of metrics presented is counts of papers published by different entities. These metrics can be viewed as output and productivity measures. They are not direct measures of research quality, although there is some threshold quality level inferred due to these papers= publication in the (typically) high caliber of journals accessed by the SCI.

4.1.1 Prolific Authors

The author field was separated from the database, and a frequency count of author appearances was made. In the SCI database results, there were 12,839 different authors, and 41,167 author listings (the occurrence of each author's name on a paper is defined as an author listing). While the average number of listings per author is about 3.2, the most prolific authors (e.g., ACHIBA Y,143; KROTO HW,121; KIKUCHI K,115; SAITO Y,112; TAYLOR R,111; SHINOHARA H,107; SMALLEY RE, 98) have listings about an order of magnitude greater than the average. There were 10,515 papers retrieved, yielding an average of 3.92 authors per paper.

Previous DT/ bibliometrics studies were conducted of the technical fields of: 1) near-earth space (NES) (7); 2) hypersonic and supersonic flow over aerodynamic bodies (HSF) (6); 3) Chemistry (JACS) (8) as represented by the Journal of the American Chemical Society; 4) Aircraft (AIR); 5) Hydrodynamic flow over surfaces (HYD). Overall parameters of these studies are shown in Table 0.

TABLE 0 - DT STUDIES OF TOPICAL FIELDS

METRIC / STUDY	FUL	JACS	NES	HYD	HSF	AIR	RIA
NUMBER OF ARTICLES	10515	2150	5481	4608	1284	4346	2300

START YEAR	1991	1994	1993	1991	1993	1991	1991
END YEAR	M-1998	1994	M-1996	M-1998	M-1996	M-1998	E-1995

TABLE 0 - DT STUDIES OF TOPICAL FIELDS

These studies yielded: 1) 3.37 authors per paper for the NES results; 2) 2.63 authors per paper for the HSF results; 3) 3.79 authors per paper for the Chemistry results; 4) 2.09 authors per paper for the AIR results; 5) 2.29 authors per paper for the HYDRO results. A previous study on the non-technical field of research impact assessment (RIA) yielded about 1.68 authors per paper. See Table 1 for summary statistics of these previous studies.

TABLE 1 - AUTHOR BIBLIOMETRICS - SCI

METRIC / STUDY	FUL	JACS	NES	HYD	HSF	AIR	RIA
NUMBER OF AUTHORS	12837	6535	12453	7869	2483	6619	2975
NUMBER OF AUTHOR LISTINGS	41167	8151	18474	10558	3372	9085	3868
AVERAGE NUMBER OF LISTINGS PER AUTHOR	3.2	1.2	1.5	1.3	1.38	1.4	1.3
NUMBER OF PAPERS RETRIEVED	10515	2150	5481	4608	1284	4346	2300
AVERAGE NUMBER OF AUTHOR LISTINGS PER PAPER	3.92	3.79	3.37	2.29	2.63	2.09	1.68

TABLE 1 - AUTHOR BIBLIOMETRICS - SCI

Table 1 compares the SCI author bibliometric statistics for the different studies. These studies are listed, proceeding from left to right, in approximate order of the (subjectively estimated) science/ technology ratio of the underlying field. Thus, the leftmost field listed, FUL, is estimated to be the most basic (based on the specific query used and the themes of the papers retrieved), and the rightmost technical field, AIR, is estimated as the most applied. RIA, the rightmost column, is not a technical field, and is listed for completeness only. It should be emphasized that the subjective judgements used to estimate the maturity of these technical fields were based on the SCI journal papers only, and not on other data sources such as patent databases.

In Table 1, five variables/ figures of merit are presented for each study. The number of authors represents the total number of different names contained in the author blocks, while the number of author listings is the sum over all authors of the number of times each author's name was listed in an author block. The average number of (author) listings per author is the ratio of the above two quantities. The number of papers retrieved is the total number of relevant papers that comprised the database and was used for the analyses, while the average number of author listings per paper is the number of author listings divided by the number of papers retrieved.

In all cases, the most prolific authors had listings more than an order of magnitude greater than the average number of listings per author. The average number of listings per author is remarkably

consistent except for FUL, where it is about 2.5 times the average of the other fields studied. FUL is a very young and dynamic research field, with extensive global activity, participation, and competition. Based on the SCI and EC papers examined for the present study, there is little technology development at present, at least in comparison with the other fields. Whereas the technology component of myriad fields tends to be characterized by less papers than the research component, FUL does not suffer from this limitation on its average activity. In addition, for developed S&T areas, many of the papers may not have a strict discipline focus, but may address uses of the technology. These papers could be somewhat peripheral or tangential to the central discipline, and the authors may not be heavy contributors to the discipline per se. In FUL, the papers are written by active researchers solely focused on advancing the state-of-the-art, and the peripheral authors who might contribute a paper ot two do not surface often in this topical research area.

While there is a wide range among disciplines in the number of papers retrieved, the average number of author listings per paper decreases steadily proceeding from the most basic fields to the most applied. The three most basic fields (FUL, JACS, NES) tend to be experiment-dominated, with much less effort devoted to computational modeling (as will be shown in the later DT sections). In many cases, these experiments require expensive equipment and large teams of researchers because of their complexity, and this is reflected in the large numbers of authors on the papers produced.

Figure 1 shows the distribution function of author listing frequency for the fullerene, NES, JACS, HSF, AIR, and HYDRO databases. The abcissa is the number of author listings n, and the ordinate is the number of authors f(n) who have author listing n. In each case, the distribution function has been normalized to the number of authors who have one listing in the respective databases. The graph is plotted on a semi-log scale to stretch the lower ordinate region.

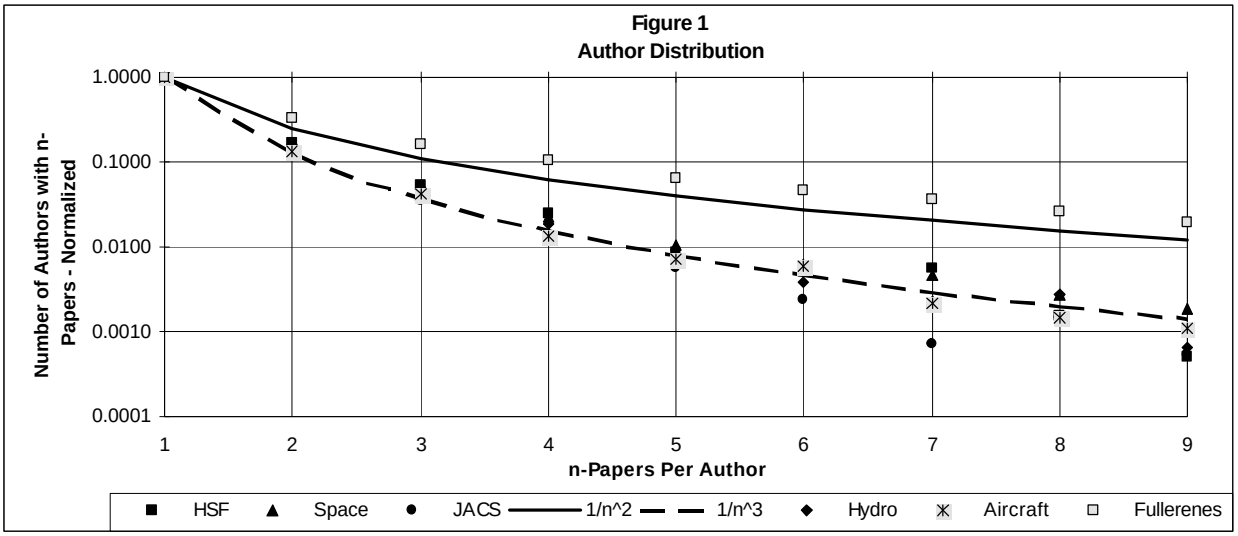


FIGURE 1 - AUTHOR FREQUENCY

The solid line on Figure 1 is the nominal >1/n^2' Lotka's Law (12) distribution. With the exception of the FUL data, all of the experimental data decline much steeper than the >1/n^2' Law predicts, centering about a >1/n^3' distribution. In the studies reported in the present document, the base of journals has been widened relative to what was available to Lotka. More journals of all types are available through the SCI. Also, because of the S&T scope of the present studies, more technology and applications - oriented journals of peripheral relation to the core science disciplines are included. As the base of journals is widened, and more non-core journals are included in the source database, a larger diversity of authors is also included in the source database. These additional authors, who are less prolific and recognized in the discipline than the core authors, will populate the lower regions of the distribution function, and will effectively skew the distribution function toward larger gradients relative to the Lotka distribution.

In the anomalous FUL case, the discipline is sufficiently young and mainly in the basic research phase that the widening of the journal base has not yet occurred. As the next section on journal bibliometrics shows, even though FUL has twice the numbers of papers relative to any of the other fields examined in this study, the total number of journals in which FUL authors publish is no larger than any of the other fields. The research authors want to establish their reputations in the core research journals, and therefore have a higher number of papers per journal as also shown in the next section. In addition, the more sporadic nature of publication in the discipline-peripheral technology and applications oriented journals has not yet occurred. The FUL case matches most closely the discipline structure used in Lotka's work, and the FUL distribution matches the nominal Lotka Law distribution most closely.

In summary, the nominal Lotka distribution can be viewed as most applicable to core discipline authors associated with the core discipline literature, while the present method reported in this paper is more focused on studying the technical discipline from a broader perspective. In this sense, the specific form of Lotka's Law that applies then becomes a function of how one defines the literature and core journals in a field, as well as the development status of the discipline.

4.1.2 Journals Containing Most Fullerene Papers

A similar process was used to develop a frequency count of journal appearances. In the SCI database, there were 680 different journals represented, with an average of 15.5 papers per journal. The journals containing the most fullerene papers (e.g., CHEMICAL PHYSICS LETTERS,800; PHYSICAL REVIEW B-CONDENSED MATTER,780; JOURNAL OF PHYSICAL CHEMISTRY,390; SYNTHETIC METALS,341; FULLERENE SCIENCE AND TECHNOLOGY,332; JOURNAL OF THE AMERICAN CHEMICAL SOCIETY,302) had in some cases an order of magnitude more papers than the average.

TABLE 2 - JOURNAL BIBLIOMETRICS - SCI

METRIC / STUDY	FUL	JACS	NES	HYD	HSF	AIR	RIA
NUMBER OF PAPERS RETRIEVED	10515	2150	5481	4608	1284	4346	2300
NUMBER OF JOURNALS	680	1	628	675	277	713	645

AVERAGE NUMBER OF PAPERS PER JOURNAL	15.46	2150	8.73	6.83	4.6	6.10	3.57
BRADFORD'S LAW - RATIO BETWEEN GROUPS	2.2		2	1.5	3	3.1	

TABLE 2 - JOURNAL BIBLIOMETRICS - SCI

Table 2 compares the SCI journal bibliometric statistics for the different studies. Four variables/ figures of merit are presented for each study. The number of journals represents the total number of different journal names contained in the source blocks. The average number of papers per journal is the ratio of total papers retrieved to total number of journals. The Bradford's Law (13) metric derives from the following definition/ re- statement of the Law: if the journals for a bibliography are grouped in order of decreasing publications, such that each group of journals contains the same number of papers, then the ratio of number of journals in each successive group will be a constant greater than unity. The Bradford's Law metric in Table 2 is this ratio between journal groups.

In all of the studies performed, the journals containing the most papers had an order of magnitude more papers than the average number of papers per journal. One unexpected finding is the closeness of the magnitudes of number of journals for the different studies. Of the seven different topics studied, using different experts and different queries and different versions of the SCI and having different science/ technology ratios, the total number of journals for five of those topics is within about ten percent of 650. In fact, for four of those five journals, the total number of journals is within about five percent of 650. There are two outliers, JACS and HSF. The JACS study used one year's issues from the Journal of the American Chemical Society, and HSF is a much narrower and more limited field than the other broader fields studied. The question arises, why would the total number of journals across diverse fields be so similar, especially since the total number of papers differed by about a factor of five for the five fields of interest? No obvious answer emerges.

The average number of papers per journal decreases as the topical areas become more applied. This reflects the reality that technology-oriented papers tend to be published in a greater variety of journals that have a smaller concentration about any single research discipline, whereas research-oriented papers tend to be published in a smaller group of journals that are heavily discipline focused. Before discussing the Bradford's Law results for Table 2, examples of how the Bradford's Law ratios are computed for HSF and FUL are presented below.

For the HSF database, the first journal group selected contained one journal with 231 papers (AIAA JOURNAL); the second group had 3 journals with 237 papers; third group 9 journals with 229 papers; fourth group 25 journals with 229 papers; and fifth group 70 journals with 229 papers. The ratio of numbers of journals per group between successive groups was approximately three, in excellent agreement with Bradford's Law.

For the FUL database, the first group selected contained two journals with 1,580 papers (CHEMICAL PHYSICS LETTERS,800; PHYSICAL REVIEW B-CONDENSED MATTER,780); the second group had 5 journals with 1,627 papers; third group 10 journals with 1,642 papers; fourth

group 21 journals with 1,584 papers; fifth group 47 journals with 1,572 papers. The ratio of numbers of journals per group between successive groups is approximately 2.2, again in agreement with Bradford's law.

For the Bradford's Law results of Table 2, the basic fields tend to have a ratio of about two, while the more applied fields have a ratio of about three. This means that in the basic fields there are more core discipline-oriented journals in which researchers would be motivated to publish relative to those in the applied fields. This conclusion is substantiated further by a more detailed examination of the numbers presented in the FUL and HSF examples. For the first three journal groups, the ratio of the cumulative number of journals to the total number of journals for the topical area is .025 for FUL and .047 for HSF. Since the first two or three journal groups tend to be the core topical groups, this result means that there is more depth in the FUL core than in the HSF core. The journals in which researchers are motivated to publish penetrates much deeper into the total FUL journal body relative to the total HSF body. In other words, there are more good basic research journals available for publication in FUL than there are in HSF.

Figure 2 shows the distribution function of journal frequency for the fullerene, AIR, HYDRO, HSF, NES, and RIA databases. The JACS database was derived from one journal only, The Journal of the American Chemical Society, and therefore was not applicable to this chart. The abcissa is the number of papers n from the relevant database published in a given journal, and the ordinate is the number of journals which contain n papers. In each case, the distribution function has been normalized to the number of journals that contain one relevant paper. Again, because of the strong initial gradients, the graph is plotted on a semi-log scale.

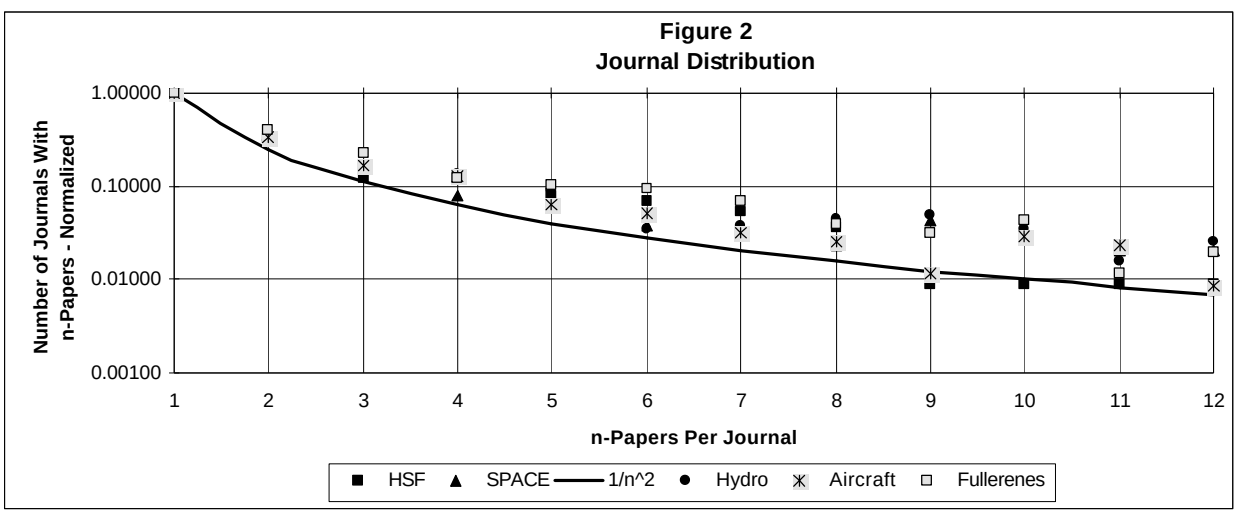


FIGURE 2 - JOURNAL FREQUENCY

The solid line in Figure 2 is a >1/n^2' distribution, and represents a lower bound of all the experimental data. On average, the FUL data again appear to have the shallowest gradients. The

rationale follows that of the previous section, and need not be repeated here.

4.1.3 Institutions Producing Most Fullerene Papers

A similar process was used to develop a frequency count of institutional address appearances. It should be noted that many different organizational components may be included under the single organizational heading (e.g., Harvard Univ could include the Chemistry Department, Biology Department, Physics Department, etc.). Lack of space precluded printing out the components under the organizational heading.

There were 2,168 different organizations listed in the SCI author address organizations, with an average of 4.85 papers per organization. The institutions producing most fullerene papers (e.g., RUSSIA,RUSSIAN ACAD SCI,602; USA,RICE UNIV,467; USA,UNIV PENN,314; USA,UNIV CALIF SANTA BARBARA,264; UK,UNIV SUSSEX,248; USA,MIT,221; JAPAN,TOKYO METROPOLITAN UNIV,217; JAPAN,TOHOKU UNIV,207; PEOPLES R CHINA,CHINESE ACAD SCI,206) were greater than an order of magnitude more productive than the average. In aggregate, the University of California campuses are the most productive of any o f the institutions in terms of papers published (~700), although no statements can be made about their production efficiency, since research expenditures were not included in this study. The top position of the Russian Academy of Sciences and the high ranking of some Japanese universities and that of the Chinese Academy has to be considered remarkable.

TABLE 3 - INSTITUTION BIBLIOMETRICS - SCI

METRIC / STUDY	FUL	JACS	NES	HYD	HSF	AIR	RIA
NUMBER OF PAPERS RETRIEVED	10515	2150	5481	4608	1284	4346	2300
NUMBER OF INSTITUTIONS	2168	750	10435	1905	661	1484	1125
AVERAGE NUMBER OF PAPERS PER INSTITUTION	4.85	2.9	0.53	2.42	1.94	2.93	2
AVERAGE NUMBER OF AUTHORS PER INSTITUTION	5.92	8.7	1.19	4.13	3.76	4.46	2.64

TABLE 3 - INSTITUTION BIBLIOMETRICS - SCI

Table 3 compares the SCI institutional bibliometric statistics for the different studies. Four variables/ figures of merit are presented for each study. The number of institutions represents the total number of different institution names contained in the address blocks. The average number of papers per institution is the ratio of total papers retrieved to total number of institutions. The average number of authors per institution is the ratio of total number of authors to total number of institutions.

In all topical areas examined, the institutions producing the most papers were greater than an order of magnitude more productive than the average institution. The total number of institutions producing papers differs substantially for the different topical areas, with the NES number of institutions appearing as a major outlier. The average number of papers per institution does not

follow any discernible trend, at least with respect to the science/ technology ratio of the discipline. The NES average papers number is much lower than for the other topical areas. Combining the average author listings per paper result from Table 1 with the average papers per institution from Table 3, the NES picture is one of many diverse participants per study from myriad institutions.

For the near-earth space focus of the NES study, that centered mainly about unmanned satellites and the manned orbiting platforms, the space vehicle tends to serve as a 'truck' or 'bus', which transports the science experiments and scientists. Thus, the central NES component is not so much a technical research discipline as it is the vehicle that enables the research to be accomplished. The actual research performed is not focused on the vehicle, and is spread among many very diverse areas and performers and institutions.

At the other extreme in Table 3, the number of papers per institution for FUL appears to be substantially greater than for the other studies. The dominant cause appears to derive from the large number of papers per author for FUL shown in Table 1. FUL is a young dynamic field with a number of centers containing strong efforts in this topical area (see last metric in Table 3), and the combination of high critical mass fractions per center with high productivity per author produces the large number of papers per institution.

There appear to be no discernible trends in Table 3 for the final metric, average number of authors per institution. Again, the NES value of 1.19 is substantially lower than that of the other studies, for the same reason that the number of papers per institution was lower. And again, using the NES EC results (7) of 14,036 authors and 2,000-2,700 institutions, the EC average of ~6.5 authors per institution is much more in line with the results of other studies in Table 3.

Figure 3 shows the distribution function of institution frequency for the fullerene, HSF, NES, JACS, AIR, and HYDRO databases. The abcissa is the number of papers n in the database produced by a given institution, and the ordinate is the number of institutions that produced n relevant papers. In each case, the distribution function has been normalized to the number of institutions that produced one relevant paper.

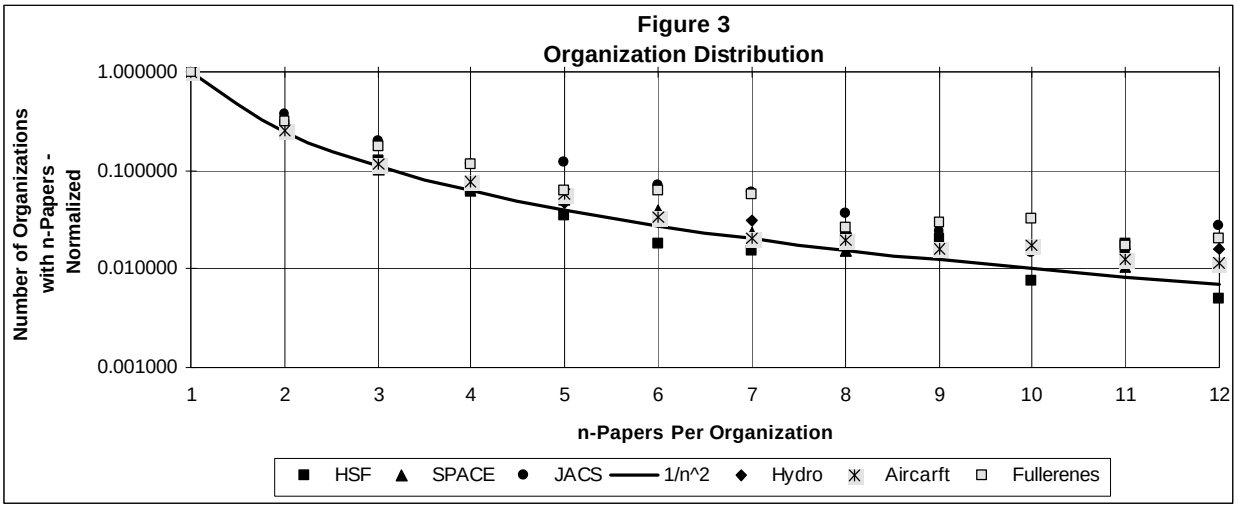


FIGURE 3 - INSTITUTION FREQUENCY

The data center around a >1/n^2' distribution remarkably well, although the FUL data exhibit the shallowest gradients again, for the same reasons as mentioned above. For a >1/n^2' distribution, the number of organizations that generate three papers is about eleven percent of the organizations that generate one paper only. Also, integrating this distribut ion function shows that more than 67% ofthe papers result from organizations that produce three or less papers.

4.1.4 Countries Producing Most Fullerene Papers

There were 64 different countries listed in the SCI results. The dominance of a handful of countries was clearly evident (e.g., USA, 5,861; JAPAN, 2,840; GERMANY, 1,500; PEOPLES R CHINA, 1,363; RUSSIA, 1,177; FRANCE, 1,117; UK,1001) but a series of small countries (SWITZERLAND, TAIWAN, BELGIUM, ISRAEL, SWEDEN, AUSTRIA, HUNGARY, THE NETHERLANDS) are also quite remarkably productive.

The UNITED STATES is more than twice as prolific as its nearest competitor (JAPAN), and is as prolific as its major competitors combined (JAPAN, GERMANY, PEOPLES REP OF CHINA). A 1997 study (14) listed the papers contributed by the top 50 nations to the world science literature; i.e., numbers of publications in the SCI. The top performers are in line with the bibliometric results of the seven DT studies.

4.2 Citation Statistics on Authors, Papers, and Journals

The second group of metrics presented is counts of citations to papers published by different entities. While citations are ordinarily used as impact or quality metrics (15), much caution needs to be exercised in their frequency count interpretation, since there are numerous reasons why authors cite or do not cite particular papers (16,17,18).

The citations in all the SCI papers were aggregated, the authors, specific papers, years, journals, and countries cited most frequently were identified, and were presented in order of decreasing frequency. A small percentage of any of these categories received large numbers of citations. From the citation year results, the most recent papers tended to be the most highly cited. This reflected rapidly evolving fields of research.

4.2.1 Most Cited Authors

The citations in all 10,515 SCI papers were aggregated into a file of 263,844 entries, yielding an average of 25.1 references per paper. There were 33,579 different authors cited, with an average of 7.85 citations per cited author. A relatively few percent received large numbers of citations (e.g., KROTO HW, 4,328; KRATSCHMER W, 3,472; IIJIMA S, 1,787; TAYLOR R, 1,721; HADDON RC, 1,711; HEBARD AF, 1,563). However, in all the studies, the most cited authors, while prolific, are not the most prolific authors (except in one anomolous case, KROTO, in the FUL study), and vice versa. For example, the three most highly cited authors (KROTO-HW, KRATSCHMER-W and IIJIMA-S) ranked numbers 2, 36, 161, respectively, in the prolific authors list. The three most prolific authors (ACHIBA-Y, KROTO-HW, KIKUCHI-K) ranked numbers 197, 1, 28, respectively, in cite-ability. Part of this difference may be due to the time lag between the highly cited authors' productivity at the time their highly cited papers were written and their productivity today, as well as the phase in their career of the prolific authors. Another partial explanation may be the intrinsic nature of the papers; the large numbers of papers produced may reflect more applied papers, which lend themselves more to shorter-term production line type output. Stated differently, the time required to produce a fundamental seminal highly cited paper probably does not allow overly high volumes of papers to be produced.

TABLE 4 - CITED AUTHOR BIBLIOMETRICS - SCI

METRIC / STUDY	FUL	JACS	NES	HYD	HSF	AIR	RIA
NUMBER OF PAPERS RETRIEVED	10515	2150	5481	4608	1284	4346	2300
NUMBER OF CITATIONS	263844	85000+	140662	82395	26768	45744	37000+
AVERAGE NUMBER OF CITATIONS PER PAPER	25.1	39.5	25.7	17.9	20.9	10.5	16.1
NUMBER OF AUTHORS CITED	33579	32450	42094	26322	11138	21868	18140
AVERAGE NUMBER OF CITATIONS PER AUTHOR CITED	7.86	2.62	3.34	3.13	2.4	2.09	2
NUMBER OF AUTHORS	12837	6535	12453	7869	2483	6619	2975
AVERAGE NUMBER OF CITATIONS PER AUTHOR	20.6	13	11.3	10.5	10.8	6.9	12.4

TABLE 4 - CITED AUTHOR BIBLIOMETRICS - SCI

Table 4 compares the bibliometric statistics for the different studies. Seven variables/ figures of merit are presented for each study. The number of citations represents the total numbers of references in all papers retrieved. The average number of citations per paper is the ratio of total number of citations to total number of papers retrieved. The number of authors cited is the total

number of different first authors cited. The average number of citations per author cited is the ratio of total number of citations to total number of authors cited. The average number of citations per author is the ratio of references to authors.

From Table 4, there appears to be a difference between the more basic and applied areas in the average number of citations per paper. The more basic papers have more references than the applied papers. The basic papers tend to be more research-literature oriented, and are dependent on published documents, whereas the applied papers tend to be technology-product oriented, with a reduced dependence on literature precedents and acknowledgements.

FUL clearly stands out in both average number of citations per author cited and average number of citations per author. FUL appears to be a young basic research field with a modest-sized core group of active researchers citing another modest-sized core group of active researchers, with much overlap between the two groups. Because the citations are focused on the modest-sized field of basic researchers, and not more broadly-based as in the more mature technological fields, there is a substantial number of citations per author cited. Because of the breadth of research activity in FUL, paper authors are motivated to document this activity as extensively as possible. Both of these latter two metrics tend to decrease with increasing technical field maturity.

JACS is somewhat of an outlier to this trend in average number of citations per author cited. It should be remembered that JACS is far less focused than FUL, since JACS covers all of Chemistry, and therefore would be expected to generate citations for a much broader group of authors than the more focused FUL. This dilution over many Chemistry sub-disciplines leads to less citations per author cited for JACS relative to FUL.

Figure 4 shows the distribution function of author citation frequency for the fullerene, NES, HSF, JACS, AIR, and HYDRO databases. The abcissa is the total number of citations n received by a given author, and the ordinate is the number of authors that received n total citations. In each case, the distribution function has been normalized to the number of authors that received one citation.

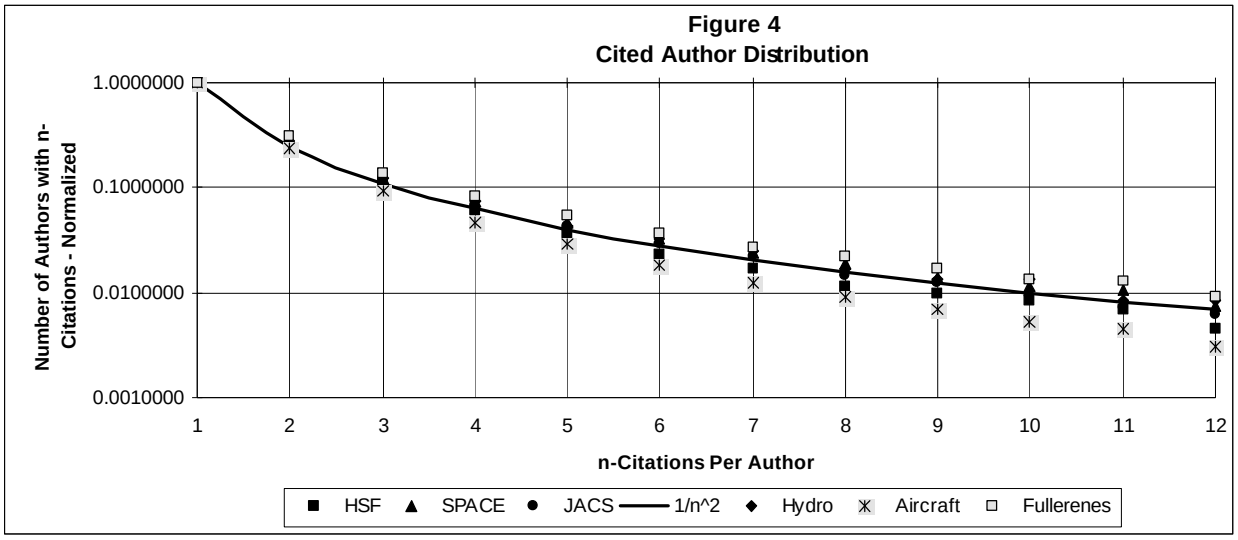


FIGURE 4 - AUTHOR CITATION FREQUENCY

The data cluster very closely around a >1/n^2' distribution, making this distribution far more universal than the somewhat discipline-dependent author publishing distribution. The FUL data are slightly above the curve, and exhibit the shallowest gradients. This relationship between the FUL data and the other discipline data occurs in all the citation distribution functions, and will be discussed in more detail in the next section on paper citation distributions.

Integration of this >1/n^2' distribut ion function shows that over 67% of the citations are from authors cited three times or less. Some caveats are in order at this point. The citation data for Figures 4, 5, 6 represents citations generated only by the specific records in each database. It does not represent all the citations received by the references in those records; these references in the database records could have been cited additionally by papers in other technical disciplines. In addition, since very recent papers are included in the references, there is probably some skewing of the distribution function toward lower numbers of citations in these figures relative to distribution functions that don't include very recently published references. Recent papers don't have sufficient time to accumulate more than a small number of citations.

Conversely, the sample studies referenced in the next section do not have the two limitations described in the above paragraph. In the sample study, a small number of papers was selected. All citations to those papers from all fields were included, and a 4-5 year time interval between date of publication and the present was chosen to allow reasonable numbers of citations to accumulate.

4.2.2 Most Cited Papers

Table 5 compares the bibliometric statistics for the different studies. Four variables/ figures of merit are presented for each study. The number of different papers cited is the total number of different papers referenced by the papers in the database. The average number of citations per cited paper is

the ratio of number of citations to number of different papers cited. The average number of papers cited per author cited is the ratio of total papers cited to total authors cited.

TABLE 5 - CITED PAPER BIBLIOMETRICS - SCI

METRIC / STUDY	FUL	JACS	NES	HYD	HSF	AIR	RIA
NUMBER OF CITATIONS	263844	85000+	140662	82395	26768	45744	37000+
NUMBER OF DIFFERENT PAPERS CITED	75890	64800	93194	57618	20950	38792	30400
AVERAGE NUMBER OF CITATIONS PER CITED PAPER	3.48	1.31	1.51	1.43	1.27	1.18	1.22
AVER. NUMBER OF PAPERS CITED PER AUTHOR CITED	2.26	2	2.21	2.19	1.88	1.77	1.68

TABLE 5 – CITED PAPER BIBLIOMETRICS – SCI

There were 75,890 different papers cited, with an average of 3.48 citations per cited paper. Relatively few papers were highly cited (e.g., KRATSCHMER W 1990 NATURE V347, 2,773; KROTO HW 1985 NATURE V318, 2,319; HEBARD AF 1991 NATURE V350, 1,177; IIJIMA S 1991 NATURE V354, 816). Relative to the other disciplines studied, the most highly cited FUL papers have larger numbers of citations (in some cases, orders of magnitude larger), and more recent publication dates. This reflects the more intensive FUL research activity, and the young rapidly evolving nature of the field.

From Table 5, there appears to be a trend in average number of citations per cited paper, with this metric decreasing with increasing technical field maturity. This trend reflects the decreased dependence of the product-oriented applied papers on the research-oriented published literature, paralleling the conclusion reached in the previous section. FUL stands out on this metric, again as a result of the concentration of the modest-sized community of citing researchers on the modest- sized community of active focused researchers.

4.2.2.1 Aggregate Distribution Functions

Figure 5 shows the distribution function of paper citation frequency for the fullerene, NES, HSF, JACS, AIR, and HYDRO databases. The abcissa is the total number of citations n received by a given paper, and the ordinate is the number of papers that received n total citations. In each case, the distribution function has been normalized to the number of papers that received one citation.

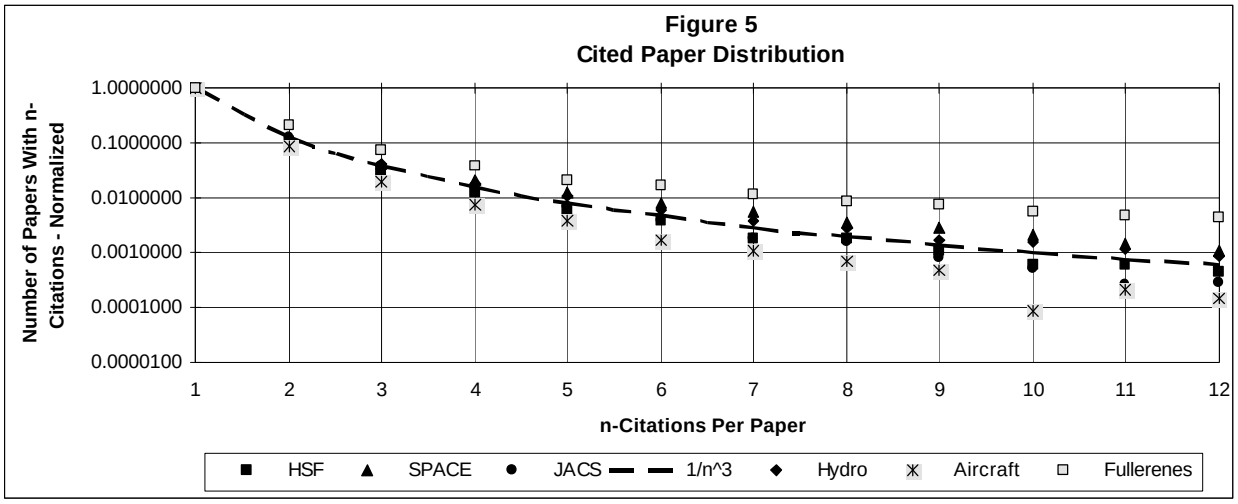


FIGURE 5 - PAPER CITATION FREQUENCY

For five of the six topical fields presented, the data follow a >1/n^3' distribution very closely, as contrasted with the >1/n^2' distribution for author citations. Examination of the five topical studies that produced the five sets of data showed that each of the highly cited authors had a wide range of citations for his/ her different papers. For any given highly cited author, most papers will receive few citations. It is the infusion of numbers of lowly cited papers from the highly cited authors which expands the pool of lowly cited papers in Figure 5, and results in the conversion of the >1/n^2' distribution of Figure 4 to the >1/n^3' distribution of Figure 5. This effect appears to transcend the five different science and technology topical fields, and to be almost universal based on the limited data presented for the six topical science and technology fields. The resulting relation among the distribution functions, the Kostoff-Eberhart-Toothman (KET) Law (6), can be re-stated as follows: for a topical science and technology field, the ratio of the normalized number of authors with n citations per author to the normalized number of papers with n citations per paper is n, for low to moderate values of n.

The FUL distribution from Figure 5 is between a >1/n^3' and >1/n^2' distribution. Its apparent modest deviation from the KET Law prediction, however, is somewhat muted by the FUL author distribution from Figure 4 also lying slightly above the >1/n^2' average of the other five disciplines. In Figure 5, the AIR distribution function exhibits the highest gradient, and the FUL distribution function exhibits the lowest one. The differences between these two distribut ions reflect the intrinsic differences of the maturity of the underlying disciplines. Aircraft S&T has been an established topical area for many years. The technology/science ratio is perhaps the highest of all the six disciplines studied. Fullerenes were discovered in the mid-1980s. As the DT analyses will show in the later sections of this paper, fullerenes S&T is essentially at the basic research experimentally - focused stage, based on the published journal literature. Its technology/science ratio is the lowest of the six disciplines studied. The other five disciplines have established an equilibrium between science and technology, whereas fullerenes are still following a start-up transient toward this equilibrium.

As shown in recent S&T data mining studies (6,7), the more basic papers tend to receive more citations than the applied papers, and the more basic journals consequently receive more citations than the more applied journals. Thus, in an S&T field such as Aircraft, that has a substantial ratio of applied to basic papers, there are fewer papers that are realistic candidates for a high number of citations. The ratio of Aircraft papers that receive a large number of citations to those receiving one citation would therefore be relatively small. Conversely, in an S&T field such as fullerenes, that has a small ratio of applied to basic papers, there are many more papers that are realistic candidates for a high number of citations. The ratio of fullerene papers that receive a large number of citations to those that receive one citation would therefore be relatively large compared to Aircraft. The data support this argument, and if/when fullerenes will advance into the technology development stage from the published literature perspective, the fullerene distribution function of Figure 5 would be expected to evolve to the distribution function predicted by the KET Law. In some sense, the KET Law can be viewed as a metric of the basic/applied balance, or equilibrated developmental maturity, of an S&T discipline.

4.2.3 Most Cited Journals

There were 13,294 different journals and other sources cited. Relatively few sources were highly cited (e.g., NATURE, 21,773; CHEM PHYS LETT, 20,735; J AM CHEM SOC, 19,534; PHYS REV B, 17,985; PHYS REV LETT, 15,482; J PHYS CHEM US, 15,120; SCIENCE, 11,801).

TABLE 6 - CITED JOURNAL BIBLIOMETRICS - SCI

METRIC / STUDY	FUL	JACS	NES	HYD	HSF	AIR	RIA
NUMBER OF CITATIONS	263844	850000+	140662	82395	26768	45744	370000+
NUMBER OF DIFFERENT JOURNALS/ SOURCES CITED	13294	6725	28740	21523	9498	21518	
AVERAGE NUMBER OF CITATIONS PER CITED JOURNAL	19.85	12.6	4.89	3.83	2.82	2.13	
NUMBER OF AUTHORS	12837	6535	12453	7869	2483	6619	2975
AVERAGE NUMBER OF JOURNALS CITED PER AUTHOR	1.04	1.03	2.31	2.74	3.83	3.25	0.00
NUMBER OF AUTHORS CITED	33579	32450	42094	26322	11138	21868	18140
AVER. NUMB. OF AUTHORS CITED PER JOURNAL CITED	2.53	4.83	1.46	1.22	1.17	1.02	

TABLE 6 - CITED JOURNAL BIBLIOMETRICS - SCI

Table 6 compares the bibliometric statistics for the different studies. Seven variables/ figures of merit are presented for each study. The number of different journals/ sources cited is the total number of different journals and other sources referenced by the papers in the database. The average number of citations per cited journal is the ratio of number of citations to number of different

journals and other sources cited. The average number of journals cited per author is the ratio of total journals and other sources cited to total authors. The average number of authors cited per journal cited is the total number of authors cited to total number of journals and other sources cited.

Fullerenes is the most basic of the six S&T areas studied with DT so far, based on the journal publications literature. It has the strongest journal correlation between high numbers of publications and citations. In the previous DT studies, some journals tended to publish many topical papers and be highly cited, some journals tended to publish many topical papers but not be highly cited, and some journals tended to publish relatively few topical papers but be highly cited. Most of the disciplines studied had a technology component along with a research component. The topical published papers tended to be slightly more applied than some of their references, and thus the journals which contained a large number of the topical published papers tended to be more applied than the journals which contained their more basic references. These more basic journals tended to rank higher in citations relative to publications, while the more applied journals tended to rank higher in publications relative to citations. Fullerenes is a relatively young topical area, and the bulk of the S&T effort is concentrated on research. Most of the papers are basic research, and the thrust of most of the journals that publish these papers is also basic.

There is a definite trend in average number of citations per cited journal, decreasing sharply from the basic fields to the applied fields. One needs to make a distinction here between the journals in which authors publish and the journals that they cite.

As the Bradford's Law results showed, there were more credible journals in which the researchers could publish in the basic fields compared to the applied fields. However, in the case of citations, there is a wider variety of journals that the researchers in the applied fields will access (both basic and applied journals) than the researchers in the basic fields will access (basic). Therefore, it would be expected that the researchers in basic fields (who cite more frequently as shown above, and who cite a narrower group of journals than their applied counterparts) would have a substantially higher value of this 'citations per cited journal' metric than their applied counterparts.

This difference in breadth of journals cited between the researchers in basic and applied fields, discussed in the previous paragraph, is substantiated and displayed most dramatically by the average number of journals cited per author metric. The metric increases sharply from the basic fields to the applied fields.

The final metric listed, average number of authors cited per journal cited, trends downward as the fields become more applied, with the lone exception of JACS. As stated previously, the researchers in the more applied fields tend to cite from a wider variety of journals than their counterparts in the more basic fields, and the denominator of this metric therefore increases as the fields become more applied. In the JACS case, the number of authors cited is slightly exaggerated because of its breadth of coverage, as shown in Table 5. This effect would tend to increase the metric numerator modestly. Probably the more pronounced effect derives from the tendency of authors in a given journal to cite that journal more frequently than would be expected on average. Since JACS was the only study in which a single journal was used, there is probably some skewing of the JACS authors toward citing

JACS papers, and hence the anomalous value of the final metric.

Figure 6 shows the distribution function of journal citation frequency for the fullerene, NES, HSF, JACS, AIR, and HYDRO databases. The abcissa is the total number of citations n received by a given journal, and the ordinate is the number of journals that received n total citations. In each case, the distribution function has been normalized to the number of journals that received one citation.

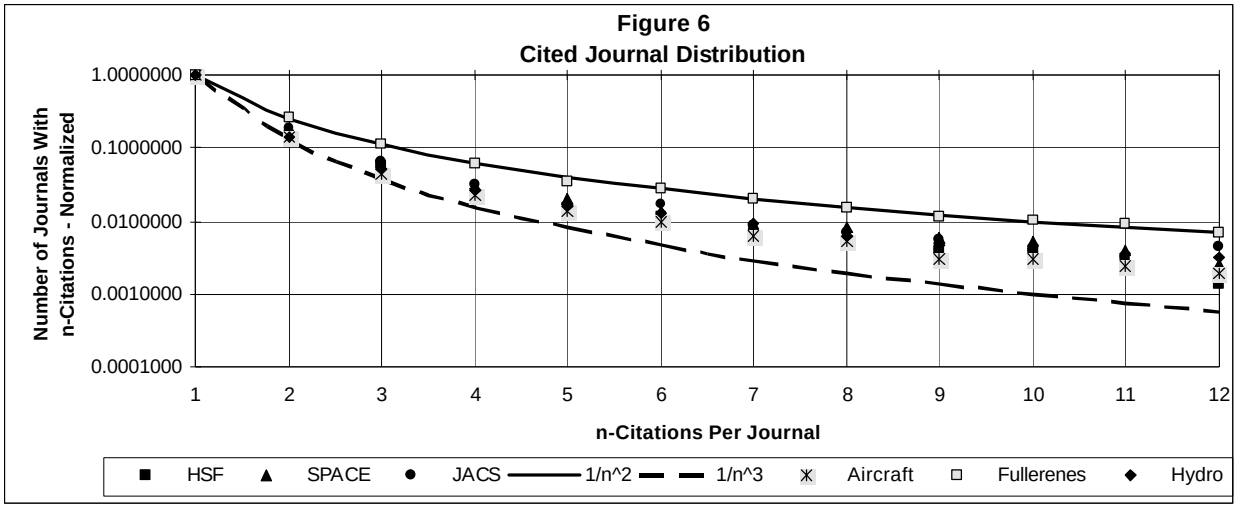


FIGURE 6 - JOURNAL CITATION FREQUENCY

The data follow approximately a >1/n^2.5' distribution. Paralleling the distributions of Figure 5, FUL exhibits the shallowest gradient, and AIR exhibits the steepest one. The reasons for these differences are identically those behind the Figure 5 differences, and need not be repeated here.

As Bradford's Law suggests, there is a concentration of papers in the higher-quality core journals. When this is coupled with the strong non-linearity of the distribution of cited papers as shown in the previous section, a further separation among journals (than the >1/n^2' average distribution of Figure 2) based on citations received would be expected. This effect is strongly muted because the wide disparity in citations per paper within a given journal is integrated out to arrive at the citations per journal for all papers published by the journal.

The authors end this bibliometrics section by recommending that the reader interested in researching the topical field of interest would be well-advised to, first, obtain the highly-cited papers listed and, second, peruse those sources that are highly cited and/or which contain large numbers of recently published topical area papers.

7. REFERENCES FOR APPENDIX 7-A

1). Kostoff, R. N. (1993). Database Tomography for Technical Intelligence. Competitive Intelligence Review. 4:1.

2). Kostoff, R.N. (1994). Database Tomography: Origins and Applications. Competitive Intelligence Review. Special Issue on Technology, 5:1.

3). Kostoff, R. N. et al (1995). System and Method for Database Tomography. U.S. Patent Number 5440481.

4). Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. (1997a). Database Tomography for Information Retrieval. Journal of Information Science. 23:4.

5). Kostoff, R. N. (1997c). The Handbook of Research Impact Assessment. Seventh Edition. DTIC Report Number ADA296021. Also, available at http://www.dtic.mil/dtic/kostoff/index.html

6). Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. (1999a). Hypersonic and Supersonic Flow Roadmaps Using Bibliometrics and Database Tomography. JASIS. 15 April.

7). Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. (1998a). Database Tomography for Technical Intelligence: A Roadmap of the Near-Earth Space Science and Technology Literature. Information Processing and Management. 34:1.

8). Kostoff, R. N., Eberhart, H. J., Toothman, D. R., and Pellenbarg, R. (1997b). Database Tomography for Technical Intelligence: Comparative Roadmaps of the Research Impact Assessment Literature and the Journal of the American Chemical Society, Scientometrics, 40:1.

9). SCI. (1999). Science Citation Index. Institute for Scientific Information. Phila., PA.

10). Kostoff, R. N. (1999b). Science and Technology Innovation. Technovation. 19.

11). EC. (1999). Engineering Compendex. Engineering Information, Inc. Hoboken, NJ.

12). Lotka, A. J. (1926). The Frequency Distribution of Scientific Productivity. Journal of the Washington Academy of Sciences. 16.

13). Bradford, S. C. (1934). Sources of Information on Specific Subjects. Engineering, 137.

14). Anwar, M. A., and Abu Bakar, A. B. (1997). Current State of Science and Technology in the Muslim World. Scientometrics. 40 (1).

15). Garfield, E. (1985) History of Citation Indexes for Chemistry - A Brief Review. JCICS. 25(3). 170-174.

16). Kostoff, R. N. (1997d). Use and Misuse of Metrics in Research Evaluation. Science and

Engineering Ethics. 3:2.

17). Kostoff, R. N. (1997e). Citation Analysis Cross-Field Normalization: A New Paradigm. Scientometrics. 39:3.

18). MacRoberts, M., and MacRoberts, B. (1996). Problems of Citation Analysis. Scientometrics. 36:3. July-August.

19). Kostoff, R. N. (1998b). The Use and Misuse of Citation Analysis in Research Evaluation. Scientometrics, 43:1, September.

APPENDIX 7-B.
DATABASE TOMOGRAPHY APPLIED TO AN AIRCRAFT SCIENCE AND TECHNOLOGY INVESTMENT STRATEGY [Kostoff, 2000d]

I. INTRODUCTION

This Appendix summarizes the results of applying Text Data Mining (TDM) techniques to Aircraft S&T records retrieved from two source technology databases for the purpose of obtaining technical intelligence on aircraft S&T. A much more detailed presentation of the results and TDM techniques is contained in the study's final report (1). Two complementary TDM techniques were used in this study: 1) bibliometrics to identify the infrastructure of Aircraft S&T (e.g., who are the performers, where are the results archived, what are the seminal papers), and 2) computational linguistics to identify the main Aircraft S&T thematic areas, the relationships of these thematic areas to each other and to the infrastructure. The source databases examined were the Science Citation Index (basic research; 1991-1998) and the Engineering Compendex (applied research/ technology; 1990-1998). Records were retrieved from these databases using an iterative query technique, and then examined using a patented software system for analyzing large amounts of textural material (2, 3).

Aircraft S&T, as defined by the authors for this study, consists of development of different aircraft/ helicopter components or technologies to improve system performance, safety or reduce costs. Use of aircraft for purposes other than platform S&T development, such as crop dusting or as an instrument platform for geophysical experiments, was typically excluded unless an extrapolation to improving military aircraft performance could be identified.

The final query used to retrieve records from the SCI contained 207 terms, and is shown in reference 1. The final query used to retrieve records from the EC contained essentially the 13 terms preceding the NOT boolean in the SCI query (aircraft or air vehicle* or helicopter* or rotorcraft or UAV or UCAV or VTOL or V/STOL or ASTOVL or STOVL or avionic* or cockpit or aircrew*). Very few abstracts that were extraneous to the focus of the study were retrieved from the EC, and the EC database did not require the same number of iterations used for the SCI database. This derives from the fact that the platform technology focus of the study is better aligned with the platform technology orientation of the EC database than the science orientation of the SCI database. In the pre-filtered SCI aircraft-related records, many records related to the use of aircraft as a platform for performing research, and the resultant SCI query had to be expanded with negation terms to excise these records from the final retrieval.

II. RESULTS

II-A. Bibliometrics

The SCI/ EC metrics are summarized in Table 1.

TABLE 1

BIBLIOMETRIC INDICATORS FOR SCI AND EC

METRIC	SCI	EC
PAPERS RETRIEVED	4346	15673
AUTHORS	6619	25586
AUTHOR LISTINGS	9085	34973
LISTINGS per AUTHOR	1.37	1.37
AUTHORS per PAPER	2.09	2.23
JOURNALS per CONF PROC	713	1876
PAPERS per JOURNAL	6.1	8.4
ORGANIZATIONS	1486	4759
PAPERS per ORGANIZATION	2.93	3.29
COUNTRIES	56	71
U.S. PAPERS	2771	8527
% U.S. PAPERS	64	54
TOTAL REFERENCES	45744	na
REFERENCES per PAPER	10.5	na
AUTHORS CITED	21868	na
CITATIONS per AUTHOR	2.09	na
PAPERS CITED	38792	na
CITATIONS per CITED PAPER	1.18	na

II-A-1. Prolific Aircraft Related Authors

II-A-1-a. **SCI** - CHOPRA, I., ATLURI, S. N., CHATTOPADHGAY, A., FORD, T., HESS, R., ERICSSON, L. E.

II-A-1-b. **EC** - CHOPRA, I; CELI, R; RAY,A.; PARKINSON, B; and SRIDHAR, B.

The presence of a moderate number of collaborators per Aircraft paper (Table 1) means that the expected large experimental research pro jects from lab and flight experiments do not dominate what

is reported to the literature, and that individual small-scale projects play an important role in Aircraft research.

II-A-2. Journals Containing Most Aircraft Related Papers

II-A-2-a. **SCI** - JOURNAL OF AIRCRAFT, AVIATION WEEK AND SPACE TECHNOLOGY, JOURNAL OF GUIDANCE CONTROL AND DYNAMICS, AIRCRAFT ENGINEERING AND AEROSPACE TECHNOLOGY, JOURNAL OF THE AMERICAN HELICOPTER SOCIETY, AIAA JOURNAL, AERONAUTICAL JOURNAL, IZVESTIYA VYSSHIKH UCHEBNYKH ZAVEDENII AVIATSIONAYA TEKHNIKA, AEROSPACE ENGINEERING, AEROSPACE AMERICA, and NOUVELLE REVUE AERONAUTIQUE ASTRONAUTIQUE

II-A-2-b. **EC** - Of the eleven highest in the in the SCI, all but three appear in the top 25 of the EC listing. They were, AIRCRAFT ENGINEERING AND AEROSPACE TECHNOLOGY (#38), AEROSPACE AMERICA (#40) and NOUVELLE REVUE AERONAUTIQUE (did not appear in the EC listing at all). This overlap between aircraft science and aircraft technology journals reflects the blurred distinction between aircraft science and technology. Much of aircraft science, like much of engineering science in general, tends to be relatively applied in an absolute scale. In the near-earth space TDM study (4), the SCI journal set was relatively independent of the EC journal set. This reflects the real-world deep stratification between space science and space technology.

II-A-3. Organizations Producing Most Aircraft Papers

II-A-3-a. **SCI** - NASA, USAF, USN, GEORGIA INST. TECH., GENERAL ELECTRIC, US

ARMY, VPI, TECHNION {ISRAEL}, BOEING, PURDUE UNIV., McDONNELL DOUGLAS, PENN STATE UNIV., DLR {GERMANY}, and the INDIAN INST. TECH. {INDIA}

II-A-3-b. **EC** - NASA, McDONNELL DOUGLAS, BOEING, LOCKHEED MARTIN, GEORGIA INST. OF TECH., GENERAL ELECTRIC, UNIV. OF MARYLAND, USAF, NORTHWESTERN POLYTECHNICAL UNIV.{CHINA}, UNIV. OF CALIFORNIA)

In both databases, the NASA Labs were the most prolific producers by far, as was the case in a similar study of the hypersonic & supersonic literature (5). Since funding levels were not examined, bibliometric productivity per dollar was not generated.

II-A-4. Countries Producing Most Aircraft Related Papers

II-A-4-a. **SCI** – U.S. (2771); U.K. (507); Germany (250); France (218); Japan (218).

II-A-4-b. **EC** – U.S. (8527); U.K. (875); China (562); Germany (468); Canada (363).

The dominance of a handful of countries is clearly evident. The UNITED STATES is five times (SCI) and ten times (EC) more prolific than its nearest competitor (UK). In both the Aircraft –SCI and EC databases, the USA is as prolific as all its competitors combined.

II-A-5. Most Cited Aircraft Related Authors

II-A-5-a. **SCI** - ERICSSON, L.E.-117; JOHNSON, W.-97; MIELE, A.-96; DOYLE, J.C.-82; and TISCHLER, M.B-80. The most cited authors, while prolific, are not the most prolific authors, and

vice versa. For example, the authors listed above (ERICSSON, JOHNSON, MIELE, DOYLE, and TISCHLER) ranked 14, 918, 87, not listed, and 35, respectively, in the prolific authors list. The five most prolific technical paper authors (CHOPRA, I.; ATLURI, S. N.; CHATTOPADHYAY, A.; FORD, T.; and HESS, R.) ranked 91, 41, 11, not listed, and 9, respectively, in citability.

Compared to a similar recent TDM analysis of "Fullerenes" (a particular construct of carbon atoms), these aircraft author citation numbers are very low (6). The most cited aircraft authors (ERICSSON-117, JOHNSON-97) were cited more than an order of magnitude less than the most cited fullerene authors (KROTO-4328, KRATSCHMER-3472). This reflects both the more applied nature of aircraft research relative to fullerenes, and the high level of fullerenes research activity relative to aircraft research activity.

II-A-6. Most Cited Aircraft Related Papers

II-A-6-a. **SCI** - JOHNSON, 1980 - 28; SNELL, 1992 - 25; DOYLE, 1989 - 23; LANE, 1988 - 22; ISIDORI, 1989 - 20).

Essentially all the highly cited papers (e.g., 13 out of the first 15) were from guidance and control related journals. The citation numbers for even the very highly cited papers are very modest in an absolute sense; none exceed thirty. This reflects the relatively low level of effort in aircraft research as contrasted with some other fields. For example, the previously cited study of "Fullerenes" (6) shows some highly cited papers receiving two orders of magnitude greater citations than the 'highly' cited aircraft papers. In addition, from the citation year results for the fullerene study, the most recent papers are the most highly cited. This reflects a rapidly evolving field of research, as well as

the newness of fullerenes. In contrast, the Aircraft-SCI database indicates that the highly cited papers were published in the 70's and 80's with only a few in the early 90's.

II-A-7. Most Cited Aircraft Related Journals

II-A-7-a. **SCI** - JOURNAL OF AIRCRAFT, AIAA JOURNAL, JOURNAL OF GUIDANCE CONTROL AND DYNAMICS, JOURNAL OF THE AMERICAN HELICOPTER SOCIETY, IEEE TRANSACTIONS IN AUTOMATIC CONTROL, JOURNAL OF SOUND AND VIBRATION, JOURNAL OF FLUID MECHANICS, VERTICA, INTERNATIONAL JOURNAL OF CONTROL, JOURNAL OF THE ACOUSTIC SOCIETY OF AMERICA, AUTOMATICA, and ASTM-STP.

There is more correlation between journals that are highly cited and contain large numbers of aircraft papers than between highly prolific and cited authors. The time span over which a journal develops and maintains a reputation for high quality is long compared to the gap between publication and citation, and one should expect that in the steady state the journals that publish many aircraft papers would also publish the higher quality papers.

Bradford's law (7) for journal publications allows journals to be grouped by primary core, secondary core, etc, where each group of journals contains the same number of papers. For the Aircraft SCI database, the first group selected contains three journals with 857 papers (JOURNAL OF AIRCRAFT, AVIATION WEEK AND SPACE TECHNOLOGY, JOURNAL OF GUIDANCE CONTROL AND DYNAMICS); the second group has 10 journals with 864 papers; etc. The ten most highly cited papers in the aircraft study were examined. It was found that only one of these ten

was contained in the first core group of three highest cited journals (based on the Bradford Law). In addition, none of the ten were found in the second core group of eight journals. One can, therefore, conclude that to research a particular aircraft technology, confining one's reading to the first one or two core journal groups will exclude many high quality documents. TDM can make the user aware of these omitted papers in the target field, and, equally important, can make the user aware of papers in disparate disciplines that could impact the target field.

IV. SUMMARY

In summary, Database Tomography (DT) and Bibliometrics would appear to be an extremely effective tool for technology program managers in the development of an investment strategy. The process allows for the development of a very focused database which can be used for a variety of searches permitting the program manager to query the state-of-the-art in a given technology (over the time span of database articles). In addition, through bibliometric analysis, the techniques allow for the determination of the most active and prolific researchers and organizations in the technical area. Highly cited authors, organizations and journals can be determined, all of which will greatly assist the program manager as he or she develops a new program plan by identifying and allowing for the possible interaction with the best talent in a given technology. Linchpin papers for a specific technology area can be identified as those most highly cited and will rapidly provide a current perspective on the state-of-the-technology. One of the most powerful tools is the ability, through Phrase Frequency Analysis, to summarize, categorize, and quantify large amounts of textural technical information so that a global picture or perspective emerges. Lastly, through the use of DT, closely related themes to a given technology can be identified and pursued.

V. REFERENCES FOR APPENDIX 7-B.

1. Kostoff R.N., Green K.A., Toothman, D. R., and Humenik, J. A., “Aircraft S&T Data Mining”, Technical Report, PAX, 2000.

2. Kostoff, R.N., Eberhart, H. J., and Miles, D., “System and Method for Database Tomography,” U.S. Patent Number 5440481, 1995.

3. Kostoff, R.N, Eberhart, H. J., and Toothman, D. R., “Database Tomography for Information Retrieval,” Journal of Information Science, 23:4, 1997.

4. Kostoff, R.N., Eberhart, H. J., and Toothman, D. R., "Database Tomography for Technical Intelligence: A Roadmap of the Near-earth Space Science and Technology Literature," Information Processing and Management, 34:1, 1998.

5. Kostoff, R.N., Eberhart, H. J., and Toothman, D. R., "Hypersonic and Supersonic Flow Roadmaps using Bibliometrics and Database Tomography," Journal of the American Society for Information Science, April 15, 1999.

6. Kostoff, R.N., Braun, T., Schubert, A., Toothman, D. R., and Humenik, J. A., "Fullerene Roadmaps Using Bibliometrics and Database Tomography", Journal of Chemical Information and Computer Science, Jan-Feb, 2000.

7. Bradford, S. C., "Sources of Information on Specific Subjects," Engineering, p.137, 1934.

8. Nathman, J.B., “Science and Technology Prioritized Capabilities-Information Memorandum”, Memorandum dtd 16 Nov 98 by Director, Air Warfare Division, CNO, 1998.

APPENDIX 7-C.

SCIENCE AND TECHNOLOGY TEXT MINING: ANALYTICAL CHEMISTRY

[Kostoff, 2001i]

ABSTRACT

Text mining is the extraction of useful information from large volumes of literature. This Appendix addresses text mining in the context of the science and technology literature. It describes the major text mining components, and shows its myriad applications in support of science and technology. To show some of the text mining products, illustrative examples from diverse literatures, but (mainly) from analytical chemistry, will be presented.

BACKGROUND

The technical literature is the storage medium for science and technology (S&T) knowledge. Rapid advancement of S&T depends on the efficiency of knowledge extraction from this literature, including both infrastructure (authors, journals, institutions) and thematic (technical thrusts, relationships) information. Relative to global S&T, questions of interest center around:

- what S&T is being performed,
- who is performing the S&T,
- where is it being performed, and
- what messages and heretofore undiscovered information can be extracted from the global literature.

The expert analysts can then judge what is not being done, and recommend what should be done differently.

In the past, the technical community used the thorough but inefficient approach of visually scanning printed and electronic technical literature to identify relevant documents, then reading the relevant documents (with no decision aids) to extract the information. Now, techniques have been developed to perform the pre-selection of relevant literature semi-automatically, and to order the intrinsic technical concepts and their relationships to provide a framework for an integrated analysis. These techniques are encompassed under the umbrella of S&T text mining.

This article defines text mining, describes its major components, and shows its myriad applications to support all types of S&T functions. Text mining can benefit S&T performers, managers, sponsors, administrators, evaluators, and oversight organizations. It can serve as a catalyst to enhance peer review, metrics, road-mapping, and other decision aids. It could allow comprehensive roadmaps for strategic planning to be constructed, and thereby serve as a

foundation for international policy assessment. Text mining can support workshops and S&T reviews by identifying the key performers in disciplines related to those being evaluated. It can identify productive sites to be visited in global S&T evaluations. It can identify new information groupings, to provide novel technical insights that could lead to discovery and innovation. In parallel, this could lead to promising new S&T opportunities, and new research directions. To illustrate some of the text mining products, illustrative examples from diverse literatures, but (mainly) analytical chemistry, will be presented.

DEFINITIONS

S&T text mining is the extraction of information from technical literature. There are three major components under our definition: 1) Information Retrieval; 2) Information Processing; 3) Information Integration.

Information retrieval is the extraction of records from the source technical literatures. High quality information retrieval produces both comprehensive and highly relevant records. It is the foundational step in text mining. The most sophisticated information processing cannot compensate for insufficient core records retrieved.

Information processing is the extraction of patterns from the retrieved records. Our definition includes three components: 1) Bibliometrics; 2) Computational Linguistics; 3) Clustering. For multi-field structured records, with some free-text fields (such as paper Abstracts), *bibliometrics* is the extraction of the technical discipline infrastructure (authors, journals, organizations) as represented by the core records. *Computational linguistics* is the computer-based extraction of technical themes and their relationships. Computational linguistics is complex for technical literature analysis, because the technical phraseology appears as a foreign language to the computer. *Clustering* is the grouping of common technical themes, and could be executed as phrase pattern groupings or actual document groupings.

Information integration is the synergistic combination of the information processing computer output with the reading of the retrieved relevant records. The information processing output serves as a framework for the analysis, and the insights from reading the records enhance the skeleton structure to provide a logical integrated product.

More detailed descriptions of text mining can be found in (1) and (2).

APPLICATIONS

A few of the myriad existing and potential S&T text mining applications will be summarized.

1) RETRIEVE DOCUMENTS

Text mining can substantially improve the comprehensiveness and relevance of records retrieved from databases. There are many approaches to information retrieval. Annual conferences focus on comparing various techniques for their comprehensiveness and S/N of records retrieved (3,

4). Most high quality methods include some type of relevance feedback . This is an iterative method where a test query is generated, records are retrieved, and then patterns from the relevant and non-relevant records are used to modify the query for increased comprehensiveness and precision. These patterns are typically linguistic phrase and phrase combination patterns, but could also include infrastructure patterns such as author/ journal/ organization, etc (5).

2) IDENTIFY INFRASTRUCTURE

The infrastructure of a technical discipline consists of the authors, journals, organizations and other groups or facilities that contribute to the advancement and maintenance of the discipline. To obtain this infrastructure, scientometric studies without text mining typically assemble this literature-based information for a given discipline (e.g., 6), sometimes including temporal trends. However, text mining can identify these infrastructure elements, and in addition provide their specific relationships to the total technical discipline or to sub-discipline areas. This information is valuable for inviting the right people and discipline combinations to workshops and S&T reviews. It is also very valuable for planning a site visitation strategy for global discipline evaluations.

3) IDENTIFY TECHNICAL THEMES/ RELATIONSHIPS

Phrase pattern analyses through computational linguistics allow technical themes, their inter-relationships, their relationships with the infrastructure, and technical taxonomies to be identified. These are important for understanding the structure of a discipline, the linkages among people/ organizations/ sub-disciplines, and being able to estimate adequacies and deficiencies of S&T in sub-technology areas. Taxonomies can be generated manually from visual text analysis, or automatically through advanced text clustering techniques.

4) DISCOVERY FROM LITERATURE

Generically, literature-based discovery consists of examining relationships between linked, overlapping literatures, and discovering relationships or promising opportunities not obtainable from reading each literature separately. The general theory behind this approach, applied to two separate literatures, is based upon the following considerations (7).

Assume that two literatures can be generated, the first literature AB having a central theme "a" and sub-themes "b," and the second literature family BC having a central theme(s) "b" and sub-themes "c." From these combinations, linkages can be generated through the "b" themes which connect both literatures (e.g., AB-->BC). Those linkages that connect the disjoint components of the two literatures (e.g., the components of AB and BC whose intersection is zero) are candidates for discovery, since the disjoint themes "c" identified in literature BC could not have been obtained from reading literature AB alone.

Successful performance of this generic approach can lead to new treatments for illnesses, new materials for different applications, extrapolation of ideas from one discipline to a disparately related discipline, and identification of promising new S&T opportunities and research directions. Some studies and concept papers have been published (2, 7, 8, 9, 10, 11, 12, 13).

TECHNIQUES AND ILLUSTRATIVE EXAMPLES

This section provides illustrative examples of S&T text mining techniques. It starts with an example of a query developed for a recent Aircraft S&T study, and shows some of the lessons learned from the query development. The section then proceeds to show some bibliometrics results. Most of these are from a database of papers published recently in *Analytical Chemistry*, and the journal bibliometrics are from a Mass Spectrometry query. Computational linguistics examples are taken from a variety of sources, related to analytical chemistry where possible.

1) RECORD RETRIEVAL QUERY, AIRCRAFT TECHNOLOGY

In the typical S&T text mining analyses performed by the first author, the starting point is the generation of a record retrieval query. A query development example is provided from a recent text mining study of the Aircraft S&T literature (14) in order to illustrate an important point about query complexity.

The study's focus was the S&T of the aircraft platform. The query philosophy was to start with the term AIRCRAFT, then add terms that would expand the number of Aircraft S&T papers retrieved and would eliminate papers not relevant to Aircraft S&T. Two databases were examined, the Science Citation Index (SCI-basic research, 5300 journals accessed) and the Engineering Compendex (EC-technology development, 2600 journals accessed). The SCI record retrieval query required 207 terms (separate phrases and phrase combinations) and 3 iterations to develop, while the EC query required 13 terms and one iteration. The SCI query retrieved 4,346 relevant records, while the EC query retrieved 15,673 relevant records.

Because of the technology focus of the EC, most of the papers retrieved using an AIRCRAFT or HELICOPTER type query term focused on the S&T of the platform itself, and were aligned with the study goals. Because of the research focus of the SCI, many of the papers retrieved focused on the science that could be performed from the aircraft platform, rather than the S&T of the platform, and were not aligned with the study goals. Therefore, no adjustments were required to the EC query, whereas, with the SCI, many NOT Boolean terms were required to eliminate aircraft papers not aligned with the main study objectives. It is analogous to the selection of a mathematical coordinate system for solving a physical problem. If the grid lines are well aligned with the physical problem to be solved, the equations will be relatively simple. If the grid lines are not well aligned, the equations will contain a large number of terms required to translate between the geometry of the physical problem and the geometry of the coordinate system.

The most important message to be extracted from the aircraft and parallel studies is that ***the information retrieval query size depends on the objectives of the study, and the contents of the database relative to the study objectives.*** The query size should not be pre-determined, but should result from the attainment of the comprehensiveness and precision objectives.

Another important message is that substantial manual labor is required to examine the thousands of detailed technical phrases that result from the computational linguistics analyses of the free

text, and to make judgements about the applicability of these phrases to inclusion in the final query. Because these queries are applied to multi-discipline source databases such as the Science Citation Index, an understanding of the use of these phrases in other technical disciplines is required for successful query development. Thus, the person or team developing a query for a specific technical sub-discipline requires broader technical knowledge than in the target discipline alone.

2) BIBLIOMETRICS

-MOST PROLIFIC AUTHORS, *ANALYTICAL CHEMISTRY*
As a simple example of a bibliometrics output, records of the 2000 most recent articles (as defined by the SCI) published in the journal *Analytical Chemistry* (June 1998-August 2000) were extracted from the SCI. There were 5072 authors listed. The most prolific authors, and the number of papers on which they were listed, include: Ramsey JM (19), Smith RD (18), Wang J (17), Jacobson SC (14), Yeung ES (12), Anderson GA (11), Umezawa Y (11), Carr PW (11), Guillame YC (10), Peyrin E (10), Sweedler JV (10). These are rather impressive numbers for a two-year publication period in a prestigious journal.

The author distribution function is shown on Figure 1. Most of the authors have only one or two publications. Previous technical discipline studies

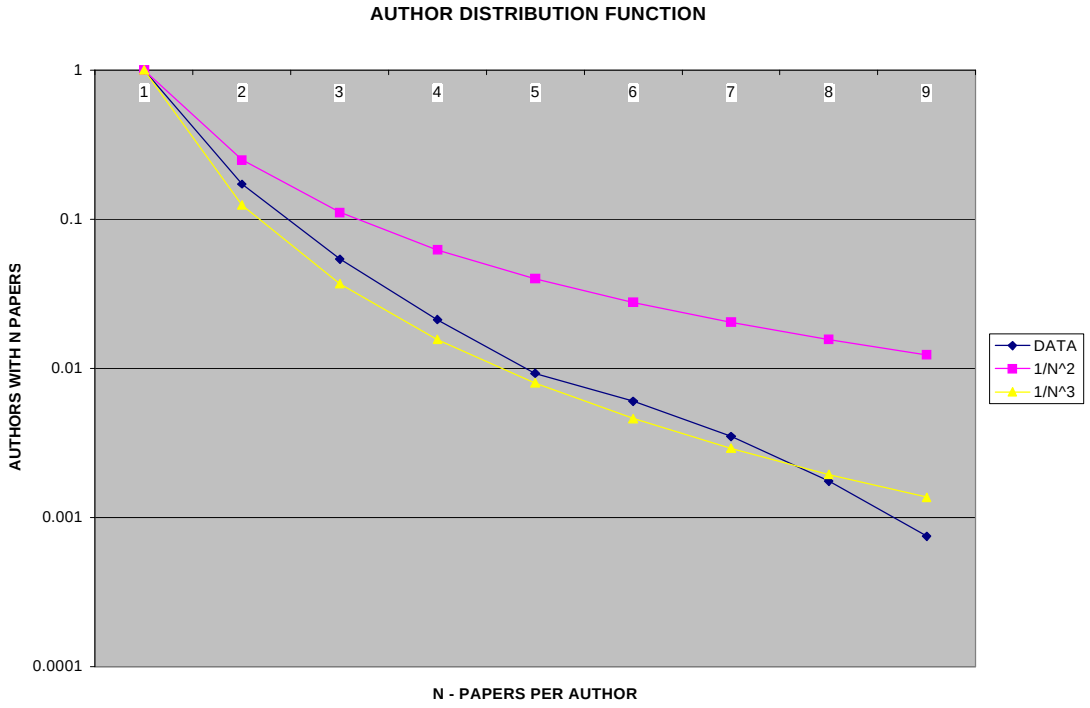


FIGURE 1

(14, 15, 16) show author distribution functions that range from 1/N^2 to 1/N^3. The present author distribution function is within that range, closer to 1/N^3.

-MOST CITED AUTHORS, *ANALYTICAL CHEMISTRY*
There were 22200 different authors cited from the same *Analytical Chemistry* database. The most cited authors include Jacobson SC (164), Giddings JC (123), Wang J (115), Bakker E (106), Grate JW (93), Bard AJ (87). There is reasonable correlation between the top 20 or so prolific authors and the top 20 cited authors, showing that many of the pioneers of present-day analytical chemistry thrust areas are still quite active. It should be re-emphasized that these integrated author citation numbers reflect only references contained in the 2000 most recent *Analytical Chemistry* articles, and an author's total citations from all sources could be substantially greater. An independent check of Bard AJ in the SCI, for example, showed tens of thousands of citations for all papers, as opposed to the 87 listed for this study.

-MOST PROLIFIC JOURNALS, MASS SPECTROMETRY
In this example, records of the 2000 most recent papers referenced in the SCI, and containing the term mass spectrometry (the highest frequency technique phrase from the 2000 records extracted from the journal *Analytical Chemistry* above) in the title were extracted. There were 377 journals listed. The journals containing the most mass spectrometry papers include *Rapid Communications In Mass Spectrometry* (224), *Journal Of Chromatography: A* (157), *Analytical Chemistry* (138), *Journal Of Mass Spectrometry* (93), *Journal Of Chromatography*: B (93), *Journal Of Analytic And Atomic Spectrometry* (75), and *Journal Of The American Society Of Mass Spectrometry* (65). The journal frequency decreases rapidly after this group. The first three journals appear to form the top core group, and the next four form the second core group.

In yhe author's standard text mining studies of a discipline, the iteratively-developed query used for the records from which the bibliometrics are derived would typically involve substantial time and effort, and contain hundreds of terms, not just one (mass spectrometry) as in this illustrative example.

-MOST CITED JOURNALS, *ANALYTICAL CHEMISTRY*
There were 6177 different journals/ sources cited by the 2000 *Analytical Chemistry* papers. The most cited journals include *Analytical Chemistry* (9107), *Journal of Chromatography: A* (1525), *Journal of Chromatography* (1427), *Journal of the American Chemical Society* (1334), *Analytic Chim Acta* (1177), *Rapid Communications in Mass Spectrometry* (901), *Journal of Electroanalytical Chemistry* (889), and *Science* (806). These rankings reflect two characteristic phenomena seen in previous studies. The journal in which the citing papers are published tends to be cited frequently, and the more fundamental journals tend to be cited with higher frequency than the applied journals.

-MOST PROLIFIC INSTITUTIONS, *ANALYTICAL CHEMISTRY*
The most prolific organizations were identified from the 2000 *Analytical Chemistry* papers database. The organization names, and the number of articles on which they were listed, include: Univ Calif (all campuses, and including LASL and LANL) (83), Oak Ridge Natl Lab (45), Univ

Michigan (36), Univ Texas (32), Univ Tokyo (31), Univ Washington (27), Iowa State Univ (27), Univ Alberta (26), Univ N Carolina (25), Indiana Univ (25), Univ Florida (23), Univ Illinois (22), Texas A&M Univ (20), Univ Lund (18), Texas Tech Univ (17), Sandia Natl Labs (17), Univ Tennessee (16), Cornell Univ (15).

This example illustrates some of the limitations of metrics in general, and bibliometrics in particular. The institutions listed tend to be large, and one would expect large numbers of outputs. There is no indication of efficiency; i.e., output per unit of resources. There is no indication of output quality, other than the papers exceeded the obviously high threshold required for publication in *Analytical Chemistry*. Because of space limitations, organizational sub-units could not be listed. Thus, the high achievements of a sub-unit may not be reflective of the institution overall.

-MOST PROLIFIC COUNTRIES, *ANALYTICAL CHEMISTRY*
The most prolific countries were identified from the 2000 *Analytical Chemistry* papers database. The country names, and the number of articles on which they were listed, include: USA (1098), Japan (156), Germany (129), Canada (118), England (96), Switzerland (62), Sweden (59), France (53), Spain (53), Netherlands (44). When all countries are included, the USA has as many listings as all other countries combined. This dominance by the USA is characteristic of total discipline study bibliometrics obtained previously, although the dominance is slightly exaggerated in *Analytical Chemistry*.

-MOST CITED PAPERS, *ANALYTICAL CHEMISTRY*
There were 35243 different papers cited by the 2000 *Analytical Chemistry* papers. The most cited papers include Jacobson SC, *Analytical Chemistry*, 1994; Fenn JB, *Science*, 1989; Harrison DJ, *Science*, 1993; Hjerten S, *Journal of Chromatography*, 1985; and Karas M, *Analytical Chemistry*, 1988. Of the ten most highly cited papers, half were in the 1980s and half were in the 1990s. This reflects a relatively dynamic field.

Again, the numbers of citations from the limited citing population do an injustice to total paper citations. The 1989 paper by Fenn JB, for example, was listed with 37 citations, but had total citations from all sources of almost 1350. Additionally, the 1980 paper by Bard AJ was listed with 25 citations, but had total citations from all sources of over 4000. Our more comprehensive discipline studies generate numbers more consonant with total citations from all sources.

BARRIERS TO S&T TEXT MINING IMPLEMENTATION

Despite the myriad potential applications of text mining to the advancement of S&T, the surface of this powerful technique has barely been scratched. There exist many barriers to its widespread implementation, and these will be outlined. These barriers include: 1) lack of incentives; 2) lack of awareness of available text mining capabilities; 3) database limitations; 4) lack of coordination in technical community; 5) text mining not integrated with business operations.

1) Lack of Incentives

A substantial effort is required to obtain high quality information retrieval and text mining. The computer can produce thousands of phrases and phrase patterns from the core text. Human expertise is required to sift out the nuggets from the large background clutter. Unfortunately, there are presently few, if any, rewards for expending the effort on high quality text mining, and there are essentially no penalties for doing low quality text mining. In addition, the 'not-invented-here' syndrome is a strong dis-incentive for expending substantial effort to determine S&T performed elsewhere.

2) Lack of Awareness of Available Text Mining Capabilities

S&T personnel are unaware of required or available processes and tools for, and subsequent potential benefits from, high quality information retrieval and text mining. How many readers of *Analytical Chemistry* had any familiarity with text mining before reading this article?

3) Database Limitations

The base data available restricts what can be obtained from text mining. There is over $500 Billion of S&T being performed globally on an annual basis. Only a very modest fraction of this S&T is documented (21). Of the S&T documented, only a modest fraction is accessed by the major S&T databases (Science Citation Index, Engineering Compendex, NTIS Technical Reports, etc). Of this accessed documented S&T, only a modest fraction is available to the user because of cost, restricted access, inclusion of data fields not uniform across databases, lack of awareness, and user unfriendliness of the software. A major factor driving this step and the previous step is that the contents of the databases are determined by the database developers, not the S&T sponsors or the users. Of the available accessed documented S&T, only a modest fraction is available to the information processing software due to poor information retrieval techniques, and poor text-to-phrase conversion techniques.

4) Lack of Coordination in Technical Community

Database development, data input quality and structure, and data dissemination require horizontal co-operation among global entities, and vertical co-operation among the full spectrum of S&T sponsors, database developers, journal publishers and editors, and research performers and managers. There is no coordinated agreement and support for the full data development and dissemination cycle. The paradox exists that co-operation among competitors is required for the common good.

5) Text Mining not Integrated with Business Operations

Organizationally, text mining and other decision aids are not treated as an integral part of the S&T strategic management process (22). Rather, it is treated as an ad hoc add-on, in isolation from other management decision aids. The downside of such an approach is that the study objectives are driven by the data available from ordinary business operations, rather than the study objectives driving the data necessary to quantify the business performance metrics.

CONCLUSIONS

Text mining comprises a system of algorithms and procedures that, when coupled with expert human analysts, can extract highly useful information from technical text. The typical iteratively-generated queries used in our studies contain a few hundred phrases/ phrase combinations. These queries are more than an order of magnitude larger than those used by the average researcher for literature searches. Queries of this length are required for comprehensive and highly relevant retrievals of the target literature, related literatures, and disparate literatures with some common thread. The quality of the retrieved literature limits the potential quality of any subsequent information processing, whether it is bibliometrics, computational linguistics, or literature-based discovery and innovation. Development of these high-quality queries requires time and some cost, and participation of both technical domain and information technology experts.

The bibliometrics analyses in our studies are useful for identifying credible experts for workshops and review panels, and for planning itineraries of productive individuals and organizations to be visited. The wide spectrum discipline database generated by the enhanced query allows more innovation-oriented workshops to be conducted (13). through identifying more related technical disciplines, and the leading experts in these disciplines.

The final benefit addressed is one that has occurred in every one of the text mining studies that have been performed, and its value cannot be stressed too strongly. From an organization's long-range strategic viewpoint, the main output from these text mining studies is the technical expert(s) who has had his/ her horizons and perspectives broadened substantially as a result of participating in the full text mining process, and who can use this expanded knowledge to better support the conduct and the management of the S&T. While the text mining tools/ processes/ protocols/ tangible products are important, they are of lesser importance to the organization's long-term strategic health relative to the expert with advanced capabilities.

Text mining has enormous potential to support the rapid advancement of S&T. High quality S&T text mining requires substantial time and effort. There exist a number of barriers to its wide-scale implementation. They all originate from the absence of serious global agreements to develop the databases, train skilled personnel in S&T text mining, develop affordable high quality text mining techniques for a variety of applications, and implement prototype demonstrations of these techniques.

REFERENCES FOR APPENDIX 7-C.

(1) Losiewicz, P.; Oard, D.; Kostoff, R. N. *Journal of Intelligent Information Systems*. 2000. 15: 2. 99-119.

(2) Hearst, M. A. *Proceedings of ACL '99, the 37th Annual Meeting of the Association of Computational Linguistics, 1999.* University of Maryland, June 20-26.

(3) PROCEEDINGS SIGIR 99, *Proceedings of the Conference on the 22nd International Conference on Research and Development in Information Retrieval*, eds., Hearst, M., Gey, F., and Tong, R., ACM Press, New York, N.Y., 1999.

(4) NIST Special Publication 500 240; The Sixth Text Retrieval Conference (TREC-6), NTIS, eds., *E. M. Voorhees and D. K. Harman*, order number PB98-148166, Department of Commerce, National Institute of Standards and Technology, 1998.
(5) Kostoff, R. N.; Eberhart, H. J.; Toothman, D. R. *Journal of Information Science* 1997. 23:4. 301-311.
(6) Braun, T.; Bujdoso, E.; Schubert, A. *Literature of Analytical Chemistry: A Scientometric Evaluation*. CRC, CRC Press, 1987.
(7) Swanson, D.R. *Perspect Biol Med.* 1986. 30: 1. 7-18.
(8) Swanson, D.R.: Smalheiser, N.R. *Artif Intell. 1997.* 91:2. 183-203.
(9) Swanson, D.R. *Abstr Pap Am Chem S. 1999.* 217. U551-U551.
(10) Smalheiser, N.R.; Swanson, D.R. *Neurosci Res Commun.* 1994. 15: (1). 1-9.

(11) Smalheiser, N.R.; Swanson, D.R. *Comput Meth Prog Bio.* 1998. 57: (3). 149-153.
(12) Smalheiser, N.R.; Swanson, D.R. *Arch Gen Psychiat.* 1998. 55 :8. 752-753.
(13) Kostoff, R. N. *Technovation*. 1999. 19: 10. 593-604.
(14) Kostoff, R. N.; Green, K. A.; Toothman, D. R.; Humenik, J. *Journal of Aircraft* 2000. 37:4. 727-730.
(15) Kostoff, R. N.; Eberhart, H. J.; Toothman, D. R. *Information Processing and Management* 1998. 34:1. 69-85.
(16) Kostoff, R. N.; Braun, T.; Schubert, A.; Toothman, D. R.; Humenik, J. *Journal of Chemical Information and Computer Science*. 2000. 40. 19-39.
(17) Gordon, M. D.; Lindsay, R. K. *JASIS.* 1996. **47**:2. 116-128.
(18) Finn, R. *The Scientist*. 1998. 12:10. 12-13.
(19) Kostoff, R. N.; Eberhart, H. J.; Toothman, D. R. *Journal of the American Society for Information Science*. 1999. 50:5. 427-447.
(20) Zamir, O. *PhD Thesis* . University of Washington, 1999.
(21) Kostoff, R. N. *The Scientist*, 2000. 14:9. 6-6.
(22) Kostoff, R. N.; Geisler, E. *Technology Analysis and Strategic Management*. 1999. 11:4. 493-525.
(23) Kostoff, R. N.; DeMarco, R. A. *Analytical Chemistry*. 2001. 73:13. 370-378A.

APPENDIX 7-D.

POWER SOURCE TEXT MINING USING BIBLIOMETRICS AND DATABASE TOMOGRAPHY [Kostoff, 2005c]

1. INTRODUCTION

The present Appendix describes use of the DT process, supplemented by literature bibliometric analyses, to derive technical intelligence from the published literature of Power Sources science and technology.

Power Sources, as defined by the author for this study, consists of systems and processes for generating and converting power, and storing energy. It is defined operationally by a query with two components: 1) a phrase-based query, obtained by the iterative technique referenced in the next paragraph; and 2) a journal-title-based query, obtained by identifying non-technology-specific power source journals from the SCI journal listing under Energy and Fuels whose articles were deemed highly relevant to the Power Sources topic. Since one of the key outputs of the present study is a query that can be used by the community to access relevant Power Sources documents, a recommended query based on this study is presented in Appendix 1. This query serves as the operational definition of Power Sources, and its development is discussed in the database generation section.

To execute the study reported in this paper, a database of relevant Power Sources articles is generated using the iterative search approach of Simulated Nucleation [4, 5]. Then, the database is analyzed to produce the following characteristics and key features of the Power Sources field: recent prolific Power Sources authors; journals that contain numerous Power Sources papers; institutions that produce numerous Power Sources papers; keywords most frequently specified by the Power

Sources authors; authors, papers and journals cited most frequently; pervasive technical themes of Power Sources; and relationships among the pervasive themes and sub-themes.

2. BACKGROUND

2.1 Overview

Recent DT/ bibliometrics studies were conducted of the technical fields of: 1) Near-earth space (NES) [6]; 2) Hypersonic and supersonic flow over aerodynamic bodies (HSF) [5]; 3) Chemistry (JACS) [7] as represented by the Journal of the American Chemical Society; 4) Fullerenes (FUL) [8]; 5) Aircraft (AIR) [9]; 6) Hydrodynamic flow over surfaces (HYD); 7) Electrochemical Power Sources (ECHEM); and 8) the non-technical field of research impact assessment (RIA) [7]. Overall parameters of these studies from the SCI database results and the current EPS study are shown in Table 1.

TABLE 1 - DT STUDIES OF TOPICAL FIELDS

TOPICAL AREA	NUMBER OF SCI ARTICLES	YEARS COVERED
1) NEAR-EARTH SPACE (NES)	5480	1993-MID 1996
2) HYPERSONICS (HSF)	1284	1993-MID 1996
3)CHEMISTRY (JACS)	2150	1994
4) FULLERENES (FUL)	10515	1991-MID 1998
5) AIRCRAFT (AIR)	4346	1991-MID 1998
6) HYDRODYNAMICS (HYD)	4608	1991-MID 1998
7) ELECTROCHEM POWER (ECHEM)	6985	1991-MID-2001
8) RESEARCH ASSESSMENT (RIA)	2300	1991-BEG 1995
9) ELECTRIC POWER SOURCES (EPS)	20835	1991 – LATE 2000

Unique Study Features

The study reported in the present Appendix is in the journal article abstract category. It differs from the previous published papers in this category [5-9] in four respects. First, the topical domain (power sources) is completely different. Second, a more rigorous technical theme clustering approach is used. Third, the phrase-based query approach has been supplemented by the journal-title-based query approach. Fourth, since estimation of relative global levels of emphasis in power sources was desired, a generic power sources query was used in both the phrase-based and journal-title-based queries (e.g., ELECTRICITY PRODUCTION), rather than using power source-specific terms (e.g., FUEL CELL). A companion study will examine the more specific sub-area of ELECTROCHEMICAL POWER SOURCES using specific terms rather than the generic terms.

3. DATABASE GENERATION

The key step in the power source literature analysis is the generation of the database. There are three key elements to database generation: the overall objectives, the approach selected, and the database used. Each of these elements is described.

3.1 Overall Study Objectives

The main objective was to identify global S&T that had both direct and indirect relations to Power Sources. One sub-objective was to estimate the overall level o f global effort in Power Sources S&T, as reflected by the emphases in the published literature. Another sub-objective was to determine whether any radically new power sources were under development.

It was believed that if known specific technical terms were used for the query, there would be three negative impacts relative to the objectives above. First, the query would be biased toward the specific technologies reflected in the query, and the records retrieved would reflect this bias. The relative global efforts devoted toward each technology would have little credibility. Second, use of specific technical terms in the query would identify advances made in existing technologies, but might not access radically new technologies. Third, the query size would have been unmanageable, and unusable in present search engines. An unpublished study of controlled fusion energy resulted in a query of hundreds of terms after only the first iteration. The companion study to the present study, on the topic of electrochemical power sources, generated a query with hundreds of terms. Summing this experience over all the source, converter, and storage technologies contained within the umbrella of power sources S&T would have generated many hundreds or thousands of query terms.

Thus, it was decided to use generic energy or power-related terms for the query, relatively independent of any specific power supply, conversion, or storage system (e.g., ELECTRICITY PRODUCTION vs LIGHT-WATER REACTOR). This approach would retrieve documents that described technologies specifically related to power production, conversion, and storage. To retrieve documents related to power production, but where the author may not have used specific terminology relating the technology to power production in the write-up, the journal-based approach was added. The concept was to identify power source journals that were generic, not source specific, and add their articles to the phrase-based query database.

However, even with the use of both approaches, one class of articles will not be retrieved. These are power source-related articles that do not contain the generic terms relating them to power sources, nor are published in a journal with a dedicated power source emphasis. Thus, an article on a new scientific phenomenon potentially related to power sources that was published in, for example, Science or Nature would not appear in this retrieval. To retrieve such articles, a detailed technology-specific query, such as the type developed in past DT studies, is required. A companion study on Electrochemical Power Sources developed such a query.

3.2 Databases and Approach

The Science Citation Index was the database used for the present study. The approach used for query development was the DT-based iterative relevance feedback concept [4].

3.2.1 Science Citation Index [10]

The database consists of selected journal records (including authors, titles, journals, author addresses, author keywords, abstract narratives, and references cited for each paper) obtained by searching the Web version of the SCI for power source articles. At the time the present paper was written, the Web version of the SCI accessed about 5600 journals (mainly in physical, engineering, and life sciences basic research).

The SCI database selected represents a fraction of the available Power Source (mainly research) literature, that in turn represents a fraction of the Power Source S&T actually performed globally

[11]. It does not include the large body of classified literature, or company proprietary technology literature. It does not include technical reports or books or patents on Power Sources. It covers a finite slice of time (1991 to late 2000). The database used represents the bulk of the peer-reviewed high quality Power Source science and techno logy documented, and is a representative sample of all Power Source science and technology in recent times.

To extract the relevant articles from the SCI, the phrase-based query and the journal-title-based query were used, and the results combined with duplications eliminated. For application of the phrase-based query, the Title, Keyword, and Abstract fields were searched using phrases relevant to power sources. The resultant Abstracts were culled to those relevant to power sources. The search was performed with the aid of two powerful DT tools (multi-word phrase frequency analysis and phrase proximity analysis) using the process of Simulated Nucleation [4].

An initial query of generic power source-related terms produced two groups of papers: one group was judged by domain experts to be relevant to the subject matter, the other was judged to be non-relevant. Gradations of relevancy or non-relevancy were not considered. An initial database of Titles, Keywords, and Abstracts was created for each of the two groups of papers. Phrase frequency and proximity analyses were performed on this textual database for each group. The high frequency single, double, and triple word phrases characteristic of the relevant group, and their boolean combinations, were then added to the query to expand the papers retrieved. Similar phrases characteristic of the non-relevant group were effectively subtracted from the query to contract the papers retrieved. The process was repeated on the new database of Titles, Keywords, and Abstracts obtained from the search. A few more iterations were performed until the number of records

retrieved stabilized (convergence). The final approximately 400 term phrase-based query used for the Power Source study is shown in Appendix 1.

For application of the journal-title-based query to the SCI database, articles contained in the 68 journals classified by the SCI under the category Energy and Fuels were sampled. Those journals that were not power-source specific, and that contained a very high fraction of articles deemed relevant to the Power Source topic, were identified, and all their articles were included in the retrieved database. The final journal title-based query used for the Power Source study identified the eleven journals shown in the Introduction.

4. RESULTS

The results from the publications bibliometric analyses are presented in section 4.1, followed by the results from the citations bibliometrics analysis in section 4.2. Results from the DT analyses are shown in section 4.3. The SCI bibliometric fields incorporated into the database included, for each paper, the author, journal, institution, and Keywords. In addition, the SCI included references for each paper.

4.1 Publication Statistics on Authors, Journals, Organizations, Countries

The first group of metrics presented is counts of papers published by different entities. These metrics can be viewed as output and productivity measures. They are not direct measures of research quality, although there is some threshold quality level inferred, since these papers are published in the

(typically) high caliber journals accessed by the SCI.

4.1.1 Author Frequency Results

There were 20825 papers retrieved, 34808 different authors, and 60493 author listings. The occurrence of each author's name on a paper is defined as an author listing. While the average number of listings per author is about 1.7, the ten most prolific authors (see Table 2) have listings more than an order of magnitude greater than the average. The number of papers listed for each author are those in the database of records extracted from the SCI using the query, not the total number of author papers listed in the source SCI database.

TABLE 2 – MOST PROLIFIC AUTHORS
(present institution listed)

AUTHOR NAME	INSTITUTION	COUNTRY	# PAPERS
WU C	U. S. NAVAL ACADEMY	USA	71
KANDIYOTI R	UNIVERSITY LONDON	UK	69
TIWARI GN	INDIAN INST TECHNOLOGY	INDIA	62
DINCER I	KING FAHD UNIV	SAUDI ARABIA	61
GARG HP	INDIAN INST TECHNOLOGY	INDIA	49
KANDPAL TC	INDIAN INST TECHNOLOGY	INDIA	48
SNAPE CE	UNIV NOTTINGHAM	UK	43
WILLIAMS A	UNIV LEEDS	UK	42
ISHIKAWA M	YAMAGUCHI UNIV	JAPAN	41
KUMAR S	INDIAN INST TECHNOLOGY	INDIA	39

Of the ten most prolific authors listed in Table 2, four are from India, three are from the UK, and one each from the USA, Japan, and Saudi Arabia. All are from universities. This country distribution differs radically from any in previous studies, with the high concentration from India. The electrochemical power sources study showed 65% of the prolific authors from the Far East, mainly

Japan and China.

Because of the nature of the query used in the present study, many traditional energy production and conversion technologies were included (solar cooking, solar drying, solar distillation, biomass, coal combustion, etc). Reading of thousands of Abstracts confirmed that much of the Power Sources S&T focused on relatively low technology traditional approaches, especially research from the developing countries. The most prolific Indian authors addressed the solar and biomass topics. Interestingly, the most prolific British authors all concentrated on coal, including combustion, properties, and gasification.

4.1.2 Journals Containing Most Power Sources Papers

There were 1422 different journals represented. This is twice the number of journals from any of the previous studies, and again reflects the multi-disciplined nature of EPS. There was an average of 14.64 papers per journal. This number is somewhat inflated compared to the journal averages from other text mining studies. In the journal-derived component of the present study, all the papers in eleven journals were used. Nevertheless, even for those journals identified by the query-derived component of the database, the journals containing the most Power Source papers had in some cases an order of magnitude more papers than the average (See Table 3).

TABLE 3 – JOURNALS FROM QUERY-DERIVED COMPONENT OF DATABASE CONTAINING MOST PAPERS

JOURNAL NAMES	# PAPERS
J. ENG. GAS. TURBINES POWER-TRANS. ASME	200

INT. J. HYDROG. ENERGY	186
J. PROPUL. POWER	140
BIOMASS BIOENERG.	134
COMBUST. SCI. TECHNOL.	121
BRENNST.-WARME-KRAFT	119
IEEE TRANS. MAGN.	108
COMBUST. FLAME	103
ENERGY POLICY	102
SOL. ENERGY	98
APPL. ENERGY	90
COMBUST. EXPLOS.	88
J. APPL. PHYS.	82
SOLID STATE ION.	75
FUSION TECHNOL.	71
J. ELECTROCHEM. SOC.	67
IEEE TRANS. ENERGY CONVERS.	62
JSME INT. J. SER. B-FLUIDS THERM. ENG.	58
APPL. THERM. ENG.	57
IEEE TRANS. POWER SYST.	55

4.1.3 Institutions Producing Most Power Sources Papers

A similar process was used to develop a frequency count of institutional address appearances. It should be noted that many different organizational components may be included under the single organizational heading (e.g., Harvard Univ could include the Chemistry Department, Biology Department, Physics Department, etc.). Identifying the higher level institutions is instrumental for these DT studies. Once they have been identified through bibliometric analysis, subsequent measures may be taken (if desired) to identify particular departments within an institution.

TABLE 4 – PROLIFIC INSTITUTIONS

INSTITUTION NAMES	COUNTRY	# PAPERS
INDIAN INST TECHNOL	INDIA	415
CSIC	SPAIN	186
PENN STATE UNIV	USA	172

RUSSIAN ACAD SCI	RUSSIA	164
TOHOKU UNIV	JAPAN	163
ARGONNE NATL LAB	USA	142
CSIRO	AUSTRALIA	137
KING FAHD UNIV PETR & MINERALS	SAUDI ARABIA	137
UNIV LEEDS	UK	127
UNIV TOKYO	JAPAN	122

Of the ten most prolific institutions, four are from the Far East, two are from Western Europe, two from the USA, one from Eastern Europe, and one from the Middle East. Five are universities, and the remaining five institutions are research institutes. Compared to previous studies, the ratios of research institutes to universities is relatively high in this study.

4.1.4 Countries Producing Most Power Sources Papers

There are 78 different countries listed in the results. The country bibliometric results are summarized in Table 5. The dominance of a handful of countries is clearly evident.

TABLE 5 – PROLIFIC COUNTRIES

COUNTRY	#PAPERS	POPULATION (MILLIONS)	GROSS DOMESTIC PRODUCT ($BILLIONS)	#PAPERS/ POPULATION	#PAPERS/ GROSS DOMESTIC PRODUCT
USA	5285	278	9963	19.01079	0.530463
JAPAN	2269	127	3150	17.86614	0.720317
ENGLAND	1358	60	1360	22.63333	0.998529
INDIA	1196	1030	2200	1.161165	0.543636
GERMANY	1141	83	1936	13.74699	0.58936
CANADA	997	31	775	32.16129	1.286452

FRANCE	813	59	1448	13.77966	0.561464
AUSTRALIA	603	19	445	31.73684	1.355056
PEOPLES R CHINA	586	1284	4500	0.456386	0.130222
ITALY	559	58	1273	9.637931	0.43912
SPAIN	498	40	720	12.45	0.691667
TURKEY	474	66	444	7.181818	1.067568
RUSSIA	464	145	1120	3.2	0.414286
SWEDEN	382	9	197	42.44444	1.939086
NETHERLANDS	353	16	388	22.0625	0.909794
SOUTH KOREA	316	48	765	6.583333	0.413072
EGYPT	294	68	247	4.323529	1.190283
POLAND	256	39	328	6.564103	0.780488
SAUDI ARABIA	248	23	232	10.78261	1.068966
GREECE	225	11	182	20.45455	1.236264

There appear to be three dominant groups in the twenty most prolific countries. The US and Japan constitute the most dominant group. England, India, Germany, Canada, and France constitute the next group, and the remaining countries constitute the third group.

Of these top twenty countries, two are from North America, five are from the Far East, nine are from Western Europe, two are from Eastern Europe, and two are from the Middle East. South America and Africa are not represented.

Weighting these regions by number of papers, the ranking is North America (6282), Western Europe (5803), Far East (4970), Eastern Europe (720), and Middle East (542). When total population and GDP are taken into account, some dramatic changes occur. For papers per unit of population in the top twenty, the top five are mainly Western European and English-speaking nations (SWEDEN, CANADA, AUSTRALIA, UK, NETHERLANDS), and the bottom five are dominated by Asia and Eastern Europe (CHINA, INDIA, RUSSIA, EGYPT, POLAND). For

papers per unit of GDP in the top twenty, the top five are mainly developed nations (SWEDEN, AUSTRALIA, CANADA, GREECE, EGYPT), and the bottom five are a more amorphous mix (CHINA, SOUTH KOREA, RUSSIA, ITALY, USA). Interestingly, for all three productivity measures, Canada and Australia rank high.

Figure 1 contains a co-occurrence matrix of the top 15 countries. In terms of absolute numbers of co-authored papers, the USA major partners are Canada, Japan, Germany, England, China, and France. Overall, countries in similar geographical regions tend to co-publish substantially, although the larger producers (e.g., USA, Japan) are universal in their co-publishing.

FIGURE 1 – COUNTRY-COUNTRY CO-OCCURRENCE MATRIX

	# Records	5285	2269	1358	1196	1141	997	813	603	586	559	498	474	464	382	353
# Records	**Country**	**USA**	**JAPAN**	**ENGLAND**	**INDIA**	**GERMANY**	**CANADA**	**FRANCE**	**Australia**	**PEOPLES R CHINA**	**ITALY**	**SPAIN**	**TURKEY**	**RUSSIA**	**SWEDEN**	**NETHERLANDS**
5285	**USA**	5285	84	59	27	62	85	47	30	56	28	25	9	20	8	29
2269	**Japan**	84	2269	14	11	11	26	10	19	19	5	2	2	5	2	3
1358	**England**	59	14	1358	6	21	7	20	11	10	14	24	16	2	8	11
1196	**India**	27	11	6	1196	8	4	2	1	1	5	1				1
1141	**Germany**	62	11	21	8	1141	10	15	7	1	10	8	6	8	9	13
997	**Canada**	85	26	7	4	10	997	13	6	10	2	2	6	3	2	2
813	**France**	47	10	20	2	15	13	813	1		17	30		14		9
603	**Australia**	30	19	11	1	7	6	1	603	11		1	1	1	3	2
586	**Peoples R China**	56	19	10	1	1	10		11	586					4	5
559	**Italy**	28	5	14	5	10	2	17			559	6	1	1	6	7
498	**Spain**	25	2	24	1	8	2	30	1		6	498		1	1	5
474	**Turkey**	9	2	16		6	6		1		1		474		2	2
464	**Russia**	20	5	2		8	3	14	1		1	1		464	2	7
382	**Sweden**	8	2	8		9	2		3	4	6	1	2	2	382	3

353	**Netherlands**	29	3	11	1	13	2	9	2	5	7	5	2	7	3	353

Figure 2 contains a Country-Time matrix, where the matrix elements are numbers of papers produced. The year 2000 results are only partially complete. Country productivity varied considerably as a function of time. For example, over the decade the USA increased number of papers by only a few percent. Japan doubled, England, India, Germany increased by about 50%, and China, South Korea, and Turkey approximately quintupled.

FIGURE 2 – COUNTRY-TIME MATRIX

RowItems	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
USA	471	456	587	532	505	566	552	521	500	433
JAPAN	132	137	154	144	267	227	363	259	270	251
ENGLAND	79	93	112	157	143	159	130	158	146	132
INDIA	119	85	94	130	111	128	113	144	124	114
GERMANY	102	95	110	106	103	107	103	148	136	83
CANADA	72	85	95	92	124	116	116	84	107	91
FRANCE	52	44	62	79	92	92	88	93	129	64
AUSTRALIA	37	54	54	55	38	73	54	60	59	73
PEOPLES R CHINA	23	22	33	29	44	70	57	106	107	79
ITALY	22	27	48	47	61	57	59	82	70	65
SPAIN	20	26	23	51	49	54	71	57	77	60
TURKEY	12	16	26	29	46	63	57	56	78	83
RUSSIA		15	32	36	43	56	61	43	64	35
SWEDEN	21	16	33	39	27	60	40	46	41	52
NETHERLANDS	14	26	35	45	34	44	37	45	32	29
SOUTH KOREA	15	13	7	11	23	24	38	42	78	53
EGYPT	16	12	27	37	27	32	39	36	23	38
SAUDI ARABIA	14	11	16	29	21	41	12	41	37	24
POLAND	9	11	20	37	29	25	23	37	28	28
GREECE	11	13	16	21	17	26	26	35	27	28
TAIWAN	12	12	13	21	18	35	26	23	18	29
ISRAEL	14	14	27	11	19	18	20	24	27	17
SCOTLAND	13	7	13	18	13	19	22	32	24	21
FINLAND	16	14	11	14	23	23	17	26	19	20
BRAZIL	3	12	5	3	6	16	23	34	33	30

Figure 3 contains a Country-Journal matrix, for the top fifteen countries and top seventeen journals. The matrix entries are expressed in decimal fraction of each country's total papers in the seventeen journals. For each country, the bulk of its papers are contained in about four of the seventeen journals (i.e., journals containing about ten percent or more of a country's total papers).

In decreasing order, the four main journals for USA papers are: ENERGY & FUELS, FUEL, J POWER SOURCES, ENERGY. The papers in Energy & Fuels focus mainly (not exclusively) on fossil fuel properties, combustion efficiencies and pollution. The papers in Fuel focus mainly (with some biomass exceptions) on fossil fuel properties, additives, and reactant product properties and utilization. The papers in Journal of Power Sources focus on electrochemical power supply, with main emphasis on batteries and fuel cells. The papers in Energy focus on energy utilization, with emphasis on increasing efficiency and alternatives to reduce pollution.

For India, the five journals are: ENERGY CONV MANAG, INT J ENERGY RES, J POWER SOURCES, RENEW ENERGY, FUEL. The papers in Energy Conversion & Management focus on energy utilization, aimed at improving energy efficiency and reducing pollutants, with balanced emphasis given to solar and biomass systems. The papers in International Journal of Energy Research focus on performance of total energy systems and components, with reasonable emphasis provided to solar energy systems. The papers in Journal of Power Sources focus on rechargeable batteries and fuel cells. The papers in Renewable Energy focus on alternative energy sources and utilization, with focus on solar, but inclusion of biomass and other renewables like wind as well. The papers in Fuel focus on properties and combustion products of (mainly) fossil fuels. While there

is overlap with the USA in technical areas studies, there appears to be much more relative emphasis in solar-based systems and alternative power supplies in India relative to the USA.

For China, the four journals are: J POWER SOURCES, FUEL, ENERGY CONV MANAG, ENERGY. The papers in Journal of Power Sources focus on batteries (mainly rechargeable lithium) and fuel cells. The papers in Fuel focus on properties, combustion, and products of (mainly) fossil fuels, and, of those, almost exclusively on coals. The papers in Energy Conversion and Management focus on analysis of energy conversion and utilization across a wide variety of systems and applications. The papers in Energy focus on analysis and modeling of energy utilization in a wide variety of systems and applications. Relative to India, China has less focus on the solar and other alternative supplies, and more on fossil fuel combustion. All the above conclusions are based on these four or five major publishing journals' contents only, for each country.

FIGURE 3 – COUNTRY-JOURNAL MATRIX

JOURNAL	USA	JAPAN	ENGLAND	INDIA	GERMANY	CANADA	FRANCE	AUSTRALIA	CHINA	ITALY	SPAIN	TURKEY	RUSSIA	SWEDEN	NETHERLANDS
Fuel	0.157	0.126	0.305	0.092	0.147	0.211	0.23	0.337	0.175	0.147	0.44	0.198	0.207	0.171	0.183
J. Power Sources	0.151	0.3	0.16	0.109	0.374	0.135	0.398	0.19	0.305	0.203	0.08	0.002	0.239	0.122	0.228
Energy Fuels	0.27	0.211	0.047	0.015	0.056	0.16	0.126	0.153	0.056	0.04	0.269	0.05	0.033	0.137	0.041
Energy Conv. Manag.	0.07	0.181	0.069	0.296	0.043	0.097	0.05	0.033	0.133	0.168	0.031	0.214	0.109	0.072	0.219
Renew. Energy	0.033	0.041	0.181	0.096	0.104	0.031	0.081	0.151	0.047	0.176	0.088	0.074	0.065	0.11	0.082
Energy	0.091	0.062	0.025	0.082	0.078	0.056	0.027	0.019	0.128	0.053	0.047	0.133	0.054	0.152	0.068
Int. J. Energy Res.	0.022	0.016	0.054	0.197	0.024	0.087	0.025	0.041	0.077	0.061	0.016	0.079	0.022	0.065	0.018
Energy Sources	0.04	0.01	0.014	0.063	0.017	0.14	0.012	0.017	0.023	0.013	0.005	0.219	0.022	0.019	0.009
J. Eng. Gas. Turbines Power-Trans. ASME	0.043	0.018	0.012	0.001	0.015	0.011	0.002	0.000	0.002	0.024	0.000	0.000	0.000	0.011	0.005

J. Inst. Energy	0.009	0.003	0.088	0.004	0.006	0.009	0.006	0.021	0.019	0.005	0.000	0.01	0.011	0.03	0.05
Int. J. Hydrog. Energy	0.016	0.008	0.003	0.017	0.047	0.027	0.002	0.000	0.009	0.032	0.003	0.002	0.109	0.011	0.005
J. Propul. Power	0.033	0.006	0.002	0.002	0.004	0.004	0.000	0.006	0.005	0.003	0.000	0.000	0.065	0.011	0.009
Biomass Bioenerg.	0.013	7E-04	0.011	0.011	0.006	0.016	0.002	0.000	0.000	0.011	0.005	0.000	0.000	0.076	0.032
Combust. Sci. Technol.	0.016	0.005	0.016	0.003	0.019	0.004	0.008	0.004	0.002	0.021	0.000	0.000	0.022	0.000	0.014
Combust. Flame	0.016	0.004	0.008	0.000	0.009	0.009	0.01	0.004	0.007	0.011	0.005	0.000	0.011	0.000	0.014
Sol. Energy	0.004	0.002	0.005	0.011	0.026	0.003	0.006	0.023	0.009	0.019	0.01	0.019	0.000	0.008	0.018
IEEE Trans. Magn.	0.017	0.007	0.001	0.001	0.026	0.001	0.014	0.002	0.002	0.011	0.000	0.000	0.033	0.004	0.005

4.2 Citation Statistics on Authors, Papers, and Journals

The second group of metrics presented is counts of citations to papers published by different entities. While citations are ordinarily used as impact or quality metrics [15], much caution needs to be exercised in their frequency count interpretation, since there are numerous reasons why authors cite or do not cite particular papers [16, 17].

The citations in all the retrieved SCI papers were aggregated, the authors, specific papers, years, journals, and countries cited most frequently were identified, and were presented in order of decreasing frequency. A small percentage of any of these categories received large numbers of citations. From the citation year results, the most recent papers tended to be the most highly cited. This reflected rapidly evolving fields of research.

4.2.1 Most Cited Authors

The most highly cited authors are listed in Table 6.

TABLE 6 – MOST CITED AUTHORS

(cited by other papers in this database only)

AUTHOR	TOPIC	INSTITUTION	COUNTRY	#CITES
SOLOMON PR	COAL PYROLYSIS	ADV FUEL RES INC	USA	510
PAVLOV D	LEAD-ACID BATTERIES	BULGARIAN ACAD SCI	BULGARIA	420
BEJAN A	THERMODYNAMICS	DUKE UNIV	USA	405
AURBACH D	LITHIUM BATTERIES	BAR ILAN UNIV	ISRAEL	367
LARSEN JW	COAL PYROLYSIS	LEHIGH UNIV	USA	355
MOCHIDA I	CARBON APPLICATIONS	KYUSHU UNIV	JAPAN	292
OHZUKU T	LITHIUM BATTERIES	OSAKA CITY UNIV	JAPAN	274
SUUBERG EM	COAL PYROLYSIS	BROWN UNIV	USA	245
NISHIOKA M	COMBUSTION	NAGOYA UNIV	JAPAN	233
WU C	THERMODYNAMICS	US NAVAL ACADEMY	USA	230
DUFFIE JA	SOLAR HEATING	UNIV WISCONSIN	USA	221
VANKREVELEN DW	POLYMERS	AKZO RES AND ENGRNG	NETHERLANDS	206
DEVOS A	THERMODYNAMICS	STATE UNIV GHENT	BELGIUM	198
SUZUKI T	COAL PYROLYSIS	KYOTO UNIV	JAPAN	196
PAINTER PC	COAL PROPERTIES	PENN STATE UNIV	USA	194
LI CZ	COAL PYROLYSIS	UNIV LONDON IMPER COLL	UK	193
SABBAH R	COMB THERMODYNAMICS	CNRS	FRANCE	190
HEROD AA	COAL COMBUSTION	UNIV LONDON IMPER COLL	UK	190
CHEN JC	THERMODYNAMICS	XIAMEN UNIV	CHINA	185
HUFFMAN GP	FOSSIL COMBUSTION	UNIV KENTUCKY	USA	184

Of the twenty most cited authors, eight are from the USA, four are from Japan, five are from Western Europe, one from Israel, one from Bulgaria, and one from China. This is a far different distribution from the most prolific authors, where half were from Asia, and ten percent from the USA. There are a number of potential reasons for this difference, including difference in quality and late entry into the research discipline. In another three or four years, when the papers from present-

day authors have accumulated sufficient citations, firmer conclusions about quality can be drawn.

Ten of the authors worked on fossil fuels (mainly coal, mainly combustion), five worked in thermodynamics, three worked on batteries (mainly lithium), one worked on solar, and one worked on polymers.

The lists of most prolific authors and most highly cited authors only had one name in common (WU, C). This phenomenon of minimal intersection has been observed in all other text mining studies performed by the first author.

Sixteen of the authors' institutions are universities, two are government-sponsored research laboratories, and two are private companies. The appearance of the companies on this list is another differentiator from the list of most prolific authors.

The citation data for authors and journals represents citations generated only by the specific records extracted from the SCI database for this study. It does not represent all the citations received by the references in those records; these references in the database records could have been cited additionally by papers in other technical disciplines.

4.2.2 Most Cited Papers

The most highly cited papers are listed in Table 7.

TABLE 7 – MOST CITED PAPERS

(total citations listed in SCI)

AUTHOR	YEAR	JOURNAL	VOLUME	SCI CITES	TOTAL CITES
CURZON FL	1975	AM J PHYS	V43	154	366
CARNOT ENGINE EFFICIENCY AT MAXIMUM POWER OUTPUT					
MILLER JA	1989	PROG ENERG COMBUST	V15	90	825
MODELING NITROGEN CHEMISTRY IN COMBUSTION					
SOLUM MS	1989	ENERG FUEL	V3	83	170
SOLID STATE NMR OF ARGONNE PREMIUM COALS					
VORRES KS	1990	ENERG FUEL	V4	82	153
ARGONNE PREMIUM COAL					
FONG R	1990	J ELECTROCHEM SOC	V137	68	346
LITHIUM INTERCALATION INTO CARBON					
LARSEN JW	1985	J ORG CHEM	V50	59	125
STRUCTURE OF BITUMINOUS COALS					
SOLOMON PR	1990	ENERG FUEL	V4	59	44
ARGONNE PREMIUM COAL ANALYSIS					
IINO M	1988	FUEL	V67	56	112
COAL EXTRACTION					
OHZUKU T	1990	J ELECTROCHEM SOC	V137	54	336
MANGANESE DIOXIDE IN LITHIUM NONAQUEOUS CELL					
NISHIOKA M	1990	ENERG FUEL	V4	51	80
AROMATIC STRUCTURES IN COALS					

The theme of each paper is shown in italics on the line after the paper listing. The order of paper listings is inverse number of citations by other papers in the extracted database analyzed. The total number of citations from the SCI paper listing, a more accurate measure of total impact, is shown in the last column on the right.

Energy and Fuels contains the most papers, four out of the ten listed. Most of the journals are fundamental science journals, and most of the topics have a fundamental science theme. Most of the

papers are from the 1989-1990 time frame. This reflects a dynamic research field, with seminal works being performed in the recent past.

Six papers focus on coal issues, one on combustion, one on thermodynamics, and two on secondary lithium battery issues. Thus, the intellectual heritage focus is on conversion to electricity with a thermal step, as opposed to direct conversion to electricity. Even though the text analysis will show later a significant effort on renewables, this level of effort is not reflected in the intellectual heritage.

4.2.3. Most Cited Journals

TABLE 8 – MOST CITED JOURNALS
(cited by other papers in this database only)

JOURNAL	TIMES CITED
FUEL	15013
J ELECTROCHEM SOC	6600
ENERG FUEL	6317
J POWER SOURCES	4238
SOL ENERGY	2957
COMBUST FLAME	2611
SOLID STATE IONICS	1922
J CHEM PHYS	1752
CARBON	1686
J APPL PHYS	1654
J PHYS CHEM-US	1652
FUEL PROCESS TECHNOL	1573
ELECTROCHIM ACTA	1558
COMBUST SCI TECHNOL	1523
J AM CHEM SOC	1511
ENERGY	1466
IND ENG CHEM RES	1426
ANAL CHEM	1412
J CATAL	1371
NATURE	1358

Fuel received almost as many citations as the next three journals combined. Most of the highly cited journals are fossil fuel/ combustion oriented or electrochemical power source oriented. These are followed by some fundamental Chemistry and Physics journals. The only renewables journal interspersed is Solar Energy. These results are fully in line with those of the most cited authors and papers, and suggest that consensus seminal works have yet to be established for many of the renewables areas.

The authors end this bibliometrics section by recommending that the reader interested in researching the topical field of interest would be well-advised to, first, obtain the highly-cited papers listed and, second, peruse those sources that are highly cited and/or contain large numbers of recently published papers.

5. SUMMARY AND DISCUSSION

A query and journal-based hybrid process was used to retrieve records from the SCI for analysis. Generic energy or power-related terms were used for the query, relatively independent of any specific power supply, conversion, or storage system (e.g., ELECTRICITY PRODUCTION vs LIGHT-WATER REACTOR). This approach would retrieve documents that described technologies specifically related to power production, conversion, and storage. To retrieve documents related to power production, but where the author may not have used specific terminology relating the

technology to power production in the write-up, the journal-based approach was added. The concept was to identify power source journals that were generic, not source specific, and add their articles to the phrase-based query database.

Even with the use of both approaches, one class of articles will not be retrieved. These are power source-related articles that do not contain the generic terms relating them to power sources, nor are published in a journal with a dedicated power source emphasis. Thus, an article on a new scientific phenomenon potentially related to power sources that was published in, for example, Science or Nature would not appear in this retrieval. To retrieve such articles, a detailed technology-specific query, such as the type developed in past DT studies, is required.

Bibliometric analyses produced the EPS technical infrastructure. The most prolific EPS authors, journals, institutions, countries, cited authors/ journals/ paper were presented. There were 133 different countries listed. The dominance of a handful of countries was clearly evident (e.g., USA, Japan, England, India, Germany, Canada, France) but a series of small countries (Turkey, South Korea, Egypt, Greece, Taiwan) are also productive. The United States is more than twice as prolific as its nearest competitor (Japan), and is as prolific as its major competitors combined.

7. APPENDIX 1 TO APPENDIX 7-D - POWER SOURCES QUERY

Phrase-Based Component

(BIOMASS ENERGY OR CONVENTIONAL ENERGY OR DISTRICT HEATING OR

ELECTRICAL ENERGY OR ENERGY CONSUMED OR ENERGY RECOVERY OR ENERGY RESOURCE* OR ENERGY STORAGE OR HEAT ENGINE* OR HYBRID ENERGY OR MAGNETIC ENERGY OR POWER CONVERSION OR RENEWABLE SOURCE* OR SUSTAINABLE ENERGY OR (COGENERATION SAME (POWER OR HEAT)) OR (COMBUSTION SAME (ENERGY OR FUEL* OR POWER)) OR (ELECTRIC POWER SAME (RESEARCH OR TECHNOLOGY OR TURBOGENERATOR)) OR (ELECTRIC SAME (ENERGY CONSUMPTION OR FOSSIL FUEL* OR OUTPUT POWER OR POWER GENERATION OR POWER PRODUCTION OR TURBINE)) OR (ELECTRICAL SAME (EFFICIENCY OR ELECTRON MEDIATOR OR ENERGY SUPPLY OR FUEL* OR HEAT OR POWER DENSITY OR POWER GENERATION)) OR (ELECTRICITY SAME (BIOMASS OR ENERGY CONVERSION OR ENERGY SUPPLY OR ENERGY SYSTEM OR ENERGY TECHNOLOG* OR HEAT OR MICROBIAL FUEL* OR POWER GENERATION OR RENEWABLE ENERGY OR THERMAL)) OR (ENERGY CONSUMPTION SAME (BIOMASS OR POWER OR RENEWABLE ENERGY)) OR (ENERGY CONVERSION SAME RENEWABLE ENERGY) OR (ENERGY DISTRIBUTION SAME (ENERGY SOURCE* OR RENEWABLE ENERGY)) OR (ENERGY EFFICIENCY SAME POWER) OR (ENERGY SOURCE* SAME (ENERGY CONVERSION OR MOTOR* OR POWER GENERATION OR RENEWABLE ENERGY)) OR (ENERGY SYSTEM SAME POWER) OR (ENERGY TECHNOLOG* SAME (BIOMASS OR POWER OR RENEWABLE ENERGY)) OR (ENGINE SAME (ENERGY OR FUEL* OR POWER GENERATION OR POWER SYSTEM)) OR (FUEL* SAME (CYCLE OR ELECTRIC OR ELECTRIC ENERGY OR ELECTRIC POWER OR ELECTRON MEDIATOR OR ENERGY CONSUMPTION OR ENERGY SOURCE* OR ENERGY SYSTEM OR HEAT RECOVERY OR ION

CONDUCTIVITY OR POWER DENSITY OR POWER GENERATION OR POWER PLANT* OR POWER PRODUCTION OR RENEWABLE ENERGY OR RESEARCH AND DEVELOPMENT OR STORAGE OR THERMAL ENERGY OR VEHICLE OR BIOMASS OR COMBUSTION OR ENERGY SOURCE* OR RENEWABLE ENERGY OR TURBINE)) OR (HEAT RECOVERY SAME POWER) OR (POWER DENSITY SAME ION CONDUCTIVITY) OR (POWER GENERATION SAME (COMBINED CYCLE OR EFFICIENCY OR ENERGY CONVERSION OR HEAT OR PLANT* OR RESEARCH OR TECHNOLOGIES)) OR (POWER PLANT* SAME (COMBINED CYCLE OR EFFICIENCY OR ELECTRIC OR ENERGY OR POWER GENERATION)) OR (RENEWABLE ENERGY SAME (BIOMASS OR CONVERSION OR POWER GENERATION OR RESEARCH OR SUSTAINABLE DEVELOPMENT)) OR (THERMAL ENERGY SAME (POWER OR RENEWABLE ENERGY OR RESEARCH AND DEVELOPMENT))) NOT (ACBL OR ACCIDENT OR ACCIDENTS OR ACOUSTICALLY OR ACTA METALLURGICA INC OR ACTINIDE* OR ACTIVATION ENERGY ASYMPTOTICS OR ADIABATIC SATURATION COOLING OR AEROSOL OR AGE OR AIDS OR ANIMALS OR ANNEALED OR ANTISOLVENT OR AQUIFERS OR ASH-CONCRETE OR ASHES OR ATHENS OR BANDWIDTH OR BEAMS OR BENIGN OR BIT OR BODY OR CABLES OR CALIBRATION OR CANCER OR CAPITA OR CCA OR CELLULAR OR CEMENT OR CENT OR CHLORIDE OR CHLOROPHYLL OR CHROMOPHORE OR CIRCULATION OR CLAD OR CLOUD OR CLOUDS OR CONTAMINATION OR CORIOLIS OR CORONAL OR CRYOSTAT OR CURE OR CURING OR DAILY PEAK POWER OR DC DC CONVERTERS OR DEFORMATION OR DEICING OR DESALINATION OR DESALTING OR DESICCANT OR DETECTORS OR DISEASE OR DISTRICT HEATING SYSTEMS OR

DRUG OR DUMP OR EHL OR ELASTIC ENERGY STORAGE OR ELPI OR EROSION OR EXCIMER OR FACTORY OR FAT OR FATE OR FATIGUE OR FEEDFORWARD OR FERMION OR FIREBALL OR FISH OR FLARES OR FLUXES OR FOOT OR FRACTAL OR FREE FATTY ACIDS OR FREEBOARD OR FUMIGATION OR FUZZY OR GALAXIES OR GATE OR GEOLOGIC OR GLASSY OR HAND AND FOOT OR HANDPIECE OR HEAL OR HEALTH OR HEAR OR HEAT PIPE HEAT OR HEAT TRANSFER EQUATION OR HEAT TREATMENT TEMPERATURE OR HMX OR HYDRAULIC OR HYDRAZINE OR HYPERSONIC CRUISE TRAJECTORIES OR ILL OR INCOME OR INJURY OR INSTRUMENTS OR INTERNET OR INVERTER OR ISFSI OR JUICE OR KERNEL OR KILN OR LABOR OR LAKE OR LAMBDA OR LAMP OR LANDER OR LEPTIN OR LIMESTONE OR LINE CONTROL SYSTEM OR LINGUISTIC OR LOGIC OR LUBRICANT OR LUNCH OR MAGNESIUM OR MANTLE OR MBMS OR MEAL OR MERCURY OR MESOPORES OR MILE OR MILK OR MINERALS OR MLO OR MMA OR MODULATION OR MONETARY OR MONEY OR MONOTONIC OR MOTHER OR MSF OR MUSCLE OR NEEDLES OR NERVE OR NEURAL OR NFL OR NITRIC OR NITROUS OR NOISE OR NORMAL SPECTRAL EMISSIVITY OR NTT OR NUMBER OF MULTIPLEXERS OR OPERATORS OR ORBITAL OR PAIN OR PARASITIC OR PATIENTS OR PCB OR PIPING OR PLUME OR POLICIES OR PONDS OR POOL OR PROTEIN OR PROTEINS OR RADIO OR RAT OR RATS OR RECONNECTION OR REPRODUCTIVE OR RETROFIT OR RIVER OR ROAD OR ROSE OR SAUTER MEAN DIAMETER OR SEDIMENTS OR SHEET OR SIGNATURES OR SILICA OR SKELETON OR SLAG OR SOFTWARE OR SOIL OR SOILS OR SOLVENTS OR SPATIAL OR SPAWNING OR STALAGMITE OR STAR OR STOVE OR STOVES OR SURVEY OR TAX

OR THEORIES OR TIRES OR TISSUE OR TISSUES OR TRAFFIC OR TRANSFORMER OR TROPOSPHERE OR URBAN OR VITRO OR WELDING OR WOMEN OR WORKERS OR COMBUSTION DUST OR COMBUSTION MINERAL OR COMBUSTION SMOLDER OR (CONVERSION EFFICIENCY SAME LASERS) OR (ELECTRIC POWER SAME LIFE) OR (ELECTRICAL SAME (ANNEALING OR CIRCUIT OR ETCHING OR GROSS OR LIGHTING OR SPECIFIC OR WIDER)) OR (ELECTRICAL ENERGY SAME (CONCENTRATION OR POLLUTANT)) OR (ELECTRICITY SAME RECYCLING) OR (ENERGY SAME (ACCELERATION OR CONTROLLERS OR DISTURBANCE OR EQUIPARTITION OR FATTY OR FLAME OR HEART OR ISOTROPIC OR NETWORK OR NSPUDT OR PAYBACK OR PEI OR PENALTY OR SECTOR OR TREATMENT OR VELOCITY OR WAVES)) OR (ENERGY CONSUMPTION SAME PROGRAM) OR (ENERGY STORAGE SAME VIBRATIONAL) OR (ENERGY SUPPLY SAME (BOUNDARY OR DISTILLATION OR STORAGE)) OR (ENGINE SAME (ALGORITHM OR MODELS OR STABILIZATION)) OR (FUEL SAME (AEROSOL OR ALGORITHM OR HUMAN OR LEGISLATION OR NUMERICAL MODEL OR PAH OR PARTICULATE MATTER OR PLIF OR SIGNALS OR TROPOSPHERIC OR VIBRATION)) OR (FUELS SAME BUILDING) OR (HEAT STORAGE SAME HEAT PUMP) OR (POWER SAME (ABSORPTION OR ASH OR BUNDLE OR DOSE OR ECONOMY OR FAULT OR LASER OR LEAKAGE OR LINE OR LOGIC OR MINOR OR MONITORING OR POLICY OR PROBABILISTIC OR RECTIFIER OR SMES OR SWITCHES)) OR (POWER GENERATION SAME (FRACTION OR HEAT RECOVERY OR PROBLEMS OR SELF-TUNING OR SIEMENS OR STAGE)) OR (POWER PLANTS SAME (CORROSION OR MECHANICAL OR PFBC OR SEPARATION OR SIMULATION)) OR (POWER SUPPLY

SAME (CIRCUIT OR CIRCUITS OR SWITCHING)) OR (RENEWABLE ENERGY SAME FINANCIAL) OR (THERMAL ENERGY SAME (MEDIA OR PEAK OR PERCENT)))

Journal Title Component

FUEL

ENERGY FUELS

J. POWER SOURCES

ENERGY

ENERGY CONV. MANAG.

INT. J. ENERGY RES.

RENEW. ENERGY

J. INST. ENERGY

ENERGY SOURCES

PROG. ENERGY COMBUST. SCI.

RERIC INT. ENERGY J.

8. REFERENCES FOR APPENDIX 7-D

[1]. Kostoff RN. Database Tomography for technical intelligence. Competitive Intelligence Review 1993; 4(1): 38-43.

[2]. Kostoff RN. Database Tomography: origins and applications. Co mpetitive Intelligence Review 1994; Special Issue on Technology, 5(1).

[3]. Kostoff RN, et al. System and method for Database Tomography. U.S. Patent Number 5440481 1995.

[4]. Kostoff RN, Eberhart HJ, and Toothman DR. Database Tomography for information retrieval. Journal of Information Science 1997; 23(4): 301-311.

[5]. Kostoff RN, Eberhart HJ, and Toothman DR. Hypersonic and supersonic flow roadmaps using bibliometrics and Database Tomography. JASIS 15 April 1999; 50(5): 427-447.

[6]. Kostoff RN, Eberhart HJ, and Toothman DR. Database Tomography for technical intelligence: a roadmap of the near-earth space science and technology literature. Information Processing and Management 1998; 34(1): 69-85.

[7]. Kostoff RN, Eberhart HJ, Toothman DR, and Pellenbarg R. Database Tomography for technical intelligence: comparative roadmaps of the research impact assessment literature and the journal of the American chemical society, Scientometrics 1997; 40(1): 103-138.

[8] Kostoff RN, Braun T, Schubert A, Toothman DR, and Humenik J. Fullerene roadmaps using bibliometrics and Database Tomography. Journal of Chemical Information and Computer Science 2000; 40(1): 19-39.

[9] Kostoff RN, Green KA, Toothman DR, and Humenik J. Database Tomography applied to an aircraft science and technology investment strategy. Journal of Aircraft 2000; 37(4): 727-730.

[10]. SCI. Science Citation Index. Institute for Scientific Information. Phila., PA. 1999.

[11] Kostoff RN, The underpublishing of science and technology results. The Scientist 1 May 2000; 14(9): 6-6.

[12]. Kostoff RN. Science and technology innovation. Technovation 1999; 19(10): 593-604.

[13] Swanson DR, Smalheiser NR. An interactive system for finding complementary literatures: a stimulus to scientific discovery. Artificial Intelligence 1997; 91(2): 183-203.

[14] Hearst MA. Untangling text data mining". Proceedings of ACL 99, the 37th Annual Meeting of the Association for Computational Linguistics. University of Maryland. June 20-26, 1999. 1-9.

[15]. Garfield E. History of citation indexes for chemistry - a brief review. JCICS. 1985; 25(3): 170-174.

[16]. Kostoff RN. The use and misuse of citation analysis in research evaluation. Scientometrics 1998; 43:1: 27-43.

[17]. MacRoberts M, and MacRoberts B. Problems of citat ion analysis. Scientometrics 1996; 36(3): 435-444.

APPENDIX 7-E.

ELECTROCHEMICAL POWER: MILITARY REQUIREMENTS AND LITERATURE STRUCTURE [Kostoff et al, 2003d]

Electrochemical Power, as defined by the author for this study, is the generation and conversion of power, and the storage of energy, using electrochemical processes. Since one of the key outputs of the present study is a query that can be used by the community to access relevant Electrochemical Power documents, a recommended query based on this study is presented in total. This query serves as the operational definition of Electrochemical Power, and its development is discussed in the database generation section.

ELECTROCHEMICAL POWER QUERY

(fuel cell* or sofc* or pemfc* or dmfc* or ultracapacitor* or supercapacitor* or pseudocapacitor* or (capacitor* same (electrochemical or electrolyte* or double-layer)) or ((battery or batteries) same (lithium or li or electrode* or anode* or cathode* or capacity or material* or electrochemical or charge or charging or discharge* or discharging or rechargeable or electrolyte* or lithium or li or lithium-ion or nickel or metal hydride* or lead-acid or alloy*)) or ((lithium or li) same (electrochemical or discharge* or discharging or electrode* or liclo4 or rechargeable or cycling or reversible or insertion or mah or intercalation)) or (electrochemical same (discharge* or discharging or hydrogen storage or mah)) or (hydrogen storage same (alloy* or electrode*)) or (limn2o4 same electrode*) or (lipf6 same electrolyte*) or (charge-discharge same electrode*) or ((discharge capacity or metal hydride*) same electrode*) or (electrolyte* same lsgm) or (hydrogen same storage alloy*) or (nafion same polymer*) or (ptru same co) or (ruo2 same electrode*)) NOT(((electrode* or hydrogen or discharge*) same plasma*) or (discharge* same gas) or dna or assay* or biosensor* or rats or blood or capillary or protein* or mercury or clinical or amino or hydrogen peroxide or paste or corona or tissue* or helium or ascorbic acid or receptor* or chromium or radiation or bacteria* or plant* or extracellular or antenna* or magnetron or drug* or vivo or hydrolysis or ml or amperometric or care or cd or buffer or silicon or stress or sensor* or rf or filter* or switching or detection limit* or inhibition* or ar or ms or electrostatic or phi or monolayer* or gate* or sheath* or gc or depletion or combustion or serum* or toxicity or converter* or chromatography or radical* or oil* or generator* or target* or gap* or excitation* or environmental or glow* or ring or rings or diet* or pretreatment* or space charge* or amine* or ultrasound or lamp* or scan rate* or health* or solar or fe2 or reflection* or electromagnetic or carboxylic or deep or diode* or synthetic* or acetic acid or collision* or moiety or dimeric or titanate* or carbon steel* or curvature* or lithium chloride or coercive field or network* or hydrodynamic* or tris or mutant* or backbone* or decay* or monomer* or outcome* or driving or contamination or spatial or cmos or mediator* or excited or led or self-assembled or nitric oxide or i-v or array* or mmol or dt or waste* or aromatic or epitaxial or atomic force microscopy or differential pulse or viscosity or sorption or pk or native or shifts or recording* or adhesion* or dye* or surfactants)

To execute the study reported in this paper, a database of relevant Electrochemical Power articles is generated using the iterative search approach of Simulated Nucleation (5, 6). Then, the database is analyzed to produce the following characteristics and key features of the Electrochemical Power field: recent prolific Electrochemical Power authors; jour nals that contain numerous Electrochemical Power papers; institutions that produce numerous Electrochemical Power papers; keywords most frequently specified by the Electrochemical Power authors; authors, papers and journals cited most frequently; pervasive technical themes of Electrochemical Power; and relationships among the pervasive themes and sub-themes.

2. BACKGROUND

2.1 Military Requirements for Energy and Power

Fundamental to the operation of all advanced modern militaries is availability of energy and power supplies that will remove roadblocks to successful conduct of strategic and tactical missions. Different missions require far different power supplies, with different operating characteristics.

To compare the diversity of available and potential power supplies with the myriad mili tary missions and operations possible, some type of taxonomic scheme is required. One categorization revolves around whether humans are located in proximity of the power supply during the mission. Another is by geospatial location (space, atmosphere, land, sea, sub-surface) of the power supply during the mission. A third categorization is by the technology that uses the power supply (e.g., propulsion, communications, heating). A fourth categorization is by the type of fuel source (e.g., fossil, solar, nuclear, wind, etc). A fifth type of categorization is by the type of converter (e.g., heat cycle, direct conversion). Because of space limitations, this section will concentrate on the first two taxonomies.

The first taxonomy is power supplies in remote missions (where humans are not involved in-situ) and in direct missions (where humans are involved in-situ). Remote operations (e.g., space, underwater, underground, and land/ air-based robotic systems) can be further sub-divided into short-term (typically weapons launches) and long-term (typically surveillance, communications nodes). Long-term remote missions need supplies that are highly reliable (no maintenance required), long-lived, and retain performance over many cycles. While cost and efficiency are important, especially where numerous detectors with large data outputs are required, cost and efficiency could be traded off for reliability, and absence of moving parts is usually considered a positive factor. Safet y issues, such as environmental hazards, are less important for remote operations than where humans are involved in-situ. Long-term space missio ns require supplies that are lightwei ght (because oflaunch costs), launch survivable, low-G compliant, and survivable in the unique space environment (high radiation bands, large temperature swings, potential low pressure operation). Long-term buried or covert supplies (e.g., for detectors) do not have the critical weight limitation of space systems, but could be subject to harsh environmental conditions (e.g., corrosion-generating), and could have more stringent reduced signature requirements (thermal, acoustic, magnetic). Short-term remote applications (e.g., smart munitions) might have long shelf life requirements, high stress operation requirements (e.g., high-G, high temperature swings, high pressure, high vibrat ion, high shock, high

radiation, high magnetic fields), and high power density requirements, but long cycle repetition requirements would be reduced substantially. For direct operations, safety and hazard reduction considerations increase substantially, and high stress environments decrease, sometimes drastically.

The second categorization of missions discussed is geo-spatial. For space missions, power is used for vehicle and weapons propulsion, pulsed weapons, communications, surveillance, and housekeeping. Vehicle and weapons propulsion tend to be moderate/ short term high power density, pulsed weapons tend to be very high power very short term, and communications and surveillance are relatively low power and long term (with operating cycles that can range from short to long term). Other criteria for space operations were presented above.

For atmospheric missions, power is used for many of the same generic applications as space, with the major additions of combat and transport of people and materiel. Missions can be remote or direct. For both atmospheric and space missions, weight and size assume more importance than for terrestrial missions, with the exception of man-portable systems.

For stationary land-based direct missions, power is used for base maintenance operations (heating, cooling, lights, appliances, etc), communications, surveillance, local vehicle propulsion, and supply. For stationary land-based remote missions, power is used mainly for surveillance and communications, and for propulsion of robotic systems. For mobile land-based direct missions, power is used for propulsion, communications, and surveillance. For the specific case of the individual land-based warrior, power is generically required for the computer/ radio subsystem, the software subsystem, the integrated helmet assembly subsystem, and the weapon subsystem. For mobile land-based remote missions, power is used for weapons propulsion, guidance, surveillance, and communications. In the above, power production on-board a flying weapon is considered mobile remote.

For sea surface and undersea applications, the types of power requirements are comparable to those for a combination of air and land-based systems (e,g., combat, troop and materiel transport, short pulsed high power weapons, moderate pulse weapons), but the operating environment tends to be somewhat harsher (e.g., especially saline corrosion). In addition, long-term manned undersea missions tend to have higher reliability requirements more approximating those of space missions, while at the same time experiencing the constraints required for direct missions.

In general, evolving military applications require decreases in size and weight, especially for space, aircraft, and individual soldier or small team applications. For large volumes of power supply applications, such as munitions and radios, reduced cost becomes an important factor. For either weight or size reduction, or increased mission longevity, increase in energy and power density becomes important. Where people are involved, increased safety is important, and for long-term operations, environmental compliance is important. High reliability is of importance, especially where maintenance is not possible during the course of the mission (space, weapons flight, covert surveillance). Where maintenance is possible, ease of maintenance and supportability are important power supply considerations. In some militaries, limitations are placed on the types of fuels that can be used (e.g., diesel, JP-type fuels). The trend is also toward faster vehicles and weapons.

Aerodynamics dictates power requirements will increase nonlinearly with speed, and for fixed size vehicles, larger power supplies will be required.

2.2 Characteristics of Electrochemical Energy and Power

There are three main electrochemical source/ converter/ storage systems: batteries, fuel cells, and capacitors. Relative to heat engines, they have far fewer moving parts, eliminate the need for a thermal conversion step, and tend to be more reliable with lower acoustic and thermal signatures. Relative to renewable sources, they have higher energy and power densities (excluding fission or fusion as renewable sources).

2.3 Electrochemical Energy and Power for Military Applications

Batteries can be used as components of the many military applications listed above. They tend to support guidance and control, communications, propulsion, surveillance and detection, fusing, arming, and backup power. Military research is focused on more efficient, smaller, lighter, safer, cheaper, higher power and energy, more reliable, higher longevity, and more safely disposable, batteries.

Fuel cells have the same generic development targets and can potentially be used in many of the same applications as batteries, but they are not as far along in development or implementation. Fuel cells have the potential to be attractive battery replacements, because their energy storage capability is significantly greater than batteries. Very high power fuel cells are being developed for ship propulsion and ship service power; high power fuel cells are being developed for base stationary power; moderate power fuel cells are being developed for mobile electric power, auxiliary power units, and robotic vehicles; and low power fuel cells are being developed for so ldier systems (radios, cooling, heating, weapon systems), battery charging, small robotic vehicles, and remote power. These low power fuel cells have the potential to extend soldier mission times by hours, or possibly days.

Super- or ultra-capacitors are niche storage components. They have higher energy densities than conventional dielectric capacitors, but lower energy densities than batteries or fuel cells. They have higher power densities than fuel cells or batteries, but lower power densities than conventional dielectric capacitors. They are viewed as potentially competitive candidates for modern digital communication devices, which are pulsed and time shared, and involve packet transmission techniques. In their optimal operational frequency range, they can smooth the loads on batteries, thereby increasing capacity and decreasing battery costs and hazards. Their potential ruggedness and reliability are important features.

2.4 Text Mining Overview

Recent DT/ bibliometrics studies were conducted of the technical fields of: 1) Near-earth space (NES) (8); 2) Hypersonic and supersonic flow over aerodynamic bodies (HSF) (7); 3) Chemistry (JACS) (9) as represented by the Journal of the American Chemical Society; 4) Fullerenes (FUL)

(10); 5) Aircraft (AIR) (11); 6) Hydrodynamic flow over surfaces (HYD); 7) Electri c power sources (EPS); and 8) the non-technical field of research impact assessment (RIA). Overall parameters of these studies from the SCI database results and the current electrochemical study are shown in Table 1.

TABLE 1 - DT STUDIES OF TOPICAL FIELDS

TOPICAL AREA	**NUMBER OF SCI ARTICLES**	**YEARS COVERED**
1) NEAR-EARTH SPACE (NES)	5480	1993-MID 1996
2) HYPERSONICS (HSF)	1284	1993-MID 1996
3)CHEMISTRY (JACS)	2150	1994
4) FULLERENES (FUL)	10515	1991-MID 1998
5) AIRCRAFT (AIR)	4346	1991-MID 1998
6) HYDRODYNAMICS (HYD)	4608	1991-MID 1998
7) ELECTRIC POWER SOURCES (EPS)	20835	1991-BEG 2000
8) RESEARCH ASSESSMENT (RIA)	2300	1991-BEG 1995
9) ELECTROCHEMICAL POWER SOURCES (ECHEM)	6985	1993 – MID 2001

3. DATABASE GENERATION

The key step in the Electrochemical Power literature analysis is the generation of the database to be used for processing. There are three key elements to database generation: the overall objectives, the approach selected, and the database used. Each of these elements is described.

3.1 Overall Study Objectives

The main objective was to identify global S&T that had both direct and indirect relations to Electrochemical Power. A sub-objective was to estimate the overall level of global effort in Electrochemical Power S&T, as reflected by the emphases in the published literature.

3.2 Databases and Approach

For the present study, the SCI database was used. The approach used for query devel opment was the DT-based iterative relevance feedback concept (5).

3.2.1 Science Citation Index (12)

The database consists of selected journal records (including authors, titles, journals, author addresses, author keywords, abstract narratives, and references cited for each paper) obtained by searching the web version of the SCI for Electrochemical Power articles. At the time the data was extracted for the present paper (mid-2001), the version of the SCI used accessed about 5600 j ournals

(mainly in physical, engineering, and life sciences basic research).

The SCI database selected represents a fraction of the available Electrochemical Power (mainly research) literature, that in turn represents a fraction of the Electrochemical Power S&T actually performed globally (13). It does not include the large body of classified literature, or company proprietary technology literature. It does not include technical reports or books or patents on Electrochemical Power. It covers a finite slice of time (1991 to mid-2001). The database used represents the bulk of the peer-reviewed high quality Electrochemical Power research, and is a representative sample of all Electrochemical Power research in recent times.

To extract the relevant articles from the SCI, the Title, Keyword, and Abstract fields were searched using Keywords relevant to Electrochemical Power, although different procedures were used to search the Title and Abstract fields (5). The resultant Abstracts were culled to those relevant to Electrochemical Power. The search was performed with the aid of two powerful DT tools (multi-word phrase frequency analysis and phrase proximity analysis) using the process of Simulated Nucleation (5).

4. RESULTS

The results from the publications bibliometric analyses are presented in section 4.1, followed by the results from the citations bibliometrics analysis in section 4.2. Results from the DT analyses are shown in section 4.3. The SCI bibliometric fields incorporated into the database included, for each paper, the author, journal, institution, and Keywords. In addition, the SCI included references for each paper.

4.1 Publication Statistics on Authors, Journals, Organizations, Countries

The first group of metrics presented is counts of papers published by different entities. These metrics can be viewed as output and productivity measures. They are not direct measures of research quality, although there is some threshold quality level inferred, since these papers are published in the (typically) high caliber journals accessed by the SCI.

4.1.1 Author Frequency Results

There were 6985 papers retrieved, 11051 different authors, and 25465 author listings. The occurrence of each author's name on a paper is defined as an author listing. While the average number of listings per author is about 2.3, the twenty most prolific authors (see Table 2) have listings more than an order of magnitude greater than the average. The number of papers listed for each author are those in the database of records extracted from the SCI using the query, not the total number of author papers listed in the source SCI database.

TABLE 2 – MOST PROLIFIC AUTHORS

(present institution listed)

AUTHOR NAME	INSTITUTION	COUNTRY	# PAPERS
DAHN, JR	DALHOUSIE UNIV	CANADA	67
TARASCON, JM	UNIV PICARDIE	FRANCE	53
WANG, QD	ZHEJIANG UNIV	CHINA	51
LEI, YQ	ZHEJIANG UNIV	CHINA	46
LIU, HK	UNIV WOLLONGONG	AUSTRALIA	44
DOU, SX	UNIV WOLLONGONG	AUSTRALIA	44
SCROSATI, B	UNIV ROMA LA SAPIENZA	ITALY	43
LEE, JY	NATIONAL UNIV SINGAPORE	SINGAPORE	42
KUMAGAI, N	IWATE UNIV	JAPAN	41
YAMAMOTO, O	AICHI INST TECHNOLOGY	JAPAN	40
YOSHIO, M	SAGA UNIV	JAPAN	40
AURBACH, D	BAR ILAN UNIV	ISRAEL	38
UCHIDA, I	TOHOKU UNIV	JAPAN	37
WATANABE, M	UNIV YAMANASHI	JAPAN	37
CHEN, LQ	CHINESE ACAD SCIENCE	CHINA	36
TAKEDA, Y	MIE UNIV	JAPAN	36
PASSERINI, S	ENEA	ITALY	35
TIRADO, JL	UNIV CORDOBA	SPAIN	33
IWAKURA, C	UNIV OSAKA PREFECTURE	JAPAN	32
WHITE, RE	UNIV SOUTH CAROLINA	USA	32

Of the twenty most prolific authors listed in Table 2, seven are from Japan. In fact, thirteen are from the Far East, four are from Europe (Western), two are from North America, and one is from the Middle East. Eighteen are from universities, and two are from research institutes. Total publications listed in the SCI for each of these twenty authors were scanned visually, and, on average, these authors were rarely listed as first authors. For example, in their 100 most recent papers, DAHN JR was listed as first author five times, and TARASCON JM was listed as first author six times.

4.1.2 Journals Containing Most Electrochemical Power Papers

There were 587 different journals represented, with an average of 11.90 papers per journal. The journals containing the most power-related electrochemistry papers (see Table 3) had more than an order of magnitude more papers than the average.

TABLE 3 – JOURNALS CONTAINING MOST PAPERS

JOURNAL NAMES	# OF PAPERS
J. POWER SOURCES	1240
J. ELECTROCHEM. SOC.	771
SOLID STATE ION.	546
ELECTROCHIM. ACTA	403
J. ALLOY. COMPD.	290

DENKI KAGAKU	198
J. APPL. ELECTROCHEM.	167
J. ELECTROANAL. CHEM.	138
ELECTROCHEM. SOLID STATE LETT.	119
INT. J. HYDROG. ENERGY	112
RUSS. J. ELECTROCHEM.	100
ELECTROCHEMISTRY	86
J. MATER. CHEM.	81
J. SOLID STATE CHEM.	72
CHEM. MAT.	70
J. NEW MAT.ELECTROCHEM. SYST.	60
ELECTROCHEM. COMMUN.	56
SYNTH. MET.	55
BULL. ELECTROCHEM.	54
J. PHYS. CHEM. B	50

The majority of the journals are electrochemistry, with the remainder divided between chemistry and materials. There appear to be three primary groups at the top layer. The Journal of Power Sources, an international journal devoted to the science and technology of electrochemical energy systems, contains the most articles by far. This is not surprising, since its stated mission is fully aligned with the main objective of the present study. While many of its articles were retrieved by the query, essentially all of its articles are relevant to the topic of the present study.

The next group consists of the Journal of the Electrochemical Society (JES) and Solid State Ionics (SSI). The JES focuses on solid-state and electrochemical science and technology, while SSI is devoted to the physics, chemistry and materials science of diffusion, mass transport, and reactivity of solids. While these journals include aspects of electrochemistry/ electrochemical power sources in their charters, they include other aspects of chemistry (and physics) as well. The next five journals listed constitute the third group.

4.1.3 Institutions Producing Most Electrochemical Power Papers

A similar process was used to develop a frequency count of institutional address appearances. It should be noted that many different organizational components may be included under the single organizational heading (e.g., Harvard Univ could include the Chemistry Department, Biology Department, Physics Department, etc.). Identifying the higher level institutions is instrumental for these DT studies. Once they have been identified through bibliometric analysis, subsequent measures may be taken (if desired) to identify particular departments within an institution.

TABLE 4 – PROLIFIC INSTITUTIONS

INSTITUTION NAMES	COUNTRY	# OF PAPERS
CHINESE ACAD SCI	CHINA	118

KYOTO UNIV	JAPAN	108
CNRS	FRANCE	104
KOREA ADV INST SCI & TECHNOL	KOREA	90
RUSSIAN ACAD SCI	RUSSIA	89
ZHEJIANG UNIV	CHINA	85
ARGONNE NATL LAB	USA	79
UNIV CALIF BERKELEY	USA	78
TOHOKU UNIV	JAPAN	73
MIT	USA	66
CNR	ITALY	63
CENT ELECTROCHEM RES INST	INDIA	60
SEOUL NATL UNIV	KOREA	60
TOKYO INST TECHNOL	JAPAN	55
CSIC	SPAIN	55
KFA JULICH GMBH	GERMANY	54
UNIV S CAROLINA	USA	54
OSAKA NATL RES INST	JAPAN	52
UNIV TOKYO	JAPAN	51
DELFT UNIV TECHNOL	NETHERLANDS	51

Of the twenty most prolific institutions, ten are from Asia, five are from Western Europe, four from the USA, and one from Eastern Europe. Twelve are universities, and the remaining institutions are research institutes.

4.1.4 Countries Producing Most Electrochemical Power Papers

There are 78 different countries listed in the results. The country bibliometric results are summarized in Table 5. The dominance of a handful of countries is clearly evident.

TABLE 5 – PROLIFIC COUNTRIES

COUNTRY NAMES	# OF PAPERS
JAPAN	1552
USA	1318
FRANCE	558
PEOPLES R CHINA	499
SOUTH KOREA	380
GERMANY	341
CANADA	318
ENGLAND	285
ITALY	250
INDIA	249
RUSSIA	206

SPAIN	151
SWEDEN	126
AUSTRALIA	121
SWITZERLAND	113
NETHERLANDS	97
TAIWAN	90
BRAZIL	83
ISRAEL	78
POLAND	73

There appear to be three dominant groups in the twenty most prolific countries. The US and Japan constitute the most dominant group, and were the only two countries to have published more than 1000 papers on power-related electrochemistry during the past 8 years. France and China constitute the next group, but had less papers combined than either member of the first group. The next seven countries constitute the third group.

Interestingly, unlike all previous DT studies, the United States (US) was not the most prolific country. Japan had more published papers (nearly 18% more) than the US. Overall, Eastern Asia (Japan, China, South Korea, Taiwan), Northern North America (US, Canada), and Western Europe (France, Germany, UK) accounted for most of the electrochemistry research activity.

Figure 1 contains a co-occurrence matrix of the top 15 countries. In terms of absolute numbers of co-authored papers, the USA major partners are Japan, France, Italy, Canada, and South Korea. Overall, countries in similar geographical regions tend to co-publish substantially, the US being a moderate exception.

FIGURE 1 – COUNTRY CO-OCCURRENCE MATRIX

	# Records	1552	1318	558	499	380	344	341	318	250	249	206	151	126	121	113
# Records	**COUNTRY**	**JAPAN**	**USA**	**FRANCE**	**CHINA**	**SOUTH KOREA**	**UK**	**GERMANY**	**CANADA**	**ITALY**	**INDIA**	**RUSSIA**	**SPAIN**	**SWEDEN**	**AUSTRALIA**	**SWITZERLAND**
1552	**JAPAN**	1552	52	15	17	16	14	8	3	5	4	0	0	3	3	5
1318	**USA**	52	1318	36	6	17	10	9	24	27	9	6	5	5	2	4
558	**FRANCE**	15	36	558	3	4	10	13	9	9	5	4	31	3	0	4
499	**CHINA**	17	6	3	499	1	2	5	5	1	0	0	1	8	0	0
380	**SOUTH**	16	17	4	1	380	0	4	1	5	2	1	0	0	2	1

	KOREA															
344	**UK**	14	10	10	2	0	344	9	7	2	5	2	0	4	2	3
341	**GERMANY**	8	9	13	5	4	9	341	3	7	5	6	0	2	0	13
318	**CANADA**	3	24	9	5	1	7	3	318	0	0	1	0	2	1	1
250	**ITALY**	5	27	9	1	5	2	7	0	250	5	2	1	2	0	2
249	**INDIA**	4	9	5	0	2	5	5	0	5	249	0	1	0	4	2
206	**RUSSIA**	0	6	4	0	1	2	6	1	2	0	206	2	2	0	1
151	**SPAIN**	0	5	31	1	0	0	0	0	1	1	2	151	0	0	0
126	**SWEDEN**	3	5	3	8	0	4	2	2	2	0	2	0	126	0	1
121	**AUSTRALIA**	3	2	0	0	2	2	0	1	0	4	0	0	0	121	0
113	**SWITZERLAND**	5	4	4	0	1	3	13	1	2	2	1	0	1	0	113

4.2 Citation Statistics on Authors, Papers, and Journals

The second group of metrics presented is counts of citations to papers published by different entities. While citations are ordinarily used as impact or quality metrics (14), much caution needs to be exercised in their frequency count interpretation, since there are numerous reasons why authors cite or do not cite particular papers (15, 16).

The citations in all the retrieved SCI papers were aggregated, the authors, specific papers, years, journals, and countries cited most frequently were identified, and were presented in order of decreasing frequency. A small percentage of any of these categories received large numbers of citations. From the citation year results, the most recent papers tended to be the most highly cited. This reflected rapidly evolving fields of research.

4.2.1 Most Cited Authors

The most highly cited authors are listed in Table 6.

TABLE 6 – MOST CITED AUTHORS

(cited by other papers in this database only)

AUTHOR NAMES	INSTITUTIONS	COUNTRIES	TIMES CITED
OHZUKU, T	OSAKA CITY UNIV	JAPAN	1066
THACKERAY, MM	ARGONNE NAT'L LAB	USA	845
AURBACH, D	BAR ILAN UNIV	ISRAEL	808
TARASCON, JM	UNIV PICARDIE	FRANCE	755
DAHN, JR	DALHOUSIE UNIV	CANADA	698
WATANABE, M	UNIV YAMANASHI	JAPAN	601

ABRAHAM, KM	COVALENT ASSOCIATES	USA	461
GUMMOW, RJ	CSIR	SOUTH AFRICA	455
DELMAS, C	CNRS	FRANCE	429
SAKAI, T	OSAKA NAT'L RES INST	JAPAN	412
PISTOIA, G	CNR	ITALY	391
MINH, NQ	ALLIED SIGNAL AERO	USA	381
GOODENOUGH, JB	UNIV TEXAS	USA	379
ISHIHARA, T	OITA UNIV	JAPAN	370
STEELE, BCH	UNIV LONDON IMPERIAL	ENGLAND	351
REIMERS, JN	MOLI ENERGY	CANADA	345
PELED, E	TEL AVIV UNIV	ISRAEL	335
GUYOMARD, D	UNIV NANTES	FRANCE	332
MIZUSAKI, J	TOHOKU UNIV	JAPAN	324
APPLEBY, AJ	TEXAS A&M	USA	300

Of the twenty most cited authors, five are from Japan, five from the USA, five from Europe (Western), two from Canada, two from Israel, and one from Africa. This is a far different distribution from the most prolific authors, where thirteen were from the Far East. There are a number of potential reasons for this difference, including difference in quality and late entry into the research discipline. In another three or four years, when the papers from present-day authors have accumulated sufficient citations, firmer conclusions about quality can be drawn.

The lists of twenty most prolific authors and twenty most highly cited authors only had four names in common (AURBACH, TARASCON, DAHN, WATANABE). This phenomenon of minimal intersection has been observed in all other text mining studies performed by the first author.

Thirteen of the authors' institutions are universities, four are government-sponsored research laboratories, and three are private companies. The appearance of the companies on this list is another differentiator from the list of most prolific authors.

The citation data for authors and journals represents citations generated only by the specific records extracted from the SCI database for this study. It does not represent all the citations received by the references in those records; these references in the database records could have been cited additionally by papers in other technical disciplines.

4.2.2 Most Cited Papers

The most highly cited papers are listed in Table 7.

TABLE 7 – MOST CITED PAPERS

(total citations listed in SCI)

AUTHOR NAME	YEAR	JOURNAL	VOLUME	SCI CITES
TARASCON JM	1991	J ELECTROCHEM SOC	V138	272
(LIMN204 SPINEL PHASE AS SECONDARY LITHIUM CELL CATHODE)				
MINH NQ	1993	J AM CERAM SOC	V76	476
(CERAMIC FUEL CELLS - REVIEW)				
OHZUKU T	1993	J ELECTROCHEM SOC	V140	217
(SYNTHESIS OF LINIO2 FOR SECONDARY LITHIUM CELL)				
GUMMOW RJ	1994	SOLID STATE IONICS	V69	281
(IMPROVED RECHARGEABLE CAPACITY OF LIMN2O4 CATHODES)				
OHZUKU T	1990	J ELECTROCHEM SOC	V137	314
(ELECTROCHEMISTRY OF MNO2 IN LITHIUM CELLS)				
MIZUSHIMA K	1980	MATER RES BULL	V15	392
(LIXCOO2 FOR HIGH-ENERGY DENSITY BATTERY CATHODES)				
GUYOMARD D	1992	J ELECTROCHEM SOC	V139	300
(LI METAL-FREE RECHARGEABLE LIMN2O4/ CARBON CELLS)				
THACKERAY MM	1983	MATER RES BULL	V18	358
(LITHIUM INSERTION INTO MANGANESE SPINELS)				
TARASCON JM	1994	J ELECTROCHEM SOC	V141	247
(LITHIUM INSERTION INTO THE SPINEL LIMN2O4)				
FONG R	1990	J ELECTROCHEM SOC	V137	334
(LITHIUM INTERCALATION INTO CARBON USING NON-AQUEOUS CELLS)				
REIMERS JN	1992	J ELECTROCHEM SOC	V139	227
(LITHIUM INTERCALATION IN LIXCOO2)				
COURTNEY IA	1997	J ELECTROCHEM SOC	V144	147
(LITHIUM REACTION WITH TIN OXIDE COMPOSITES IN LITHIUM ION CELL)				
SATO K	1994	SCIENCE	V254	221
(LITHIUM STORAGE IN DISORDERED CARBONS)				
THACKERAY MM	1992	J ELECTROCHEM SOC	V139	202
(SPINEL ELECTRODES FROM LIMNO SYSTEM FOR SECONDARY BATTERIES)				
THACKERAY MM	1984	MATER RES BULL	V19	235
(ELECTROCHEMICAL EXTRACTION OF LITHIUM FROM LIMN2O4)				
ISHIHARA T	1994	J AMER CHEM SOC	V116	201
(DOPED LAGO3 OEROVSKITE OXIDE IONIC CONDUCTOR)				
SHANNON RD	1976	ACTA CRYSTALLOGR A	V32	10254
(IONIC-RADII AND INTERATOMIC DISTANCES IN HALIDES AND CHALCOGENIDES)				
WILLEMS JJG	1984	PHILLIPS J RESEARCH	V39	285
(METAL HYDRIDE ELECTRODES FOR RECHARGEABLE BATTERY)				
ABRAHAM KM	1990	J ELECTROCHEM SOC	V137	202
(LI+-CONDUCTIVE SOLID POLYMER ELECTROLYTES WITH LIQ-LIKE CONDUCT)				
OHZUKU T	1993	ELECTROCHIMICA ACTA	V38	139
(LI-N-CO OXIDES FOR SECONDARY LITHIUM CELLS				

The theme of each paper is shown in italics on the line after the paper listing. The order of paper listings is by number of citations by other papers in the extracted database analyzed. The total number of citations from the SCI paper listing, a more accurate measure of total impact, is shown in the last column on the right.

The Journal of the Electrochemical Societ y contains the most papers, twelve out of the twenty listed. Most of the journals are fundamental science journals, and most of the topics have a fundamental science theme. Most of the papers are from the 1990s, with four being from the 1980s, and one extremely highly cited paper being from 1976. This reflects a dynamic research field, with seminal works being performed in the recent past.

Sixteen of the papers address issues related to lithium secondary batteries, with the dominant issue theme being lithium insertion/ intercalation to avoid free-metal formation. Two of the papers address issues related to ceramic fuel cells, with the dominant issue theme being solid oxides for high ionic conductivity. One paper addresses issues related to nickel metal hydride rechargeable batteries.

Thus, the major intellectual emphasis of cutting edge electrochemical power sources research, as evidenced by the most cited papers, is well aligned with the intellectual heritage and performance emphasis, as will be evidenced by the clustering approaches.

4.2.3. Most Cited Journals

TABLE 8 – MOST CITED JOURNALS

(cited by other papers in this database only)

JOURNAL NAMES	TIMES CITED
J ELECTROCHEM SOC	22363
SOLID STATE IONICS	9782
J POWER SOURCES	8265
ELECTROCHIM ACTA	5994
J ELECTROANAL CHEM	4607
J SOLID STATE CHEM	2364
J ALLOY COMPD	2269
J APPL ELECTROCHEM	2008
MATER RES BULL	1811
PHYS REV B	1672
J AM CHEM SOC	1491
J PHYS CHEM-US	1470
J AM CERAM SOC	1417
J LESS-COMMON MET	1399
DENKI KAGAKU	1157
SYNTHETIC MET	1041
CHEM MATER	969

ELECTROCHEMICAL SOC	851
SCIENCE	841

The Journal of the Electrochemical Society received as many citations as the next three journals combined. Most of the highly cited journals are electrochemistry, some are materials, some chemistry, with one physics journal represented. Based on all the citation results, there is little evidence that disciplines outside the tightly knit electrochemistry-materials groups relevant to the specific applications are being accessed.

4. REFERENCES FOR APPENDIX 7-E

1). Kostoff, R. N. (1999). Science and Technology Innovation. Technovation. 19: 10.

2). Kostoff, R. N. (1993). Database Tomography for Technical Intelligence. Competitive Intelligence Review. 4:1.

3). Kostoff, R.N. (1994). Database Tomography: Origins and Applications. Competitive Intelligence Review. Special Issue on Technology, 5:1.

4). Kostoff, R. N. et al (1995). System and Method for Database Tomography. U.S. Patent Number 5440481.

5). Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. (1997). Database Tomography for Information Retrieval. Journal of Information Science. 23:4.

6). Kostoff, R. N. (1997). The Handbook of Research Impact Assessment. Seventh Edition. DTIC Report Number ADA296021.

7). Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. (1999). Hypersonic and Supersonic Flow Roadmaps Using Bibliometrics and Database Tomography. JASIS. 50:5. 15 April.

8). Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. (1998). Database Tomography for Technical Intelligence: A Roadmap of the Near-Earth Space Science and Technology Literature. Information Processing and Management. 34:1.

9). Kostoff, R. N., Eberhart, H. J., Toothman, D. R., and Pellenbarg, R. (1997). Database Tomography for Technical Intelligence: Comparative Roadmaps of the Research Impact Assessment Literature and the Journal of the American Chemical Society, Scientometrics, 40:1.

10) Kostoff, R. N., Braun, T., Schubert, A., Toothman, D. R., and Humenik, J. (2000). Fullerene Roadmaps Using Bibliometrics and Database Tomography. Journal of Chemical Information and Computer Science. Jan-Feb.

11) Kostoff, R. N., Green, K. A., Toothman, D. R., and Humenik, J. A. (2000). Database Tomography Applied to an Aircraft Science and Technology Investment Strategy. Journal of Aircraft, 37:4, July-August.

12). SCI. (1999). Science Citation Index. Institute for Scientific Information. Phila., PA.

13) Kostoff, R. N. (2000). The Underpublishing of Science and Technology Results. The Scientist. 1 May.

14). Garfield, E. (1985) History of Citation Indexes for Chemistry - A Brief Review. JCICS. 25(3). 170-174.

15). Kostoff, R. N. (1998). The Use and Misuse of Citation Analysis in Research Evaluation. Scientometrics, 43:1, September.

16). MacRoberts, M., and MacRoberts, B. (1996). Problems of Citation Analysis. Scientometrics. 36:3. July-August.

APPENDIX 7-F.

NONLINEAR DYNAMICS TEXT MINING USING BIBLIOMETRICS AND DATABASE TOMOGRAPHY [Kostoff et al, 2004a]

OVERVIEW

The present Appendix describes use of the DT process, supplemented by literature bibliometric analyses, to derive technical intelligence from the published literature of Nonlinear Dynamics science and technology.

Nonlinear Dynamics, as defined by the author for this study, is that class of motions in deterministic physical and mathematical systems whose time evolution has a sensitive dependence on initial conditions. Since one of the key outputs of the present study is a query that can be used by the community to access relevant Nonlinear Dynamics documents, a recommended query based on this study is presented in total. This query serves as the operational definition of Nonlinear Dynamics, and its development is discussed in detail in the database generation section.

NONLINEAR DYNAMICS QUERY

((CHAO* AND (SYSTEM* OR DYNAMIC* OR PERIODIC* OR NONLINEAR OR BIFURCATION* OR MOTION* OR OSCILLAT* OR CONTROL* OR EQUATION* OR FEEDBACK* OR LYAPUNOV OR MAP* OR ORBIT* OR ALGORITHM* OR HAMILTONIAN OR LIMIT* OR QUANTUM OR REGIME* OR REGION* OR SERIES OR SIMULATION* OR THEORY OR COMMUNICATION* OR COMPLEX* OR CONVECTION OR CORRELATION* OR COUPLING OR CYCLE* OR DETERMINISTIC OR DIMENSION* OR DISTRIBUTION* OR DUFFING OR ENTROPY OR EQUILIBRIUM OR FLUCTUATION* OR FRACTAL* OR INITIAL CONDITION* OR INVARIANT* OR LASER* OR LOGISTIC OR LORENZ OR MAGNETIC FIELD* OR MECHANISM* OR MODES OR NETWORK* OR ONSET OR TIME OR FREQUENC* OR POPULATION* OR STABLE OR ADAPTIVE OR CIRCUIT* OR DISSIPAT* OR EVOLUTION OR EXPERIMENTAL OR GROWTH OR HARMONIC* OR HOMOCLINIC OR INSTABILIT* OR OPTICAL)) OR (BIFURCATION* AND (NONLINEAR OR HOMOCLINIC OR QUASIPERIODIC OR QUASI-PERIODIC OR DOUBLING OR DYNAMICAL SYSTEM* OR EVOLUTION OR INSTABILIT* OR SADDLE-NODE* OR MOTION* OR OSCILLAT* OR TRANSCRITICAL OR BISTABILITY OR LIMIT CYCLE* OR POINCARE OR LYAPUNOV OR ORBIT*)) OR (NONLINEAR AND (PERIODIC SOLUTION* OR OSCILLAT* OR MOTION* OR HOMOCLINIC)) OR (DYNAMICAL SYSTEM* AND (NONLINEAR OR STOCHASTIC OR NON-LINEAR)) OR ATTRACTOR* OR PERIOD DOUBLING* OR CORRELATION DIMENSION* OR LYAPUNOV EXPONENT* OR PERIODIC ORBIT* OR NONLINEAR DYNAMICAL) NOT (CHAO OR CHAOBOR* OR CHAOTROP* OR CAROTID OR ARTERY OR STENOSIS OR PULMONARY OR

VASCULAR OR ANEURYSM* OR ARTERIES OR VEIN* OR TUMOR* OR SURGERY)

To execute the study reported in this paper, a database of relevant Nonlinear Dynamics articles is generated using the iterative search approach of Simulated Nucleation [Kostoff et al, 1997a, 2001]. Then, the database is analyzed to produce the following characteristics and key features of the Nonlinear Dynamics field: recent prolific Nonlinear Dynamics authors; journals that contain numerous Nonlinear Dynamics papers; institutions that produce numerous Nonlinear Dynamics papers; keywords most frequently specified by the Nonlinear Dynamics authors; authors, papers and journals cited most frequent ly; pervasive technical themes o f Nonlinear Dynamics; and relationships among the pervasive themes and sub-themes.

Recent DT/ bibliometrics studies were conducted of the technical fields of: 1) Near-earth space (NES) [Kostoff et al, 1998]; 2) Hypersonic and supersonic flow over aerodynamic bodies (HSF) [Kostoff et al, 1999]; 3) Chemistry (JACS) [Kostoff et al, 1997b] as represented by the Journal of the American Chemical Society; 4) Fullerenes (FUL) [Kostoff et al; 2000a] 5) Aircraft (AIR) [Kostoff et al, 2000b]; 6) Hydrodynamic flow over surfaces (HYD); 7) Electric Power Sources (EPS); 8) Electrochemical Power Sources (ECHEM) [Kostoff et al, 2002] and 9) the non-technical field of research impact assessment (RIA) [Kostoff et al, 1997b]. Overall parameters of these studies from the SCI database results and the current Nonlinear Dynamics study are shown in Table 1.

TABLE 1 - DT STUDIES OF TOPICAL FIELDS

TOPICAL AREA	NUMBER OF SCI ARTICLES	YEARS COVERED
1) NEAR-EARTH SPACE (NES)	5480	1993-MID 1996
2) HYPERSONICS (HSF)	1284	1993-MID 1996
3)CHEMISTRY (JACS)	2150	1994
4) FULLERENES (FUL)	10515	1991-MID 1998
5) AIRCRAFT (AIR)	4346	1991-MID 1998
6) HYDRODYNAMICS (HYD)	4608	1991-MID 1998
7) ELECTRIC POWER SOURCES (EPS)	20835	1991-BEG 2000
8) ELECTROCHEMICAL POWER SOURCES (ECHEM)	6985	1993 – MID 2001
9) RESEARCH ASSESSMENT (RIA)	2300	1991-BEG 1995
10) NONLINEAR DYNAMICS (NONLIN)	6118 (2001)	1991, 2001

2.2 Unique Study Features

The study reported in the present Appendix differs from the previous published papers in this category [Kostoff, 1999; Kostoff et al, 1998, 1997b, 2000a, 2000b, 2002] in five respects. First, the topical domain (Nonlinear Dynamics) is completely different. Second, a much more rigorous statistically-based technical theme clustering approach is used. Third, bibliometric clustering is

presented for two database fields: authors and countries. Fourth, a combination of fuzzy logic and manual aggregation was used in phrase selection to consolidate similar phrases, thereby allowing additional phrases to be used in the clusters and increase the scope of the clusters. Finally, the marginal utility algorithm was applied for the first time, allowing only the highest payo ff terms to be included in the final query, and resulting in an efficient query.

3. DATABASE GENERATION

The key step in the Nonlinear Dynamics literature analysis is the generation of the database to be used for processing. There are three key elements to database generation: the overall objectives, the approach selected, and the database used. Each of these elements is described.

3.1 Overall Study Objectives

The main objective was to identify global S&T that had both direct and indirect relations to Nonlinear Dynamics. A sub-objective was to estimate the overall level of global effort in Nonlinear Dynamics S&T, as reflected by the emphases in the published literature.

3.2 Databases and Approach

For the present study, the SCI database (including both the Science Citation Index and the Social Science Citation Index) was used. The approach used for query development was the DT-based iterative relevance feedback concept [Kostoff et al, 1997a].

3.2.1 Science Citation Index/ Social Science Citation Index (SCI) [SCI, 2002]

The retrieved database used for analysis consists of selected journal records (including the fields of authors, titles, journals, author addresses, author keywords, abstract narratives, and references cited for each paper) obtained by searching the Web version of the SCI for Nonlinear Dynamics articles. At the time the final data was extracted for the present paper (early 2002), the version of the SCI used accessed about 5600 journals (mainly in physical, engineering, and life sciences basic research) from the Science Citation Index, and over 1700 journals from the Social Science Citation Index. There is some overlap among the journals. For example, for 2001, there were 999620 total articles in the Science Citation Index, 149672 articles in the Social Sciences Citation Index, and 1104275 articles in the combined databases. Thus, 45017 articles were shared by both databases, f our percent of the total, but thirty percent of the Social Science Citation Index.

The SCI database selected represents a fraction of the available Nonlinear Dynamics (mainly research) literature, that in turn represents a fraction of the Nonlinear Dynamics S&T actually performed globally [Kostoff, 2000]. It does not include the large body of classified literature, or company proprietary technology literature. It does not include technical reports or books or patents on Nonlinear Dynamics. It covers a finite slice of time (1991, 2001). The database used represents the bulk of the peer-reviewed high quality Nonlinear Dynamics research literature, and is a representative sample of all Nonlinear Dynamics research in recent times.

In order to generate an efficient final query, a new process termed Marginal Utility was applied. At the start of the final iteration, a modified query Q1 was inserted into the SCI, and records were retrieved. A sample of these records was then categorized into relevant and non-relevant. Each term in Q1 was inserted into the Marginal Utility algorithm, and the marginal number of relevant and non-relevant records in the sample that the query term would retrieve was computed. Only those terms that retrieved a high ratio of relevant to non-relevant records were retained. Since (by design) each query term had been used to retrieve records from the SCI as part of Q1, the marginal ratio of relevant to non-relevant records from the sample would represent the marginal ratio of relevant to non-relevant records from the SCI. The final efficient query Q2, consisting of the highest marginal utility terms, was shown in the Introduction.

In the Marginal Utility algorithm, terms that co-occur strongly in records with previously-selected terms are essentially duplicative from the retrieval perspective, and can be eliminated. Thus, the order in which terms are selected becomes important. An automated query term selection algorithm using Marginal Utility is being developed that will examine all ordering combinations, in order to identify the most efficient query.

The authors believe that queries of these magnitudes and complexities are required when necessary to provide a tailored database of relevant records that encompasses the broader aspects of target disciplines. In particular, if it is desired to enhance the transfer of ideas across disparate disciplines, and thereby stimulate the potential for innovation and discovery from complementary literatures [Kostoff, 1999], then even more complex queries using Simulated Nucleation may be required.

However, even with queries of this magnitude, not all records will be retrieved. As a point of reference, there were 204 articles with Abstracts published in the International Journal of Bifurcation and Chaos in 2001, of which 164 (~80%) were retrieved for this study. This was the highest fraction retrieved for any journal examined. For all the journals examined, some records had insufficient verbiage in their text fields, or had very non-standard verbiage relative to the main topical themes. Either of these problems precluded the query's accessing the record(s). To retrieve records with non-standard very low frequency terminology from all the journals accessed would require queries that contain thousands of terms. The reader should think about how many fewer Nonlinear Dynamics records would have been accessed with the typical search queries containing about a half dozen terms, and how author and journal citation rates are negatively impacted by the combination of deficient queries and insufficient verbiage in the record text fields.

4. RESULTS

The results from the publications bibliometric analyses are presented in section 4.1, followed by the results from the citations bibliometrics analysis in section 4.2. Results from the DT analyses are shown in section 4.3. The SCI bibliometric fields incorporated into the database included, for each paper, the author, journal, institution, Keywords, and references.

4.1 Publication Statistics on Authors, Journals, Organizations, Countries

The first group of metrics presented is counts of papers published by different entities. These metrics can be viewed as output and productivity measures. They are not direct measures of research quality, although there is some threshold quality level inferred, since these papers are published in the (typically) high caliber journals accessed by the SCI.

4.1.1 Author Frequency Results

For 2001, there were 6118 papers retrieved, 12136 different authors, and 16370 author listings. The occurrence of each author's name on a paper is defined as an author listing. While the average number of listings per author is about 1.34, the nineteen most prolific authors (see Table 2A) have listings more than an order of magnitude greater than the average. The number of papers listed for each author are those in the database of records extracted from the SCI using the query, not the total number of author papers listed in the source SCI database.

TABLE 2A – MOST PROLIFIC AUTHORS - 2001
(present institution listed)

AUTHOR	INSTITUTION	COUNTRY	#PAPERS
CHEN--GR	CITY UNIV HONG KONG	CHINA	24
LAI--YC	ARIZONA STATE	USA	21
NAYFEH--AH	VPI	USA	16
HU--G	CHINA CTR ADV S&T	CHINA	15
MOSEKILDE--E	TECH UNIV	DENMARK	15
XU--JX	XIAN JIAOTONG UNIV	CHINA	14
AIHARA--K	UNIV TOKYO	JAPAN	13
GASPARD--P	FREE UNIV BRUSSELS	BELGIUM	12
ZHENG--ZG	BEIJING NORMAL UNIV	CHINA	11
ALI--MK	UNIV LETHBRIDGE	CANADA	10
HU--BB	HONG KONG BAPTIST UNIV	CHINA	10
LLIBRE--J	UNIV AUTONOMA BARCELONA	SPAIN	10
GREBOGI--C	UNIV SAO PAULO	BRAZIL	9
KIM--SY	KANGWEON NATIONAL UNIV	SOUTH KOREA	9
KURTHS--J	UNIV POTSDAM	GERMANY	9
KUZNETSOV--SP	RUSSIAN ACADEMY OF SCIENCES	RUSSIA	9
LIU--JM	UCLA	USA	9
LIU--ZR	YUNNAN UNIV	CHINA	9
OTT--E	UNIV MARYLAND	USA	9

Of the nineteen most prolific authors listed in Table 2A, six are from China. In fact, eight are from the Far East, four are from Western Europe, one is from Eastern Europe, five are from North America, and one is from South America. Seventeen are from universities, and two are from research institutes.

To determine the trends in this regional mix of prolific authors, the same query was applied to 1991 only. Table 2B lists the most prolific authors for 1991.

TABLE 2B – MOST PROLIFIC AUTHORS - 1991

AUTHOR	INSTITUTION	COUNTRY	#PAPERS
OTT--E	UNIV MARYLAND	USA	13
GRAHAM--R	UNIV ESSEN GESAMTHSCH	GERMANY	12
PARISI--J	UNIV TUBINGEN	GERMANY	9
YORKE--JA	UNIV MARYLAND	USA	9
VAVRIV--DM	AM GORKII STATE UNIVERSITY	UKRAINE	8
SHEPELYANSKY--DL	NOVOSIBIRSK NUCL PHYS INST	SIBERIA	7
GREBOGI--C	UNIV MARYLAND	USA	6
MANDEL--P	UNIV LIBRE BRUXELLES	BELGIUM	6
SCOTT--SK	UNIV LEEDS	ENGLAND	6
STOOP--R	UNIV ZURICH	SWITZERLAND	6
SWINNEY--HL	UNIV TEXAS	USA	6
TEMAM--R	UNIV PARIS	FRANCE	6
ASHOURABDALLA--M	UCLA	USA	5
BADII--R	LAUSANNE UNIV	SWITZERLAND	5
BUCHNER--J	UCLA	USA	5
CASATI--G	UNIV MILAN	ITALY	5
ELNASCHIE--MS	CORNELL UNIV	USA	5
EPSTEIN--IR	BRANDEIS UNIV	USA	5
ERTL--G	MAX PLANCK GESELL	GERMANY	5

The regional mix of authors has some major differences from the 2001 results. Of the nineteen most prolific authors listed in Table 2B, none are from the Far East, eight are from the USA, nine are from Western Europe, and two are from Eastern Europe. Eighteen are from universities, and one is from a research institute.

Only two names were common to both lists, Ott and Grebogi. However, some researchers can have an off year for a number of reasons, so individual comparisons over two years, especially two widely separated years, may not be overly important. More important are country comparisons, and maybe institutional comparisons to some extent. These entities integrate over many individuals, and their performance would be more reflective of national policy. In this regard, the aggregate shift of prolific performers from the NATO countries in 1991 to those of the Far East in 2001 stands out.

4.1.2 Journals Containing Most Nonlinear Dynamics Papers

For 2001, there were 1151 different journals represented, with an average of 11.90 papers per

journal. The journals containing the most Nonlinear Dynamics papers (see Table 3A) had more than an order of magnitude more papers than the average.

TABLE 3A – JOURNALS CONTAINING MOST PAPERS - 2001

JOURNAL	# PAPERS
PHYS. REV. E	489
PHYS. REV. LETT.	175
INT. J. BIFURCATION CHAOS	164
PHYS. LETT. A	125
PHYSICA D	113
CHAOS SOLITONS FRACTALS	104
NONLINEAR ANAL.-THEORY METHODS APPL.	100
IEEE TRANS. CIRCUITS SYST. I-FUNDAM. THEOR. APPL.	92
PHYSICA A	85
PHYS. REV. B	84
J. PHYS. A-MATH. GEN.	73
PHYS. REV. A	72
J. FLUID MECH.	56
ACTA PHYS. SIN.	52
PHYS. PLASMAS	51
PHYS. REV. D	51
J. CHEM. PHYS.	48
J. SOUND VIBR.	45
PHYS. SCR.	45
ASTROPHYS. J.	45

The majority of the journals are physics, with the remainder divided between mathematics and electronics. Phys Rev E is the Physical Review journal assigned to chaos, while Phys Rev letters receives important papers for rapid publishing. Many (not all) of the other journals do not focus on nonlinear topics, but include papers in their specialties that also involve nonlinear aspects.

To determine the trends in journals containing the most Nonlinear Dynamics papers, the results from 1991 are examined. Table 3B contains the top twenty journals.

TABLE 3B – JOURNALS CONTAINING MOST PAPERS - 1991

JOURNAL	# PAPERS
PHYS. REV. A	176
PHYS. LETT. A	98
PHYSICA D	97
PHYS. REV. LETT.	77

J. FLUID MECH.	49
J. CHEM. PHYS.	48
EUROPHYS. LETT.	37
PHYS. REV. B-CONDENS MATTER	37
NONLINEARITY	37
J. PHYS. A-MATH. GEN.	32
GEOPHYS. RES. LETT.	28
J. STAT. PHYS.	28
ASTROPHYS. J.	24
EUR. J. MECH. B-FLUIDS	24
OPT. COMMUN.	23
NONLINEAR ANAL.-THEORY METHODS APPL.	20
PHYS. REV. D	19
LECT. NOTES MATH.	19
INT. J. NON-LINEAR MECH.	18
J. PHYS. CHEM.	17

While the most prolific authors could be expected to change over a decade, for a number of reasons, the most prolific journals should be more stable. Comparison of Tables 3A and 3B shows this to be true. Of the nineteen most prolific journals, eleven are in common. For 2001, two journals were added devoted solely to chaos and closely related topics (CHAOS SOLITONS FRACTALS, INTERNATIONAL JOURNAL OF BIFURCATION AND CHAOS). Perhaps the largest change is the drop of Physical Review A from first in 1991 to twelfth in 2001, and the appearance of Physical Review E as first in 2001. Phys Rev E was split from Phys Rev A during the past decade, and received the Physical Review assignment for papers in chaos.

4.1.3 Institutions Producing Most Nonlinear Dynamics Papers

A similar process was used to develop a frequency count of institutional address appearances. It should be noted that many different organizational components may be included under the single organizational heading (e.g., Harvard Univ could include the Chemistry Department, Biology Department, Physics Department, etc.). Identifying the higher level institutions is instrumental for these DT studies. Once they have been identified through bibliometric analysis, subsequent measures may be taken (if desired) to identify particular departments within an institution.

TABLE 4A – PROLIFIC INSTITUTIONS – 2001

INSTITUTION	COUNTRY	# PAPERS
RUSSIAN ACAD SCI	RUSSIA	165
CHINESE ACAD SCI	CHINA	72
UNIV TOKYO	JAPAN	68
UNIV CALIF SAN DIEGO	USA	67

UNIV MARYLAND	USA	61
UNIV CALIF BERKELEY	USA	53
ARIZONA STATE UNIV	USA	48
UNIV CALIF LOS ANGELES	USA	47
FREE UNIV BRUSSELS	BELGIUM	47
CORNELL UNIV	USA	43
UNIV TEXAS	USA	43
UNIV HOUSTON	USA	41
UNIV ILLINOIS	USA	41
GEORGIA INST TECHNOL	USA	40
PRINCETON UNIV	USA	40
INDIAN INST TECHNOL	INDIA	39
MIT	USA	38
CNRS	FRANCE	37
IST NAZL FIS NUCL	ITALY	36
MAX PLANCK INST PHYS KOMPLEXER SYST	GERMANY	36
TECHNION ISRAEL INST TECHNOL	ISRAEL	36
BEIJING NORMAL UNIV	CHINA	36
MOSCOW MV LOMONOSOV STATE UNIV	RUSSIA	36
NORTHWESTERN UNIV	USA	36
UNIV SAO PAULO	BRAZIL	34
TECH UNIV DENMARK	DENMARK	34
UNIV WASHINGTON	USA	34
UNIV PARIS 06	FRANCE	33
CITY UNIV HONG KONG	CHINA	33
UNIV CAMBRIDGE	ENGLAND	33

For 2001, of the thirty most prolific institutions, fourteen are from the USA, seven are from Western Europe, five are from Asia, two are from Eastern Europe, one is from Latin America, and one is from the Middle East. Twenty-five are universities, and the remaining institutions are research institutes. The most prolific institutions for Nonlinear Dynamics papers correlate well with institutions that have Centers for Nonlinear Dynamics.

To determine the trends in institutions containing the most Nonlinear Dynamics papers, the results from 1991 were examined. Table 4B contains the top thirty institutions.

TABLE 4B – PROLIFIC INSTITUTIONS – 1991

INSTITUTION	COUNTRY	# PAPERS
ACAD SCI USSR	USSR	49
UNIV TEXAS	USA	35

MIT	USA	33
UNIV MARYLAND	USA	31
UNIV CAMBRIDGE	ENGLAND	29
USN	USA	29
UNIV CALIF LOS ANGELES	USA	28
CORNELL UNIV	USA	27
UNIV CALIF SAN DIEGO	USA	26
CALTECH	USA	25
ACAD SCI UKSSR	USSR	25
UNIV ILLINOIS	USA	25
UNIV CALIF LOS ALAMOS SCI LAB	USA	24
UNIV ARIZONA	USA	23
UNIV TORONTO	CANADA	22
UNIV CALIF BERKELEY	USA	22
UNIV MINNESOTA	USA	21
UNIV PARIS 11	FRANCE	21
NASA	USA	21
NORTHWESTERN UNIV	USA	20
UNIV LEEDS	ENGLAND	20
GEORGIA INST TECHNOL	USA	19
UNIV ESSEN GESAMTHSCH	GERMANY	19
UNIV HOUSTON	USA	19
UNIV TOKYO	JAPAN	18
MV LOMONOSOV STATE UNIV	USSR	18
UNIV PARIS 06	FRANCE	18
PRINCETON UNIV	USA	17
BROWN UNIV	USA	16
UNIV COLORADO	USA	16

Of the thirty most prolific institutions in 1991, twenty are from the USA, five are from Western Europe, three are from Eastern Europe, one is from Asia, and one is from Canada. The major shift is substitution of Asian institutions for USA institutions. In addition, twenty-five institutions are universities, and five are research institutes.

There are at least five factors that underlay the quality and quantity of Nonlinear Dynamics research. First, Nonlinear Dynamics is on the cutting edge of physics research, and has applicability to many different S&T disciplines. It is a prime research area for an institution's academic expansion.

Second, advances in Nonlinear Dynamics requires people who are intelligent and well-trained in physics and mathematics. Asian countries have large populations, and large numbers of researchers, well trained in physics, mathematics, and other fundamental disciplines. They tend to score well in international scientific education competitions. They have the educational foundations for becoming

major contributors.

Third, much of Nonlinear Dynamics requires the extensive use of computers, to perform and display results of theoretical computations, and support analysis of experimental data. The growth of affordable personal computers, mainly in the decade of the 90s, has allowed poor third-world countries to acquire modern computational facilities, and compete as almost equals in this area.

Fourth, there is a strong theoretical component, that requires substantial intellect and minimal funding. This provides poorer countries with a large supply of well-educated professionals the opportunity to gain high visibility in theoretical studies of Nonlinear Dynamics.

Fifth, there is a strong data analysis component, with three aspects to the data analysis: 1) the ease in obtaining the data; 2) the ability to analyze the data; 3) the tools needed to support the analysis. Item 2) requires well-trained professionals, and the proliferation of such people in Asian countries was addressed previously. Item 3) involves modern computers, and the recent proliferation of these facilities in Asian countries was also addressed previously. Item 1) depends on the data source. For data that requires expensive laboratory or field or flight tests to acquire, the poorer countries are at a distinct dis-advantage relative to the developed countries. For example, in the China/ USA comparison presented later, it is shown that China has very little effort in disciplines such as space phenomena analysis or controlled fusion plasma analysis. This is undoubtedly related to the high costs of acquiring data in these areas, and China's lack of a substantial experimental effort in these areas. However, there is much data that can be analyzed wit h the techniques of Nonlinear Dynamics that does not require expensive facilities, and the less affluent Asian countries can focus substantial efforts in these areas.

4.1.4 Countries Producing Most Nonlinear Dynamics Papers

There are 78 different countries listed in the results for 2001. The country bibliometric results are summarized in Table 5A and shown graphically in Figure 1. The dominance of a handful of countries is clearly evident.

TABLE 5A – PROLIFIC COUNTRIES - 2001

COUNTRY	# PAPERS
USA	1797
PEOPLES R CHINA	588
GERMANY	585
JAPAN	470
FRANCE	426
ENGLAND	415
RUSSIA	394
ITALY	338
SPAIN	260

CANADA	242
BRAZIL	173
INDIA	157
NETHERLANDS	141
ISRAEL	127
POLAND	123
AUSTRALIA	118
TAIWAN	110
SOUTH KOREA	109
MEXICO	101
BELGIUM	99
UKRAINE	79
GREECE	74
SWEDEN	71
ARGENTINA	70
DENMARK	60
SCOTLAND	55
SWITZERLAND	53
AUSTRIA	47
HUNGARY	47
EGYPT	35

There appear to be two dominant groupings. The first group is the USA. It has as many papers as the members of the second group, People's Republic of China, Germany, and Japan.

To determine the trends in countries containing the most nonlinear dynamics papers, the results from 1991 were examined. Table 5B summarizes results from the top twenty countries, and Figure 2 displays these results graphically.

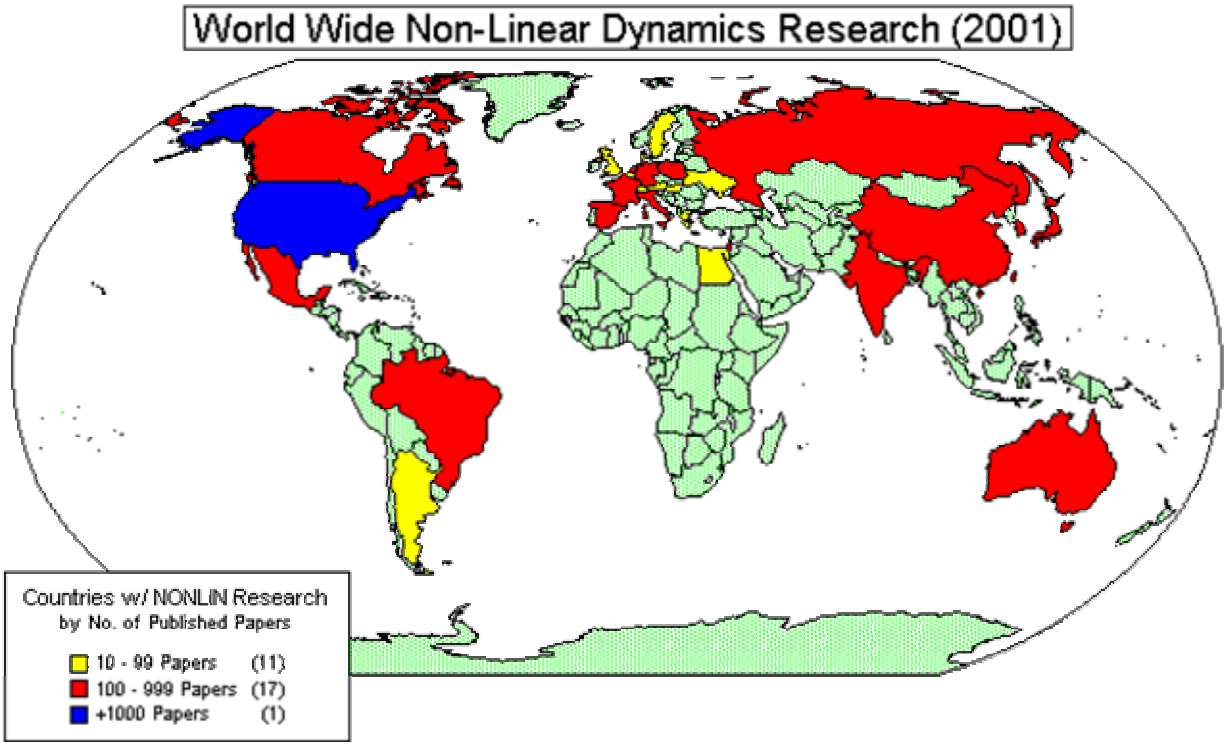


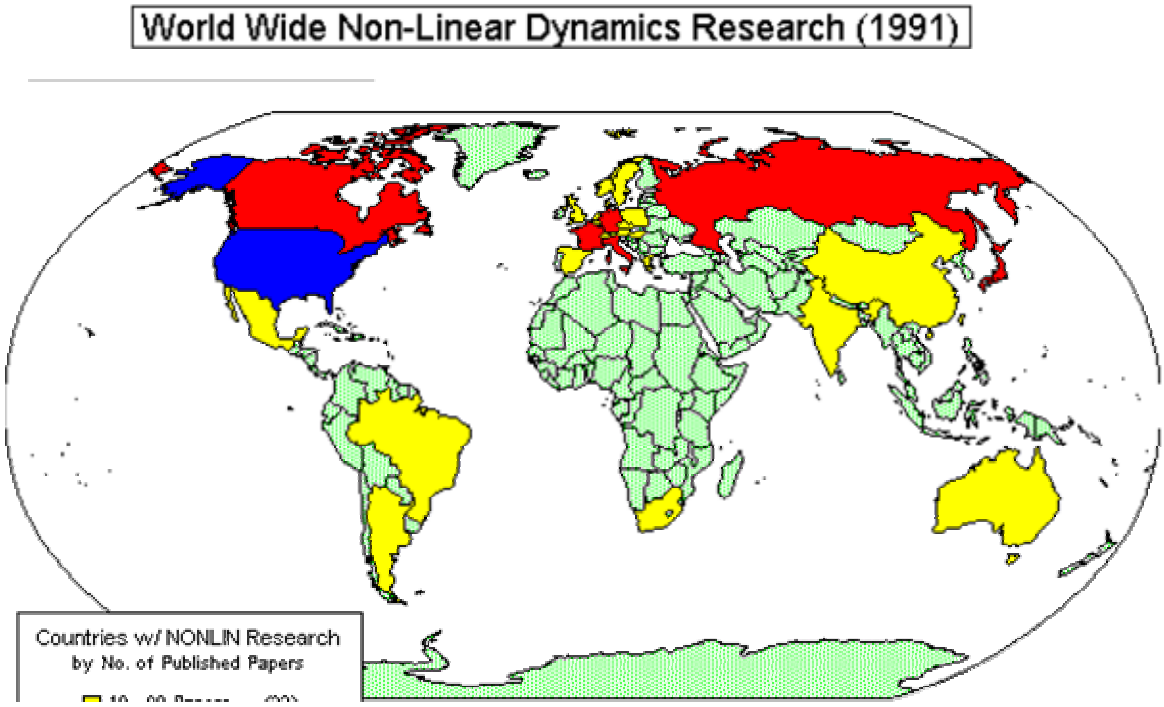
FIGURE 1 – COUNTRIES WITH THE MOST NONLINEAR DYNAMICS PAPERS – 2001

TABLE 5B – PROLIFIC COUNTRIES - 1991

COUNTRY	# PAPERS
USA	1031
GERMANY	247
USSR	207
ENGLAND	162
FRANCE	158
JAPAN	154
CANADA	118
ITALY	117
INDIA	65
POLAND	65
PEOPLES R CHINA	63
ISRAEL	52
AUSTRALIA	43
NETHERLANDS	43
SWITZERLAND	40
SPAIN	38

BELGIUM	27
BRAZIL	26
GREECE	25
DENMARK	22
HUNGARY	22
SCOTLAND	22
TAIWAN	22
CZECHOSLOVAKIA	17
SWEDEN	16
AUSTRIA	13
ARGENTINA	11
SOUTH AFRICA	11
MEXICO	10
NORWAY	10

FIGURE 2 – COUNTRIES WITH THE MOST NONLINEAR DYNAMICS PAPERS – 1991



The major shift is the increased ranking of People's Republic of China from 11th in 1991 to 2nd in 2001, and the concomitant increase in numbers of papers from 63 to 584. To place China's increase in Nonlinear Dynamics papers in perspective, it is compared to China's overall increase in SCI papers from 1991 to 2001. In 1991, China had 8174 entries in the SCI, and in 2001, China had 36765 entries in the SCI. Thus, while China's papers in Nonlinear Dynamics in the SCI increased by a factor of ~9.25 from 1991 to 2001, China's overall increase in SCI papers from 1991 to 2001 was a factor of ~4.5. Thus, China's Nonlinear Dynamics papers outpaced its average growth of SCI papers by a factor of ~ 2.

Figure 3 contains a co-occurrence matrix of the top 15 countries. In terms of absolute numbers of co-authored papers, the USA major partners are Germany, China, France, Canada, and England. Interestingly, the USA is China's dominant major partner, having four times the number of co-authored papers with China (72) as China's next larger partner, Canada (18). Overall, countries in similar geographical regions tend to co-publish substantially, the US being a moderate exception.

FIGURE 3 – COUNTRY CO-OCCURRENCE MATRIX

Items	Brazil	Canada	England	France	Germany	India	Israel	Italy	Japan	Netherlands	Peoples R China	Poland	Russia	Spain	USA
BRAZIL	173	0	4	10	5	0	1	4	0	1	6	1	3	4	29
CANADA	0	242	14	11	10	1	3	5	5	1	18	1	5	3	62
ENGLAND	4	14	415	20	28	4	5	9	5	12	10	4	19	11	55
FRANCE	10	11	20	426	28	4	3	27	8	7	0	4	21	11	62
GERMANY	5	10	28	28	585	3	19	18	8	21	13	16	44	12	74
INDIA	0	1	4	4	3	157	0	1	1	1	0	0	2	3	16
ISRAEL	1	3	5	3	19	0	127	1	4	2	0	4	4	2	37
ITALY	4	5	9	27	18	1	1	338	6	8	4	4	11	15	47
JAPAN	0	5	5	8	8	1	4	6	470	5	14	1	7	3	45
NETHERLANDS	1	1	12	7	21	1	2	8	5	141	1	1	12	6	27
PEOPLES R CHINA	6	18	10	0	13	0	0	4	14	1	588	0	3	5	72
POLAND	1	1	4	4	16	0	4	4	1	1	0	123	5	3	21
RUSSIA	3	5	19	21	44	2	4	11	7	12	3	5	394	13	26
SPAIN	4	3	11	11	12	3	2	15	3	6	5	3	13	260	39
USA	29	62	55	62	74	16	37	47	45	27	72	21	26	39	1797

10 - 99 Papers (22)
100 - 999 Papers (6)
1000 Papers (1)

4.2 Citation Statistics on Authors, Papers, and Journals

The second group of metrics presented is counts of citations to papers published by different entities. While citations are ordinarily used as impact or quality metrics [Garfield, 1985], much caution needs to be exercised in their frequency count interpretation, since there are numerous reasons why authors cite or do not cite particular papers [Kostoff, 1998; MacRoberts and MacRoberts, 1996].

The citations in all the retrieved SCI papers were aggregated, the authors, specific papers, years, journals, and countries cited most frequently were identified, and were presented in order of decreasing frequency. A small percentage of any of these categories received large numbers of citations. From the citation year results, the most recent papers tended to be the most highly cited. This reflected rapidly evolving fields of research.

4.2.1 Most Cited Authors

The most highly cited authors from the 2001 database are listed in Table 6.

TABLE 6 – MOST CITED AUTHORS
(cited by other papers in this database only)

AUTHOR	INSTITUTION	COUNTRY	# CITES
OTT E	UNIV MARYLAND	USA	399
GRASSBERGER P	KFA JULICH GMBH	GERMANY	329
PECORA LM	USN	USA	323
GUCKENHEIMER J	CORNELL	USA	305
NAYFEH AH	VPI	USA	296
KANEKO K	UNIV TOKYO	JAPAN	247
BERRY MV	UNIV BRISTOL	ENGLAND	235
ARNOLD VI	RUSSIAN ACADEMY OF SCIENCE	RUSSIA	230
TAKENS F	UNIV GRONINGEN	NETHERLANDS	212
GASPARD P	FREE UNIV BRUSSELS	BELGIUM	199
GUTZWILLER MC	IBM	USA	194
THEILER J	LOS ALAMOS NATIONAL LAB	USA	194
ABARBANEL HDI	UNIV CAL SAN DIEGO	USA	193
GREBOGI C	UNIV SAO PAULO	BRAZIL	192
LAI YC	ARIZONA STATE	USA	187
ECKMANN JP	UNIV GENEVA	SWITZERLAND	185
LORENZ EN	MIT	USA	174
PIKOVSKY AS	UNIV POTSDAM	GERMANY	172
PRESS WH	HARVARD UNIV	USA	163
CASATI G	UNIV INSUBRIA	ITALY	163

Of the twenty most cited authors, ten are from the USA, seven from Western Europe, one from

Russia, one from Japan, and one from Latin America. This is a far different distribution from the most prolific authors of 2001, where eight of nineteen were from the Far East. This distribution of most cited authors more closely resembles the distribution of most prolific authors from 1991, where none were from the Far East.

There are a number of potential reasons for this difference between most prolific and cited authors in 2001. The most prolific may not be the highest quality, or many of the most prolific authors could be relatively recent, and insufficient time has elapsed for their citations to accumulate. In another three or four years, when the papers from present-day authors have accumulated sufficient citations, firmer conclusions about quality can be drawn.

The lists of nineteen most prolific authors from 2001 and twenty most highly cited authors only had five names in common (OTT, NAYFEH, GASPARD, GREBOGI, LAI). This phenomenon of minimal intersection has been observed in all other text mining studies performed by the first author.

Fifteen of the authors' institutions are universities, four are government-sponsored research laboratories, and one is a private company.

The citation data for authors and journals represents citations generated only by the specific records extracted from the SCI database for this study. It does not represent all the citations received by the references in those records; these references in the database records could have been cited additionally by papers in other technical disciplines.

4.2.2 Most Cited Papers

The most highly cited documents from the 2001 database are listed in Table 7.

TABLE 7 – MOST CITED DOCUMENTS

(total citations listed in SCI)

AUTHOR NAME	YEAR	JOURNAL	VOLUME / PAGE	# SCI CITES
PECORA LM	1990	PHYS REV LETT	V64,P821	938
(SYNCHRONIZATION IN CHAOTIC SYSTEMS)				
GUCKENHEIMER J	1983	NONLINEAR OSCILLATIONS		
(MATHEMATICAL STUDIES OF BIFURCATIONS)				
OTT E	1990	PHYS REV LETT	V64,P1196	1274
(CONTROLLING CHAOS)				
LORENZ EN	1963	J ATMOS SCI	V20,P130	2971
(DETERMINISTIC NONPERIODIC FLOW)				
CROSS MC	1993	REV MOD PHYS	V65,P851	1500
(PATTERN-FORMATION OUTSIDE OF EQUILIBRIUM)				
WOLF A	1985	PHYSICA D	V16,P285	1566

(DETERMINING LYAPUNOV EXPONENTS FROM A TIME-SERIES; INTRODUCED CHAOS)				
TAKENS F	1981	LECT NOTES MATH	V898,P366	
(MATHEMATICAL PAPER ON ANALYSIS OF CHAOTIC TIME SERIES)				
OTT E	1993	CHAOS DYNAMICAL SYST		
(CHAOS CONTROL THEORY)				
GRASSBERGER P	1983	PHYSICA D	V9,P189	1567
(MEASURING THE STRANGENESS (FRACTAL GEOMETRY) OF STRANGE ATTRACTORS)				
GUTZWILLER MC	1990	CHAOS CLASSICAL QUAN		
(QUANTUM IDEAS ON CHAOS)				
ROSENBLUM MG	1996	PHYS REV LETT	V76,P1804	241
(PHASE SYNCHRONIZATION OF CHAOTIC OSCILLATORS)				
GRASSBERGER P	1983	PHYS REV LETT	V50,P345	1369
(CHARACTERIZATION OF STRANGE ATTRACTORS IN AN OSCILLATOR'S PHASE SPACE)				
ECKMANN JP	1985	REV MOD PHYS	V57,P617	1557
(ERGODIC-THEORY OF CHAOS AND STRANGE ATTRACTORS)				
THEILER J	1992	PHYSICA D	V58,P77	568
(SURROGATE DATA TESTING FOR NONLINEARITY IN TIME-SERIES)				
NAYFEH AH	1979	NONLINEAR OSCILLATIONS		
(TEXTBOOK ON NONLINEAR MECHANICS)				
FUJISAKA H	1983	PROG THEOR PHYS	V69,P32	294
(STABILITY THEORY OF SYNCHRONOUS MOTION IN COUPLED-OSCILLATOR SYSTEM)				
WIGGINS S	1990	INTRO APPL NONLINEAR		
(APPLIED NONLINEAR DYNAMICAL SYSTEMS AND CHAOS)				
RULKOV NF	1995	PHYS REV E	V51,P980	213
(SYNCHRONIZATION OF CHAOS IN DIRECTIONALLY COUPLED CHAOTIC SYSTEMS)				
PYRAGAS K	1992	PHYS LETT A	V170,P421	512
(CONTINUOUS CONTROL OF CHAOS BY SELF-CONTROLLING FEEDBACK)				
LICHTENBERG AJ	1992	REGULAR CHAOTIC DYNA		
(CHAOTIC MOTION IN NONLINEAR DYNAMICAL SYSTEMS)				

The theme of each paper is shown in italics on the line after the paper listing. The order of paper listings is by number of citations by other papers in the extracted database analyzed. The total number of citations from the SCI paper listing, a more accurate measure of total impact, is shown in the last column on the right.

Physical Review Letters contains the most papers by far, four out of the twenty listed. Most of the

journals are fundamental science journals, and most of the topics have a fundamental science theme. The majority of the papers are from the 1990s, with seven from the 1980s, one from the 1970s, and one extremely highly cited paper being from 1963. This reflects a dynamic research field, with seminal works being performed in the recent past.

Eight of the papers address issues related to chaos, with the dominant themes being conditions for determining chaos, and properties of strange attractors Four of the papers address issues related to synchronization, with the focus on coupled chaotic oscillators. Three of the papers address issues related to control, emphasizing self-controlling feedback. One paper addresses stability-related issues, focusing on bifurcations, and one paper focuses on quantum chaos. There are three nonlinear dynamics books in the top twenty cited documents.

Thus, the major intellectual emphasis of cutting edge Nonlinear Dynamics research, as evidenced by the most cited papers, is well aligned with the intellectual heritage and performance emphasis, as will be evidenced by the clustering approaches presented later.

4.2.3. Most Cited Journals

The most highly cited journals from the 2001 database are listed in Table 8.

TABLE 8 – MOST CITED JOURNALS

(cited by other papers in this database only)

JOURNAL	TIMES CITED
PHYS REV LETT	10786
PHYS REV E	5310
PHYS REV A	3603
PHYSICA D	3579
PHYS LETT A	2308
J CHEM PHYS	2138
J FLUID MECH	2002
PHYS REV B	1969
NATURE	1911
ASTROPHYS J	1367
INT J BIFURCAT CHAOS	1279
SCIENCE	1256
PHYS REV D	1215
J PHYS A-MATH GEN	1073
PHYS FLUIDS	907
J ATMOS SCI	871
REV MOD PHYS	864
PHYS REP	813

J STAT PHYS	790
CHAOS	777

The first two groups of cited journals clearly stand out. PHYS REV LETT received almost as many cites as the three journals in the next group (PHYS REV E, PHYS REV A, PHYSICA D), or even the five journals in the following group (PHYS LETT A, J CHEM PHYS, J FLUID MECH, PHYS REV B, NATURE). PHYS REV LETT emphasizes rapid publication of 'hot' topics, and would therefore tend to establish primacy in an emerging field. Since one aspect of citations is identifying the original literature of a new topic, a credible journal with these characteristics would tend to receive large numbers of citations.

Unlike the relatively disjoint relationship between most prolific authors in 2001 and most cited authors, the relationship between most prolific journals in 2001 and most cited journals was much closer. Nine of the ten most highly cited journals were also on the list of twenty most prolific journals in 2001. The more applied journals on the most prolific list for 2001 are replaced by the more fundamental journals on the most cited list.

The authors end this bibliometrics section by recommending that the reader interested in researching the topical field of interest would be well-advised to, first, obtain the highly-cited papers listed and, second, peruse those sources that are highly cited and/or contain large numbers of recently published papers.

4. REFERENCES FOR APPENDIX 7-F

Garfield, E. [1985] History of citation indexes for chemistry - a brief review. JCICS. 25(3). 170-174.

Kostoff, R. N. [2000]. The underpublishing of science and technology results. The Scientist. 14(9). 6-6. 1 May.

Kostoff, R. N. [1993]. Database Tomography for technical intelligence. Competitive Intelligence Review. 4(1). 38-43.

Kostoff, R. N. [1998]. The use and misuse of citation analysis in research evaluation. Scientometrics. 43(1). 27-43. September.

Kostoff, R. N. [1999]. Science and technology innovation. Technovation. 19(10). 593-604.

Kostoff, R. N. et al [1995]. System and method for Database Tomography. U.S. Patent Number 5440481.

Kostoff, R. N., Braun, T., Schubert, A., Toothman, D. R., and Humenik, J. A. [2000]. Fullerene roadmaps using bibliometrics and Database Tomography. Journal of Chemical Information and Computer Science. 40(1). 19-39. Jan-Feb.

Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. [1997]. Database Tomography for information retrieval. Journal of Information Science. 23(4). 301-311.

Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. [1998]. Database Tomography for technical intelligence: a roadmap of the near-earth space science and technology literature. Information Processing and Management. 34(1). 69-85.

Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. [1999]. Hypersonic and supersonic flow roadmaps using bibliometrics and Database Tomography. JASIS. 50(5). 427-447. 15 April.

Kostoff, R. N., Eberhart, H. J., Toothman, D. R., and Pellenbarg, R. [1997]. Database Tomography for technical intelligence: comparative roadmaps of the research impact assessment literature and the Journal of the American Chemical Society, Scientometrics. 40(1). 103-138.

Kostoff, R. N., Green, K. A., Toothman, D. R., and Humenik, J. A. [2000]. Database Tomography applied to an aircraft science and technology investment strategy. Journal of Aircraft. 37(4). 727-730. July-August.

Kostoff, R. N., Toothman, D. R., Eberhart, H. J., and Humenik, J. A. [2001]. Text mining using Database Tomography and bibliometrics: a review. Technology Forecasting and Social Change. 68(3). 223-253.

Kostoff, R. N., Tshiteya, R., Pfeil, K. M., and Humenik, J. A. [2002]. Electrochemical power source roadmaps using bibliometrics and Database Tomography. Journal of Power Sources. 110(1). 163-176.

Kostoff, R.N. [1994]. Database Tomography: origins and applications. Competitive Intelligence Review. Special Issue on Technology, 5:1. 48-55.

MacRoberts, M. H., and MacRoberts, B. R. [1996]. Problems of citation analysis. Scientometrics. 36(3). 435-444. July-August.

Proceeding of the 1st Experimental Chaos Conference. [1992]. Editors, S. Vohra, M. Spano, M. Shlesinger, L. Pecora, W. Ditto. World Scientific Pub. [Singapore]

Proceeding of the 2nd Experimental Chaos Conference. [1995]. Editors, W. Ditto, L. Pecora, M. Shlesinger, M. Spano, S. Vohra, World Scientific Pub. [Singapore]

Proceeding of the 3rd Experimental Chaos Conference. [1996]. Editors, R. Harrison, W. Lu, W. Ditto, L. Pecora, M. Spano, S. Vohra. World Scientific Pub. [Singapore]

Proceeding of the 4th Experimental Chaos Conference. [1998]. Editors, M. Ding, W. Ditto, L. Pecora, M. Spano, S. Vohra, World Scientific Pub. [Singapore]

Proceeding of the 5th Experimental Chaos Conference. [2001]. Editors, M. Ding, W. Ditto, L. Pecora, M. Spano, S. Vohra, World Scientific Pub. [Singapore]

Proceeding of the 6th Experimental Chaos Conference. [2002]. Editors, S. Bocaletti, B. Gluckman, J. Kurths, L. Pecora, , M. Spano. American Institute of Physics Conference Proceedings Vol. 622 [Melville, NY]

SCI. [2002]. Science Citation Index. Institute for Scientific Information. Phila., PA.

APPENDIX 7-G.

FRACTALS TEXT MINING USING BIBLIOMETRICS AND DATABASE TOMOGRAPHY (Kostoff et al, 2004c)

The present appendix describes use of the DT process, supplemented by literature bibliometric analyses, to derive technical intelligence from the published literature of Fractals science and technology.

Fractals, as defined by the authors for this study, are geometric structures (e.g., Mandelbrot set, percolation clusters, diffusion-limited aggregates) or dynamical processes (e.g., f ractional Brownian motion, avalanches, turbulent intermittency) that possess features on many scales related through a power law relationship. Since one of the key outputs of the present study is a query that can be used by the community to access relevant Fractals documents, a recommended query based on this study is presented in total. This query serves as the operational definition of Fractals, and its development is discussed in detail in the database generation section.

FRACTALS QUERY

FRACTAL* OR SELF-SIMILAR* OR SELF-ORGANIZED CRITICALITY OR MULTIFRACTAL OR ANOMALOUS DIFFUSION OR SCALE INVARIANT OR HAUSDORFF DIMENSION OR DIFFUSION LIMITED AGGREGATION OR FRACTIONAL BROWNIAN MOTION OR MANDELBROT OR LACUNARITY OR CANTOR SET OR NONFRACTAL OR MONOFRACTAL NOT FRACTALKINE*

To execute the study reported in this Appendix, a database of relevant Fractals articles is generated using the iterative search approach of Simulated Nucleation [Kostoff et al, 1997a, 2001]. Then, the database is analyzed to produce the following characteristics and key features of the Fractals field: recent prolific Fractals authors; journals that contain numerous Fractals papers; institutions that produce numerous Fractals papers; keywords most frequently specified by the Fractals authors; authors, papers and journals cited most frequently; pervasive technical themes of Fractals; and relationships among the pervasive themes and sub-themes.

2. BACKGROUND

Recent DT/ bibliometrics studies were conducted of the technical fields of: 1) Near-earth space (NES) [Kostoff et al, 1998]; 2) Hypersonic and supersonic flow over aerodynamic bodies (HSF) [Kostoff et al, 1999]; 3) Chemistry (JACS) [Kostoff et al, 1997b] as represented by the Journal of the American Chemical Society; 4) Fullerenes (FUL) [Kostoff et al; 2000a] 5) Aircraft (AIR) [Kostoff et al, 2000b]; 6) Hydrodynamic flow over surfaces (HYD); 7) Electric Power Sources (EPS); 8) Electrochemical Power Sources (ECHEM) [Kostoff et al, 2002] 9) the non-technical field of

research impact assessment (RIA) [Kostoff et al, 1997b], and 10) NonLinear Dynamics (NONLIN) [Kostoff et al, In Press]. Overall parameters of these studies from the SCI database results and the current Fractals study are shown in Table 1.

TABLE 1 - DT STUDIES OF TOPICAL FIELDS

TOPICAL AREA	NUMBER OF SCI ARTICLES	YEARS COVERED
1) NEAR-EARTH SPACE (NES)	5480	1993-MID 1996
2) HYPERSONICS (HSF)	1284	1993-MID 1996
3)CHEMISTRY (JACS)	2150	1994
4) FULLERENES (FUL)	10515	1991-MID 1998
5) AIRCRAFT (AIR)	4346	1991-MID 1998
6) HYDRODYNAMICS (HYD)	4608	1991-MID 1998
7) ELECTRIC POWER SOURCES (EPS)	20835	1991-BEG 2000
8) ELECTROCHEMICAL POWER SOURCES (ECHEM)	6985	1993 – MID 2001
9) RESEARCH ASSESSMENT (RIA)	2300	1991-BEG 1995
10) NONLINEAR DYNAMICS (NONLIN)	6118 (2001)	1991, 2001
11) FRACTALS (FRACT)	4454 (2001-02); 4211(1991-93)	1991-93; 2001-02

2.2 Unique Study Features

The study reported in the present Appendix is in the journal article abstract category. It differs from the previous published papers in this category [Kostoff, 1999; Kostoff et al, 1998, 1997b, 2000a, 2000b, 2002] in five respects. First, the topical domain (Fractals) is completely different. Second, a document clustering technique for theme categorization, based on Greedy String Tiling for text similarity, was developed and included, to complement the word/ concept clustering approach. Third, bibliometric clustering is presented for two database fields: authors and countries. Fourth, factor matrix filtering was developed and used to select context-dependent words for input to the clustering algorithm, thereby leading to more sharply defined clusters. Finally, the marginal utility algorithm was applied, allowing only the highest payoff terms to be included in the final query, and resulting in an efficient query.

3. DATABASE GENERATION

The key step in the Fractals literature analysis is the generation of the database to be used for processing. There are three key elements to database generat ion: the overall objectives, the approach selected, and the database used. Each of these elements is described.

3.1 Overall Study Objectives

The main objective was to identify global S&T that had both direct and indirect relations to Fractals. A sub-objective was to estimate the overall level of global effort in Fractals S&T, as reflected by the emphases in the published literature.

3.2 Databases and Approach

For the present study, the SCI database (including both the Science Citation Index and the Social Science Citation Index) was used. The approach used for query development was the DT-based iterative relevance feedback concept [Kostoff et al, 1997a].

3.2.1 Science Citation Index/ Social Science Citation Index (SCI) [SCI, 2002]

The retrieved database used for analysis consists of selected journal records (including the fields of authors, titles, journals, author addresses, author keywords, abstract narratives, and references cited for each paper) obtained by searching the Web version of the SCI for Fractals articles. At the time the final data was extracted for the present paper (Fall 2002), the version of the SCI used accessed about 5600 journals (mainly in physical, engineering, and life sciences basic research) from the Science Citation Index, and over 1700 journals from the Social Science Citation Index.

The SCI database selected represents a fraction of the available Fractals (mainly research) literature, that in turn represents a fraction of the Fractals S&T actually performed globally [Kostoff, 2000]. It does not include the large body of classified literature, or company proprietary technology literature. It does not include technical reports or books or patents on Fractals. It covers a finite slice of time (1991-93, 2001-02). The database used represents the bulk of the peer-reviewed high quality Fractals research literature, and is a representative sample of all Fractals research in recent times.

In order to generate an efficient final query, a new process termed Marginal Utility was applied. At the start of the final iteration, a modified query Q1 was inserted into the SCI, and records were retrieved. A sample of these records was then categorized into relevant and non-relevant. Each term in Q1 was inserted into the Marginal Utility algorithm, and the marginal number of relevant and non-relevant records in the sample that the query term would retrieve was computed. Only those terms that retrieved a high ratio of relevant to non-relevant records were retained. Since (by design) each query term had been used to retrieve records from the SCI as part of Q1, the marginal ratio of relevant to non-relevant records from the sample would represent the marginal ratio of relevant to non-relevant records from the SCI. The final efficient query Q2, consisting of the highest marginal utility terms, was shown in the Introduction.

In the Marginal Utility algorithm, terms that co-occur strongly in records with previously-selected terms are essentially duplicative from the retrieval perspective, and can be eliminated. Thus, the order in which terms are selected becomes important. An automated query term selection algorithm using Marginal Utility is being developed that will examine all ordering combinations, in order to identify the most efficient query.

The author believes that queries of these magnitudes and complexities are required when necessary to provide a tailored database of relevant records that encompasses the broader aspects of target disciplines. In particular, if it is desired to enhance the transfer of ideas across disparate disciplines, and thereby stimulate the potential for innovation and discovery from complementary literatures [Kostoff, 1999], then even more complex queries using Simulated Nucleation may be required.

However, even with queries of this magnitude, not all records will be retrieved. As a point of reference, there were 39 articles with Abstracts published in the present journal in 2001, of which 31 (~80%) were retrieved for this study. This was the highest fraction retrieved for any journal examined. For all the journals examined, some records had insufficient verbiage in their text fields, or had very non-standard verbiage relative to the main topical themes. Either of these problems precluded the query's accessing the record(s). To retrieve records with non-standard very low frequency terminology from all the journals accessed would require queries that contain thousands of terms. **The reader should think about how many fewer Fractals records would have been accessed with the typical search queries containing about a half dozen terms, and how author and journal citation rates are negatively impacted by the combination of deficient queries and insufficient verbiage in the record text fields.**

4. RESULTS

The results from the publications bibliometric analyses are presented in section 4.1, followed by the results from the citations bibliometrics analysis in section 4.2. Results from the DT analyses are shown in section 4.3. The SCI bibliometric fields incorporated into the database included, for each paper, the author, journal, institution, Keywords, and references.

4.1 Publication Statistics on Authors, Journals, Organizations, Countries

The first group of metrics presented is counts of papers published by different entities. These metrics can be viewed as output and productivity measures. They are not direct measures of research quality, although there is some threshold quality level inferred, since these papers are published in the (typically) high caliber journals accessed by the SCI.

4.1.2 Author Frequency Results

For 2001-02, there were 4464 papers retrieved (4380 of which had Abstracts), 9403 different authors, and 12780 author listings. The occurrence of each author's name on a paper is defined as an author listing. While the average number of listings per author is about 1.36, the nineteen most prolific authors (see Table 2A) have listings more than an order of magnitude greater than the average. The number of papers listed for each author are those in the database of records extracted from the SCI using the query, not the total number of author papers listed in the source SCI database.

TABLE 2A – MOST PROLIFIC AUTHORS – 2001-02

(present institution listed)

AUTHOR	INSTITUTION	COUNTRY	#PAPERS
STANLEY--HE	BOSTON UNIV	USA	15
HUIKURI--HV	UNIV OULU	FINLAND	14
WU--ZQ	UNIV SCI AND TECH	CHINA	13
ZASLAVSKY--GM	NYU	USA	12
JIN--ZZ	WUHAN UNIV	CHINA	11
MAKIKALLIO--TH	UNIV OULU	FINLAND	11
SIDHARTH--BG	BM BIRLA SCIENCE CENTER	INDIA	11
ZOU--XW	WUHAN UNIV	CHINA	11
HAVLIN--S	BAR-ILAN UNIV	ISRAEL	10
LAU--KS	CHINESE UNIV HONG KONG	CHINA	10
MENDES--RS	UNIV ESTADUAL MERINGA	BRAZIL	10
TAN--ZJ	WUHAN UNIV	CHINA	10
TSALLIS--C	CTR BRASILEIRO PESQUISAS FIS	BRAZIL	10
BERSHADSKII--A	ICAR	ISRAEL	9
FUJITA--H	HYOGO PREF INST IND RES	JAPAN	9
LAPENNA--V	CNR	ITALY	9
SUN--X	UNIV SCI AND TECH CHINA	CHINA	9
VELTRI--P	UNIV CALABRIA	ITALY	9

Of the eighteen most prolific authors listed in Table 2A, six are from China. In fact, six are from the Far East, two are from the East, two ares from the Mid East, two are from Western Europe, two are from Northern Europe, two are from North America, and two are from South America. Thirteen are from universities, and five are from research institutes.

To determine the trends in this regional mix of prolific authors, the same query was applied to 1991-93 only. Table 2B lists the most prolific authors for 1991-93.

TABLE 2B – MOST PROLIFIC AUTHORS – 1991-93

AUTHOR	INSTITUTION	COUNTRY	# PAPERS
MEAKIN--P	UNIV OSLO	NORWAY	24
STANLEY--HE	BOSTON UNIV	USA	23
HAVLIN--S	BAR-ILAN UNIV	ISRAEL	20
VLAD--MO	KFA JULICH GMBH	GERMANY	19
NAGATANI--T	SHIZUOKA UNIV	JAPAN	18
BALANKIN--AS	FE DZERZHINSKII MIL ACADEMY	RUSSIA	17
PIETRONERO--L	UNIV ROME LA SAPIENZA	ITALY	16
FEDER--J	UNIV OSLO	NORWAY	15

JOSSANG--T	UNIV OSLO	NORWAY	14
SALVAREZZA--RC	NATL UNIV LA PLATA	ARGENTINA	13
ARVIA--AJ	NATL UNIV LA PLATA	ARGENTINA	12
PROCACCIA--I	WEIZMAN INST SCI	ISRAEL	12
SORNETTE--D	UNIV NICE SOPHIA ANTIPOLIS	FRANCE	12
BRAS--RL	MIT	USA	11
GIONA--M	UNIV ROME LA SAPIENZA	ITALY	11
MILOSEVIC--S	UNIV BELGRADE	YUGOSLAVIA	11
MOSOLOV--AB	POLITECNIC TURIN	ITALY	11
SAPOVAL--B	ECOLE POLYTECHNIQUE	FRANCE	11

The regional mix of authors has some major differences from the 2001 results. Of the eighteen most prolific authors listed in Table 2B, one is from the Far East, two are from the Mid East, two are from North America, two are from South America, six are from Western Europe, three are from Northern Europe, and two are from Eastern Europe. Seventeen are from universities, and one is from a research institute.

Only two names were common to both lists, Stanley and Havlin, and they co-author to a reasonable extent. However, some researchers can have an off year for a number of reasons, so individual comparisons over two years, especially two widely separated years, may not be overly important. More important are country comparisons, and maybe institutional comparisons to some extent. These entities integrate over many individuals, and their performance would be more reflective of national policy. In this regard, the aggregate shift of prolific performers from the European countries in 1991-93 to those of the East/ Far East in 2001-02 stands out.

4.1.2 Journals Containing Most Fractals Papers

For 2001-02, there were 1238 different journals represented, with an average of 3.61 papers per journal. The journals containing the most Fractals papers (see Table 3A) had more than an order of magnitude more papers than the average.

TABLE 3A – JOURNALS CONTAINING MOST PAPERS – 2001-02

JOURNAL	# PAPERS
PHYSICAL REVIEW E	314
PHYSICA A	151
CHAOS SOLITONS & FRACTALS	100
PHYSICAL REVIEW LETTERS	91
PHYSICAL REVIEW B	82
FRACTALS-COMPLEX GEOMETRY PATTERNS AND SCALING IN NATURE AND SOCIETY	60

ASTROPHYSICAL JOURNAL	55
PHYSICS LETTERS A	49
PHYSICAL REVIEW D	44
LANGMUIR	38
JOURNAL OF COLLOID AND INTERFACE SCIENCE	37
JOURNAL OF PHYSICS A-MATHEMATICAL AND GENERAL	36
EUROPHYSICS LETTERS	34
ASTRONOMY & ASTROPHYSICS	33
JOURNAL OF FLUID MECHANICS	31
JOURNAL OF STATISTICAL PHYSICS	29
EUROPEAN PHYSICAL JOURNAL B	28
MONTHLY NOTICES OF THE ROYAL ASTRONOMICAL SOCIETY	28
PHYSICS OF PLASMAS	26

Essentially all of the journals are physics, ranging in mission from dedication to fractals (FRACTALS) to sub-branches of physics that include fractal analyses (PHYSICS OF PLASMAS).

To determine the trends in journals containing the most Fractals papers, the results from 1991-93 are examined. Table 3B contains the top twenty journals.

TABLE 3B – JOURNALS CONTAINING MOST PAPERS – 1991-93

JOURNAL	# PAPERS
PHYSICA A	213
PHYSICAL REVIEW A	174
PHYSICAL REVIEW LETTERS	173
PHYSICAL REVIEW B-CONDENSED MATTER	115
PHYSICAL REVIEW E	86
ASTROPHYSICAL JOURNAL	86
PHYSICS LETTERS A	85
JOURNAL OF PHYSICS A-MATHEMATICAL AND GENERAL	77
JOURNAL OF STATISTICAL PHYSICS	73
PHYSICA D	57
EUROPHYSICS LETTERS	52
PHYSICS OF FLUIDS A-FLUID DYNAMICS	50
PHYSICS LETTERS B	50
PHYSICAL REVIEW D	44
JOURNAL OF PHYSICS-CONDENSED MATTER	43
GEOPHYSICAL RESEARCH LETTERS	40
JOURNAL OF CHEMICAL PHYSICS	35

JOURNAL OF NON-CRYSTALLINE SOLIDS	33
JOURNAL OF THE PHYSICAL SOCIETY OF JAPAN	32
JOURNAL OF FLUID MECHANICS	32

While the most prolific authors could be expected to change over a decade, for a number of reasons, the most prolific journals should be more stable. Comparison of Tables 3A and 3B shows this to be true. Of the twenty most prolific journals, eleven are in common.

The journals in the top twenty in 1991-93 that were not included in the top twenty from 2001-02 tended to be the more traditional discipline-oriented physics journals (JOURNAL OF PHYSICS-CONDENSED MATTER, GEOPHYSICAL RESEARCH LETTERS, JOURNAL OF CHEMICAL PHYSICS, JOURNAL OF NON-CRYSTALLINE SOLIDS, PHYSICS OF FLUIDS-FLUID DYNAMICS, ETC). The journals in the top twenty in 2001-02 that were not included in the top twenty from 1991-93 tended to be the more generic non-discipline oriented physics journals (FRACTALS, CHAOS SOLITONS AND FRACTALS, LANGMUIR, JOURNAL OF COLLOID AND INTERFACE SCIENCE, ETC).

4.1.3 Institutions Producing Most Fractals Papers

A similar process was used to develop a frequency count of institutional address appearances. It should be noted that many different organizational components may be included under the single organizational heading (e.g., Harvard Univ could include the Chemistry Department, Biology Department, Physics Department, etc.). Identifying the higher level institutions is instrumental for these DT studies. Once they have been identified through bibliometric analysis, subsequent measures may be taken (if desired) to identify particular departments within an institution.

TABLE 4A – PROLIFIC INSTITUTIONS – 2001-02

INSTITUTION	COUNTRY	# PAPERS
RUSSIAN ACAD SCI	RUSSIA	135
CHINESE ACAD SCI	CHINA	65
MIT	USA	54
UNIV CAMBRIDGE	UK	47
UNIV PARIS	FRANCE	46
CNRS	FRANCE	43
BOSTON UNIV	USA	42
CNR	ITALY	40
UNIV SCI & TECHNOL CHINA	CHINA	38
UNIV CALIF LOS ANGELES	USA	37
UNIV TOKYO	JAPAN	35
UNIV CALIF BERKELEY	USA	34
HARVARD UNIV	USA	31

KYOTO UNIV	JAPAN	31
ECOLE POLYTECH	FRANCE	31
CORNELL UNIV	USA	29
POLISH ACAD SCI	POLAND	29
CHINESE UNIV HONG KONG	CHINA	28
TSING HUA UNIV	CHINA	28
PENN STATE UNIV	USA	28

For 2001, of the twenty most prolific institutions, seven are from the USA, five are from Western Europe, six are from Asia, and two are from Eastern Europe. Fifteen are universities, and the remaining institutions are research institutes.

To determine the trends in institutions containing the most Fractals papers, the results from 1991-93 were examined. Table 4B contains the top twenty institutions.

TABLE 4B – PROLIFIC INSTITUTIONS – 1991-93

INSTITUTION	COUNTRY	# PAPERS
RUSSIAN ACAD SCI	RUSSIA	110
TEL AVIV UNIV	ISRAEL	51
IBM CORP	USA	49
CORNELL UNIV	USA	48
NASA	USA	47
KFA JULICH GMBH	GERMANY	47
MIT	USA	47
UNIV CHICAGO	USA	45
UNIV CAMBRIDGE	UK	45
UNIV ILLINOIS	USA	45
ACAD SINICA	TAIWAN/ CHINA	44
UNIV MARYLAND	USA	44
UNIV TOKYO	JAPAN	42
UNIV CALIF SAN DIEGO	USA	40
UNIV ROME LA SAPIENZA	ITALY	39
UNIV CALIF BERKELEY	USA	38
BOSTON UNIV	USA	35
UNIV MICHIGAN	USA	34
PRINCETON UNIV	USA	34
ECOLE POLYTECH	FRANCE	33

Of the twenty most prolific institutions in 1991-93, twelve are from the USA, four are from Western Europe, one is from Eastern Europe, one is from the mid East, and one is from Taiwan/ China. The major shift is substitution of Asian institutions for USA institutions. In addition, sixteen institutions

are universities, four are research institutes, and one is industrial research.

4.1.4 Countries Producing Most Fractals Papers

There are 90 different countries listed in the results for 2001-02. The country bibliometric results are summarized in Table 5A. The dominance of a handful of countries is clearly evident.

TABLE 5A – PROLIFIC COUNTRIES – 2001-02

COUNTRY	# PAPERS
USA	1223
FRANCE	464
PEOPLES R CHINA	398
GERMANY	373
JAPAN	340
RUSSIA	329
ENGLAND	299
ITALY	277
SPAIN	172
CANADA	167
BRAZIL	156
POLAND	137
INDIA	112
ISRAEL	112
AUSTRALIA	110
NETHERLANDS	84
GREECE	71
TAIWAN	69
SWEDEN	68
SOUTH KOREA	63
ARGENTINA	60
SWITZERLAND	57
HUNGARY	56
BELGIUM	51
FINLAND	49
UKRAINE	47
DENMARK	43
SCOTLAND	42
MEXICO	41
AUSTRIA	37
NEW ZEALAND	29

There appear to be two dominant groupings. The first group is the USA. It has half as many papers as the members of the second group combined, France, People's Republic of China, Germany, Japan, Russia, England, and italy.

To determine the trends in countries containing the most Fractals papers, the results from 1991-93 were examined. Table 5B summarizes results from the top twenty countries.

TABLE 5B – PROLIFIC COUNTRIES – 1991-93

COUNTRY	# PAPERS
USA	1596
FRANCE	475
GERMANY	442
JAPAN	331
ENGLAND	257
ITALY	244
CANADA	226
USSR	202
PEOPLES R CHINA	152
ISRAEL	132
INDIA	117
RUSSIA	113
SPAIN	94
NETHERLANDS	88
SWITZERLAND	83
POLAND	75
AUSTRALIA	70
NORWAY	53
DENMARK	48
SWEDEN	43
BRAZIL	40
BELGIUM	38
GREECE	38
SCOTLAND	35
HUNGARY	31
ARGENTINA	30
AUSTRIA	29
TAIWAN	27
CZECHOSLOVAKIA	26
SOUTH KOREA	25

The countries of the Former Soviet Union had 337 papers in aggregate in the 1991-93 time frame, and 402 in aggregate in the 2001-02 time frame. The major shift is the increased ranking of People's Republic of China from 9th in 1991-93 to third (or fourth, depending on whether the Former Soviet Union is aggregated, or not) in 2001-02, and the concomitant increase in numbers of papers from 152 to 399.

Figure 1 contains a co-occurrence matrix of the top 15 countries for 2001-02. In terms of absolute numbers of co-authored papers, the USA major partners are France, Germany, Canada, England, Japan, and italy. Interestingly, the USA is China's dominant major partner, having 2.5 times the number of co-authored papers with China (30) as China's next larger partner, Germany (12). Overall, countries in similar geographical regions tend to co-publish substantially, the US being a moderate exception.

Figure 2 contains a co-occurrence matrix of the top 15 countries for 1991-93. In terms of absolute numbers of co-authored papers, the USA major partners are France, Germany, Israel, Italy, and Canada. Again, the USA was China's major partner, having slightly more co-authored papers with China (10) than China's next larger partners, Germany (8) and Italy (7).

FIGURE 1 – COUNTRY CO-OCCURRENCE MATRIX – 2001-02

Items	Australia	Brazil	Canada	Englnd	France	Germany	India	Israel	Italy	Japan	Peoples R China	Poland	Russia	Spain	USA
AUSTRALIA	110	0	4	8	5	4	0	1	1	7	8	0	2	2	25
BRAZIL	0	156	2	4	11	3	2	2	8	0	1	0	5	5	26
CANADA	4	2	167	8	17	8	2	0	5	3	7	0	4	3	49
ENGLAND	8	4	8	299	16	13	3	4	16	9	1	5	10	11	46
FRANCE	5	11	17	16	464	21	8	5	36	15	5	6	17	14	72
GERMANY	4	3	8	13	21	373	6	9	18	13	12	18	27	8	50
INDIA	0	2	2	3	8	6	112	0	2	2	0	0	4	1	10
ISRA	1	2	0	4	5	9	0	112	3	4	1	2	5	0	30

EL															
ITALY	1	8	5	16	36	18	2	3	277	2	1	4	11	12	37
JAPAN	7	0	3	9	15	13	2	4	2	340	7	3	15	4	32
PEOPLES R CHINA	8	1	7	1	5	12	0	1	1	7	398	0	1	1	30
POLAND	0	0	0	5	6	18	0	2	4	3	0	137	0	2	14
RUSSIA	2	5	4	10	17	27	4	5	11	15	1	0	329	6	21
SPAIN	2	5	3	11	14	8	1	0	12	4	1	2	6	172	19
USA	25	26	49	46	72	50	10	30	37	32	30	14	21	19	1223

FIGURE 2 – COUNTRY CO-OCCURRENCE MATRIX – 1991-93

COUNTRY	CANADA	ENGLAND	FRANCE	GERMANY	INDIA	ISRAEL	ITALY	JAPAN	NETHERLANDS	PEOPLES R CHINA	RUSSIA	SPAIN	SWITZERLAND	USA	USSR
CANADA	226	3	25	4	2	3	2	3	2	1	0	3	0	32	3
ENGLAND	3	257	8	6	4	1	7	4	3	2	3	7	2	25	3
FRANCE	25	8	475	23	0	12	23	2	10	0	3	5	10	79	5
GERMANY	4	6	23	442	1	15	11	10	5	8	7	1	19	54	5
INDIA	2	4	0	1	117	0	0	2	0	0	0	0	0	7	0
ISRAEL	3	1	12	15	0	132	3	0	1	1	6	2	9	44	3
ITALY	2	7	23	11	0	3	244	1	4	7	4	0	5	34	2

JAPAN	3	4	2	10	2	0	1	331	4	4	1	4	1	26	0
NETHERLANDS	2	3	10	5	0	1	4	4	88	3	1	1	1	17	0
PEOPLES R CHINA	1	2	0	8	0	1	7	4	3	152	1	0	0	10	0
RUSSIA	0	3	3	7	0	6	4	1	1	1	113	0	0	6	1
SPAIN	3	7	5	1	0	2	0	4	1	0	0	94	1	12	0
SWITZERLAND	0	2	10	19	0	9	5	1	1	0	0	1	83	8	0
USA	32	25	79	54	7	44	34	26	17	10	6	12	8	1596	16
USSR	3	3	5	5	0	3	2	0	0	0	1	0	0	16	202

4.2 Citation Statistics on Authors, Papers, and Journals

The second group of metrics presented is counts of citations to papers published by different entities. While citations are ordinarily used as impact or quality metrics [Garfield, 1985], much caution needs to be exercised in their frequency count interpretation, since there are numerous reasons why authors cite or do not cite particular papers [Kostoff, 1998; MacRoberts and MacRoberts, 1996].

The citations in all the retrieved SCI papers were aggregated, the authors, specific papers, years, journals, and countries cited most frequently were identified, and were presented in order of decreasing frequency. A small percentage of any of these categories received large numbers of citations. From the citation year results, the most recent papers tended to be the most highly cited. This reflected rapidly evolving fields of research.

4.2.1 Most Cited Authors

The most highly cited authors from the 2001-02 database are listed in Table 6. Many of these highly cited authors worked at a variety of institutions throughout their careers, and the institution listed was their residence when some of the highly cited work was performed.

TABLE 6 – MOST CITED AUTHORS – 2001-02

(cited by other papers in this database only)

AUTHOR	INSTITUTION	COUNTRY	# CITES
MANDELBROT BB	IBM	USA	1172
BAK P	BROOKHAVEN NATL LAB	USA	614
FALCONER KJ	UNIV BRISTOL	UK	331
MEAKIN P	DUPONT	USA	291
TSALLIS C	CTR BRASILEIRO PESQUISAS FIS	BRAZIL	290

GRASSBERGER P	UNIV WUPPERTAL	GERMANY	221
FEDER J	UNIV OSLO	NORWAY	203
WITTEN TA	EXXON RES & ENG	USA	187
HALSEY TC	UNIV CHICAGO	USA	170
FRISCH U	CNRS	FRANCE	158
TURCOTTE DL	CORNELL UNIV	USA	158
VICSEK T	EOTVOS LORAND UNIV	HUNGARY	157
AVNIR D	HEBREW UNIV	ISRAEL	156
METZLER R	UNIV ULM	GERMANY	146
KOLMOGOROV AN	LOMONOSOV STATE UNIV	RUSSIA	145
STAUFFER D	KFA JULICH GMBH	GERMANY	144
PFEIFER P	UNIV BIELEFELD	GERMANY	142
ELNASCHIE MS	CORNELL UNIV	USA	136
BENZI R	UNIV ROME TOR VERGATA	ITALY	131
ZASLAVSKY GM	ACAD SCI USSR	RUSSIA	128

Of the twenty most cited authors, seven are from the USA, eight from Western Europe, three from Eastern Europe, one from the Mid East, and one from Latin America. This is a far different distribution from the most prolific authors of 2001-02, where eight of nineteen were from the East/ Far East. This distribution of most cited authors more closely resembles the distribution of most prolific authors from 1991-93, where only one was from the Far East.

There are a number of potential reasons for this regional difference between most prolific and cited authors in 2001-02. The most prolific may not be the highest quality, or many of the most prolific authors could be relatively recent, and insufficient time has elapsed for their citations to accumulate. In another three or four years, when the papers from present-day authors have accumulated sufficient citations, firmer conclusions about quality can be drawn.

The lists of nineteen most prolific authors from 2001-02 and twenty most highly cited authors only had two names in common (ZASLAVSKY, TSALLIS). This phenomenon of minimal intersection has been observed in all other text mining studies performed by the first author. In addition, the lists of eighteen most prolific authors from 1991-93 and twenty most highly cited authors only had one name in common (MEAKIN). This disconnect is more disconcerting, since adequate time has accumulated in the past decade for these 1991-93 papers to gather citations. A more detailed examination of all these papers would be required to resolve this dilemma, and that is beyond the scope of the present paper.

Twelve of the most cited authors' institutions are universities, five are government-sponsored research laboratories, and three are private companies.

The citation data for authors and journals represents citations generated only by the specific records extracted from the SCI database for this study. It does not represent all the citations received by the references in those records; these references in the database records could have been cited additionally by papers in other technical disciplines.

4.2.2 Most Cited Papers

The most highly cited documents from the 2001-02 database are listed in Table 7.

TABLE 7 – MOST CITED DOCUMENTS
(total citations listed in SCI)

DOCUMENT	# CITES
MANDELBROT BB, 1982, FRACTAL GEOMETRY NAT	5107
FRACTAL GEOMETRY OF NATURE	
BAK P, 1987, PHYS REV LETT, V59, P381	1731
SELF-ORGANIZED CRITICALITY	
MANDELBROT BB, 1983, FRACTAL GEOMETRY NAT	2942
FRACTAL GEOMETRY OF NATURE	
FEDER J, 1988, FRACTALS	2057
GENERAL FRACTALS	
BAK P, 1988, PHYS REV A, V38, P364	1279
SELF-ORGANIZED CRITICALITY	
WITTEN TA, 1981, PHYS REV LETT, V47, P1400	2181
DIFFUSION-LIMITED AGGREGATION	
HALSEY TC, 1986, PHYS REV A, V33, P1141	1505
FRACTAL MEASURES AND THEIR SINGULARITIES	
MANDELBROT BB, 1968, SIAM REV, V10, P422	876
FRACTIONAL BROWNIAN MOTIONS AND NOISES	
FALCONER K, 1990, FRACTAL GEOMETRY MAT	415
MATHEMATICAL FOUNDATIONS OF FRACTAL GEOMETRY	
TSALLIS C, 1988, J STAT PHYS, V52, P479	641
GENERALIZATION OF BOLTZMANN-GIBBS STATISTICS	
VICSEK T, 1992, FRACTAL GROWTH PHENO	478
FRACTAL GROWTH PHENOMENA	
LELAND WE, 1994, IEEE ACM T NETWORK, V2, P1	371
SELF-SIMILAR NATURE OF ETHERNET TRAFFIC	
BARABASI AL, 1995, FRACTAL CONCEPTS SUR	1026
FRACTAL CONCEPTS IN SURFACE GROWTH	
HAVLIN S, 1987, ADV PHYS, V36, P695	918
DIFFUSION IN DISORDERED MEDIA	
BOUCHAUD JP, 1990, PHYS REP, V195, P127	702
ANOMALOUS DIFFUSION IN DISORDERED MEDIA	
HENTSCHEL HGE, 1983, PHYSICA D, V8, P435	920
GENERALIZED DIMENSIONS OF FRACTALS AND STRANGE ATTRACTORS	

MANDELBROT BB, 1974, J FLUID MECH, V62, P331 686
INTERMITTENT TURBULENCE IN SELF-SIMILAR CASCADES
HUTCHINSON JE, 1981, INDIANA U MATH J, V30, P713 470
FRACTALS AND SELF SIMILARITY
MANDELBROT BB, 1984, NATURE, V308, P721 547
FRACTAL CHARACTER OF FRACTURE SURFACES OF METALS
SAMORODNITSKY G, 1994, STABLE NONGAUSSIAN R 393
STABLE NONGAUSSIAN RANDOM PROCESSES

The theme of each paper is shown in italics on the line after the paper listing. The order of paper listings is by number of citations by other papers in the extracted database analyzed. The total number of citations from the SCI paper listing, a more accurate measure of total impact, is shown in the last column on the right.

Physical Review Letters contains the most papers, two out of the twenty listed. There are a substantial number of books listed (about 1/3), noticeably larger than in other topics studied. Reasons for this are unclear.

Most of the journals are fundamental science journals, and most of the topics have a fundamental science theme. The majority of the papers are from the 1980s, with seven from the 1990s, and one paper from 1968.

There are three Fractals books in the top twenty cited documents. Several of the most cited papers are review articles. Otherwise the most cited papers appear in physics journals focused on fractal motions, growth of fractal shapes, fractal noise, and fractal measures.

The list of most cited includes general books by Mandelbrot, and Feder, covering many fractals topics. The paper of Bak is a theory called "self-organized criticality" of why natural objects can wind up as fractal shapes. The other themes cited are mostly **fractal motions** or fractal random processes (mostly generalizations on Brownian motion but with different scaling properties), or random walks called Levy flights with jump sizes on all scales. Another theme is **fractal noise**, i.e., fluctuations that are wild and fractal. A third theme is **fractal growth.** How can particle or clusters of particle aggregate into fractal shapes. How can fractal biological shapes, like the branching in the lung, grow, or how can shapes break down (dissolve, weather etc) leaving fractal shapes behind. A fourth theme is **fractal measures**. How can fractal objects be characterized? One way is with a fractal dimension. Another is to treat the fractal dimension as a variable and get a distribution of fractal dimensions to describe fractal objects. It should be noted that fractals are a condition that can arise within physical theories, to obtain fractal motions or fractal shapes under certain conditions.

Thus, the major intellectual emphasis of cutting edge Fractals research, as evidenced by the most cited papers, is well aligned with the intellectual heritage and performance emphasis, as will be evidenced by the clustering approaches presented later.

4.2.3. Most Cited Journals

The most highly cited journals from the 2001-02 database are listed in Table 8.

TABLE 8 – MOST CITED JOURNALS
(cited by other papers in this database only)

JOURNAL	# CITES
PHYS REV LETT	7048
PHYS REV E	3602
ASTROPHYS J	3068
PHYS REV B	2395
NATURE	1754
PHYS REV A	1609
PHYSICA A	1335
J FLUID MECH	1208
J PHYS A-MATH GEN	1122
J CHEM PHYS	1061
SCIENCE	1001
PHYS REV D	992
PHYSICA D	976
MON NOT R ASTRON SOC	875
PHYS LETT A	851
J COLLOID INTERF SCI	847
ASTRON ASTROPHYS	782
J STAT PHYS	753
PHYS FLUIDS	686
WATER RESOUR RES	665

Three main groups of cited journals may be discerned. PHYS REV LETT received almost as many cites as the three journals in the next group (PHYS REV E, ASTROPHYS J, PHYS REV B), or even the first five journals in the following group (NATURE, PHYS REV A, PHYSICA A, J FLUID MECH, J PHYS A, J CHEM PHYS, SCIENCE). PHYS REV LETT emphasizes rapid publication of 'hot' topics, and would therefore tend to establish primacy in an emerging field. Since one aspect of citations is identifying the original literature of a new topic, a credible journal with these characteristics would tend to receive large numbers of citations.

Unlike the relatively disjoint relationship between most prolific authors in 2001-02 and most cited authors in 2002-02, the relationship between most prolific journals in 2001-02 and most cited journals in 2001-02 is much closer. Thirteen of the twenty most highly cited journals in 2001-02 are also on the list of nineteen most prolific journals in 2001-02. The more applied journals on the most prolific list for 2001-02 are replaced by the more fundamental journals on the most cited list for 2001-02. In addition, thirteen of the twenty most highly cited journals in 1991-93 are also on the list of twenty most prolific journals in 1991-93. In fact, all of the top ten most prolific journals from

1991-93 are on the list of twenty most highly cited journals of 2001-02. The more applied journals on the most prolific list for 1991-93 are replaced by the more fundamental journals on the most cited list for 2001-02.

The authors end this bibliometrics section by recommending that the reader interested in researching the topical field of interest would be well-advised to, first, obtain the highly-cited papers listed and, second, peruse those sources that are highly cited and/or contain large numbers of recently published papers.

4. DISCUSSION AND CONCLUSIONS

The author bibliometrics comparison of 2001-02 and 1991-93 showed a substantial regional shift from Europe to Asia over the past decade, and a more moderate shift from universities to research institutes. The regional shift has been noted in other recent text mining studies, and reflects to a large extent the increase in publications output reported by China.

The journal bibliometrics reflected a stronger concentration of Fractals publications in physics journals, with a slight shift in emphasis over the past decade from the more traditional discipline-oriented physics journals to the more generic non-discipline-oriented physics journals. The institutional bibliometrics reflected the shift from European to Asian institutions over the past decade noted under the author bibliometrics, although the shift from universities to research institutes noted under the author bibliometrics was not evident in the institutional bibliometrics results. The country bibliometrics trend over the past decade reflected the regional trend noted above. In addition, US co-authorship with China tripled over the past decade, while China's co-authorship with its second largest partner in 1991-93 (Germany) increased by 50%, and China's co-authorship woth its third largest partner in 1991-93 (Italy) decreased by 80%.

The most cited authors from 2001-02 have a far different regional distribution from that of the most prolific authors for the same time period. The regional distribution of most cited authors for 2001-02 resembles more closely the distribution of most prolific authors from 1991-93. More disconcerting, the list of eighteen most prolific authors from 1991-93 and twenty most highly cited authors had only one name in common. This raises the issue of whether an intrinsic incompatibility exists between producing large numbers of papers and producing seminal papers.

The most cited document is a twenty year old book by Mandelbrot. This is the first time that a book has been the most cited document in the first author's text mining studies. In fact, the ten most highly cited documents were published more than a decade ago! The focus of these documents is on Fractals fundamentals. The highly cited documents in the top twenty list that were published in the mid-1990s reflect the Fractals applications as much, or more, than intrinsic Fractals fundamentals. These observations suggest a study area whose intrinsic fundamental advances peaked about a decade or two ago, and which has now evolved into an applications focus. This data-based conclusion correlates well with the intuitive conclusion one draws when reading thousands of Fractals Abstracts from the last decade.

Finally, the most cited journal (Physical Review Letters) emphasizes rapid publication of 'hot' topics, and would therefore tend to establish primacy in an emerging field. Since one aspect of citations is identifying the original literature of a new topic, a credible journal with these characteristics would tend to receive large numbers of citations. This result should send a clear message to the editors of traditional journals, whose present practices involve long review and publication times, but who wish to improve their Journal Impact Factors.

5. REFERENCES FOR APPENDIX 7-G

Cutting, D. R., Karger, D. R, Pedersen, J. O. and Tukey, J. W. (1992). Scatter/Gather: A cluster-based approach to browsing large document collections. In *Proceedings of the 15th International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR'92). 318-329.

Garfield, E. [1985] History of citation indexes for chemistry - a brief review. JCICS. 25(3). 170-174.

Guha, S., Rastogi, R. and Shim, K. (1998). CURE: An efficient clustering algorithm for large databases. In *Proceedings of the ACM-SIGMOD 1998 International Conference on Management of Data* (SIGMOD'98). 73-84.

Hearst, M. A. (1998). The use of categories and clusters in information access interfaces. In T. Strzalkowski (ed.), Natural Language Information Retrieval. Kluwer Academic Publishers.

Karypis, G.; Han, E.-H.; and Kumar, V. (1999).. Chameleon: A hierarchical clustering algorithm using dynamic modeling. IEEE Computer: Special Issue on Data Analysis and Mining 32(8). 68--75.

Kostoff, R. N. [2000]. The underpublishing of science and technology results. The Scientist. 14(9). 6-6. 1 May.

Kostoff, R. N. [1993]. Database Tomography for technical intelligence. Competitive Intelligence Review. 4(1). 38-43.

Kostoff, R. N. [1998]. The use and misuse of citation analysis in research evaluation. Scientometrics. 43(1). 27-43. September.

Kostoff, R. N. [1999]. Science and technology innovation. Technovation. 19(10). 593-604.

Kostoff, R. N. et al [1995]. System and method for Database Tomography. U.S. Patent Number 5440481.

Kostoff, R. N., Braun, T., Schubert, A., Toothman, D. R., and Humenik, J. A. [2000]. Fullerene roadmaps using bibliometrics and Database Tomography. Journal of Chemical Information and Computer Science. 40(1). 19-39. Jan-Feb.

Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. [1997]. Database Tomography for information retrieval. Journal of Information Science. 23(4). 301-311.

Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. [1998]. Database Tomography for technical intelligence: a roadmap of the near-earth space science and technology literature. Information Processing and Management. 34(1). 69-85.

Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. [1999]. Hypersonic and supersonic flow roadmaps using bibliometrics and Database Tomography. JASIS. 50(5). 427-447. 15 April.

Kostoff, R. N., Eberhart, H. J., Toothman, D. R., and Pellenbarg, R. [1997]. Database Tomography for technical intelligence: comparative roadmaps of the research impact assessment literature and the Journal of the American Chemical Society, Scientometrics. 40(1). 103-138.

Kostoff, R. N., Green, K. A., Toothman, D. R., and Humenik, J. A. [2000]. Database Tomography applied to an aircraft science and technology investment strategy. Journal of Aircraft. 37(4). 727-730. July-August.

Kostoff, R. N., Toothman, D. R., Eberhart, H. J., and Humenik, J. A. [2001]. Text mining using Database Tomography and bibliometrics: a review. Technology Forecasting and Social Change. 68(3). 223-253.

Kostoff, R. N., Tshiteya, R., Pfeil, K. M., and Humenik, J. A. [2002]. Electrochemical power source roadmaps using bibliometrics and Database Tomography. Journal of Power Sources. 110(1). 163-176.

Kostoff, R.N. [1994]. Database Tomography: origins and applications. Competitive Intelligence Review. Special Issue on Technology, 5:1. 48-55.

MacRoberts, M. H., and MacRoberts, B. R. [1996]. Problems of citation analysis. Scientometrics. 36(3). 435-444. July-August.

Rasmussen, E. *Clustering Algorithms*. In W. B. Frakes and R. Baeza-Yates (eds.) Information Retrieval Data Structures and Algorithms, Prentice Hall, N. J., 1992.

SCI. [2002]. Science Citation Index. Institute for Scientific Information. Phila., PA.

Steinbach, M.; Karypis, G.; and Kumar, V. (2000). . A comparison of document clustering techniques. Technical Report #00--034. Department of Computer Science and Engineering. University of Minnesota.

Willet, P. (1988). Recent trends in hierarchical document clustering: A critical review. *Information Processing and Management*, 24:577-597.

Zamir, O. and Etzioni, O. (1998). Web document clustering: A feasibility demonstration. In: *Proceedings of the 19th International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR'98). 46-54.

6. APPENDIX TO APPENDIX 7-G– GREEDY STRING TILING (GST) CLUSTERING

Greedy String Tiling clustering is a method of grouping text or text character documents (files) by similarity. All documents to be grouped are placed in a database. Each pair of documents is compared by GST, an algorithm originally used to detect plagiarism [Wise, 1993; Prechelt et al, 2002], and a similarity score is assigned to the pair. Then, hierarchical aggregation clustering (Rasmussen, 1992; Steinbach, 2000) is performed on all the documents, using the similarity score for group assignment.

Greedy String Tiling computes the similarity of a pair of documents in two phases. First, all documents to be compared are parsed, and converted into token strings (words or characters). Second, these token strings are compared in pairs for determining the similarity of each pair. During each comparison, the GST algorithm attempts to cover one token string (document) with sub-strings ('tiles') taken from the other string. These sub-strings are not allowed to overlap, resulting in a one to one mapping of tokens. The attribute "greedy" stems from the fact that the algorithm matches the longest sub-strings first to find the most relevant sequences first.

A number of similarity metrics can be defined once the tiling is completed. One similarity metric is the percentage of both token strings that is covered. Another similarity metric is the absolute number of shared tokens. A third similarity metric is the mutual information index. Depending on the purpose of the matching, additional weightings can be used for the similarity matrix to increase the ranking precision. For example, if plagiarism is one study objective, additional weighting could be given to shared string length. All similarity metrics have positive and negative features, and the choice of metric is somewhat influenced by the study objectives and the structure of the database.

Once the document similarity matrix has been generated, myriad clustering techniques can be used to produce a classification scheme (taxonomy). In the present study, multi-link hierarchical aggregation was used. Three clustering variants were actually generated, although the extension to other clustering schemes is straight-forward. Single-link, average-link, and complete-link variants are implemented. The variants differ in how the decision of merging to clusters is made. Single-link requires that the similarity of at least two documents is higher than a certain threshold, while complete-link requires that the similarity between all documents in both clusters beeing higher than a threshold. Average-link requires that the average pair-wise similarity between the documents of both clusters exceed the threshold. For the present study, average-link appeared to give good results, and was the clustering method used.

APPENDIX 7-H

SCIENCE AND TECHNOLOGY TEXT MINING: CITATION MINING OF DYNAMIC GRANULAR SYSTEMS (Kostoff et al, 2001b; Del Rio et al, 2002)

I. ABSTRACT

Background: Research sponsors, evaluators, managers, and performers have strong motivations in insuring that their research products reach the intended audience. Further, it is important to understand the infrastructure characteristics of the specific audience reached (names, organizations, countries). Because of the many direct and indirect pathways through which fundamental research can impact applications, identifying the user audience and the research impacts can be very complex and time consuming.

Objective: The purpose of this appendix is to describe a novel approach for identifying the pathways through which research can impact other research, technology development, and applications, and to identify the technical and infrastructure characteristics of the user population.

Approach: Citation Mining, a novel literature-based approach that integrates citation bibliometrics with text mining (extraction of useful information from text), was developed to identify the user community and its characteristics. Citation Mining starts with a group of core papers whose impact is to be examined, retrieves the papers that cite these core papers, and then analyzes the bibliometrics characteristics of the citing papers as well as their linguistic and thematic characteristics. The Science Citation Index is used as the source database for the core and citing papers, since its citation-based structure enables the capability to perform citation studies easily. The user community is characterized by the papers in the SCI that 1) cite the original research papers, and 2) cite the succeeding generations of these papers as well. Text mining is performed on the citing papers to identify the technical areas impacted by the research, the relationships among these technical areas, and relationships among the technical areas and the infrastructure (authors, journals, organizations). A key component of text mining, concept clustering, was used to provide both a taxonomy of the citing papers' technical themes and further technical insights based on theme relationships arising from the grouping process. Bibliometrics is performed on the citing papers to profile the user characteristics. In a specific example, Citation Mining is applied to the ~300 first generation citing papers of a fundamental physics paper on the dynamics of vibrating sand-piles.

Results: Most of the ~300 citing papers were basic research whose main themes were aligned with those of the cited paper. There were three main findings from a temporal analysis of the citing papers. First, the tail of total annual citation counts is very long, and shows little sign of abating. This is one characteristic feature of a seminal paper.

Second, the fraction of extra-discipline basic research citing papers to total citing papers ranges from about 15-25% annually, with no latency period evident. This lag-free extra-disciplinary diffusion may have been due to the combination of intrinsic broad-based applicability of the subject matter and publication of the paper in a high-circulation science journal with very broad-based readership. The text mining alone identified the intra-discipline applications and extra-discipline impacts and applications; this was confirmed by detailed reading of the ~300 abstracts.

Third, a four-year latency period exists prior to the emergence of the higher development category citing papers. This correlates with the results from the bibliometrics component. From the present study, it is not possible to differentiate the reasons for this important result. The latency could have been due to the inability of the technology community to *immediately* recognize the potential applications of the science. Or, it could have been due to the information remaining in the basic research journals, and not reaching the applications community. Or, the time that an application needs to be developed in this discipline is of the order of four years. Thus, the basic science publication feature that may have contributed heavily to extra-discipline citations may also have limited higher development category citations for the latency period.

Conclusions: The combination of citation bibliometrics and text mining provides a synergy unavailable with each approach taken independently. Furthermore, text mining is a REQUIREMENT for a feasible comprehensive research impact determination. The integrated multi-generation citation analysis required for broad research impact determination of highly cited papers will produce thousands or tens or hundreds of thousands of citing paper Abstracts. Text mining allows the impacts of research on advanced development categories and/ or extra-discipline categories to be obtained without having to read all these citing paper Abstracts. The multi-field bibliometrics provide multiple documented perspectives on the users of the research, and indicate whether the documented audience reached is the desired target audience.

II. BACKGROUND

Identification of diverse research impacts is important to research managers, evaluators, and sponsors, and ultimately to performers. They are interested in the types of people and organizations citing the research outputs, and whether the citing audience is the target audience. Also, they are interested in whether the development categories and technical disciplines impacted by the research outputs are the desired targets. Since fundamental research can evolve along myriad paths, tracking diverse impacts becomes complex.

Presently, there are three generic approaches to tracking the impact of research: qualitative, semi-quantitative, and quantitative (Kostoff, 1997). Qualitative approaches are variants of peer review. Panels of experts are assembled, and impacts are identified based on the participants' knowledge, and usually personal experiences. The results are usually long on subjectivity, and short on independent documentation.

Semi-quantitative approaches are probably the most widely used for tracking impact (Kostoff, 1994). They include retrospective studies such as Hindsight (DOD, 1969) and Traces (IITRI, 1968), and

various types of research sponsor accomplishment books such as those from DOE (DOE, 1983, 1986) and DARPA (IDA, 1991). A detailed treatment is contained in (Kostoff, 1997). Semi-quantitative approaches tend to be grounded in corporate memory of the participants, although some studies (Narin, 1989) follow the citation trail for supplementation. Their focus is detailed examination of a few high impact cases, rather than a wide-scale identification of many diverse impacts. As in the peer review approach, semi-quantitative approaches also have a high subjective component.

Quantitative approaches are also widely used for impact tracking (Kostoff, 1994, 1997). They tend to be divided between economic methods such as cost-benefit and internal rate-of-return (Averch, 1994; Tassey, 1999), and S&T indicators such as publications and patents (Narin, 1994), and their citations. They are the most objective of the three generic methods for tracking and quantifying research impact. However, many assumptions related to cost and benefit allocation are required for the economic studies (Kostoff, 1997). Additionally, many assumptions are required to accept correlation between numerical indicator values and degree of impact.

Thus, one of the gaps of all these impact tracking techniques is objective identification of the full scope of impacts produced by the research. These impacts include both the directly identifiable research impacts and the indirect impacts. For that fraction of performed research that is documented in the technical literature, tracking of direct and indirect research impacts on intermediate and final useful products becomes possible through tracking of generations of citations to the original research. If this wide scale impact information were obtained, then the in-depth studies performed by the semi-quantitative methods could cover an expanded range, or the roadmap of impacts could be presented as a self-contained valuable finding.

Even though the premier database for citation tracking, the Science Citation Index (SCI), contains a number of data fields abstracted from the full-text published papers, past citation-based studies using the SCI have focused almost exclusively on citation counts as an impact metric. Reviews of these citation studies can be found in (De Solla Price, 1986; Braun, 1987; Egghe, 1990). The potential impact of citation counts on decision-making is small, since the information content of citation counts alone is very limited. However, these citing records contain a wealth of information in their two main categories of diverse fields. The non-free-text fields, such as Author, Journal, Address, etc, describe the infrastructure characteristics of the citing community. The free-text fields, such as Title, Abstract, and Keywords (Keywords is not strictly a free-text field, but has sufficient technical characteristics to be included in this grouping), describe the technical characteristics of the impacted research, development, and applications areas.

Use of the SCI non-free-text fields for citing paper bibliometric analysis has been published on a very sporadic basis, and typically only for one or two data fields (Steele, 2000; Herring, 1999; Davidse, 1997). The focus of most of these studies has been on relating citations or citation rates to the few field variables examined. There do not appear to have been any citation studies performed for the specific purpose of user population profiling, where <u>many</u> of the available fields are examined in an integrated manner.

Recently, scientists have addressed the problem of citation in scientific research from a different perspective: looking for a topological description of citations (Bilke and Peterson, 2001), from power laws in citation networks (Redner, 1998), or power laws in number of cites received by journals according with their number of published papers (Katz, 2000) and finally trying to find some universal classes (Amaral et al. 2001). To overcome the limitations of these techniques, a phenomenological approach to deal with the information available and obtain a more detailed description of this complex system is presented in this paper.

Use of the SCI free-text fields for coupled trans-citation citing paper/ cited paper text mining analysis has not been published, although text mining studies of SCI and other database free-text fields have been reported (e.g., Kostoff et al, 2000a, 2000b, 2002, 2003).

III. OBJECTIVES

The objectives of the present paper are:

i) Demonstrate the feasibility of tracking the myriad impacts of research on other research, development, and applications, using the technical literature.
ii) Demonstrate the feasibility of identifying a broad range of research product user characteristics, using the technical literature.
iii) Relate thematic characteristics of citing papers to their cited papers.

IV. APPROACH

The present paper describes a novel process, Citation Mining (Kostoff et al, 2001a, Del Rio at al, 2002), that uses the best features of citation bibliometrics and text mining to track and document the impact of basic research on the larger R&D community across many generations. In Citation Mining, text mining (Kostoff et al, 2000a, 2000b, 2002, 2003; Losiewicz, 2000) of the cited and citing papers (trans-citation) supplements the information derived from the semi-structured field bibliometric analyses. Text mining illuminates the trans-citation thematic relationships, and provides insights of knowledge diffusion to other intra-discipline research, advanced intra-discipline development, and extra-discipline research and development. The addition of text mining to citation bibliometrics makes feasible the large-scale multi-generation citation studies that are necessary to display the full impacts of research.

A proof-of-principle demonstration of Citation Mining for user population profiling and research impact was performed on four sets of cited papers. The papers were selected based on the authors' technical interests, rather than a random representative sample. It was desired to have one group of papers representative of basic research, and another group representative of applied research. Two of the sets were selected Mexican and U. S. applied photo-voltaic research papers, and two of the sets were selected British and U. S. fundamental vibrating sand-pile research papers.

This paper presents the bibliometrics of those papers that cited all four sets of papers mentioned above, then focuses on the trans-citation coupled citing paper/ cited paper text mining results for one

of the sets, a highly cited U. S. vibrating sand-pile paper (Jaeger, 1992). Vibrating sand-piles are important in their own right, since they model the behavior of granular systems used in agriculture (seeds, grains), geo logy (avalanches, so il mechanics), constructi on (gravel, sand), and manufacturing (powders, lubricants, sand-blasting). The underlying phenomena exhibited in their static and dynamic states can be found in many disparate applications, such as fusion confinement, geological formations, self-assembly of materials, thin film structure ordering, shock-wave statistics, and crowded airspace. Statistically, the sand-pile paper selected has sufficient ci ting papers for adequate text mining statistics. It covers an exciting area of physics research, and its technical sub-themes have potential for extrapolation to other technical disciplines.

The analyses performed were of two types: bibliometrics and text mining. The text mining was subdivided into two components, manual concept clustering and statisti cal concept clustering. These different types of analyses are described in the following sections.

IV-A. Bibliometrics Analysis

The citing paper summaries (records) were retrieved from the SCI. Analyses of the different non-free-text fields in each record were performed, to identify the infrastructure characteristics of the citing papers (authors, journals, institutions, countries, technical disciplines, etc).

This section starts by identifying the types of data contained in the SCI (circa early 2000), and the types of analyses that will be performed on this information (see Table 1).

FIGURE 1 – SAMPLE SCI RECORD

Textual data mining to support science and technology management
Losiewicz P, Oard DW, **Kostoff RN**
JOURNAL OF INTELLIGENT INFORMATION SYSTEMS
15 (2): 99-119 SEP 2000

Document type: Article **Language:** English **Cited References:** 30 **Times Cited:** 7

Abstract:
This paper surveys applications of data mining techniques to large text collections, and illustrates how those techniques can be used to support the management of science and technology research. Specific issues that arise repeatedly in the conduct of research management are described, and a textual data mining architecture that extends a classic paradigm for knowledge discovery in databases is introduced. That architecture integrates information retrieval from text collections, information extraction to obtain data from individual texts, data warehousing for the extracted data, data mining to discover useful patterns in the data, and visualization of the resulting patterns. At the core of this architecture is a broad view of data mining-the process of discovering patterns in large collections of data-and that step is described in some detail. The final section of the paper illustrates how these ideas can be applied in practice, drawing upon examples from the recently completed first phase of the textual data mining program at the Office of Naval Research. The paper concludes by identifying some research directions that offer significant potential for improving the utility of textual data mining for research management applications.

Author Keywords:
text data mining, information retrieval, knowledge discovery in databases, bibliometrics, computational linguistics

Addresses:
Losiewicz P, ACS Def Inc, Rome, NY USA
ACS Def Inc, Rome, NY USA
Univ Maryland, Coll Lib & Informat Serv, College Pk, MD 20742 USA
Off Naval Res, Arlington, VA 22217 USA

Publisher:
KLUWER ACADEMIC PUBL, DORDRECHT

IDS Number:
333GP

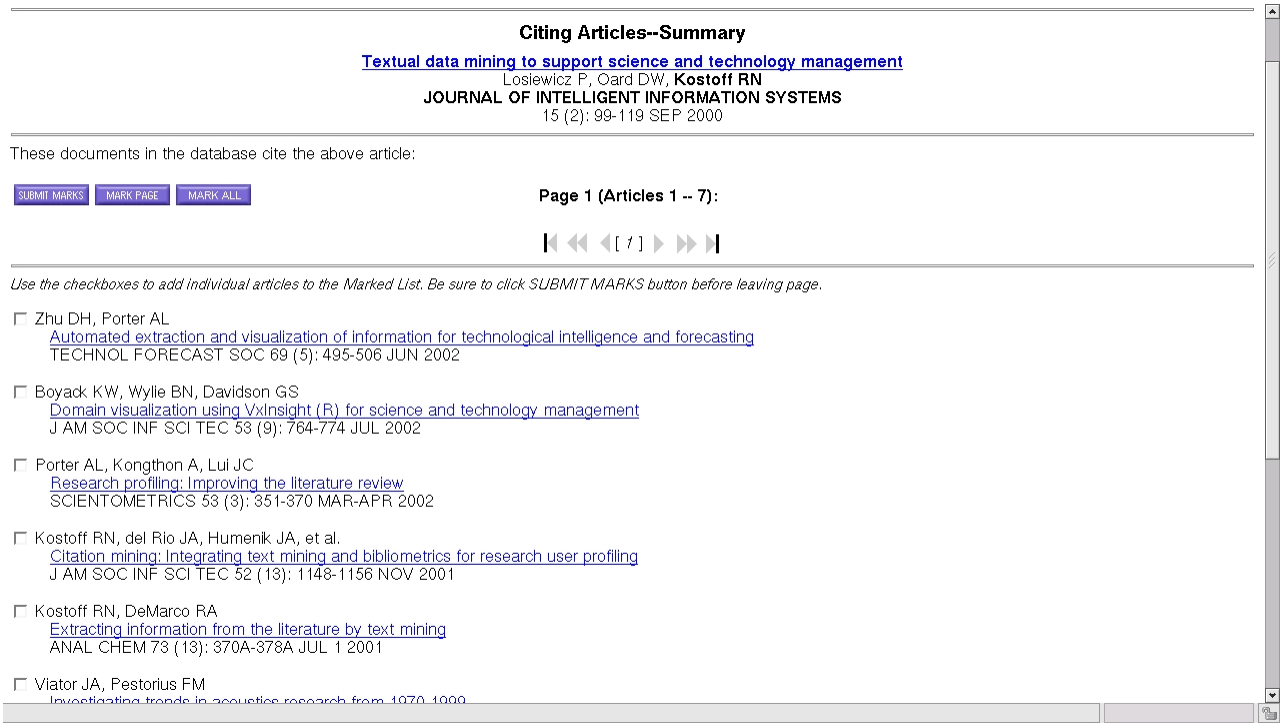
ISSN:
0925-9902

Figure 1 shows a sample record from the SCI. The actual paper that it represents is referred in the following description as the 'full paper'. Starting from the top, the individual fields are described in Table 1:

TABLE 1 – SCI RECORD FIELDS

1) Title - the complete title of the full paper.
2) Authors - all the authors of the full paper.
3) Source - journal name (e.g., Journal of Intelligent Information Systems).
4) Issue/ Page(s)/ Publication Date
5) Document Type - (e.g., Article, note, review, letter).
6) Language – the language of the full text document.
7) Cited References - the number and names of the references cited in the full paper
8) Times Cited - the number and names of the papers (whose records are contained in the SCI) that cited the full paper (see Figure 2). Thus, the number shown in this field is a lower bound.
9) Related Records – records that share one or more references (not shown).
10) Abstract - the complete Abstract from the full paper.
11) Author Keywords - keywords supplied by the author. In this example, no Keywords were supplied by the indexer, but the SCI contains a field for indexer Keywords, if supplied.
12) Addresses - organizational and street addresses of the authors. For multiple authors, this can be a difficult field to interpret accurately. Different authors from the same organizational unit may describe their organizational level differently. Different authors may abbreviate the same organizational unit differently.
13) Publisher

FIGURE 2. LIST OF CITING PAPERS OF ARTICLE SHOWN IN FIGURE 1.



Citing Articles--Summary

Textual data mining to support science and technology management
Losiewicz P, Oard DW, **Kostoff RN**
JOURNAL OF INTELLIGENT INFORMATION SYSTEMS
15 (2): 99-119 SEP 2000

These documents in the database cite the above article:

SUBMIT MARKS | MARK PAGE | MARK ALL

Page 1 (Articles 1 -- 7):

[1]

Use the checkboxes to add individual articles to the Marked List. Be sure to click SUBMIT MARKS button before leaving page.

- Zhu DH, Porter AL
  Automated extraction and visualization of information for technological intelligence and forecasting
  TECHNOL FORECAST SOC 69 (5): 495-506 JUN 2002
- Boyack KW, Wylie BN, Davidson GS
  Domain visualization using VxInsight (R) for science and technology management
  J AM SOC INF SCI TEC 53 (9): 764-774 JUL 2002
- Porter AL, Kongthon A, Lui JC
  Research profiling: Improving the literature review
  SCIENTOMETRICS 53 (3): 351-370 MAR-APR 2002
- Kostoff RN, del Rio JA, Humenik JA, et al.
  Citation mining: Integrating text mining and bibliometrics for research user profiling
  J AM SOC INF SCI TEC 52 (13): 1148-1156 NOV 2001
- Kostoff RN, DeMarco RA
  Extracting information from the literature by text mining
  ANAL CHEM 73 (13): 370A-378A JUL 1 2001
- Viator JA, Pestorius FM

How can the above fields be used in Citation Mining? In this paper, a phenomenological method to analyze the total information available in SCI database is proposed, as follows:

Title field is used in text mining together with the other unstructured text fields, Abstracts and Keywords, to perform the correlation analysis of the themes in the cited paper to those of the citing papers. Computational linguistics analysis is then performed.

Author field is used to obtain multi-author distribution profiles (e.g., number of papers with one author, number with two authors, etc).

Counts in Source field can lead to journal name distributions, theme distributions, and development level distributions.

Document Type register allows distributions o f different document types to be co mputed (e.g., three articles, four conference proceedings, etc.).

Language field allows distributions over languages to be computed.

Cited References allows a historical analysis of the problem to be performed, and this field can be used to analyze the interrelations among different groups working on related problems.

Times Cited register would be important if the citing papers are of sufficient vintage. Then, their multiplier effect would be of interest, and could be computed. The distribution profile of times cited of the citing papers would be generated.

The Addresses register allows distributions of names and types of institutions, and countries, to be generated. Institution and country combinations would be of special interest, and could be correlated with author combination distributions.

The present demonstration of citation mining includes a comparison of a cited research unit from a developing country with a cited research unit from a developed country. It also compares a cited unit from a basic research field with a cited unit from an applied research field. Specifically, the technique is being demonstrated using selected papers from a Mexican semiconductor applied research group (MA), a United States semiconductor applied research group (UA), a British fundamental research group (BF), and a United States fundamental research group (UF) (see Table 2). These papers were selected based on the authors' familiarity with the topical matter, and the desire to examine papers that are reasonably cited. Sets of papers having at least 50 external cites were selected for analysis in order to have a good phenomenological description.

Table 2 – Cited Papers Used for Study

GROUP	Times Cited	PAPERS
MA	59	Nair P.K. Sem. Sc. Tech. 3 (1988) 134-145 Nair P.K. J Phys D - Appl Phys, 22 (1989) 829-836 Nair M.T.S. Sem. Sc. and Tech. 4 (1989) 191- 199 Nair M. T. S. J Appl Phys, 75 (1994) 1557-1564
UF	307	Jaeger HM, 1992, Science, V255, P1523
BF	119	Mehta A, 1989, Physica A, V157, P1091 Mehta A, 1991, Phys Rev Lett, V67, P394 Barker GC, 1992, Phys Rev A, V45, P3435 Mehta A, 1996, Phys Rev E, V53, P92
UA	89	Tuttle, Prog. Photovoltaic v3, 235 (1995) Gabor, Appl. Phys. Lett. v65, 198 (1994) Tuttle, J. Appl. Phys. v78, 269 (1995) Tuttle, J. Appl. Phys. v77, 153 (1995) Nelson, J. Appl. Phys. v74 5757 (1993)

In addition, selection and banding of variables are key aspects of the bibliometric study. While specific variable values are of interest in some cases (e.g., names of specific citing institutions), there tends to be substantial value in meta-level groupings (e.g., institution class, such as government, industry, academia). Objectives of the study are to demonstrate important variables, types of meta-level groupings providing the most information and insight, and those conditions under which non-dimensionalization become useful. However, two analyses at the micro-level are presented involving specific correlations between both citing author and references for BF and UF papers. This latter analysis is directly important for the performers of scientific research. In addition, text mining could

be performed on the text fields (mainly the Abstract, but including the Title and Keywords) to supplement the analysis on the semi-structured and structured fields (see Kostoff et al., 2000a, 2000b, 2001b, 2002, 2003).

IV-B. Manual Concept Clustering

The purpose of the manual concept clustering was to generate a taxonomy (technical category classification scheme) of the database from the quantified technical phrases extracted from the free-text record fields. To generate the database, the citing papers' Abstracts were aggregated. Computational linguistics analyses were then performed on the aggregate. Technical phrases were extracted using the Database Tomography process (Kostoff et al, 1995, 2000a, 2000b; Losiewicz et al, 2000). An algorithm extracted all single, adjacent double, and adjacent triple word phrases from the text, and recorded the occurrence frequency of each phrase. While phrases containing trivial/ stop words at their beginning or end were eliminated by the algorithm, extensive manual processing was required to eliminate the low technical content phrases. Then, a taxonomy of technical sub-categories was generated by manually grouping these phrases into cohesive categories. Intra-discipline applications, and extra-discipline impacts and applications were identified from visual inspection of the phrases.

IV-C. Statistical Concept Clustering

The purpose of the statistical concept clustering was to generate taxonomies of the database semi-automatically, again from the quantified technical phrases extracted from the free-text record fields. The clustering analysis further used quantified information about the relationships among the phrases from co-occurrence data (the number of times phrases occur together in some bounded domain). The statistical clustering analyses results complemented those from the manual concept clustering, and offered added perspectives on the thematic structure of the database.

After the phrase frequency analyses were completed, co-occurrence matrices of Abstract words and phrases (each matrix element Mij is the number of times phrase or word i occurs in the same record Abstract as phrase or word j) were generated using the TechOasis phrase extraction and matrix generation software. As in the phrase frequency analysis, the phrases extracted by the TechOasis natural language processor required detailed manual examination, to eliminate the low technical content phrases. The co-occurrence matrices were input to the WINSTAT statistical clustering software, where clusters (groups of related phrases based on co-occurrence frequencies) based on both single words and multi-word phrases were generated.

Two types of statistical clustering were performed, high and low level. The high level clustering used only the highest frequency technical phrases, and resulted in broad category descriptions. The low level clustering used low frequency phrases related to selected high frequency phrases, and resulted in more detailed descriptions of the contents of each broad category.

IV-C-1. High Level Clustering

The TechOasis phrase extraction from the citing Abstracts produced two types of lists. One list contained all single words (minus those filtered with a stop word list), and the other list contained similarly filtered phrases, both single and multi-word. Both lists required further manual clean-up,

to insure that relatively high technical content material remained. The highest frequency items from each list were input separately to the TechOasis matrix generator, and two co-occurrence matrices, and resulting factor matrices, were generated.

The co-occurrence matrices were copied to an Excel file, and the matrix elements were non-dimensionalized. To generate clusters defining an overall taxonomy category structure for the citing papers, the Mutual Information Index was used as the dimensionless quantity. This indicator, the ratio of: the co-occurrence frequency between two phrases squared (Cij^2) to the product of the phrase occurrence frequencies (Ci*Cj), incorporates the co-occurrence of each phrase relative to its occurrence in the total text. The co-occurrence matrix row and column headings are arranged in order of decreasing frequency, with the highest frequency phrase occurring at the matrix origin. Based on the intrinsic nature of word and phrase frequencies, the row and column heading frequencies decrease rapidly with distance from the matrix origin. With increasing distance from the origin, the matrix becomes more and more sparse, although the phrases themselves have higher but more focused technical content. In parallel, the Mutual Information Index's values decrease rapidly as the distance from the matrix origin increases. Thus, the Mutual Information Index is useful for relating the highest frequency terms only, and for providing the top-level structural description of the taxonomy categories.

IV-C-2. Low Level Clustering

To obtain a more detailed technical understanding of the clusters and their contents, the lower frequency phrases in each cluster need to be identified. A different matrix element non-dimensional quantity is required, one whose magnitudes remain relatively invariant to distance from the matrix origin. In addition, a different approach for clustering the low frequency phrases in the sparse matrix regions is required, one that relates the very detailed low frequency phrases to the more general high frequency phrases that define the cluster structure. In this way, the low frequency phrases can be placed in their appropriate cluster taxonomy categories.

The method chosen to identify the lower frequency phrases is as follows. Start with the cluster taxonomy structure defined by grouping the higher frequency phrases using the Average Neighbor agglomoration technique and the Mutual Information Index. Then, for each high frequency phrase in each cluster, find all phrases whose value of the Inclusion Index Ii exceeds some threshold. Ii is the ratio of Cij to Ci (the frequency of occurrence of phrase i in the total text), where phrase i has the lower frequency of the matrix element pair (i,j). A threshold value of 0.5 for Ii was used. The resultant lower frequency phrases identified by this method will occur rarely in the text, but when they do occur, they will be in close physical (and thematic) proximity to the higher frequency phrases.

V. RESULTS

V-A. Citation Bibliometrics

Figure 3
Authors Distribution Function

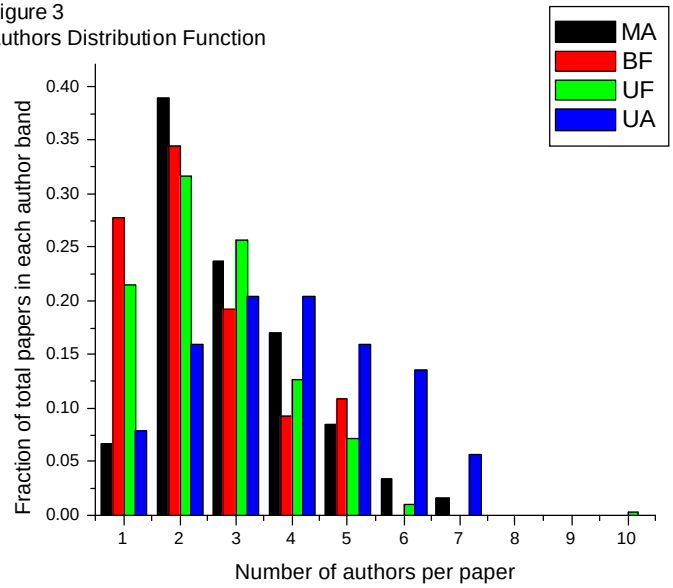
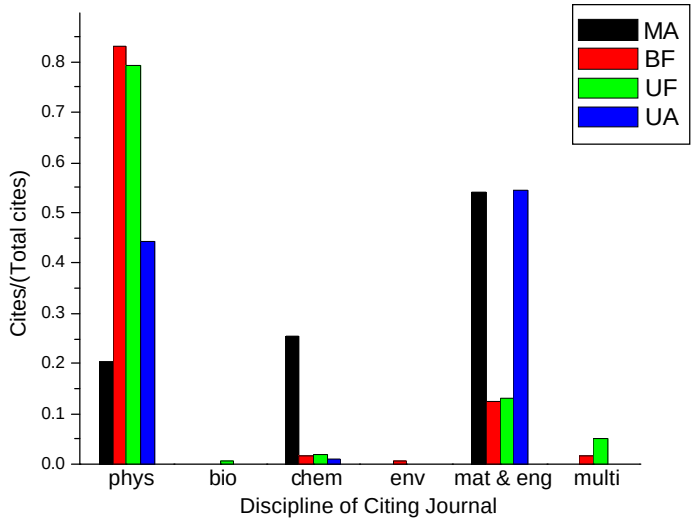


Figure 3 contains a bar graph of multi-author distribution for the four sets analyzed. The ordinate represents the fraction of total papers published in each author band, and the abscissa represents the number of authors per paper. The most striking feature of this graph is the behavior at the wings. The papers citing basic research dominate the low end (single author), while the papers citing applied research dominate the high end (6-7 authors). The papers citing basic research (BF and UF) have a similar number of authors per paper, with a maximum in the frequency distribution at two authors per paper. The UA citing papers show gaussian-like authorship distribution with three and four authors per paper, while the MA group citing papers show a distribution similar to the groups citing fundamental research papers but with fewer single-author papers. These four sets show author distributions where 90% of the papers had less than six authors. These results confirm the diversity of collaborative group compositions over different disciplines and levels of development.

Generally, as projects become more applied, they tend to become larger and more expensive, and require more resources. They also usually require the integration of multiple disciplines. Both these characteristics typically result in larger research groups, and hence in more contributors to a project and its resulting documents. Experimental work usually invo lves larger teams than theoretical work, while modeling and simulation activities tend to allow more individual efforts. The strong experimental emphasis of the two applied semiconductor groups, with little evidence of computer simulation shown, results in large teams on average. The more balanced theory/ experiment combination of the basic research group tends to suppress larger team efforts in favor of more individualized research. In addition, the intrinsic nature of sandpile vibrat ion research, as opposed to elementary particle or fusion research, does not require large facilities and large research teams.

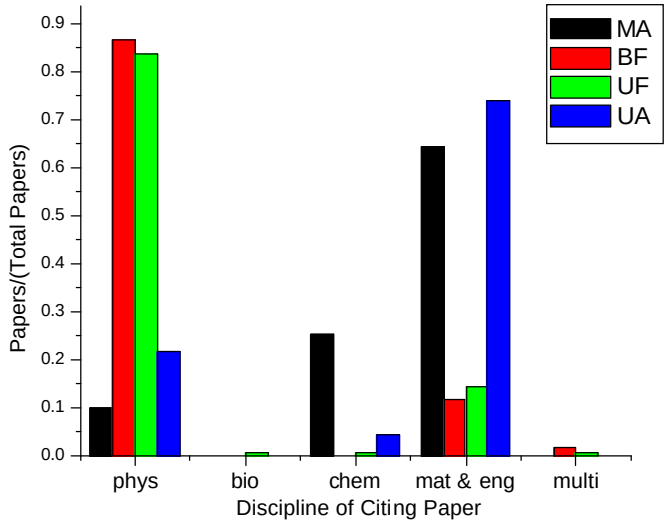
The citing journal discipline frequency is shown in Figure 4. Clearly, each paper set has defined its main discipline well. Also, there is a symmetry in the cross citing disciplines. UF and BF groups were cited more than 80% in fundamental journals and close to 10% in applied journals. Similarly, MA and UA groups were cited close to 50% in applied journals and 45% in fundamental journals. These journal discipline results suggest that the applications developed by the MA group have a strong impact on chemical journals, while the applications developed by the UA group strongly impact physics journals. A point to be stressed is that only the fundamental papers received cites in journals clearly outside of their disciplines.

Figure 4
Citing Journal Theme Distribution



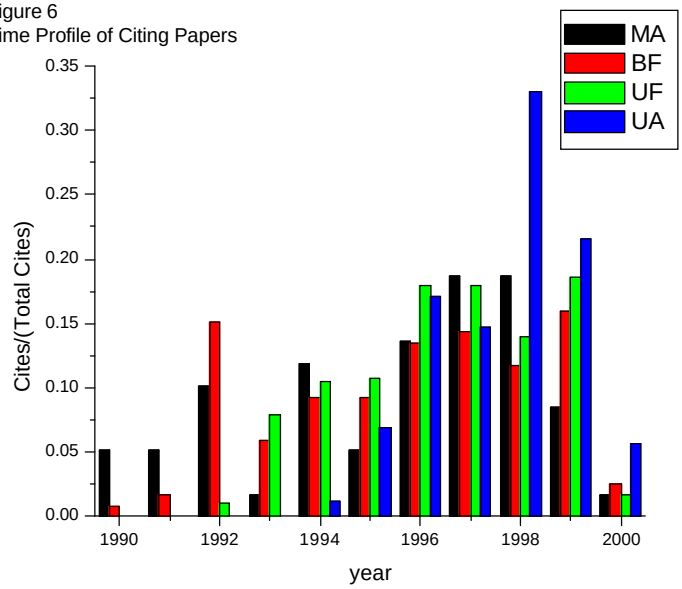
The discipline distribution of the citing papers, produced by analyzing the papers' Abstracts and Titles, is shown in Figure 5. It is slightly different from Figure 4. As concluded in the text mining, these free-text fields provide far more precise information than can be obtained from the journal discipline. Multi-disciplinary journals can publish uni-disciplinary papers from many different disciplines. Also, the journal categories, determined by ISI, are not a unique reflection of specific contents (e.g., an environmental journal can accept engineering papers, a materials journal can accept physics papers, etc). However, the chemical nature of the papers/ journals impacted by the MA group is confirmed.

Figure 5
Citing Paper Theme Distribution



In three of the four sets analyzed, the component papers were published in different years. The MA set was published from 1989 to 1994, UA from 1994 to 1995, BF from 1989 to 1996, while UF includes only one paper published in 1992. Figure 6 shows a clear oscillating behavior of UA and BF, due partly to the different dates of paper publication. Also, most of the sets have between 10% and 20% of cites per year, while the UA set received 38% of the cites in 1998.

Figure 6
Time Profile of Citing Papers



The single highly-cited paper feature of the UF set allows additional analyses and perspectives. In Figure 6a, the UF citing paper disciplines are shown as a function of time. As time evolves, citing papers from disciplines other than those of the cited paper emerge. An important point is the four-year delay of the systematic appearance of the more applied engineering and materials science citing papers.

Figure 6a
Time Profile of Citing Papers

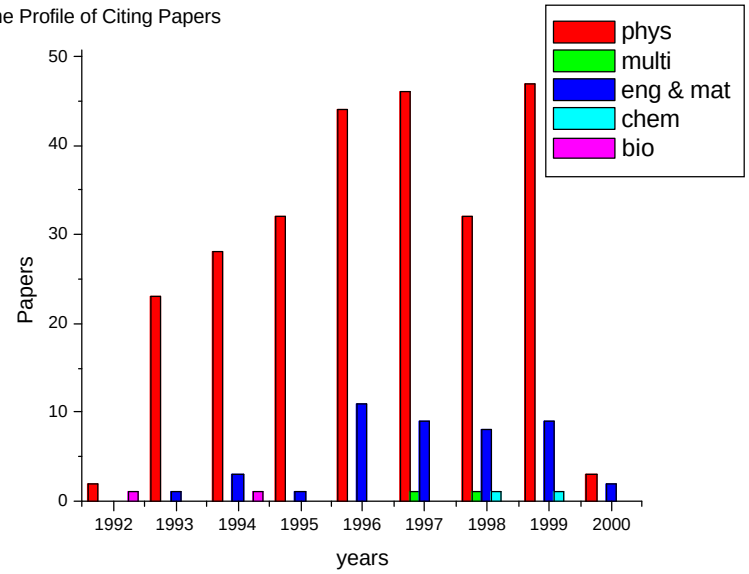


Figure 7 shows that most cites appear in articles. The four analyzed sets are cited in review articles and letters. This indicates the relevance of the analyzed papers. One important point is that only the fundamental papers are cited in notes, and only the UF paper was cited in an editorial document.

Figure 7
Citing Paper Type Profile

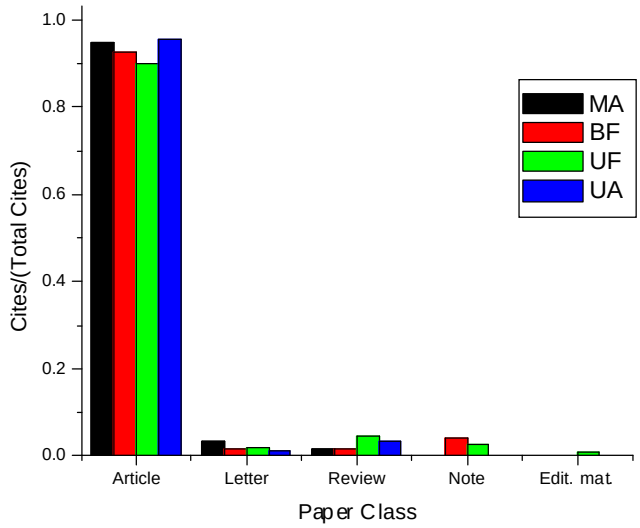


Figure 8 shows that English is the dominant language of all the paper sets analyzed. However, the surprising appearance of a significant number of citing papers written in Romanian for the MA set indicates that MA's work is important for at least one developing country. Also, there are no papers in Spanish.

Figure 8
Citing Paper Language Profile

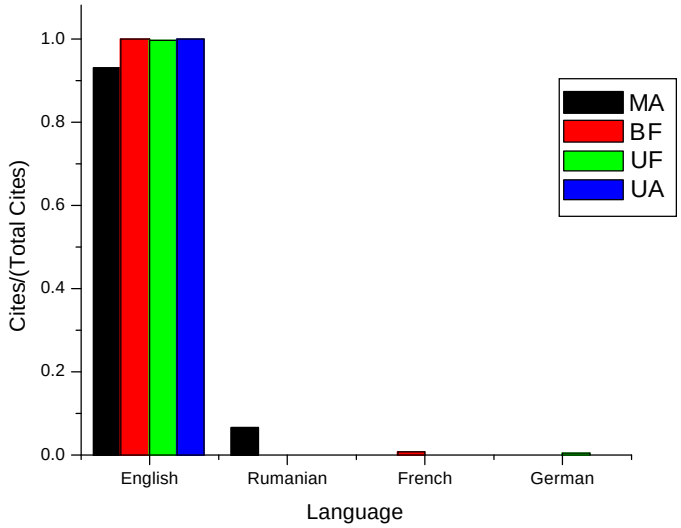
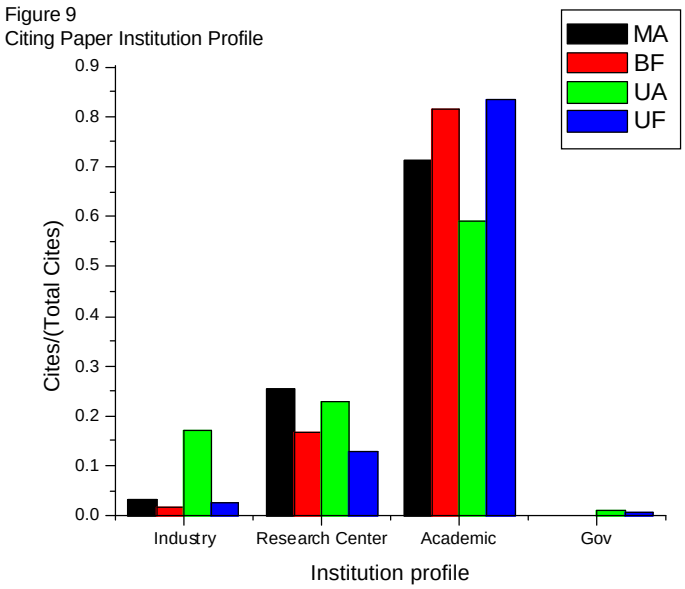


Figure 9 shows the profile of the citing institutions. Clearly, academia has the highest citing rates. Industry publications cite the advances in high-technological developments, but are not citing the advances in fundamental research. Research Centers fo llow applied and fundamental research about equally. Direct government participation is not significant in the fields studied. Government/ national laboratories were classified under research centers.



There are 44 countries represented in the citing paper sets analyzed. Figure 10 shows only those countries with at least 10% of the citations for a set. USA has the most cites in aggregate. India has the largest cites of the MA set; Japan has the largest cites of the UA set. This fact is due to the different nature of the applied technology developed by MA and UA. The UA set contains work related to high technology, and the MA set is dedicated to explore low-cost technology. Therefore, this last set is cited by the less affluent countries of India, Romania and Mexico. India and Mexico also cite fundamental research, but not Romania. It is important to stress that if no low-cost technology papers were considered, these latter countries would not appear in this graph, and only developed countries would appear. Another point is that England does not cite UA works.

Figure 10
Citing Paper Country Profile

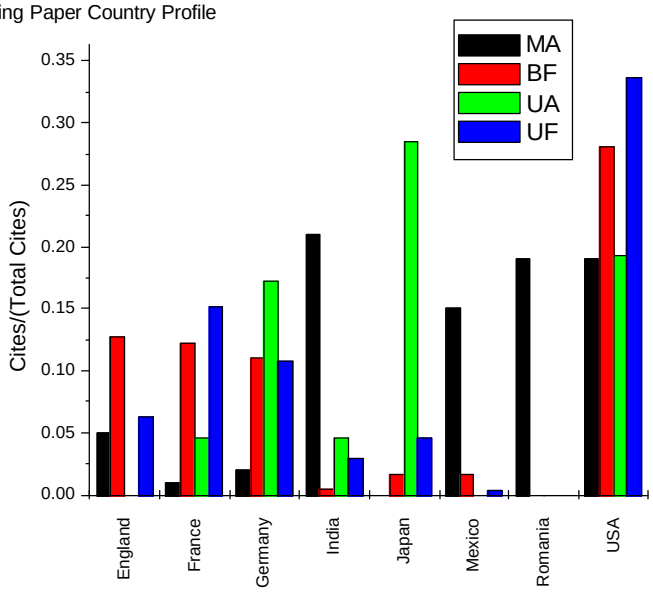
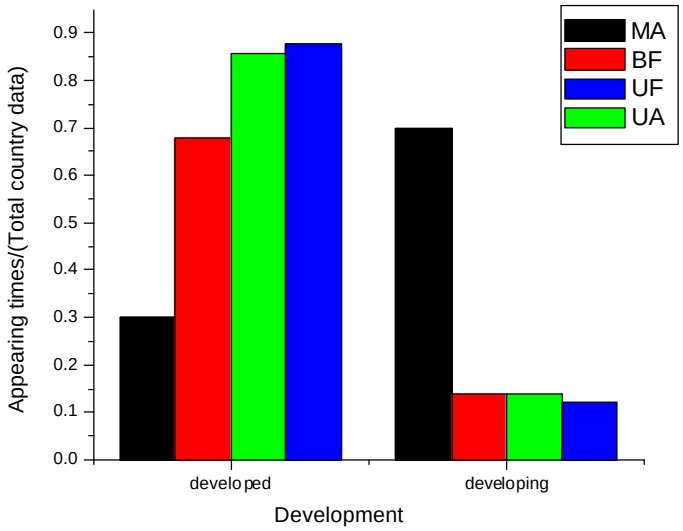


Figure 11 shows clearly that the low-cost technology papers are cited by developing countries. Developed countries cite the mostly high-technology papers.

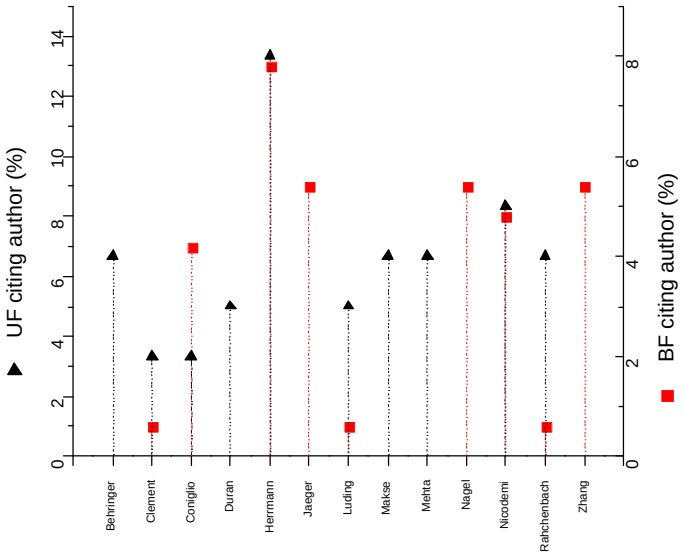
Figure 11
Citing Country Development Phase Profile



The analysis of the most common citing authors is presented in figures 12 and 13 where the frequency of an author citing UF (triangle) or BF (square) is plotted. Figure 12 shows that there is a close relation between the citing authors for both BF and UF groups. There is a common citing

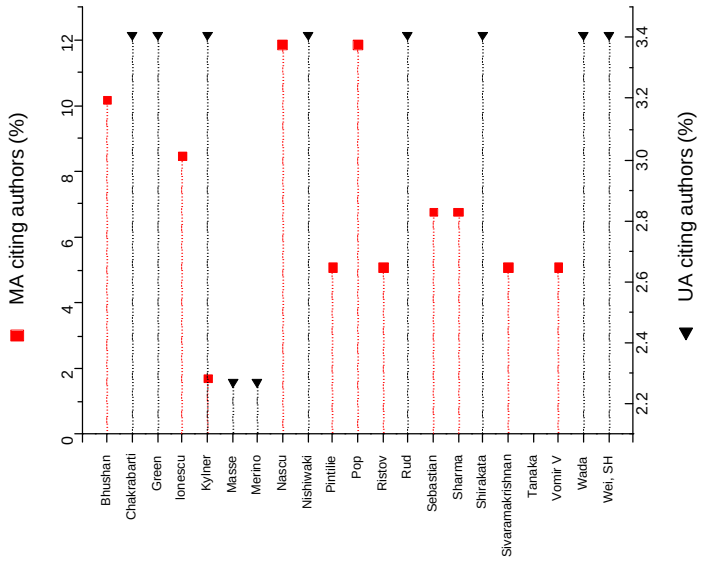
author who occupied the highest position in the frequency plot in both sets (Hermann, HJ). Three of the highest citing authors are not shared between the citing sets of UF and BF. Jaeger and Nagel are the authors of the UF paper and Mehta is one of the authors of BF paper. They maintain awareness of each other's work.

Figure 12



In contradistinction, Figure 13 shows that MA and UA have no intersection between their topics (low cost photovoltaic thin films and high efficient photovoltaic cells, respectively), from the perspective of the highest citing authors. Previous citation results have shown that applied research authors tend to cite more fundamental research, along relatively stratified lines. In Figure 13, it is clear that the maximum citing author of the MA group is a Romanian researcher.

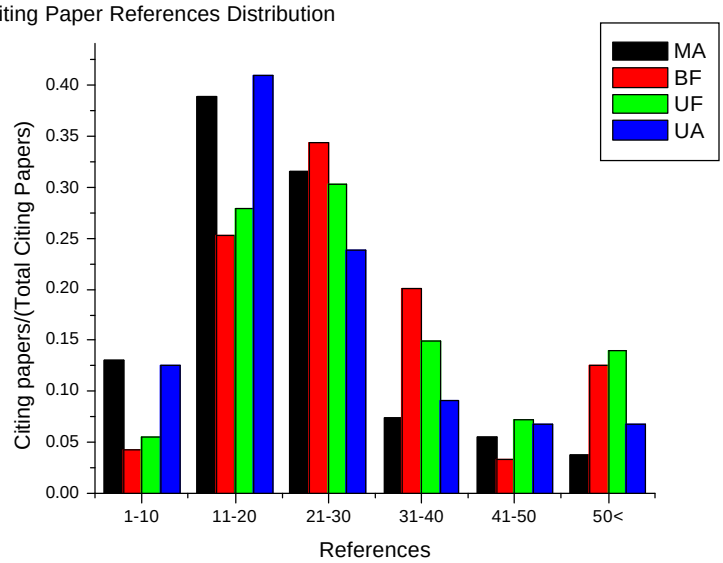
Figure 13



Tables A1 and A2 in the Appendix present the numerical data.

In Figure 14, it is clear that there are common features in the number of references in those papers that cite the core applied and fundamental papers, but there are also some differences. For instance, at the lower end of the spectrum (0-20), the applied papers' citing papers dominate. At the higher end of the spectrum (21-50+), the fundamental papers' citing papers dominate, with the exception of the BF anomaly at 41-50.

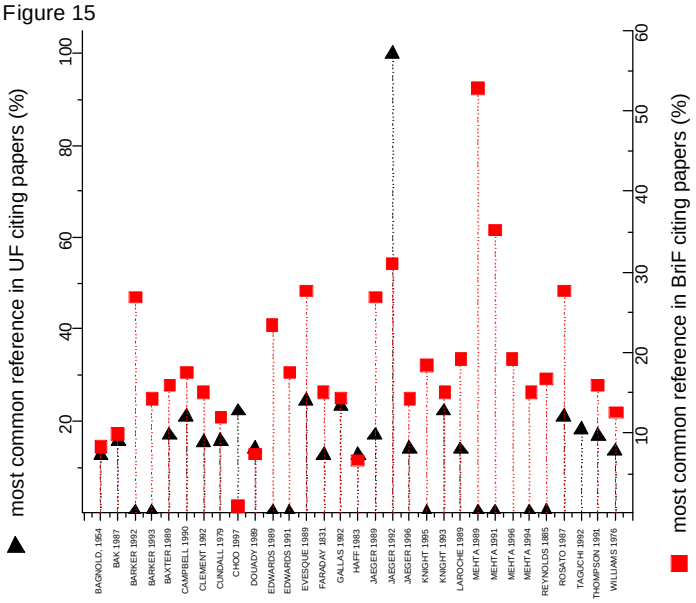
Figure 14
Citing Paper References Distribution



There are many possible reasons for these differences, and separating out the effects is complex. There are two different technical disciplines, and each one has its citing culture and traditions. Also, each technical discipline has a different level of research activity, and this could influence the magnitude of citations generated. Basic researchers tend to document more, and therefore produce a larger literature to cite. Finally, there may be different citing practices in basic and applied research.

Frequency analysis of the most common references in the citing papers provides insight to co-cited papers, and allows a historical perspective to be obtained. The reference-frequency for the UF and BF citing papers is shown in Figure 15. This figure shows clearly that the fundamental papers dealing with sand-piles are actually correlated.

Figure 15

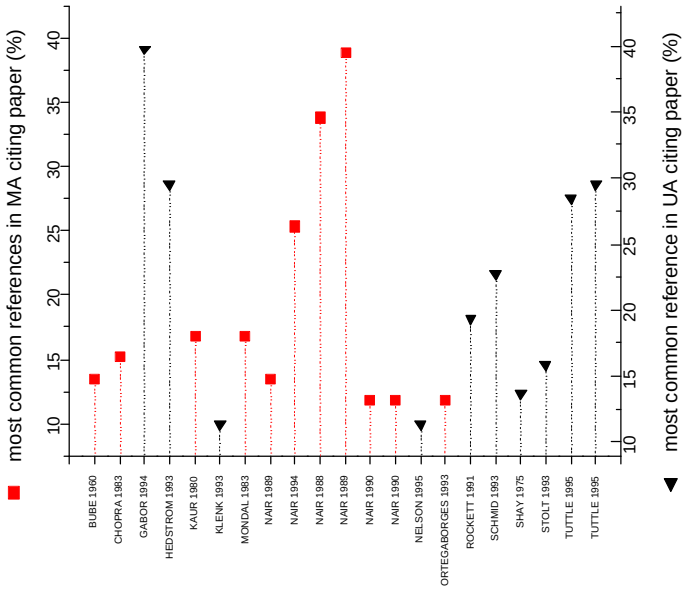


In this figure, Faraday's work (1831) appears within the twenty papers most cited in the UF and BF citing papers. This indicates the fundamental and seminal character of the experimental work performed by Faraday. Also, Reynolds' work (1885) appears within the twenty most cited papers in the references of the BF set. These two references also indicate the longevity of the unsolved problems tackled by the UF and BF groups.

The highest frequency co-cited papers have three interesting characteristics. They are essentially all in the same general physics area, they are all published in fundamental science journals (mainly physics), and they are all relatively recent, indicating a dynamic research area with high turnover. The detailed table is presented in the appendix.

The corresponding analysis of the most common references in the applied MA and UA groups is presented in figure 16. This figure shows clearly that these two groups have no correlations. However, in the detailed correlation analysis, there is one paper in the intersection of these two groups.

Figure 16



SUMMARY AND CONCLUSIONS

The first two objectives of this study were to demonstrate the feasibility of tracking the myriad impacts of research on other research, development, and applications, using the technical literature, and demonstrate the feasibility of identifying a broad range of research product user characteristics, using the technical literature. Both of these objectives were accomplished, along with some interesting technical insights about vibrating sandpile dynamics and temporal characteristics of information diffusion from research to applications. This wide range of results leads to the following conclusions.

Exploitation of the other types of information contained in the SCI and associated with the citation process offers the potential for providing R&D sponsors information that can help guide future directions of their R&D. In addition, the complete Citation Mining process described in the present paper has the potential to objectively document the breadth of impact of basic research on the R&D community. The addition of text mining to citation bibliometrics will make feasible the large-scale multi-generation citation studies that are necessary to display the full impacts of research.

Text mining is a requirement for making the total Citation Mining possible. Without text mining, either an overly general automated technique, such as journal classificat ion, must be used to identify application areas, or tens or hundreds of thousands of Abstracts must be read. Text mining can locate small numbers of extra-discipline phrases (small signals) from large numbers of intra-

discipline phrases (large clutter), and allow only those Abstracts of specific interest to be selected and read.

A substantial amount of human judgement and labor is required for all aspects of Citation Mining. For the bibliometric component of Citation Mining reported in this paper, classifying the results in groupings where judgement is required (e.g., Abstract technical theme, or applications theme) necessitates substantial work. For the text mining component described in detail in this paper, thousands of technical phrases must be examined. Judgements must be made as to their alignment with the main themes of the cited paper(s). Some of the bibliometric components conceivably could be automated (e.g., all the SCI journals could be classified by technical theme beforehand, then the alignment of the cited journal theme to the citing journal theme could be generated automatically). It is not clear how the selection of extra-discipline phrases could be automated, given the intense expert judgement required.

This study referred to, but did not examine details of, second or higher generation citations. The authors believe they are valid measures or indicators of influence and impact, but the actual method of impact quantification remains an open question. More research is required to understand the principles of allocating impact among a paper's references.

Finally, there is a very important message that emerges from the results of the present study relative to the sponsorship of basic research. Over the past decade, the trend in industry and government has been toward requirements-driven research (e.g., the term 'strategic research' is becoming used more widely in government agencies, and corporately-funded industrial research has strongly evolved into profit-center sponsored research). While this may be beneficial to the sponsoring organization from a short-term tactical perspective, the long-term strategic perspective may suffer. Would fundamental sand-pile research receive funding from Tokamak, air traffic control, or materials programs, even though sand-pile research could impact these or many other types of applications, as shown in this paper? *It is necessary to stress that sponsorship of some unfettered research must be protected, for the strategic long-term benefits on global technology and applications!*

REFERENCES FOR APPENDIX 7-H

Amaral, LAN, Gopikrishnan, P., Matia, K., Plerou, V. and Stanley E.H. (2001). Application of statistical physics methods and concepts to the study of science & technology systems, Scientometrics, 51, 3

Averch, H., (1994) "Economic Approaches to the Evaluation of Research", Evaluation Review, 18:1, February, p. 77-88.

Bilke, S. and Peterson, C. (2001). Topological properties of citation and metabolic networks, Phys. Rev. E 64, 036106

Braun, T. et al, (1987), "Literature of Analytical Chemistry. A Scientometric Evaluation", CRC Press.

Davidse, R. J., and Van Raan, A. F. J., (1997). “Out of Particles: Impact of CERN, DESY, and SLAC Research to Fields other than Physics”, Scientometrics, 40:2. P. 171-193.

De Solla Price, D. J., (1986) “Little Science, Big Science and Beyond”, Columbia University Press.

Del Río, J. A., Kostoff, R. N., García, E. O., Ramírez, A. M., and Humenik, J. A., (2000). “Citation Mining: Citing Population Profiling using Bibliometrics and Text Mining”. Centro de Investigación en Energía, Universidad Nacional Autonoma de Mexico http://www.cie.unam.mx/xml/tc/ft/arp/citing.xml.

Del Rio, J. A., Kostoff, R. N., Garcia, E. O., Ramirez, A. M., and Humenik, J. A. (2002). “Phenomenological Approach to Profile Impact of Scientific Research: Citation Mining.” Advances in Complex Systems. 5:1. 19-42.

DOD, (1969) Project Hindsight, Office of the Director of Defense Research and Engineering, Wash., D. C., DTIC No. AD495905, October.

DOE, (1983) "Health and Environmental Research: Summary of Accomplishments", Office of Energy Research, Office of Program Analysis, Report No. DOE/ER-0194, May; DOE, (1986) "Health and Environmental Research: Summary of Accomplishments", Office of Energy Research, Office of Program Analysis, Report No. DOE/ER-0275, August.

Egghe, L., and Rousseau, R., (1990) “Introduction to Informetrics”, Elsevier.

Herring, S. D., (1999). “The Value of Inter-disciplinarity: A Study Based on the Design of Internet Search Engines”, JASIS, 1 April, p. 358-365.

IDA, (1991) "DARPA Technical Accomplishments", Volume I, IDA Paper P-2192, February 1990; Volume II, IDA Paper P-2429, April 1991; Volume III, IDA Paper P-2538, July 1991, Institute for Defense Analysis.

IITRI, (1968) "Technology in Retrospect and Critical Events in Science", Illinois Institute of Technology Research Institute Report, December.

Jaeger, H. M., and Nagel, S. R. (1992). “Physics of the Granular State”. Science. 256. 20 March, p. 1523-1531.

Katz, J.S. (2000) “Scale-independent Indicators and Research Evaluation” Science and Public Policy. 27. I. 23-36.

Kostoff, R. N., (1994) "Assessing Research Impact: US Government Retrospective and Quantitative Approaches", Science and Public Policy, 21:1, February.

Kostoff, R. N., Eberhart, H. J., and Miles, D. (1995). "System and Method for Database Tomography", U. S. Patent Number 5440481, August 8.

Kostoff, R. N., (1997) "The Handbook of Research Impact Assessment", DTIC Report Number ADA296021, Summer. Also, see http://www.dtic.mil/dtic/kostoff/index.html.

Kostoff, R. N., Braun, T., Schubert, A., Toothman, D. R., and Humenik, J. (2000a). "Fullerene Roadmaps Using Bibliometrics and Database Tomography". Journal of Chemical Information and Computer Science. 40:1. Jan-Feb. p. 19-39.

Kostoff, R. N., Green, K. A., Toothman, D. R. Humenik, J. A., (2000b) "Database Tomography Applied to an Aircraft Science and Technology investment Strategy", Journal of Aircraft, 37:4, p. 727-730.

Kostoff, R. N., Del Rio, J. A., García, E. O., Ramírez, A. M., and Humenik, J. A. (2001a). "Citation Mining: Integrating Text Mining and Bibliometrics for Research User Profiling". JASIST. 52:13. 1148-1156. 52:13. November.

Kostoff, R. N., and DeMarco, R. A. (2001b). "Science and Technology Text Mining". Analytical Chemistry. 73:13. 370-378A. 1 July.

Kostoff, R. N., Tshiteya, R., Pfeil, K. M., and Humenik, J. A. (2002). "Electrochemical Power Source Roadmaps using Bibliometrics and Database Tomography". Journal of Power Sources. 110:1. 163-176.

Kostoff, R. N., Shlesinger, M., and Malpohl, G. (2003). "Fractals Roadmaps using Bibliometrics and Database Tomography". Fractals. December.

Losiewicz, P., Oard, D., and Kostoff, R. N. (2000). "Textual Data Mining to Support Science and Technology Management". Journal of Intelligent Information Systems. 15:2, p. 99-119.

Narin, F., (1989) "The Impact of Different Modes of Research Funding", in: Evered, David and Harnett, Sara, Eds., The Evaluation of Scientific Research, John Wiley and Sons, Chichester, UK, p. 120-140.

Narin, F., Olivastro, D., and Stevens, K. A., (1994) "Bibliometrics -Theory, Practice, and Problems", Evaluation Review, 18:1, February, p. 65-76.

Redner, S. (1998). "How Popular is your Paper? An Emperical Study of the Citation Distribution." Eur. Phys. J. B 4, 131.

Steele, T. W., (2000). "The Impact of Interdisciplinary Research in the Environmental Science: A Forestry Case Study", JASIS, 15 March, p. 476-484.

Tassey, G., (1999) "Lessons Learned About The Methodology Of Economic Impact Studies: The NIST Experience," Evaluation and Program Planning, **22, p.** 113—119.

VI. APPENDIX TO APPENDIX 7-H

Tables A1 to A4 contain the most frequent citing authors for the four sets of papers.

TABLE A1 – BF CITING AUTHORS

BF Citing Authors		
Citing Author	Citing Times	Percentage
Herrmann, HJ	16	13
Jaeger, HM	11	9
Nagel, SR	11	9
Zhang, ZP	11	9
Nicodemi, M	10	8

TABLE A2 – UF CITING AUTHORS

UF Citing Authors		
Citing Author	Citing Times	Percentage
Herrmann, HJ	24	8
Nicodemi, M	14	5
Rahchenbach, J	11	4
Mehta, A	11	4
Makse, HA	11	4
Behringer, RP	11	4
Duran, J	10	3
Luding, S	9	3
Coniglio, A	8	2
Clement, E	8	2

TABLE A3 – MA CITING AUTHORS

MA Citing Authors		
Citing Author	Citing Times	Percentage
Nascu C	7	0.12
Pop I	7	0.12
Bhushan S	6	0.10
Ionescu V	5	0.08

TABLE A4 UA Citing Authors

UA Citing Authors		
Citing Author	Citing Times	Percentage
Rud, VY	8	9
Wada, T	8	9
Negami, T	7	8
ZUNGER, A	6	7
Kohara, N	5	6
Schock, HW	5	6
Tanaka, T	5	6
Yamaguchi, T	5	6
Yoshida, A	5	6

Tables A5 to A8 contain frequencies of most cited papers in the citing papers of the four different sets.

TABLE A5 – FREQUENCIES OF REFERENCES IN BF CITING PAPERS

Frequencies of References in BF Citing Papers		
Paper	**Times**	
MEHTA A, 1989, PHYSICA A, V157, P1091	**63**	**52.9%**
MEHTA A, 1991, PHYS REV LETT, V67, P394	**42**	**35.3%**
JAEGER HM, 1992, SCIENCE, V255, P1523	**37**	**31.1%**
EVESQUE P, 1989, PHYS REV LETT, V62, P44	**33**	**27.7%**
ROSATO A, 1987, PHYS REV LETT, V58, P1038	**33**	**27.7%**
BARKER GC, 1992, PHYS REV A, V45, P3435	**32**	**26.9%**
JAEGER HM, 1989, PHYS REV LETT, V62, P40	**32**	**26.9%**
EDWARDS SF, 1989, PHYSICA A, V157, P1080	**28**	**23.5%**
LAROCHE C, 1989, J PHYS-PARIS, V50, P699	**23**	**19.3%**

MEHTA A, 1996, PHYS REV E, V53, P92	23	19.3%
KNIGHT JB, 1995, PHYS REV E, V51, P3957	22	18.5%
CAMPBELL CS, 1990, ANNU REV FLUID MECH, V22, P57	21	17.6%
EDWARDS SF, 1991, J STAT PHYS, V62, P889	21	17.6%
REYNOLDS O, 1885, PHILOS MAG 5, V20, P469	20	16.8%
BAXTER GW, 1989, PHYS REV LETT, V62, P2825	19	16.0%
THOMPSON PA, 1991, PHYS REV LETT, V67, P1751	19	16.0%
CLEMENT E, 1991, EUROPHYS LETT, V16, P133	18	15.1%
FARADAY M, 1831, PHIL T R SOC LONDON, V52, P299	18	15.1%
KNIGHT JB, 1993, PHYS REV LETT, V70, P3728	18	15.1%
MEHTA A, 1994, GRANULAR MATTER	18	15.1%
BARKER GC, 1993, PHYS REV E, V47, P184	17	14.3%
GALLAS JAC, 1992, PHYS REV LETT, V69, P1371	17	14.3%
JAEGER HM, 1996, REV MOD PHYS, V68, P1259	17	14.3%

TABLE A6 – FREQUENCIES OF REFERENCES IN UA CITING PAPERS

Frequencies of References in UA Citing Papers		
Paper	Times	
GABOR AM, 1994, APPL PHYS LETT, V65, P198	35	39.8%
HEDSTROM J, 1993, P 23 IEEE PHOT SPEC, P364	26	29.5%
TUTTLE JR, 1995, PROG PHOTOVOLTAICS, V3, P383	26	29.5%
TUTTLE JR, 1995, J APPL PHYS, V77, P153	25	28.4%
SCHMID D, 1993, J APPL PHYS, V73, P2902	20	22.7%
ROCKETT A, 1991, J APPL PHYS, V70, PR81	17	19.3%
STOLT L, 1993, APPL PHYS LETT, V62, P597	14	15.9%
SHAY JL, 1975, TERNARY CHALCOPYRITE	12	13.6%
KLENK R, 1993, ADV MATER, V5, P144	10	11.4%
NELSON AJ, 1995, J APPL PHYS, V78, P269	10	11.4%
BOEHNKE UC, 1987, J MATER SCI, V22, P1635	9	10.2%
CONTRERAS MA, 1994, PROG PHOTOVOLTAICS R, V2, P287	9	10.2%
FEARHEILEY ML, 1986, SOL CELLS, V16, P91	9	10.2%
JAFFE JE, 1984, PHYS REV B, V29, P1882	8	9.1%
NELSON AJ, 1993, J APPL PHYS, V74, P5757	8	9.1%
TUTTLE JR, 1996, MATER RES SOC SYMP P, V426, P143	8	9.1%

TABLE A7 – FREQUENCIES OF REFERENCES IN MA CITING PAPERS

Frequencies of References in MA Citing Papers		
Paper	**Times**	
NAIR PK, 1989, J PHYS D APPL PHYS, V22, P829	**23**	**25.84%**
NAIR PK, 1988, SEMICOND SCI TECH, V3, P134	**20**	**22.47%**
NAIR MTS, 1994, J APPL PHYS, V75, P1557	**15**	**16.85%**
KAUR I, 1980, J ELECTROCHEM SOC, V127, P943	**10**	**11.24%**
MONDAL A, 1983, SOL ENERG MATER, V7, P431	**10**	**11.24%**
CHOPRA KL, 1983, THIN FILM SOLAR CELL	**9**	**10.11%**
BUBE RH, 1960, PHOTOCONDUCTIVITY SO	**8**	**8.99%**
NAIR MTS, 1989, SEMICOND SCI TECH, V4, P191	**8**	**8.99%**

TABLE A8 – FREQUENCIES OF REFERENCES IN UF CITING PAPERS

Frequencies of References in UF Citing Papers		
Paper	**Times**	
JAEGER HM, 1992, SCIENCE, V255, P1523	**307**	**100%**
EVESQUE P, 1989, PHYS REV LETT, V62, P44	**75**	**24.4%**
GALLAS JAC, 1992, PHYS REV LETT, V69, P1371	**72**	**23.4%**
CHOO K, 1997, PHYS REV LETT, V79, P2975	**68**	**22.1%**
KNIGHT JB, 1993, PHYS REV LETT, V70, P3728	**68**	**22.1%**
ROSATO A, 1987, PHYS REV LETT, V58, P1038	**64**	**20.8%**
CAMPBELL CS, 1990, ANNU REV FLUID MECH, V22, P57	**62**	**20.8%**
TAGUCHI YH, 1992, PHYS REV LETT, V69, P1367	**56**	**18.2%**
JAEGER HM, 1989, PHYS REV LETT, V62, P40	**52**	**16.9%**
BAXTER GW, 1989, PHYS REV LETT, V62, P2825	**52**	**16.9%**
THOMPSON PA, 1991, PHYS REV LETT, V67, P1751	**51**	**16.6%**
BAK P, 1987, PHYS REV LETT, V59, P381	**48**	**15.6%**
CUNDALL PA, 1979, GEOTECHNIQUE, V29, P47	**48**	**15.6%**
CLEMENT E, 1992, PHYS REV LETT, V69, P1189	**47**	**15.3%**
JAEGER HM, 1996, REV MOD PHYS, V68, P1259	**43**	**14.0%**
DOUADY S, 1989, EUROPHYS LETT, V8, P621	**43**	**14.0%**
LAROCHE C, 1989, J PHYS-PARIS, V50, P669	**42**	**13.7%**
WILLIAMS JC, 1976, POWDER TECHNOL, V15, P 245	**41**	**13.4%**
HAFF PK, 1983, J FLUID MECH, V134, P401	**38**	**12.4%**
FARADAY M, 1831, PHIL T R SOC LONDON, V52, P299	**37**	**12.5%**

BAGNOLD RA, 1954, P ROY SOC LOND A MAT, V225, P49	**37**	**12.5%**

APPENDIX 7-I

MACROMOLECULE MASS SPECTROMETRY: CITATION MINING OF USER DOCUMENTS [Kostoff et al, 2004d]

1) ABSTRACT

Identifying the users and impact of research is important for research performers, managers, evaluators, and sponsors. It is important to know whether the audience reached is the audience desired. It is useful to understand the technical characteristics of the other research/ development/ applications impacted by the originating research, and to understand other characteristics (names, organizations, countries) of the users impacted by the research. Because of the many indirect pathways through which fundamental research can impact applications, identifying the user audience and the research impacts can be very complex and time consuming.

The purpose of this Appendix is to identify the literature pathways through which two highly-cited papers of 2002 Chemistry Nobel Laureates Fenn and Tanaka impacted other research, technology development, and applications, and to identify the technical and infrastructure characteristics of the user population.

Citation Mining, an integration of citation bibliometrics and text mining, was applied to the >1600 first generation Science Citation Index (SCI) citing papers to Fenn's 1989 Science paper on Electrospray Ionization for Mass Spectrometry, and to the >400 first generation SCI citing papers to Tanaka's 1988 Rapid Communications in Mass Spectrometry paper on Laser Ionization Time-of-Flight Mass Spectrometry. Text mining was performed on the citing papers to identify the technical areas impacted by the research, the relationships among these technical areas, and relationships among the technical areas and the infrastructure (authors, journals, organizations). Bibliometrics was performed on the citing papers to profile the user characteristics.

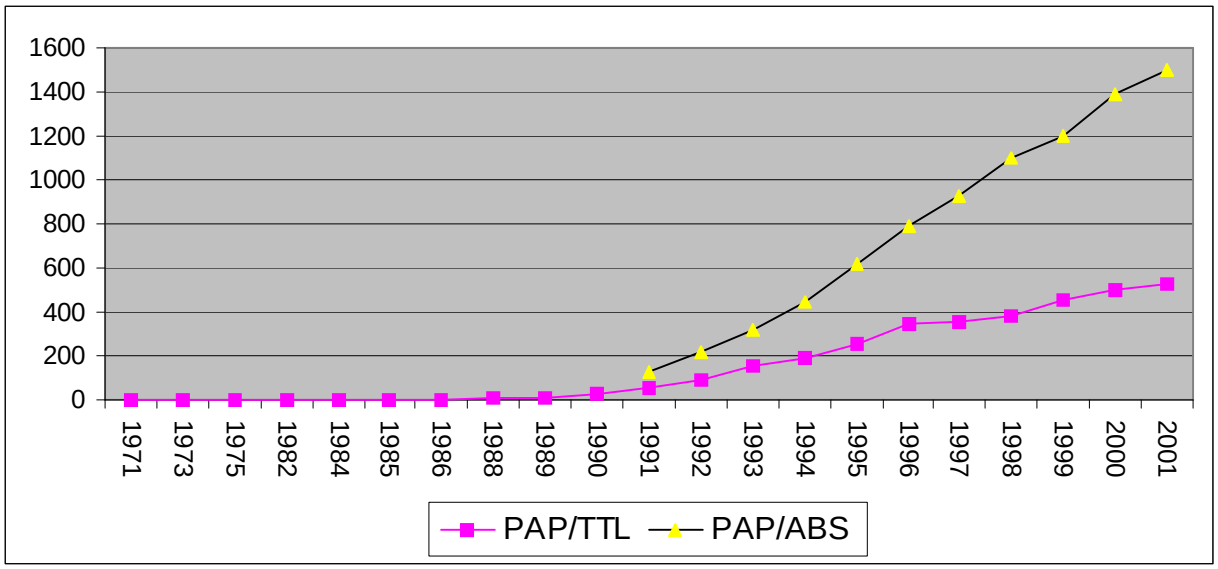
The combination of citation bibliometrics and text mining provides a synergy unavailable with each approach taken independently. Furthermore, text mining is a REQUIREMENT for a feasible comprehensive research impact determination. The integrated multi-generation citation analysis required for broad research impact determination of highly cited papers will produce thousands or tens or hundreds of thousands of citing paper Abstracts. Text mining allows the impacts of research on advanced development categories and/ or extra-discipline categories to be obtained without having to read all these citing paper Abstracts. The multi-field bibliometrics provide multiple documented perspectives on the users of the research, and indicate whether the documented audience reached is the desired target audience.

2) BACKGROUND

The 2002 Nobel Prize in Chemistry was shared by John B. Fenn, Koichi Tanaka, and Kurt Wuthrich for their work in developing methods to enable the identification and structural analysis of biological macromolecules. In particular, Fenn and Tanaka focused on soft desorption ionization methods. Fenn concentrated on electrospray ionization (1-7), and Tanaka concentrated on soft laser desorption (8-10).

The impact of these researchers can be viewed from a literature perspective. Figure 1A shows the growth in the SCI Electrospray Ionization Mass Spectrometry literature (retrieved by the query Electrospray AND (Mass OR Ion* OR Spectrometry)). The upper curve is based on papers retrieved by a query applied to all text fields (Title, Abstract, Keywords), while the lower curve is based on a query applied to the Title field only. Before 1991, Abstracts were not available for SCI papers.

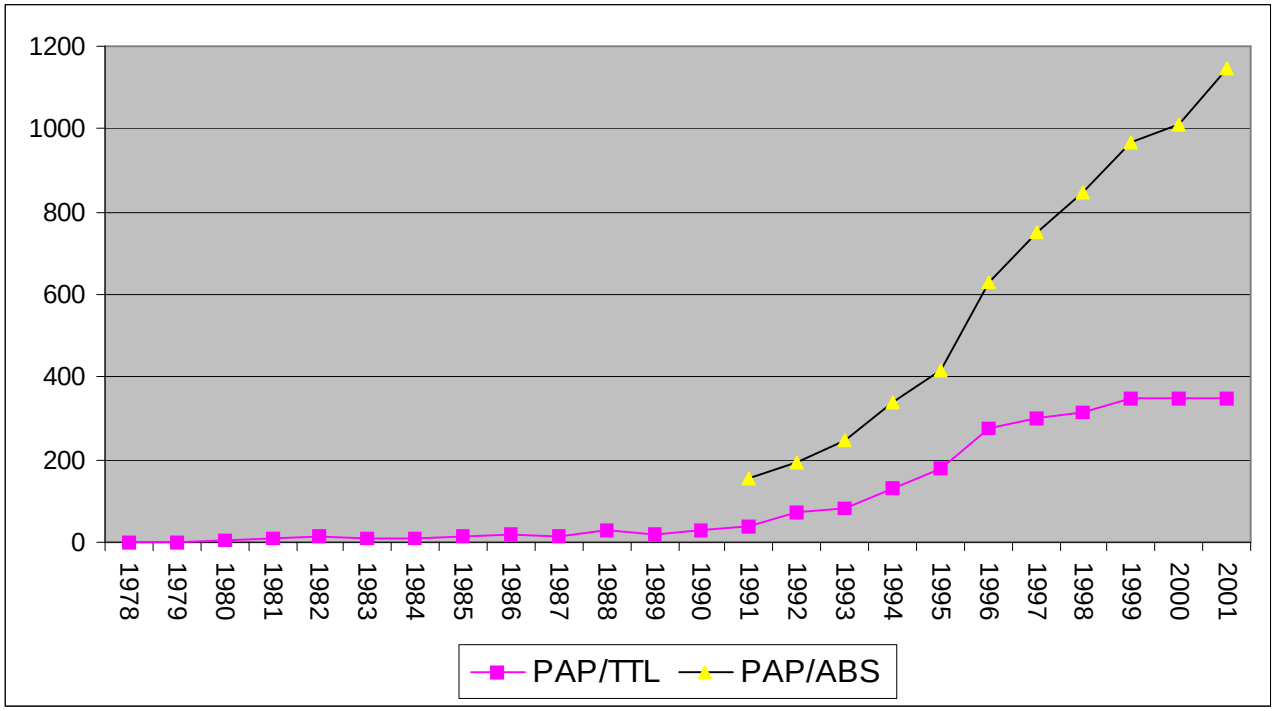
FIGURE 1A – GROWTH IN ELECTROSPRAY LITERATURE
(Papers per Year vs Time)



In the years that growth accelerated initially (1988-1990), essentially all the papers retrieved from the database cited one or more of Fenn's papers dating from 1984 (1-7). From the 'bottom-up' perspective, references 1-7 received a total of 151 citations between 1984 and 1990, of which 143 were from external groups. The top twenty of these 143 citing papers received over 150 citations apiece, with an aggregate second-generation citation total (for these top twenty alone) of 5400 citations.

Figure 1B shows the growth in the Laser Desorption Mass Spectrometry literature (retrieved by the query Laser AND Desorption AND (Ion* OR Mass Spectrometry).

FIGURE 1B – GROWTH IN SCI LASER DESORPTION LITERATURE
(Papers per Year vs Time)



In the years that growth accelerated initially (1990-1992), 145 papers were retrieved from the title search only. Of the top fifty cited papers of the 145 retrieved, ranging in citations from 983 to 33, Tanaka's 1988 paper was referenced in fifteen. Interestingly, one or more of Beavis's papers were referenced in 37 of these top fifty cited papers, and one or more of Karas' papers were referenced in 38 of these top fifty cited papers. From the 'bottom-up' perspective, reference 8 received a total of 69 citations between 1988 and 1992, of which all were from external groups. The top fourteen of these 69 citing papers received over 100 citations apiece, with an aggregate second-generation citation total (for these top fourteen alone) of 3140 citations.

References 1 to 8 have been cited highly. In particular, references 1-7 have received ~590, 210, 670, 210, 370, 1630, 890 citations respectively, by November 2002, and reference 8 has received 410 citations. The citing community can be viewed as a sub-set of the total user community. Identifying the characteristics of the citing community would provide one perspective on the diversity of *impact* that these papers have had or, more accurately, on the diversity of *citings* that these papers have had.

Citation Mining (11, 11a) is a technique developed for the purpose of characterizing the aggregate citing papers of a research unit. A research unit can consist of one paper, selected papers from an author, or selected papers from a group or technical discipline. In Citation Mining, text mining (12,13) analyses are performed on the aggregate citing papers. The bibliometrics component yields the infrastructure information (e.g., prolific authors, journals, institutions, countries, most cited authors, papers, journals, etc), and the computational linguistics component yields the pervasive technical thrusts and the relationships among the thrusts. A temporal component documents the dissemination of information to the research and user community. See (14) for an example of text mining applied to Electrochemical Power Sources.

The Science Citation Index (SCI) is a database that links papers (P1) in journals indexed by the SCI to other SCI papers (P2) that cite the original papers P1, and contains references (P3) in the original papers P1 as well. While the SCI accesses many of the premier research journals, it does not access all technical journals published. In the present study, the SCI is used to identify the citing papers to Fenn's and Tanaka's original papers. Thus, all the citing papers in the technical literature will not be identified, only those in journals accessed by the SCI.

This paper describes the application of Citation Mining to the subset of the most highly cited papers of Fenn (6) and Tanaka (8) referenced above, using the SCI as the source for citing papers. It was desired to examine papers that were cited highly, preferably with multi-discipline readership journals where possible, to obtain the broadest potential areas for application. Because the SCI did not use Abstracts until 1991, and because Abstract analysis is a key feature of Citation Mining, it was desired to examine papers published relatively close to 1991. Because temporal dissemination and impacts of the initial cited papers is also a key feature of citation mining, it was desired to limit the analysis to one paper from each researcher, in order to have a sharp starting point in time. Therefore, references (6) and (8) were selected as the seeds for the Citation Mining process.

Section 3 presents the Results, divided into a bibliometrics sub-section and a computational linguistics sub-section. Section 4 presents the Summary and Conclusions, and section 5 contains the References.

3) RESULTS

The results from the publications bibliometric analyses are presented in section 3.1, followed by the results from the citations bibliometrics analysis in section 3.2. Results from the computational linguistics analyses are shown in section 3.3. The SCI bibliometric fields incorporated into the database included, for each paper, the author, journal, institution, Keywords, and references.

3.1 Publication Statistics on Authors, Journals, Organizations, Countries

The first group of metrics presented is counts of papers published by different entities. These metrics can be viewed as output and productivity measures. They are not direct measures of research quality,

although there is some threshold quality level inferred, since these papers are published in the (typically) high caliber journals accessed by the SCI.

There were 1628 papers that cited Fenn's 1989 paper, and 410 papers that cited Tanaka's 1988 paper. Because the SCI did not start to publish Abstracts until 1991, and because not all citing papers have Abstracts, only 1433 of the Fenn citing papers in the SCI database contain Abstracts, and only 344 of the Tanaka citing papers contain Abstracts. The bibliometrics analyses are performed on the total number of citing papers, whereas the computational linguistics are performed on those papers with Abstracts.

3.1.1. Author Frequency Results

The 1628 Fenn citing papers contain 3602 different authors, and 6263 author listings, resulting in 3.8 author listings per paper. The 410 Tanaka citing papers contain 973 different authors and 1462 different author listings, resulting in 3.57 author listings per paper. The occurrence of each author's name on a paper is defined as an author listing. The number of author listings per paper is relatively high in either case, and seems to follow a trend set by earlier text mining studies. In four previous chemistry-related text mining studies (14-17), this ratio averaged over 3.5, while in three previous fluid mechanics-related text mining studies (18-20), this ratio averaged under 2.5. A high value of this ratio tends to indicate large teams characteristic of large experimental efforts, while a low value of this ratio tends to indicate small teams characteristic of individual theoretical or computational modeling efforts. The most prolific authors of the Fenn citing papers are listed in Table 1A, and the most prolific authors of the Tanaka citing papers are listed in Table 1B.

TABLE 1A – MOST PROLIFIC AUTHORS – FENN CITING PAPERS
(present institution listed)

AUTHOR	INSTITUTION	COUNTRY	# PAPERS
SMITH—RD	PACIFIC NW NATL LAB	USA	48
MCLUCKEY—SA	PURDUE UNIV	USA	43
MCLAFFERTY—FW	CORNELL UNIV	USA	42
LOO—JA	PFIZER GLOBAL R&D	USA	37
CLEMMER—DE	INDIANA UNIV	USA	34
COLTON—R	LA TROBE UNIV	AUSTRALIA	34
MANN—M	UNIV SO DENMARK	DENMARK	29
MUDDIMAN—DC	VCU	USA	26
ROEPSTORFF—P	ODENSE UNIV	DENMARK	26
TRAEGER—JC	LA TROBE UNIV	AUSTRALIA	26
WILLIAMS—ER	UNIV CAL BERKELEY	USA	22
HENION—JD	CORNELL UNIV	USA	20
MARSHALL—AG	FLORIDA STATE UNIV	USA	19
ARAKAWA—R	KANSAI UNIV	JAPAN	18
COUNTERMAN—AE	INDIANA UNIV	USA	18
STEPHENSON—JL	RES TRIANGLE INST	USA	18

VANBERKEL—GJ	OAK RIDGE NATL LAB	USA	18
CHAIT—BT	ROCKEFELLER UNIV	USA	17
LITTLE—DP	SEQUENOM, INC	USA	15
EDMONDS—CG	PACIFIC NW NATL LAB	USA	14
JOHNSON—RS	IMMUNEX R&D CORP	USA	14
SENKO—MW	FLORIDA STATE UNIV	USA	14

TABLE 1B – MOST PROLIFIC AUTHORS – TANAKA CITING PAPERS

AUTHOR	INSTITUTION	COUNTRY	# PAPERS
ZENOBI—R	SWISS FED INST TECH	SWITZERLAND	18
HILLENKAMP—F	UNIV MUNSTER	GERMANY	12
KARAS—M	UNIV FRANKFURT	GERMANY	12
COTTER—RJ	JHU	USA	11
GROTEMEYER—J	UNIV KIEL	GERMANY	9
KNOCHENMUSS—R	SWISS FED INST TECH	SWITZERLAND	9
WILKINS—CL	UNIV ARKANSAS	USA	9
DERRICK—PJ	UNIV WARWICK	UK	8
HERCULES—DM	VANDERBILT UNIV	USA	8
AMSTER—IJ	UNIV GEORGIA	USA	7
RUSSELL—DH	TEXAS A&M UNIV	USA	7
BAHR—U	JW GOETHE UNIV	GERMANY	6
BURLINGAME—AL	UNIV CAL SAN FRANCISCO	USA	6
CASTORO—JA	UNIV CAL RIVERSIDE	USA	6
DEAK—G	DEBRECEN UNIV MED	HUNGARY	6
FENSELAU—C	UNIV MARYLAND	USA	6
KEKI—S	LAJOS KOSSUTH UNIV	HUNGARY	6
KUHN—G	FED INST MAT RES & TEST	GERMANY	6
PERERA—IK	UNIV HULL	UK	6
SCHLAG—EW	TECH INST MUNCHEN	GERMANY	6
SUNDQVIST—BUR	UNIV UPPSALA	SWEDEN	6
WEIDNER—S	FED INST MAT RES & TEST	GERMANY	6
ZSUGA—M	DEBRECEN UNIV MED	HUNGARY	6

These regional distributions are very different. For the Fenn citing papers, of the 22 most prolific authors, seventeen are from the USA, two are from Australia, two are from Denmark, and one is from Japan. Fifteen are from universities, three are from research institutes, and four are from industry.

For the Tanaka citing papers, of the 23 most prolific authors, eight are from the USA, and the remainder are from Europe, mainly central Europe. Twenty are from universities, and three are from research institutes. No authors are common to the two lists of prolific citing authors. Why are there no prolific citing authors from Japan, and why are there no prolific citing authors from

industry, for Tanaka's research? This is surprising, since Tanaka is both from Japan and industry.

Two notes of caution. First, the institutions listed are typically the most recent at which the author can be found. Since many researchers have cycled through a number of institutions globally over the course of their careers, the author numbers may not compare exactly with the institution or country numbers shown later. Second, separate listing of authors does not mean that the papers are separate. For example, most, if not all, of the papers by Hillenkamp and Karas in Table 1B are co-authored.

3.1.2 Journal frequency results

There were 317 different journals represented in the Fenn citing papers, with an average of 5.14 papers per journal. There were 112 different journals represented in the Tanaka citing papers, with an average of 3.67 papers per journal. These ratios are about half the values as the previous chemistry text mining studies, but on the same order as the previous fluid mechanics text mining studies. The previous text mining studies were thematic (i.e., all the papers had the common themes of the search query), while the present aggregation of citing papers is not thematic in the same sense. Given the thematic focus of many technical journals, it is reasonable that the citing papers would be distributed over a wider group of journals, with a wider aggregate thematic base. The journals containing the most Fenn citing papers are listed in Table 2A, and the journals containing the most Tanaka citing papers are listed in Table 2B.

TABLE 2A – JOURNALS CONTAINING MOST FENN CITING PAPERS

JOURNAL	# PAPERS
ANALYTICAL CHEMISTRY	193
JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY	139
RAPID COMMUNICATIONS IN MASS SPECTROMETRY	132
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY	72
JOURNAL OF MASS SPECTROMETRY	68
ANALYTICAL BIOCHEMISTRY	37
INTERNATIONAL JOURNAL OF MASS SPECTROMETRY	33
JOURNAL OF CHROMATOGRAPHY A	29
INTERNATIONAL JOURNAL OF MASS SPECTROMETRY AND ION PROCESSES	26
BIOCHEMISTRY	25
JOURNAL OF BIOLOGICAL CHEMISTRY	23
ELECTROPHORESIS	23
INORGANICA CHIMICA ACTA	21

PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	20
PROTEIN SCIENCE	19
JOURNAL OF AEROSOL SCIENCE	19
BIOLOGICAL MASS SPECTROMETRY	19
ANALYTICA CHIMICA ACTA	18
MASS SPECTROMETRY REVIEWS	17
EUROPEAN JOURNAL OF BIOCHEMISTRY	17

TABLE 2B – JOURNALS CONTAINING MOST TANAKA CITING PAPERS

JOURNAL	# PAPERS
RAPID COMMUNICATIONS IN MASS SPECTROMETRY	70
ANALYTICAL CHEMISTRY	56
JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY	34
INTERNATIONAL JOURNAL OF MASS SPECTROMETRY AND ION PROCESSES	20
JOURNAL OF MASS SPECTROMETRY	16
MACROMOLECULES	14
ORGANIC MASS SPECTROMETRY	13
INTERNATIONAL JOURNAL OF MASS SPECTROMETRY	11
JOURNAL OF CHROMATOGRAPHY A	7
FRESENIUS JOURNAL OF ANALYTICAL CHEMISTRY	6
ANALYTICA CHIMICA ACTA	6
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY	5
BIOLOGICAL MASS SPECTROMETRY	5
EUROPEAN MASS SPECTROMETRY	5
JOURNAL OF BIOLOGICAL CHEMISTRY	5
MASS SPECTROMETRY REVIEWS	4
REVIEW OF SCIENTIFIC INSTRUMENTS	4
JOURNAL OF PHYSICAL CHEMISTRY B	4

In both cases, the most prolific journals focus on mass spectrometry, chemistry, and biology. Three journals stand out as the first tier for containing the most cited papers: ANALYTICAL CHEMISTRY, JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY, RAPID COMMUNICATIONS IN MASS SPECTROMETRY. Twelve journals are in common between the two lists. The Fenn citing journals not in common tend to focus on biology/ biochemistry (ANALYTICAL BIOCHEMISTRY, BIOCHEMISTRY, PROTEIN SCIENCE, EUROPEAN JOURNAL OF BIOCHEMISTRY), while the Tanaka citing journals not in common tend to focus on the technique/ instrumentation (REVIEW OF SCIENTIFIC INSTRUMENTS, ORGANIC MASS SPECTROMETRY, EUROPEAN MASS SPECTROMETRY).

3.1.3 Institution frequency results

A similar process was used to develop a frequency count of institutional address appearances. It should be noted that many different organizational components may be included under the single organizational heading (e.g., Harvard Univ could include the Chemistry Department, Biology Department, Physics Department, etc.). Identifying the higher level institutions is instrumental for these DT studies. Once they have been identified through bibliometric analysis, subsequent measures may be taken (if desired) to identify particular departments within an institution.

There were 801 different institutions represented in the Fenn citing papers, with an average of 2.03 papers per institution. There were 315 different institutions represented in the Tanaka citing papers, with an average of 1.3 papers per institution. The institutions producing the most Fenn citing papers are listed in Table 3A, and the institutions producing the most Tanaka citing papers are listed in Table 3B.

TABLE 3A – INSTITUTIONS PRODUCING MOST FENN CITING PAPERS

INSTITUTION	COUNTRY	# PAPERS
CORNELL UNIV	USA	66
OAK RIDGE NATL LAB	USA	52
BATTELLE MEM INST	USA	47
VIRGINIA COMMONWEALTH UNIV	USA	41
YALE UNIV	USA	38
INDIANA UNIV	USA	38
UNIV WASHINGTON	USA	36
LA TROBE UNIV	AUSTRALIA	35
ODENSE UNIV	DENMARK	33
OSAKA UNIV	JAPAN	29
NATL RES COUNCIL CANADA	CANADA	26
UNIV ALBERTA	CANADA	25
PURDUE UNIV	USA	25
UNIV CALIF SAN FRANCISCO	USA	25
UNIV CALIF BERKELEY	USA	22
FLORIDA STATE UNIV	USA	22
UNIV MICHIGAN	USA	18
ROCKEFELLER UNIV	USA	17
NYU	USA	17
CALTECH	USA	17

TABLE 3B – INSTITUTIONS PRODUCING MOST TANAKA CITING PAPERS

INSTITUTION	COUNTRY	# PAPERS

SWISS FED INST TECH	SWITZERLAND	18
UNIV MUNSTER	GERMANY	14
JOHNS HOPKINS UNIV	USA	12
UNIV GEORGIA	USA	11
TECH UNIV MUNICH	GERMANY	9
UNIV CALIF RIVERSIDE	USA	9
UNIV WARWICK	UK	9
UNIV PITTSBURGH	USA	7
UNIV CALIF SAN FRANCISCO	USA	6
UNIV UPPSALA	SWEDEN	6
UNIV VIENNA	AUSTRIA	6
INDIANA UNIV	USA	6
UNIV ILLINOIS	USA	6
CNR	ITALY	6
LOUISIANA STATE UNIV	USA	5
ROHM & HAAS CO	USA	5
ARIZONA STATE UNIV	USA	5
TEXAS A&M UNIV	USA	5
ROCKEFELLER UNIV	USA	5
OSAKA UNIV	JAPAN	5

Of the twenty institutions producing the most Fenn citing papers, seventeen are from North America, one from Europe, and two from the Far East. Eighteen are universities, and two are research institutes. Of the twenty institutions producing the most Tanaka citing papers, twelve are from the USA, seven are from Europe, and one is from Japan. Eighteen are universities, one is a research institute, and one is from industry. Four institutions are in common between the two lists: UNIV CAL SAN FRANCISCO, INDIANA UNIV, ROCKEFELLER UNIV, OSAKA UNIV.

3.1.4 Country frequency results

There are 51 different countries listed in the Fenn citing papers, and 36 different countries listed in the Tanaka citing papers. The countries producing the most Fenn citing papers are listed in Table 4A, and the countries producing the most Tanaka citing papers are listed in Table 4B. The dominance of a handful of countries is clearly evident.

TABLE 4A – COUNTRIES PRODUCING THE MOST FENN CITING PAPERS

COUNTRY	# PAPERS
USA	917
CANADA	119
GERMANY	115
JAPAN	102
ENGLAND	83

FRANCE	80
AUSTRALIA	69
DENMARK	42
NETHERLANDS	36
SWEDEN	35
SWITZERLAND	35
PEOPLES R CHINA	28
ITALY	26
BELGIUM	22
SPAIN	15
RUSSIA	12
SCOTLAND	12
HUNGARY	11
NEW ZEALAND	10
TAIWAN	8

TABLE 4B – COUNTRIES PRODUCING THE MOST TANAKA CITING PAPERS

COUNTRY	# PAPERS
USA	193
GERMANY	48
ENGLAND	33
JAPAN	31
CANADA	23
SWITZERLAND	23
NETHERLANDS	12
FRANCE	11
SWEDEN	10
HUNGARY	8
ITALY	8
AUSTRALIA	6
AUSTRIA	6
SCOTLAND	6
BELGIUM	5
PEOPLES R CHINA	5
ISRAEL	4
RUSSIA	4

The USA clearly dominates. The next tier is high on both lists (GERMANY, ENGLAND, JAPAN, CANADA), with Switzerland appearing high on the Tanaka citing list. Thus, while Japan is not very visible in terms of prolific citing authors or institutions, especially with respect to Tanaka's paper, it has reasonable representation in terms of country citations. This implies a diverse group of citing authors in Japan, with the exception of the group at Osaka University.

Figure 1A contains a co-occurrence matrix of the top 15 countries listed in the Fenn citing papers, and Figure 1B contains a co-occurrence matrix of the top 15 countries listed in the Tanaka citing papers.

In terms of absolute numbers of co-authored papers, the USA major partners are Canada, Japan, Germany, England, and France. Additionally, the USA is the major partner for ten of the countries, the exceptions being Australia, Belgium, Holland, and China.

FIGURE 1A – COUNTRY CO-OCCURRENCE MATRIX FOR FENN CITING PAPERS

COUNTRY	AUSTRAL	BELGIUM	CANADA	DENMARK	ENGLAND	FRANCE	GERMANY	ITALY	JAPAN	HOLLAND	CHINA	SPAIN	SWEDEN	SWITZER	USA
AUSTRALIA	69	0	1	0	3	1	2	0	0	0	0	0	0	0	1
BELGIUM	0	22	0	0	0	3	1	0	0	1	0	1	0	0	1
CANADA	1	0	119	1	4	8	1	0	0	0	1	0	0	0	20
DENMARK	0	0	1	42	0	3	4	0	1	1	0	0	1	0	4
ENGLAND	3	0	4	0	83	4	3	1	1	4	0	0	0	3	15
FRANCE	1	3	8	3	4	80	2	1	1	1	0	0	1	3	11
GERMANY	2	1	1	4	3	2	115	0	4	1	0	0	1	4	15
ITALY	0	0	0	0	1	1	0	26	0	0	0	0	1	1	2
JAPAN	0	0	0	1	1	1	4	0	102	0	1	0	0	0	16
HOLLAND	0	1	0	1	4	1	1	0	0	36	0	1	1	1	0
CHINA	0	0	1	0	0	0	0	0	1	0	28	0	2	0	0
SPAIN	0	1	0	0	0	0	0	0	0	1	0	15	0	0	2
SWEDEN	0	0	0	1	0	1	1	1	0	1	2	0	35	0	5
SWITZERLAND	0	0	0	0	3	3	4	1	0	1	0	0	0	35	5
USA	1	1	20	4	15	11	15	2	16	0	0	2	5	5	917

FIGURE 1B – COUNTRY CO-OCCURRENCE MATRIX FOR TANAKA CITING PAPERS

COUNTRY	AUSTRAL	AUSTRIA	BELGIUM	CANADA	ENGLAND	FRANCE	GERMANY	HUNGARY	ITALY	JAPAN	HOLLAND	CHINA	SCOTLAND	SWEDEN	SWITZER	USA
AUSTRALIA	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AUSTRIA	0	6	0	0	0	1	1	0	0	0	0	0	0	0	0	0
BELGIUM	0	0	5	0	0	0	0	1	0	0	0	0	0	0	0	1

CANADA	0	0	0	23	1	0	1	0	0	0	0	0	0	0	0	6
ENGLAND	0	0	0	1	33	0	1	1	0	0	1	0	1	1	2	4
FRANCE	0	1	0	0	0	11	1	0	0	0	0	0	0	0	0	1
GERMANY	0	1	0	1	1	1	48	1	0	0	0	0	0	0	0	7
HUNGARY	0	0	1	0	1	0	1	8	0	0	0	0	0	0	0	1
ITALY	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	1
JAPAN	0	0	0	0	0	0	0	0	0	31	0	0	0	0	0	3
HOLLAND	0	0	0	0	1	0	0	0	0	0	12	0	0	0	0	0
CHINA	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	1
SCOTLAND	0	0	0	0	1	0	0	0	0	0	0	0	6	0	2	0
SWEDEN	0	0	0	0	1	0	0	0	0	0	0	0	0	10	0	2
SWITZERLAND	0	0	0	0	2	0	0	0	0	0	0	0	2	0	23	0
USA	0	0	1	6	4	1	7	1	1	3	0	1	0	2	0	193

In terms of absolute numbers of co-authored Fenn-citing papers, the USA major partners are Canada, Japan, Germany, England, and France. Additionally, the USA is the major partner for ten of the countries, the exceptions being Australia, Belgium, Holland, and China.

In terms of absolute numbers of co-authored Tanaka-citing papers, the USA major partners are Germany, Canada, England, and Japan. Additionally, the USA is the major partner for nine of the countries, the exceptions being Australia, Austria, Holland, Scotland, and Switzerland.

3.2 Citation Statistics on Authors, Papers, and Journals

The second group of metrics presented is counts of citations to papers published by different entities. While citations are ordinarily used as impact or quality metrics [Garfield, 1985], *much caution needs to be exercised in their frequency count interpretation, since there are numerous reasons why authors cite or do not cite particular papers* [Kostoff, 1998; MacRoberts and MacRoberts, 1996].

The citations in all the retrieved SCI papers were aggregated, the authors, specific papers, years, journals, and countries cited most frequently were identified, and were presented in order of decreasing frequency. A small percentage of any of these categories received large numbers of citations.

3.2.1 Author citation frequency results

The most highly cited authors in the Fenn citing papers are listed in Table 5A, and the most highly cited authors in the Tanaka citing papers are listed in Table 5B. These represent the authors who are highly co-cited with Fenn and Tanaka, respectively. Only the first authors of the cited papers in the Fenn citing papers are listed.

TABLE 5A – MOST CITED AUTHORS IN FENN CITING PAPERS
(cited by other papers in this database only)

AUTHOR	INSTITUTION	COUNTRY	# CITES
FENN JB	VCU	USA	1982
SMITH RD	PACIFIC NW NATL LAB	USA	1134
LOO JA	PFIZER GLOBAL R&D	USA	875
KARAS M	UNIV FRANKFURT	GERMANY	600
MCLUCKEY SA	PURDUE UNIV	USA	541
MANN M	UNIV SO DENMARK	DENMARK	450
BIEMANN K	MIT	USA	343
CHOWDHURY SK	SANOFI WINTHROP INC	USA	302
COVEY TR	SCIEX LTD	CANADA	297
KATTA V	AMGEN INC	USA	287
YAMASHITA M	TOKAI UNIV	JAPAN	285
HUNT DF	UNIV VIRGINIA	USA	279
VANBERKEL GJ	OAK RIDGE NATL LAB	USA	266
COLTON R	LA TROBE UNIV	AUSTRALIA	258
MARSHALL AG	FLORIDA STATE UNIV	USA	252
MCLAFFERTY FW	CORNELL UNIV	USA	239
HILLENKAMP F	UNIV MUNSTER	GERMANY	235
GANEM B	CORNELL UNIV	USA	217
BRUINS AP	UNIV GRONINGEN	NETHERLANDS	211
WILM M	EUROPEAN MOL BIOL LAB	GERMANY	203
BEAVIS RC	NYU	USA	202

TABLE 5B – MOST CITED AUTHORS IN TANAKA CITING PAPERS
(cited by other papers in this database only)

AUTHOR	INSTITUTION	COUNTRY	# CITES
KARAS M	UNIV FRANKFURT	GERMANY	659
BEAVIS RC	NYU	USA	422
TANAKA K	SHIMADZU CORP	JAPAN	410
HILLENKAMP F	UNIV MUNSTER	GERMANY	242
SPENGLER B	UNIV GIESSEN	GERMANY	201
DANIS PO	ROHM AND HAAS CO	USA	143
MONTAUDO G	UNIV PISA	ITALY	134
COTTER RJ	JHU	USA	114
VERTES A	GWU	USA	111
FENN JB	VCU	USA	102
NELSON RW	INTRINS BIOPROBES INC	USA	97
BARBER M	UMIST	UK	94
OVERBERG A	UNIV MUNSTER	GERMANY	89
SMITH RD	PACIFIC NW NATL LAB	USA	82
BOESL U	TECH UNIV MUNICH	GERMANY	75

JUHASZ P	PERCEPT BIOSYS	USA	70
STRUPAT K	UNIV MUNSTER	GERMANY	69
CHAIT BT	ROCKEFELLER UNIV	USA	69
GROTEMEYER J	UNIV KIEL	GERMANY	64
LI L	UNIV ALBERTA	CANADA	61
BENNINGHOVEN A	UNIV MUNSTER	GERMANY	61

In the Fenn citing papers, Fenn is cited almost twice as much as the next ranked author. This is due to the citation of Fenn's other first-authored papers between 1984 and 1989, in addition to the citation of the Science article. The next tier, Smith and Loo, was a very prolific and highly cited group working on different mass spectrometry techniques, including electrospray ionization.

In the Tanaka citing papers, Tanaka actually ranks third in number of first-author citations. Karas of Frankfurt ranks first. This is due to two factors. In 1985, Karas, in conjunction with Hillenkamp of Munster, showed that an absorbing matrix could be used to vaporize small molecules without chemical degradation. Additionally, in 1988, Karas and Hillenkamp reported a MALDI approach applied to proteins shortly after Tanaka's paper was published. Thus, the papers that cite Tanaka's paper also tend to cite the groundwork papers of Karas as well as his large molecule mass determination papers. Additionally, due to a series of highly-cited papers by Beavis in the early 1990s on laser desorption mass spectrometry, many of the papers that cite Tanaka tend to multiply cite Beavis. This large co-citation of Karas and Beavis with Tanaka was alluded to in the Introduction. It was shown that, of the top fifty cited laser desorption mass spectrometry papers produced in the early high growth years, Tanaka's paper was referenced in fifteen, while Beavis's papers were referenced in 37 and Karas's papers were referenced in 38. Additionally, since Karas and Hillenkamp tended to publish jointly in the papers listed here, the above statements about Karas should apply equally well to Hillenkamp.

There are five names in common between the two lists (FENN, SMITH, KARAS, BEAVIS, HILLENKAMP). This reflects the broad interests in, and contributions these individuals have made to, mass spectrometry.

Of the 21 most cited authors in the Fenn citing papers, fourteen are from universities, three are from research institutions, and four are from industry. Of the 21 most cited authors in the Tanaka citing papers, sixteen are from universities, one is from a research institute, and four are from industry. This relatively high fraction (~20%) of cited papers from industry suggests relatively applied citing papers. The validity of this assumption is confirmed in the section on temporal citing patterns.

Finally, while Central Europe plays a modest role in the reference source for the Fenn list, it continues to play a much stronger role for the Tanaka list.

The citation data for authors and journals represents citations generated only by the specific records extracted from the SCI database for this study. It does not represent all the citations received by the

references in those records; these references in the database records could have been cited additionally by papers in other technical disciplines.

3.2.2 Document citation frequency results

The most highly cited documents in the Fenn citing papers are listed in Table 6A, and the most highly cited documents in the Tanaka citing papers are listed in Table 6B.

TABLE 6A – MOST CITED DOCUMENTS IN FENN CITING PAPERS

(total citations listed in SCI)

AUTHOR	YEAR	JOURNAL	VOLUME	TOT CITES
FENN JB	1989	SCIENCE	V246,P64	1628
ELECTROSPRAY IONIZATION FOR MASS-SPECTROMETRY OF LARGE BIOMOLECULES				
SMITH RD	1990	ANAL CHEM	V62,P882	854
BIOCHEMICAL MASS-SPECTROMETRY - ELECTROSPRAY IONIZATION				
KARAS M	1988	ANAL CHEM	V60,P2299	1329
LASER DESORPTION IONIZATION OFLARGE PROTEINS				
FENN JB	1990	MASS SPECTROM REVIEW	V9,P37	879
ELECTROSPRAY IONIZATION				
SMITH RD	1991	MASS SPECTROM REVIEW	V10,P359	482
ELECTROSPRAY IONIZATION MASS SPECTROMETRY FOR LARGE POLYPEPTIDES				
COVEY TR	1988	RAPID COMM MASS SPEC	V2,P249	486
PROTEIN MOLECULAR WEIGHTS BY ION SPRAY MASS SPECTROMETRY				
YAMASHITA M	1984	J PHYS CHEM	V88,P4451	576
ELECTROSPRAY ION-SOURCE - FREE-JET THEME				
WHITEHOUSE CM	1985	ANAL CHEM	V57,P675	653
ELECTROSPRAY INTERFACE FOR LIQUID CHROMATOGRAPHS AND MASS SPECTROMETERS				
HILLENKAMP F	1991	ANAL CHEM	V63,PA1193	983
MATRIX-ASSISTED LASER DESORPTION IONIZATION MASS-SPECTROMETRY OF BIOPOLYMERS				
MANN M	1989	ANAL CHEM	V61,P1702	361
MASS-SPECTRA OF MULTIPLY CHARGED IONS				
BRUINS AP	1987	ANAL CHEM	V59,P2642	619
LIQUID CHROMATOGRAPHY/ATMOSPHERIC PRESSURE IONIZATION MASS-SPECTROMETRY				
DOLE M	1968	J CHEM PHYS	V49,P2240	357
MOLECULAR BEAMS OF MACROIONS				
ROEPSTORFF P	1984	BIOMED MASS SPECTROM	V11,P601	1058
COMMON NOMENCLATURE FOR SEQUENCE IONS IN MASS-SPECTRA OF PEPTIDES				
CHOWDHURY SK	1990	J AM CHEM SOC	V112,P9012	230
PROBING CONFORMATIONAL-CHANGES IN PROTEINS BY MASS-SPECTROMETRY				
CHOWDHURY SK	1990	RAPID COMM MASS SPEC	V4,P81	223
ELECTROSPRAY-IONIZATION MASS-SPECTROMETER				
WILM MS	1994	INT J MASS SPECTROM	V136,P167	286
ELECTROSPRAY AND TAYLOR-CONE THEORY, DOLES BEAM OF MACROMOLECULES				
GANEM B	1991	J AM CHEM SOC	V113,P6294	248
DETECTION OF NONCOVALENT RECEPTOR LIGAND COMPLEXES BY MASS-SPECTROMETRY				
HUNT DF	1986	P NATL ACAD SCI USA	V83,P6233	530
PROTEIN SEQUENCING BY TANDEM MASS-SPECTROMETRY				

IRIBARNE JV 1976 J CHEM PHYS V64,P2287 313
EVAPORATION OF SMALL IONS FROM CHARGED DROPLETS

TABLE 6B – MOST CITED DOCUMENTS IN TANAKA CITING PAPERS

(total citations listed in SCI)

AUTHOR	YEAR	JOURNAL	VOLUME	TOT CITES
TANAKA K	1988	RAPID COMM MASS SPEC	V2,P151	410
LASER IONIZATION TIME-OF-FLIGHT MASS SPECTROMETRY				
KARAS M	1988	ANAL CHEM	V60,P2299	1329
LASER DESORPTION IONIZATION OFLARGE PROTEINS				
KARAS M	1987	INT J MASS SPECTROM	V78,P53	574
MATRIX-ASSISTED ULTRAVIOLET-LASER DESORPTION OF NONVOLATILE COMPOUNDS				
HILLENKAMP F	1991	ANAL CHEM	V63,PA1193	983
MATRIX-ASSISTED LASER DESORPTION IONIZATION MASS-SPECTROMETRY OF BIOPOLYMERS				
BEAVIS RC	1989	RAPID COMM MASS SPEC	V3,P233	233
ULTRAVIOLET LASER DESORPTION OF PROTEINS				
BEAVIS RC	1990	ANAL CHEM	V62,P1836	276
PROTEIN MOLECULAR MASS USING MATRIX-ASSISTED LASER DESORPTION MASS-SPECTROMETRY				
BEAVIS RC	1989	RAPID COMM MASS SPEC	V3,P432	357
CINNAMIC ACID DERIVATIVES MATRICES FOR UV LASER DESORPTION MASS SPECTROMETRY				
FENN JB	1989	SCIENCE	V246,P64	1628
ELECTROSPRAY IONIZATION FOR MASS-SPECTROMETRY OF LARGE BIOMOLECULES				
BEAVIS RC	1991	CHEM PHYS LETT	V181,P479	217
VELOCITY DISTRIBUTIONS OF INTACT HIGH MASS POLYPEPTIDE MOLECULE IONS PRODUCED BY MATRIX ASSISTED LASER DESORPTION				
BAHR U	1992	ANAL CHEM	V64,P2866	270
MASS-SPECTROMETRY OF SYNTHETIC-POLYMERS BY UV MATRIX-ASSISTED LASER DESORPTION IONIZATION				
STRUPAT K	1991	INT J MASS SPECTROM	V111,P89	263
LASER DESORPTION/ IONIZATION MASS SPECTROMETRY				
SPENGLER B	1990	ANAL CHEM	V62,P793	115
ULTRAVIOLET-LASER DESORPTION IONIZATION MASS-SPECTROMETRY OF LARGE PROTEINS BY PULSED ION EXTRACTION TIME-OF-FLIGHT ANALYSIS				
DANIS PO	1992	ORG MASS SPECTROM	V27,P843	158
ANALYSIS OF WATER-SOLUBLE POLYMERS BY MATRIX-ASSISTED LASER DESORPTION TIME-OF-FLIGHT MASS-SPECTROMETRY				
FENN JB	1990	MASS SPECTROM REV	V9,P37	879
ELECTROSPRAY IONIZATION				
OVERBERG A	1990	RAPID COMM MASS SPEC	V4,P293	113
INFRARED MATRIX-ASSISTED LASER DESORPTION/ IONIZATION MASS SPECTROMETRY				
BEAVIS RC	1990	P NATL ACAD SCI USA	V87,P6873	225
ANALYSIS OF PROTEIN MIXTURES BY MASS-SPECTROMETRY				
DANIS PO	1993	ORG MASS SPECTROM	V28,P923	133
SAMPLE PREPARATION FOR THE ANALYSIS OF SYNTHETIC ORGANIC POLYMERS BY MATRIX-ASSISTED LASER-DESORPTION IONIZATION				
BARBER M	1981	J CHEM SOC CHEM COMM	P325	1024
FAST ATOM BOMBARDMENT OF SOLIDS (FAB) -				

A NEW ION-SOURCE FOR MASS-SPECTROMETRY
WILEY WC 1955 REV SCI INSTRUM V26,P1150 1537
TIME-OF-FLIGHT MASS SPECTROMETER WITH IMPROVED RESOLUTION
CASTRO JA 1992 RAPID COMM MASS SPEC V6,P239 115
MATRIX-ASSISTED LASER DESORPTION IONIZATION OF HIGH-MASS MOLECULES BY FOURIER-TRANSFORM MASS-SPECTROMETRY

The theme of each paper is shown in italics on the line after the paper listing. The order of paper listings is by number of citations by other papers in the extracted database analyzed. The total number of citations from the SCI paper listing, a more accurate measure of total impact, is shown in the last column on the right.

For the Fenn citing papers, Analytical Chemistry contains the most highly cited documents (six), while for the Tanaka citing papers, both Analytical Chemistry and Rapid Communications in Mass Spectrometry each contain five.

All of the journals are fundamental science journals, and most of the topics have a fundamental science theme. Of the most highly cited documents in the Fenn citing papers, nine are from the 80s, eight are from the 90s, and one each from the 70s and 60s. Of the most highly cited documents in the Tanaka citing papers, twelve are from the 90s, seven are from the eighties, and one is from the 50s. These numbers reflect dynamically evolving disciplines, with many of the seminal works coming from recent times.

From Table 6A, about thirty percent of the papers address the phenomena underlying electrospray (ION SOURCE-FREE JET, ELECTROSPRAY INTERFACE, MULTIPLY-CHARGED IONS, MACROION BEAMS, CHARGED DROPLET ION EVAPORATION), about twenty five percent address the electrospray technique (ELECTROSPRAY IONIZATION, HYBRID MASS SPECTROMETRY), about thirty percent address applications (LARGE POLYPEPTIDES, PROTEINS, RECEPTOR LIGAND COMPLEXES), and a few address laser desorption. From Table 6B, about fifteen percent of the papers address the laser desorption approach and associated phenomena, about ten percent address the electrospray technique, and the remainder address applications (LARGE PROTEINS, NONVOLATILE COMPOUNDS, BIOPOLYMERS, LARGE BIOMOLECULES, SYNTHETIC POLYMERS), mainly using the MALDI technique. The relatively large numbers of cited papers related to applications are consistent with the observation in the previous section that a relatively substantial number of highly cited authors were from industrial organizations.

3.2.3. Journal citation frequency results

The most highly cited journals in the Fenn citing papers are listed in Table 7A, and the most highly cited journals in the Tanaka citing papers are listed in Table 7B.

TABLE 7A – MOST CITED JOURNALS IN FENN CITING PAPERS
(cited by other papers in this database only)

JOURNAL	# CITES
ANAL CHEM	8699
J AM CHEM SOC	4550
RAPID COMMUN MASS SP	3888
J AM SOC MASS SPECTR	3371
SCIENCE	3006
INT J MASS SPECTROM	2010
J BIOL CHEM	1809
P NATL ACAD SCI USA	1701
BIOCHEMISTRY-US	1305
MASS SPECTROM REV	1231
ANAL BIOCHEM	1141
J MASS SPECTROM	1076
ELECTROPHORESIS	1069
J PHYS CHEM-US	1020
J CHEM PHYS	965
J CHROMATOGR	965
ORG MASS SPECTROM	935
NATURE	888
METHOD ENZYMOL	607
J CHROMATOGR A	550

TABLE 7B – MOST CITED JOURNALS IN TANAKA CITING PAPERS

JOURNAL	# CITES
ANAL CHEM	2895
RAPID COMMUN MASS SP	2471
INT J MASS SPECTROM	1082
J AM SOC MASS SPECTR	652
J AM CHEM SOC	556
ORG MASS SPECTROM	488
J BIOL CHEM	454
SCIENCE	309
BIOMED ENVIRON MASS	293
MACROMOLECULES	285
MASS SPECTROM REV	273
P NATL ACAD SCI USA	257
CHEM PHYS LETT	244
J MASS SPECTROM	225
J CHEM PHYS	213
J PHYS CHEM-US	211
ANAL BIOCHEM	191

BIOL MASS SPECTROM	177
BIOCHEMISTRY-US	152
J CHROMATOGR	134

Sixteen of the top twenty most highly cited journals are in common between the two lists. Those not in common from Table 7A are: ELECTROPHORESIS, NATURE, METHODS ENZYMOLOGY, JOURNAL OF CHROMATOGRAPHY A. Those not in common from Table 7B are: BIOMEDICAL ENVIRONMENTAL MASS, MACROMOLECULES, CHEM PHYS LETTERS, BIOLOGICAL MASS SPECTROMETRY.

The journals containing the most Fenn citing papers (Table 2A) and the most cited journals in the Fenn citing papers (Table 7A) had thirteen journals in common. The journals containing the most Tanaka citing papers (Table 2B) and the most cited journals in the Tanaka citing papers (Table 7B) also had thirteen journals in common.

3. SUMMARY AND CONCLUSIONS

The papers that cited Fenn's 1989 Science paper and Tanaka's 1988 Rapid Communications in Mass Spectrometry paper were analyzed.

For the Fenn citing papers, of the 22 most prolific authors, seventeen are from the USA, two are from Australia, two are from Denmark, and one is from Japan. Fifteen are from universities, three are from research institutes, and four are from industry.

For the Tanaka citing papers, of the 23 most prolific authors, eight are from the USA, and the remainder are from Europe, mainly central Europe. Twenty are from universities, and three are from research institutes. No authors are common to the two lists of prolific citing authors. Why are there no prolific citing authors from Japan, and why are there no prolific citing authors from industry, for Tanaka's research?

In both cases, the most prolific journals focus on mass spectrometry, chemistry, and biology. Three journals stand out as the first tier for containing the most cited papers: ANALYTICAL CHEMISTRY, JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY, RAPID COMMUNICATIONS IN MASS SPECTROMETRY. Twelve journals are in common between the two lists. The Fenn citing journals not in common tend to focus on biology/ biochemistry (ANALYTICAL BIOCHEMISTRY, BIOCHEMISTRY, PROTEIN SCIENCE, EUROPEAN JOURNAL OF BIOCHEMISTRY), while the Tanaka citing journals not in common tend to focus on the technique/ instrumentation (REVIEW OF SCIENTIFIC INSTRUMENTS, ORGANIC MASS SPECTROMETRY, EUROPEAN MASS SPECTROMETRY).

Of the twenty institutions producing the most Fenn citing papers, seventeen are from North America, one from Europe, and two from the Far East. Eighteen are universities, and two are

research institutes. Of the twenty institutions producing the most Tanaka citing papers, twelve are from the USA, seven are from Europe, and one is from Japan. Eighteen are universities, one is a research institute, and one is from industry. Four institutions are in common between the two lists: UNIV CAL SAN FRANCISCO, INDIANA UNIV, ROCKEFELLER UNIV, OSAKA UNIV.

The USA clearly dominates in country output. The next tier is high on both lists (GERMANY, ENGLAND, JAPAN, CANADA), with Switzerland appearing high on the Tanaka citing list. Thus, while Japan was not very visible in terms of prolific citing authors or institutions, especially with respect to Tanaka's paper, it has reasonable representation in terms of country citations. This implies a diverse group of citing authors in Japan, with the exception of the group at Osaka University.

In terms of absolute numbers of co-authored papers, the USA major partners are Canada, Japan, Germany, England, and France. Additionally, the USA is the major partner for ten of the countries, the exceptions being Australia, Belgium, Holland, and China.

In the Fenn citing papers, Fenn is cited almost twice as much as the next ranked author. This is due to the citation of Fenn's other first-authored papers between 1984 and 1989, in addition to the citation of the Science article. The next tier, Smith and Loo, was a very prolific and highly cited group working on different mass spectrometry techniques, including electrospray ionization.

In the Tanaka citing papers, Tanaka actually ranks third in number of first-author citations. Karas of Frankfurt ranks first. This is due to two factors. In 1985, Karas, in conjunction with Hillenkamp of Munster, showed that an absorbing matrix could be used to vaporize small molecules without chemical degradation. Additionally, in 1988, Karas and Hillenkamp reported a MALDI approach applied to proteins shortly after Tanaka's paper was published. Thus, the papers that cite Tanaka's paper also tend to cite the groundwork papers of Karas as well as his large molecule mass determination papers. Additionally, due to a series of highly-cited papers by Beavis in the early 1990s on laser desorption mass spectrometry, many of the papers that cite Tanaka tend to multiply cite Beavis. This large co-citation of Karas and Beavis with Tanaka was alluded to in the Introduction. It was shown that, of the top fifty cited laser desorption mass spectrometry papers produced in the early high growth years, Tanaka's paper was referenced in fifteen, while Beavis's papers were referenced in 37 and Karas's papers were referenced in 38.

There are five names in common between the two lists (FENN, SMITH, KARAS, BEAVIS, HILLENKAMP). This reflects the broad interests in, and contributions these individuals have made to, mass spectrometry.

Of the 21 most cited authors in the Fenn citing papers, fourteen are from universities, three are from research institutions, and four are from industry. Of the 21 most cited authors in the Tanaka citing papers, sixteen are from universities, one is from a research institute, and four are from industry. This relatively high fraction (~20%) of cited papers from industry suggests

relatively applied citing papers. The validity of this assumption was confirmed in the section on temporal citing patterns.

Finally, while Central Europe plays a modest role in the reference source for the Fenn list, it continues to play a much stronger role for the Tanaka list.

For the Fenn citing papers, the journal Analytical Chemistry contains the most highly cited documents (six), while for the Tanaka citing papers, both Analytical Chemistry and Rapid Communications in Mass Spectrometry each contain five.

All of the journals that contain the most highly cited documents are fundamental science journals, and most of the topics have a fundamental science theme. Of the most highly cited documents in the Fenn citing papers, nine are from the 80s, eight are from the 90s, and one each from the 70s and 60s. Of the most highly cited documents in the Tanaka citing papers, twelve are from the 90s, seven are from the eighties, and one is from the 50s. These numbers reflect dynamically evolving disciplines, with many of the seminal works coming from recent times.

From the lists of references in the Fenn citing papers, about thirty percent of the papers address the phenomena underlying electrospray (ION SOURCE-FREE JET, ELECTROSPRAY INTERFACE, MULTIPLY-CHARGED IONS, MACROION BEAMS, CHARGED DROPLET ION EVAPORATION), about twent y five percent address the electrospray technique (ELECTROSPRAY IONIZATION, HYBRID MASS SPECTROMETRY), about thirty percent address applications (LARGE POLYPEPTIDES, PROTEINS, RECEPTOR LIGAND COMPLEXES), and a few address laser desorption. From the lists o f references in the Tanaka cit ing papers, about fifteen percent ofthe papers address the laser desorption approach and associated phenomena, about ten percent address the electrospray technique, and the remainder address applications (LARGE PROTEINS, NONVOLATILE COMPOUNDS, BIOPOLYMERS, LARGE BIOMOLECULES, SYNTHETIC POLYMERS), mainly using the MALDI technique. The relatively large numbers of cited papers related to applications are consistent with the observation in the previous section that a relatively substantial number of highly cited authors were from industrial organizations.

Sixteen of the top twenty most highly cited journals are in common between the two lists. Those not in common from the journals referenced in the Fenn citing papers are: ELECTROPHORESIS, NATURE, METHODS ENZYMOLOGY, JOURNAL OF CHROMATOGRAPHY A. Those not in common from the journals referenced in the Tanaka citing papers are: BIOMEDICAL ENVIRONMENTAL MASS, MACROMOLECULES, CHEM PHYS LETTERS, BIOLOGICAL MASS SPECTROMETRY.

The journals containing the most Fenn citing papers and the most cited journals in the Fenn citing papers had thirteen journals in common. The journals containing the most Tanaka citing papers and the most cited journals in the Tanaka citing papers also had thirteen journals in common.

In aggregate, the Tanaka cit ing papers have a moderately greater concentrati on in basic research than the Fenn citing papers. The Tanaka citing papers have a greater concentration in the most non-

aligned category than the Fenn citing papers. These two findings corroborate the most prolific authors bibliometrics results, which showed almost twenty percent of the most prolific Fenn citing authors were from industry, whereas none of the most prolific Tanaka citing authors were from industry.

The temporal evolution shows that about a decade is required before the applied technology citing papers become evident. It should be stressed that these are the directly citing technology papers, i.e., papers that cited the original Fenn or Tanaka papers. It is possible that indirectly citing technology papers (i.e., papers that did not cite Fenn or Tanaka's original paper, but rather cited other papers that had cited the Fenn or Tanaka original papers) appeared earlier, but this higher generation bibliometric analysis was beyond the scope of the present study.

One other citation mining study has been performed (11, 11A). Emphasized in that study, and comparable in spirit to the present study, was a detailed analysis of the 1992 Science paper of Jaeger and Nagel on dynamic granular systems. That paper was a very fundamental research paper focused on the basic physics of flowing granular systems. Relative to the Fenn and Tanaka citing papers, the Jaeger and Nagel citing papers have a substantially higher basic research fraction in aggregate. There was a four-year lag time before any applied citing papers emerged. Beyond what the numbers portray, the Jaeger and Nagel citing papers reached a wider variety of more extreme non-aligned categories than the Fenn or Tanaka citing papers (e.g., earthquakes, avalanches, traffic congestion, war games, flow immunosensors, shock waves, nanolubrication, thin film ordering). Chi-tests confirmed the validity of the differences between the Fenn-Tanaka citing papers and the Jaeger and Nagel citing papers, and between the Fenn and Tanaka citing papers as well.

REFERENCES FOR APPENDIX 7-I

1. Yamashita M, Fenn JB. Electrospray Ion-Source - Another Variation on the Free-Jet Theme. Journal of Physical Chemistry. 88 (20): 4451-4459. 1984.
2. Yamashita M, Fenn JB. Negative-Ion Production with the Electrospray Ion-Source. Journal of Physical Chemistry. 88 (20): 4671-4675. 1984.
3. Whitehouse CM, Dreyer RN, Yamashita M, Fenn JB. Electrospray Interface for Liquid Chromatographs and Mass Spectrometers. Analytical Chemistry. 57 (3): 675-679. 1985.
4. Wong SF, Meng CK, Fenn JB. Multiple Charging in Electrospray Ionization of Poly(Ethylene Glycols). Journal Of Physical Chemistry. 92 (2): 546-550 Jan 28. 1988.
5. Mann M, Meng Ck, Fenn Jb. Interpreting Mass-Spectra Of Multiply Charged Ions. Analytical Chemistry. 61 (15): 1702-1708 Aug 1 1989.
4. Fenn Jb, Mann M, Meng Ck, Wong Sf, Whitehouse Cm. Electrospray Ionization For Mass-Spectrometry Of Large Biomolecules. Science. 246 (4926): 64-71 Oct 6 1989.
5.Fenn Jb, Mann M, Meng Ck, Wong Sf, Whitehouse Cm. Electrospray Ionization-Principles And Practice. Mass Spectrometry Reviews. 9 (1): 37-70 Jan 1990.
8. Tanaka, K., Waki, H., Ido, Y., Akita, S., and Yoshida, Y. Protein and polymer analysis up to m/zx 100000 by laser ionisation time-of-flight mass spectrometry. Rapid communications in Mass Spectrometry. 2. 1988. 151-153.
9.Tanaka, K., Ido, Y., Akita, S., Yoshida, Y., and Yoshida, T. Proceedings Second Japan-China Joint

symposium on Mass spectrometry. Editors Matsuda, H. and Xiao-tian L. (Osaka, Japan, 15-18 Septembe 1987). 185-188.
10.Yoshida, T., Tanaka, K., Ido, Y., Akita, S., and Yoshida, Y. Mass Spectroscopy (Japan). 36. 1988. 59.
11.Kostoff, R. N., Del Rio, J. A., García, E. O., Ramírez, A. M., and Humenik, J. A. Citation Mining: Integrating Text Mining and Bibliometrics for Research User Profiling. JASIST. 52:13. 1148-1156. 52:13. November 2001.
11A. del Río, J.A. Kostoff, R.N. García, E. O., Ramírez, A. M., and Humenik, J. A. Phenomenological approach to profile impact of scientific research. Adv. Complex Syst. 5. 19-42 (2002).
12. Kostoff, R. N., and DeMarco, R. A. Science and Technology Text Mining. Analytical Chemistry. 73:13. 370-378A. 1 July 2001.
13. Kostoff, R. N., Toothman, D. R., Eberhart, H. J., and Humenik, J. A. Text Mining Using Database Tomography and Bibliometrics: A Review. Technology Forecasting and Social Change. 68:3. November 2001.
14. Kostoff, R. N., Tshiteya, R., Pfeil, K. M., and Humenik, J. A. Electrochemical Power Source Roadmaps using Bibliometrics and Database Tomography. Journal of Power Sources. 110:1. 163-176. 2002.
15. Kostoff, R. N., Braun, T., Schubert, A., Toothman, D. R., and Humenik, J. Fullerene Roadmaps Using Bibliometrics and Database Tomography. Journal of Chemical Information and Computer Science. 40:1. 19-39. Jan-Feb 2000.
16. Braun, T., Schubert, A. P., and Kostoff, R. N. Growth and Trends of Fullerene Research as Reflected in its Journal Literature. Chemical Reviews. 100:1. 23-27. January 2000.
17. Kostoff, R. N., Database Tomography for Technical Intelligence: Comparative Analysis of the Research Impact Assessment Literature and the Journal of the American Chemical Society. Scientometrics. 40:1. 1997.
18. Kostoff, R. N., Green, K. A., Toothman, D. R., and Humenik, J. Database Tomography Applied to an Aircraft Science and Technology Investment Strategy. Journal of Aircraft, 37:4. 727-730. July-August 2000.
19. Kostoff, R. N., Eberhart, H. J., and Toothman, D. R. Hypersonic and Supersonic Flow Roadmaps Using Bibliometrics and Database Tomography. Journal of the American Society for Information Science. 50:5. 427-447. 15 April 1999.
20. Kostoff, R. N., Coder, D., Wells, S., Toothman, D. R., and Humenik, J. Surface Hydrodynamics Roadmaps Using Bibliometrics and Database Tomography. To be Submitted for Publication.
21. Garfield, E. History of Citation Indexes for Chemistry - A Brief Review. JCICS. 25(3). 170-174. 1985.
22. Kostoff, R. N. The Use and Misuse of Citat ion Analysis in Research Evaluat ion. Scientometrics, 43:1, September, 1998.
23. MacRoberts, M., and MacRoberts, B. Problems of Citation Analysis. Scientometrics. 36:3. July-August, 1996.

APPENDIX 8

SCIENCE AND TECHNOLOGY TRANSITIONS [Kostoff, 1997c, 2004o]

This Appendix has two parts. The first part addresses accelerating the conversion from science to technology, and the second part addresses science and technology transition metrics.

8A. Accelerating the Conversion from Science to Technology [Kostoff, 1997c]

INTRODUCTION

As the technology marketplace has become global, the efficient and timely transfer of technology has assumed paramount importance. Delays in commercializing technologies can translate into surrendering substantial market shares to national or international competitors. There is a rich literature on cross-organizational and cross-national transfer of developed technology, even though substantial improvements are required in the practical aspects of the transfer of developed technology. However, there is very little in the literature addressing the problem of how science, especially fundamental science, gets converted eventually to technology, and how the efficiency (minimization of time and other resource utilization) of this process can be improved.

This aspect of technology transfer has become a very important and timely topic of national and cross-national interest, both for the federal and state agencies which sponsor substantial research and for the United States companies which compete in the global technology market. In particular, there has been substantial criticism that foreign countries, which fund far less research than the U. S., are more effective and efficient than the U. S. in converting the products of research into commercializeable technologies. The importance of efficient science-technology conversion can also be inferred from the federal agencies and industrial organizations which have restructured their science and technology development components in large part to enhance this conversion.

The remainder of the first part of this Appendix is structured as follows. Some results and principles from past classical studies of successful transitions will be presented. Then, some personal observations relating to successful transitions, and the underlying principles, will be discussed.

RESULTS FROM PAST RETROSPECTIVE STUDIES

There are two major variants of retrospective studies which have examined the science-technology evolution process. One type starts with a successful technology or system and works backwards to identify the critical R&D events which led to the end product. The other type starts with initial research grants and traces evolution forward to identify impacts. The tracing backwards approach is favored for two reasons: 1) the data are easier to obtain, since forward tracking is essentially non-existent for evolving research; and 2) the sponsors have little interest in examining research that may have gone nowhere.

In the remainder of this summary, a few of the more widely known science-technology evolution case studies will be reviewed, and the key findings will be identified. These retrospective studies include Project Hindsight, Project TRACES and its follow-on studies, and Accomplishments of the Defense Advanced Research Projects Agency (DARPA). In addition, the results of a recent

workshop, which validated most of the results from the classical studies, will be summarized.

In the 1960s, a study named Project Hindsight was sponsored by the Department of Defense (3). Hindsight examined twenty successful military systems, and identified the critical R&D events which led to the successful systems. Hindsight examined characteristics of these critical R&D events to see whether any general principles could be extracted. While there were problems with some of the constraints placed on the Hindsight study, nevertheless, some valuable conclusions emerged (4). In particular, a major conclusion related to the science-technology conversion process was that the results of research were most likely to be used when the researcher was intimately aware of the needs of the applications engineer.

In 1967, The National Science Foundation (NSF) instituted a study (5) called TRACES to trace retrospectively key events which had led to five major technological innovations. One goal was to provide more specific information on the role of the various mechanisms, institutions, and types of R&D activity required for successful technological innovation. Similar to Project Hindsight, key 'events' in the R&D history of each innovation selected were identified, and their characteristics were examined.

The study showed that non-mission research provided the origins from which science and technology could advance toward innovations. For the cases studied, the average time from conception to demonstration of an innovation was nine years. Most non-mission research appeared **completed prior to the conception** of the innovation to which it would ultimately contribute. The tracings also revealed cases in which mission-oriented research or development efforts elicited later non-mission research which often was found to be crucial to the ultimate innovation.

In a follow-on study to TRACES, the NSF sponsored Battelle-Columbus Laboratories to perform a case study examination of the process and mechanism of technological innovation (6). For each of the ten innovations studied, the significant events (important activity in the history of an innovation) and decisive events (a significant event which provides a major and essential impetus to the innovation) which contributed to the innovation were identified. The influence of various exogenous factors on the decisive events was determined, and several important characteristics of the innovative process as a whole were obtained.

The following important exogenous factors for producing significant innovations were identified:

-The technical entrepreneur (a major driving force in the innovative process);
-Early recognition of the need;
-Government funding (more generally, availability of financial support, from whatever source);
-The occurrence of an unplanned confluence of technology (confluence of technology occurred for some innovations as a result of deliberate planning, rather than by accident);
-Most of the innovations originated outside the organization that developed them;
-Additional supporting inventions were required during the development effort for all the innovations studied to arrive at a product with consumer acceptance.

The Institute for Defense Analysis produced a document (7) describing the accomplishments of the Defense Advanced Research Projects Agency (DARPA). Of the hundreds of projects and programs funded by DARPA over its then (1988) 30 year lifetime, 49 were selected and studied in detail, and conditions for success were identified.

The qualities of DARPA-supported programs and projects that contributed to success can be

summarized:

-A need existed for what the output could do;
-There was a strong commitment by individuals to a concept;
-Bright and imaginative individuals were given the opportunity to pursue ideas with minimal bureaucratic encumbrance;
-There was an ongoing stream of technical developments and evolution;
-DARPA management gave strong, top-level management support;
-There was explicit effort, taken early, to improve acceptance by the user community.

Hindsight, TRACES, and, to some degree, the DARPA accomplishments books had some similar themes. All these methods used a historiographic approach, looked for significant research or development events in the metamorphosis of research programs in their evolution to products, and attempted to convince the reader that: (1) the significant research and exploratory development events in the development of the product or process were the ones identified; (2) typically, the organization sponsoring the study was responsible for some of the (critical) significant events; (3) the final product or process to which these events contributed was important; and (4) while the costs of the research and development were not quantified, and the benefits (typically) were not quantified, the research and development were worth the cost.

Six critical conditions for innovation were identified implicitly and explictly through analysis of these retrospective studies. The most important condition from the author's perspective implicitly appears to be **the existence of a broad pool of knowledge which minimizes critical path obstacles** and can be exploited for development purposes. The time required to overcome deficiencies in the knowledge pool is the pacing item to initiate the research exploitation process. This condition is followed in importance, from the author's perspective, by a **technical entreprenuer who sees the technical opportunity and recognizes the need for innovation**, and who is willing to champion the concept for long time periods, if necessary. **While the technical entrepreneur was viewed by some of the studies as most important to the innovative process, it does not appear (to the author) to be the critical path factor. Examination of the historiographic tracings which display the significant events chronologically for each of the innovations shows that an advanced pool of knowledge must be developed in many fields before synthesis leading to an innovation can occur. The entrepreneur can be viewed as an individual or group with the vision and ability to both recognize the downstream applications (need) for the research and to assimilate and/ or enhance this diverse information and exploit it for further development. However, once this pool of knowledge exists, there are many persons or groups with capability to exploit the information, and thus the real critical path to the innovation is more likely the knowledge pool than any particular entrepreneur. The entrepreneurs listed in the studies undoubtedly accelerated the introduction of the innovation, but they were at all times paced by the developmental level of the knowledge pool.**

The third most important condition is **early recognition of the need**, coupled with early efforts taken to improve acceptance by the user community. In many cases, these functions will be performed by the entrepreneur. Also valuable for innovation are **strong financial and management support**, and occurrence of an **unplanned confluence of technology** coupled with **many continuing inventions in different areas** to support the innovation.

One goal of all the studies presented was to identify the products of research and some of their impacts. The Hindsight, TRACES, and DARPA studies tried to identify factors which influenced the productivity and impact of research. The following conclusions about the role and impact of basic research were reached:

-The majority of basic research events which directly impacted technologies or systems were non-mission oriented and occurred many decades before the technology or system emerged;
-The cumulative indirect impacts of basic research were not accounted for by any of the retrospective approaches published;
-An advanced pool of knowledge must be developed in many fields before synthesis leading to an innovation can occur;
-Allocation of benefits among researchers, organizations, and funding agencies to determine economic returns from basic research is very difficult and arbitrary, especially at the micro level.

A recent workshop validated the conclusions of these classical studies (8), at least in the corporate environment. The moderators identified the following success factors:

*Management and Organizational Infrastructure
-An organizational model that encourages coordination between research activities and product projects
-Executive-level commitment to the transfer of ideas from research groups to development groups
-Geographic and social proximity between research and development groups

*Technology Push
-Research projects that are aligned with corporate strategy
-Research projects with people highly motivated to see their research transferred into products
-A high-level visionary who champions bringing the idea to market
-Readily demonstrable improvements over existing or related products

*Demand Pull
-A product group motivated and poised to take the technology
-A significant customer with a strong need for the technology
-An involved marketing group that tracks customers' needs and markets the ideas throughout the company

These and similar studies also identified many other factors important in the successful evolution of science to technology. Additional factors, many of which will be addressed in other papers in this special issue, include: awareness of ongoing research through diverse information sources; types of cooperative R&D agreements between researchers and developers; intellectual property issues such as disclosure, protection, marketing, negotiating and licensing; Congressional incentives to collaboration; and other legal, financial, cultural, and sociological incentives and roadblocks.

PERSONAL OBSERVATIONS

From the author's viewpoint, Project Hindsight, with all of its limitations (4), produced very

relevant findings for the science-technology conversion problem. A conceptual principle for accelerating the science-technology conversion can be abstracted from the Hindsight results, and it is important to separate the conceptual principle from the implementations of the principle. In this manner, one does not become bound by the limitations of any particular implementation. This principle, termed by the author as Heightened Dual Awareness (HDA), states that in order for the science-technology conversion to be accelerated, at least two necessary conditions must be fulfilled: 1) the researcher must be intimately aware of the needs of the applications engineer; 2) the potential user of the research, or transitionee, must be aware of the progress and results of the research. In addition, if third parties are involved in the conversion and development process, such as vendors, their awareness of both ends of the conversion cycle must be maintained as well. To the degree that each of these requirements is not fulfilled, the science-technology conversion will be retarded and delayed.

The author's personal observations of examples of science which has converted to technology rapidly have borne out the validity of the HDA principle, and of the above studies' conclusions related to evolution of research into successful systems. Some of these observations will now be described.

For years the author sponsored research at the Department of Energy (DOE) National Labs. In those cases where the departments in which the research was conducted were full spectrum S&T organizations, the researchers were often the developers as well, and in any case were well aware of the needs of the developers and users. The main motivations and incentives were to transition the research as rapidly as possible, and this in fact is what occurred. As a specific example, the Materials Department at Oak Ridge National Lab was a full spectrum materials R&D operation. Intermetallics research sponsored by the author for space applications metamorphisized into the high impact Ni3Al alloy research and development for terrestrial applications. The complete cycle from research to advanced development was conducted and completed very rapidly due to the vertically integrated materials structure at Oak Ridge.

The Oak Ridge example illustrates the most straightforward application of the HDA principle. The researchers and developers are physically contiguous, and in many cases are the same person. Thus, the dual awareness is readily effected by the intrinsic structure of the physical environment, and complex management structures are not necessary to enhance dual awareness.

At Bell Laboratories in the 1960s and 70s, the research functions were linked closely with the advanced development functions through two major approaches. First, the more applied satellite laboratories were usually located adjacent to a Western Electric development and manufacturing facility, in a quasi-vertically integrated management structure (Bell Labs was an independent corporation). As in the Hindsight case, the researchers were well aware of the developers' and users' needs, and the potential users were kept apprised of the status of the research. This allowed simultaneous technology push and demand pull, and transitions occurred smoothly and rapidly.

Second, in the more centralized facilities in which the fundamental research was conducted, such as the Murray Hill laboratory, academic freedom characteristic of universities was combined with facility and staff support characteristic of the best industrial labs, with easy access to the developers. Not only did these centralized facilities contain contiguous applied research and development components, but the technical managers tended to be career Bell System employees who were extremely knowledgeable about the technological and operational needs of many different segments of the Bell System. Management awareness of both the research status and potential and technology and system needs helped strengthen the necessary linkages between basic research and the developers. A recent article on the development of the transistor by Bell Labs (9) illustrates this point. Following the

invention of the point-contact transistor, the research director did not tell the inventor to redirect his work toward further developing and refining the product. Instead, he gave that effort to another manager, and left the inventor free to seek newer frontiers.

In the Department of the Navy, much of the research at the Warfare Centers (full spectrum R&D organizations) is sponsored through the program managed by the author, the In-House Laboratory Independent Research program. Here, the Technical Directors of the Warfare Centers select projects focused on the Centers' mission requirements. The researchers tend to work part-time in development activities, and are continuously aware of both naval Fleet requirements and the state-of-the-art in the research community. Similar to the Oak Ridge example presented previously, when the researchers operate in such an applications-aware environment, their new ideas and concepts tend to be naturally associated with the naval applications, and have a higher probability of eventual utility. Fleet and technology impacts from this program (10) have been substantial.

The HDA principle as a major driver of eventual utility is not limited to the performer and potential user; it is applicable to the research sponsor environment as well. A number of research sponsoring organizations have switched from a discipline orientation to a structure where the research is vertically integrated with technology, analogous to the vertically integrated research-technology performer environment described above.

For example, in 1993, the Office of Naval Research (ONR), a science and technology development sponsor, switched to such a structure in part for the purpose of closing the gap between science and technology, and initial indications are that this is indeed occurring. ONR's program officers (POs) are responsible for the range spanning research to advanced development, and, as in the integrated laboratory environment, are intimately aware of the needs of the users. The POs have the incentives to transition the research to development as rapidly as possible.

The general conclusion that the author has drawn is that for most effective and efficient conversion of science to technology, the researcher primarily and the sponsor secondarily need to be immersed in environments where the HDA principle is most operative, and where motivations and incentives are geared toward rapid transitioning. This type of physical environment is realized most efficiently when the researchers and developers are physically contiguous. If this type of physical environment structure is not readily possible, as may be the case with some extremely fundamental university research, then attempts should be made to simulate this optimal transitioning environment through innovative management structures. This should not be interpreted as a recommendation to substitute applied research for basic research. Far too much of this substitution has occurred in the recent past. Rather, the recommendation is that basic research be conducted in an environment where there is greater awareness of the progress and potential of the research by potential transitionees and users, and opportunities to understand the needs of the developers are made available to the researchers.

The irony is that the optimal transitioning research performer environment, from a physical structure viewpoint, exists most strongly (on average) today in two types of organizations: large corporate R&D labs and large government or national labs. Yet non-government-financed basic research has essentially disappeared from the large non-medical corporate labs (11), and the large government and national labs are being downsized. This trend can only impact the conversion of mission-oriented research negatively, and could serve to hamper the competitiveness of the United States in the 21st century.

For mission-oriented agencies, to enhance the simulation of optimal transitioning physical

structures, joint university-federal or national or corporate laboratory projects should be expanded. In parallel, as the author's personal observations have also shown, the potential user needs to become involved in the research project as early, broadly, and intensely as possible. This early involvement provides the user a sense of 'ownership', and produces a more seamless transition process. In the author's experience, incorporating the potential user from the research proposal evaluation phase is not too soon for successful downstream transitions of the research products to technology.

REFERENCES FOR PART 8A.OF APPENDIX 8

1) Kostoff, R.N., (ed.), Evaluation Review, Special Issue on Research Impact Assessment, 18:1, February 1994.
2) Kostoff, R.N., (ed.), Scientometrics, Special Issue on Performance Measures for Government Sponsored Research, 36:3, July-August 1996.
3) DOD, Project Hindsight, Office of the Director Defense Research and Engineering, Wash., DC, DTIC No. AD495905, October 1969.
4) Kostoff, R. N., "The Handbook of Research Impact Assessment", Sixth Edition, Summer 1996, DTIC Report Number ADA296021, 1996.
5) IITRI, "Technology in Retrospect and Critical Events in Science", Illinois Institute of Technology Research Institute Report, December 1968.
6) Battelle, "Interactions of Science and Technology in the Innovative Process: Some Case Studies", Final Report, Prepared for the National Science Foundation, Contract NSF-C 667, Battelle Columbus Labs, March 19, 1973.
7) IDA, DARPA Technical Accomplishments, Volume I, IDA Paper P-2192, February 1990; Volume II, IDA Paper P-2429, April 1991; Volume III, IDA Paper P-2538, July 1991, Institute for Defense Analysis.
8) Isaacs, E., and Tang, J., "Technology Transfer: So Much Research, So Few Good Products," Communications of the ACM, 39:9, 1996.
9) Heppenheimer, T. A., "What Made Bell Labs Great", Invention and Technology, Summer 1996.
10) ONR, Department of the Navy In-House Laboratory Independent Research Annual Report for FY1995, December 1996.
11) Kostoff, R.N., "Performance Measures for Government Sponsored Research: Introduction and Overview", Scientometrics, 36:3, July-August 1996.
12) Kostoff, R. N., Eberhart, H. J., and Toothman, D. R., "Database Tomography for Information Retrieval", Journal of Information Science, 23:4, 1997.
13) Kostoff, R. N., " Database Tomography: Origins and Applications", Competitive Intelligence Review, Special Issue on Technology, 5:1, Spring 1994.
14) Kostoff, R. N., Eberhart, H. J., and Toothman, D. R., "Database Tomography for Technical Intelligence: A Roadmap of the Near-Earth Space Science and Technology Literature", Submitted for Publication, 1997.
15) Kostoff, R. N., "Research Impact Quantification", R&D Management, 24:3, July 1994.

8B. Science and Technology Transition Metrics [Kostoff, 2004o]

I. OVERVIEW

On 27 October 1998, a workshop was convened by the National Institute for Occupational Safety and Health (NIOSH) to identify key metrics for NIOSH's Strategic Goals. The first NIOSH Strategic Goal (Conduct a targeted program of research to reduce morbidity, injuries, and mortality among workers in high-priority areas and high-risk sectors) was the major focus of the workshop. Its two related Objectives addressed 1) the success in implementing a research program based on its 1996 National Occupational Research Agenda (NORA) priorities (NORA is a framework to guide occupational safety and health research into the next decade, and resulted in the establishment of a list of the top 21 research priorities) and 2) success in measuring its safety and health outcomes.

The author was invited to participate as a member of the panel. This Appendix generalizes a document that the author prepared for the NIOSH workshop, and was further refined during preparation for a DOE-sponsored workshop on S&T benefits, 4-5 March 2002. The Appendix focuses on key metrics for evaluating progress in a mission-oriented research program. The results and conclusions of the analyses are sufficiently generic for applicability to any science and technology (S&T) sponsoring organization.

II. BACKGROUND

The implementation of the Government Performance and Results Act of 1993 (GPRA) signaled the codification of the use of quantitative metrics to monitor the progress of government-sponsored S&T. An open question since that time has revolved around the appropriate quantities to measure, and the appropriate metrics to use.

Typically, a major event in the life of an S&T project is its transition from one level of development (e.g., basic research) to another level of development (e.g., applied research, or technology development). Could such transitions be quantified, and used to populate performance metrics? Before this question can be addressed, different types of S&T transitions need to be identified and discussed. The following paragraphs describe transitions in the context of mission-oriented government S&T-sponsoring organizations,

Mission-oriented-government S&T sponsors have the generic mission of providing S&T information to 1) the engineering development and operational/ acquisition components of their parent organizations and/ or to 2) the engineering development components of the commercial sector, depending on their organizational structure and mission. These post-S&T developers and implementers will be referred to as the customer.

S&T information can be provided to the customer through two paths: 1) development sponsored directly by the government S&T organization, or 2) development sponsored by some other S&T organization(s). Resources expended by other S&T sponsoring organizations in a given technical interest area can be much larger cumulatively than resources available to any single S&T sponsor. Therefore, leveraging of these external resources by the customer/ S&T sponsor could have cost impacts far in excess of those resulting from directly sponsored S&T.

However, advanced technical understanding is required to identify the significance of technical advances made by other organizations. S&T sponsoring organizations tend to have the largest concentration of advanced technical personnel within the customer's management purview, and are in the best position to make the customer aware of significant technical developments globally.

Therefore, S&T sponsoring organizations have a dual role in providing S&T information to their customers: direct sponsorship of S&T targeted toward obtaining this information, and making their customers aware of significant technical advances worldwide. Given these two major missions and objectives for the S&T sponsoring organizations, management performance and metrics should focus on progress made for each of these two major roles.

III. INTRODUCTION

There are four major classes of metrics available for consideration as transition metrics:

1) Activity - measures resource expenditures (e.g., people employed, operating budgets, etc), under management control (after resources received).
2) Output – tangible products under control of management (e.g., reports produced, components built)
3) Impact – measures effects on science and technology, and typically based on external judgements (e.g., transitions, citations, awards). Typically not under management control.
4) Outcome – long-term impacts on larger societal goals (e.g., health improvement, environmental remediation, etc)

Activity metrics are used mainly to normalize productivity and impact metrics. Most output metrics are used for superficial reporting purposes by S&T sponsors. Output metrics are rarely used in practice to impact major sponsor or performer management decisions, except in isolated cases like faculty tenure evaluation. They are sometimes used for research performer bonus considerations.

Outcome metrics are useful for long-term program auditing, for retrospective studies to identify critical parameters for fostering quality S&T, and for general documenting and archival purposes. Outcome metrics become operational too far into the future to impact management decisions and performance evaluation. Government military, civilian, and commercial civilian organizations have relatively rapid turnover of their highest level management. Especially in commercial organizations, portable pension plans have increased mobility, and continual de-regulation has enhanced the role of short-term market performance in driving management decisions. Motivation of government or commercial organizational management is to show progress within time frame of highest management cognizance. Management decisions are mainly governed by this time scale.

For S&T sponsors, major metrics used operationally for management decision-making and

performance evaluation are transitions from one development level to another. These are metrics that incorporate:

- The number of transitions across development levels per unit of time
- The potential impact or benefit eventually resulting from these transitions
- The probability that each transition will eventually achieve the potential impact

The remainder of this Appendix will address the impact metric of transitions.

Transitions have two components, one under control of the S&T sponsor, and the other not under sponsor control. The first component is developing S&T to the point where it has 'positive transitionability characteristics' (e.g., potential for affordability, increased performance, lighter weight, smaller, etc). The second component is the decision by the downstream developer/ user to advance development externally based on a number of exogenous parameters (e.g., geopolitical, legal, financial, etc). To some degree, whatever transition metrics are developed and implemented should reflect this division of responsibility between S&T sponsor and customer.

The transition metrics used presently for S&T sponsor performance and evaluation do not reflect this division of responsibility. Further, they do not reflect the dual role responsibility of S&T sponsors, namely, direct S&T sponsorship and increasing customer awareness of external S&T advancements. This limited scope of present day transition metrics reflects the limited scope of strategic objectives and organizational responsibilities of S&T sponsors. In addition, transitions used presently as S&T sponsor performance and evaluation metrics are not normalized to target productivity levels, and transition efficiency can not be evaluated.

This paper proposes transition metrics be re-defined to 1) reflect transition efficiency, similar to Carnot efficiency for thermodynamic systems; 2) reflect dual responsibilities of direct science and technology sponsorship and enhanced customer awareness; 3) reflect in part shared responsibility of sponsor and customer for effecting transitions successfully. This paper shows how use of these re-defined transition metrics will enhance productivity and the role of S&T sponsors in the full product development cycle. The Appendix provides supplementary information on high quality metrics.

IV. ANALYSIS

The approach taken here to re-define appropriate transition metrics is analogous to an approach used for citations [Kostoff, 1998a]. The fundamental principle is to measure the efficiency and effectiveness with which the S&T sponsor is accomplishing its broader mission. The basic objective function that contains these efficiency and effectiveness measures is the ratio of: 1) the impact (benefits) of all actual transitions <u>enabled</u> by the S&T sponsor to 2) the research transitions that would have maximized impacts (benefits) for the American public, given the level of global S&T funding in the topical areas being examined. The term '<u>enabled</u>' is used in the ratio definition to include the dual role of the S&T sponsor discussed previously. Thus, this

definition goes beyond counting of numbers of transitions, and focuses on the downstream payoffs resulting from these transitions.

The objective function can be written in equation form as:

$$R = \sum_{i=1}^{i=n} T_i * I_i \Big/ \sum_{i=1}^{i=Z} T_i * I_i \qquad (1)$$

where:

R is the objective function,

SUM is the summation operator,

i is the dummy variable that ranges between the limits shown,

Ti is the 'i'th transition from research to application,

Ii is the probable magnitude of the impact (benefit) resulting from the 'i'th transition. Ii is the product of the magnitude of the potential impact, Mi, times the probability, Pi, that the potential impact Mi will be achieved. Ii is thus defined at the probable impact of the ‘i’th transition.

n is the actual number of transitions enabled from all sources, and

Z is the potential maximum number of high impact transitions resulting from a perfect investment strategy applied to the global funding that was expended on the topical area’s S&T.

Ii is the product of the potential benefit (resulting from the ‘i’th transition) times the probability that the ‘i’th transition will actually realize that benefit, and therefore Ii should be viewed as the expected benefit.

The stage in time at which the objective function is evaluated determines the credibility of the data. If the evaluation time is far in advance of the transition time frames, then the quantities evaluated are estimates, with all the associated uncertainties. If the evaluation time is far after the transition time frames, then the quantities evaluated are much more credible, but are now outcome metrics, and lose their operational impact for the reasons discussed previously. Thus, the sum of utility and credibility for this metric is probably optimal somewhere in the time frame of the transitions being evaluated.

Obtaining credible data to evaluate the complete objective function is very difficult. In particular, Z is a hypothetical quantity based on a perfect investment strategy. It is included in the fundamental objective function statement to counteract the case where the S&T sponsor could conceivably be investing in very low-risk low-impact safe technologies, could have a high transition efficiency (ratio of number of transitions effected to potential transitions possible), and yet be ineffective relative to what could have been accomplished with a better investment

strategy.

Equation 1 can be re-written to reflect more clearly those transitions resulting from the direct sponsorship of S&T and those transitions resulting from enhanced global data awareness.

.....i=N1..……..... .j=N2……..…i=Z1 ….j=Z2

R’ = (SUM(Ti*Ii)+SUM(Tj*Ij))/(SUM(Ti*Ii)+SUM(Tj*Ij)) (2)

.....i=1...…….....j=1…….i=1……j=1

where N1 is the transitions resulting from directly sponsored S&T of the organization being evaluated, and N2 is the transitions from other globally-sponsored S&T enabled by the awareness of the technical experts in the organization being evaluated. Z1 and Z2 are the analogous numbers for ideal investment strategy and awareness.

The following section addresses different levels of approximation to the objective function, and includes comments on the strengths and weaknesses of each level.

I) Zeroth order approximation

...…....i=N1

R0 =(SUM(Ti)) (3)

...…....i=1

This approximation applies to the S&T sponsor’s projects only. Here, the number of transitions from the sponsor’s S&T is the metric. This is the easiest metric for which data can be obtained, but is essentially useless for addressing the accountability components defined above. Unfortunately, this metric is used all too commonly in many organizations. It provides no indication of impact, and no indication of how efficiently the agency is performing its function. Further, it can be ‘gamed’, where the organization funds a large number of low-risk modest-payoff projects to inflate the transition numbers. The S&T sponsor could then be transitioning a high fraction of its potentially transitionable projects, but collectively these transitions will have low impact relative to what was possible with a better investment strategy.

2) First order approximation

...…....i=N1

R1 =(SUM(Ti*Ii)) (4)

...…....i=1

Here, the product of number of transitions from the directly-sponsored S&T times expected impact per transition is the metric. It provides an indication of actual impact, but no indication of transition efficiency. Obtaining credible data for potential impacts and benefits, and the probabilities that these potential impacts and benefits will be realized, is significantly more complicated than for the zeroth order metric, but much more insight is provided. Further, this metric overcomes the 'gaming' aspect of the previous metric to some degree, since level of payoff is included in the objective function.

3) Second order approximation

...…....i=N1……i=Z1

R2 =(SUM(Ti*Ii))/(SUM(Ti*Ii) (5)

...…....i=1……i=1

In this approximation, it is assumed that a panel of experts was convened, and identified the transitions that would have occurred from the directly sponsored S&T if an ideal investment strategy had been followed and executed. These ideal transitions are reflected in the denominator. The complexity of evaluating this metric increases considerably over the first order approximation, since judgements are now required as to how many of the sponsor's projects could have transitioned. However, this metric does offer indication of efficiency, as well as impact.

4) Third order approximation

...…....i=N1……j=N2

R3 =(SUM(Ti)+SUM(Tj)) (6)

...…....i=1……j=1

This approximation sums the number of transitions resulting from directly-sponsored S&T and the number of transitions from global S&T enabled by the global S&T awareness of the S&T sponsor. While it suffers from the types of deficiencies noted in the zeroth order approximation, it nevertheless represents a step forward through the inclusion of enabled transitions from global S&T. This metric, while still primitive, provides some indication of how well the S&T sponsor is performing its knowledge awareness function, in addition to its S&T sponsoring function. However, without impact or benefit level numbers incorporated into the objective function, this metric is subject to 'gaming'.

5) Fourth order approximation

..…....i=N1……j=N2

R4 =(SUM(Ti*Ii)+SUM(Tj*Ij)) (7)

...…....i=1……j=1

Here, the product of number of transitions from the directly-sponsored S&T and enabled S&T times impact per transition is the metric. It provides an indication of actual impact, but no indication of transition efficiency. Obtaining credible data for impacts and benefits is significantly more complicated than for the third order metric, but much more insight is provided. Further, this metric overcomes the 'gaming' aspect described previously to some degree, since level of payoff is included in the objective function.

6) Fifth order approximation

..…....i=N1….……j=N2…..….i=Z1….……j=Z2

R5 =(SUM(Ti*Ii)+SUM(Tj*Ij))/(SUM(Ti*Ii)+SUM(Tj*Ij)) (8)

...…....i=1……j=1……..i=1……j=1

In this approximation, it is assumed that a panel of experts was convened. They identified the transitions that would have occurred from a) the directly sponsored S&T if an ideal investment strategy had been followed and executed, and b) the globally enabled S&T if the technical experts had been fully aware of the relevant global S&T sponsored and the relationship of the relevant global S&T to the needs of the parent organization. These ideal transitions are reflected in the denominator. The complexity of evaluating this metric increases considerably over the fourth order approximation. Judgements are now required as to how many of the sponsor's projects could have transitioned as well as the number of other global S&T projects that could have been exploited by the S&T sponsor's parent organization. However, this metric does offer indication of efficiency, as well as impact.

V. SUMMARY AND CONCLUSIONS

Transition metrics have been defined to different levels of approximation. The are based on the rate of flow of expected benefit across a transition barrier. They range in complexity from the rate of flow of numbers of transitions to the normalized rate of flow of actual expected or realized benefits. They take into account transitions resulting from the sponsor's S&T development efforts as well as transitions enabled by the S&T sponsor's awareness of S&T performed globally.

VI. SUGGESTIONS FOR FURTHER READING FOR PART 8-B OF APPENDIX 8

Geisler, E., “The Metrics of Science and Technology”, Quorum Books, Westport, CT, 2000.

Kostoff, R. N., "Peer Review: The Appropriate GPRA Metric for Research", Science, Volume 277, 1 August 1997.

Kostoff, R. N., "The Use and Misuse of Citation Analysis in Research Evaluation", Scientometrics, 43:1, September 1998a.

Kostoff, R. N., Science and Technology Metrics, http://www.onr.navy.mil/sci_tech/special/technowatch/, 1998b.

Kostoff, R. N. “The Metrics of Science and Technology”. Scientometrics. 50:2. 353-361. February 2001.

Kostoff, R. N., and Schaller, R. R. "Science and Technology Roadmaps". IEEE Transactions on Engineering Management. 48:2. 132-143. May 2001.

Kostoff, R. N. “Normalization for Citation Analysis”. Cortex. 37. 604-606. September 2001.

Kostoff, R. N. “Citation Analysis for Research Performer Quality”. Scientometrics. 53:1. 49-71. 2002.

APPENDIX 9-A

NETWORK MODELING FOR DIRECT/ INDIRECT IMPACTS [Kostoff, 1994i]

Background

In a mission-oriented research-sponsoring organization, the selection and continuation of research programs must be made on the basis of outstanding science and potential contribution to the organization's mission. There have been increasing pressures to link science and technology programs and goals more closely and clearly to organizational as well as broader societal goals [Carnegie, 1992]. The process of estimating **potential impact** of research, especially basic research, on organizational and societal goals is complex due to the myriad of pathways by which the research product can effect its impact.

Most resource-allocation methods in the literature that incorporate organizational objectives tend to be **qualitative** when addressing **basic research**, and more **quantitative** when addressing **applied research** allocation.

-(See Logsdon [1985], OTA [1986], Hall [1990], IEEE [1974, 1983], Baker [1964], Cetron [1967], Datz [1974], Baker [1974, 1975], Winkofsky [1980] for **reviews** which compare selection methods and sort these methods into categories or classes;

-see Kostoff [1983a], Hazelrigg [1982], Helin [1974], Souder [1978], Cook [1982], Nutt [1965], Souder [1975], Van de Ven [1971], Plebani [1981], Mottley [1959], Garguilo [1981], Gear [1971], Pound [1964], Dean [1965], Moore [1969], Gustafson [1971], McGuire [1973], Paolini [1977], Cooper [1978], Ramsey [1978], Krawiec [1984], Gear [1974], Keef er [1978], Madey [1985], Liberatore [1987], Dean [1962], Cramer [1964], Vanston [1977], Bell [1967], Cochran [1971], Themelis [1976], Aaker [1978], Liberatore [1981], Silverman [1981], Menke [1983], Ellis [1984], Hertz [1964], Hespos [1965], Maher [1974], Schwartz [1977] for **benefit measurement** methods [develop quantitative measures of the benefit of performing an R&D project, then select those projects which provide greatest benefit] as defined in Hall [1990];

-see Watters [1967], Asher [1962], Beged Dov [1965], Baker [1969], Souder [1973], Keown [1979], Winkofsky [1981], Taylor [1982], Hess [1962], Rosen [1965], Atkinson [1969] for **constrained optimization** approaches [optimize some objective function subject to specified resource constraints] as defined in Hall [1990];

-see Cooper [1981], Stahl [1983], Lockett [1970], Mandakovic [1985] for **cognitive emulation** models [establish an actual model of the decision making process within an organization] as defined in Hall [1990])

Almost all of the allocation techniques in the literature are more appropriate for the applied research, or development, projects. Use of R&D project selection models falls into three categories [Roessner,

1985]:

1. A decision maker was influenced on a particular decision by the findings of a specific piece of research (instrumental use);

2. A decision maker finds that a piece of research contains ideas or information that contribute to the work of his/her organization (conceptual use);

3. A decision maker uses research to advance his/her own self-interest (partisan use).

Whether these allocation techniques are categorized according to OTA [1986] (scoring models, economic models, constrained optimization models, risk analysis models), or categorized according to Hall [1990] (constrained optimization methods, benefit measurement methods, cognitive emulation models, ad hoc methods, surveys) these techniques require, in practice, a project's **development and payoff** characteristics. These characteristics can be estimated when a project's downstream development phase can be identified, such as for some types of applied research, and for many types of development projects. For many areas of basic research, development and payoff characteristics are not obvious. There do not appear to be viable quantitative resource allocation models applicable to basic research.

This Appendix discusses a network based modeling approach which would allow estimation of the direct and indirect impacts of a research program or collection of research programs. The research program impacts would be multi-faceted, including impacts on advancing its own field, on advancing allied fields, on advancing technology, on supporting operations and mission requirements, etc. The model proposed here differs from any reported in the literature in that it reflects more accurately the different types of impact which basic research generates. A major feature of the model is **inclusion of feedback from the higher development categories (e.g., exploratory development, advanced development) on the advancement of research.**

Philosophy of Proposed Network Approach

Existing matrix-based research impact models [Dean, 1972; Ibrahim, 1984]) are most useful for applied R&D concepts and utilize a **vertical impact structure (forward diffusion of knowledge) where the impacts of research flow forward only to the more advanced development categories (e.g., research----> development----> systems)**. The proposed model uses a structure of **lateral and backward diffusion of knowledge superimposed on the vertical impact structure (e.g., research----> research----> development----> research----> development----> systems).** The proposed model accounts for the upward impacts of research (forward diffusion) allowed by the present models. It also allows one research field to impact another research field (lateral diffusion) and allows the higher development categories to impact research as well (backward diffusion).

For example, a matrix model approach could have a vertical impact structure path consisting of Physics (research) impacting Lasers (technology) impacting Beam Weapons (systems). The proposed network model would include this path, but many others as well, including Physics

(research) impacting Lasers (technology) impacting nanoelectronics (research) impacting Controls (technology) impacting Beam Weapons (systems), and including Physics (research) impacting Lasers (technology) impacting Fluid Flow Visualization (research) impacting Helicopter Blade Design (technology) impacting Helicopters (systems).

The impact of much basic research, especially on the higher development categories such as systems development, proceeds through many indirect paths. A quantitative model of impact should have the capability of identifying the paths along which impact occurs and quantifying the impact along as many paths as is possible. The existing forward diffusion matrix-based models are severely constrained on the number and types of paths along which impact occurs. These models are **not** able to account for impact along **lateral diffusion** paths (e.g., research-research) or along **backward diffusion** paths (e.g., technology-research). The proposed model allows impact to occur along any of these paths, and thus includes many types of indirect impacts as well as direct impact.

Example: Differences between Matrix and Network Approaches

A simple example will show the difference in breadth of impact allowed between the proposed model and a leading existing matrix-based model [Dean, 1972]. Assume it is desired to compute the impact of a research project R on a technology project T. In the standard methodology, it is only necessary to examine **ONE** path from R to T. This is the path of direct impact, and the value of the impact is the value of the matrix element RT.

In the proposed methodology, R and T are two nodes in a fully connected network. All possible paths between R and T are examined when computing the total impact of R on T. Thus, the overwhelming **majority** of paths which contribute to the **total impact** of R on T are the **indirect impact paths**. The total impact of R on T is the sum of the link value products along **EVERY** path connecting R to T.

Continuing the example above, R could be the Physics research node and T could be the Laser technology node. In the standard matrix approach, only the direct impact of Physics on Lasers is considered. In the proposed methodology, additional paths between Physics and Lasers, such as Physics impacting Fluid Dynamics research impacting Lasers or Physics impacting Solid State Materials research impacting Lasers, would also be considered.

For a graph with a large number of nodes N, there are approximately e*m! paths (ranging in length from 1 to N-1 links) connecting R to T, where m is N-2. In the pilot study performed to test the validity of the proposed model and overviewed in this Handbook, the graph that was used consisted of 15 research nodes and 27 technology nodes. For the pilot study graph, e*m! is approximately 10 to the 47th power.

*IN THIS SIMPLE EXAMPLE BASED ON THE **SMALL** PILOT STUDY GRID, THE PROPOSED METHOD COULD THEORETICALLY EXAMINE LINK VALUE PRODUCTS ALONG **47 ORDERS OF MAGNITUDE** MORE PATHS THAN DOES THE STANDARD METHOD.*

In the actual pilot study, link value products were computed along all paths five links or less in

length. This means that approximately m^4, or 2.5 million paths connecting R to T, were examined. This same order of magnitude differential holds between the proposed method and the other matrix-based methods which were examined before the proposed method was devised.

Of equal importance to the **quantitative** difference between the two methods is the **qualitative** difference. The proposed approach allows **full weight** to be given to those research projects which have **large indirect** impacts. Many of the fundamental research areas, such as Mathematics, Physics, etc., have substantial impacts on other research areas (as well as technologies), and these indirect impacts are not fully captured in the matrix-based methods. Since the fundamental research areas tend to have indirect impact on many research and technology areas, when the impact is summed over all research and technology areas, the **total impact of these fundamental research areas becomes substantial**.

For any organization with a substantial fraction of its budget in these fundamental research areas, a method that is able to capture the sizeable indirect impacts of basic research is important. For an advanced technology development organization, where the impacts of the work are more focused to specific technologies and requirements, the benefits of the proposed multipath approach may be less (although they will always be greater than those of the matrix approaches, since the proposed method includes all the paths in the matrix approach and others).

The remainder of this section describes the proposed method, an overview of the preliminary pilot study that was performed to test the feasibility of the method, key lessons learned from the pilot study, and recommendations for an enhanced study.

METHODOLOGY

Creating Domains and Forming the Network

The research impact quantification methodology presented here displays the value of a given research program to advancing its own field, to supporting other research areas, to supporting technology, and to supporting mission requirements. The first step in the methodology is defining a domain of potential impacts. For example, if the impact of research on other research, technology, and systems is desired, then the three-level domain for the model would be research, technology, and systems. Each of these levels is subdivided further into a number of categories.

As a specific example, in the two-level domain (research, technology) pilot study that will be overviewed, research was divided into 15 categories (math, physics, chemistry, etc.) and technology was divided into 27 categories (training, navigation, countermeasures, etc.). These categories had the property of being relatively non-overlapping, and were similar to categories being used by the Navy for management purposes at the time of the study. All 42 categories are represented as nodes in a network.

Since it is assumed that research, technology, and missions are interlocked and have mutual impacts with different strengths of connectivity, each pair of categories (nodes) can be visualized as

connected with a line (link). This schematic has the form of a graph, or network in which all node pairs are connected. The lines, or links, which connect each pair of nodes, are allowed to have two values, depending on direction between the nodes. This allows any research, technology, or missions area at the lowest category breakdown level to impact any other research, technology, or missions area with a specified strength.

Since one of the desired outputs of the proposed procedure is impact of research, and since research, technology, and missions are assumed to have mutual impacts, then the generic computational problem is to obtain the impact of one node of the network on any other node in the network. Three interrelated types of impact (**DIRECT IMPACT, IMPACT, TOTAL IMPACT**) of one node on any other node will now be described.

In this multi-node network, assume 'a' is one node, 'b' is a second node, and 'x' is a third node. The **DIRECT IMPACT** of node 'a' on node 'b', or more specifically, the direct importance of results from node 'a' to the achievement of objectives of node 'b', is the value (L ab) of the link directed from node 'a' to node 'b'. Thus, if 'a' represents a research node (partial differential equations, for example), and 'b' represents a technology node (short wavelength lasers, for example), then (L ab) would represent the direct importance (or **DIRECT IMPACT**) of research results in partial differential equations to the achievement of development objectives of short wavelength lasers. The scale of (L ab) ranges from 0% importance, which means results from node 'a' have no impact on achievement of objectives of node 'b', to 100 % importance, which means results from node 'a' are absolutely crucial to the achievement of objectives of node 'b'.

The **IMPACT** of node 'a' on node 'b', along any multi-link path connecting node 'a' to node 'b', is defined as the product of the link values (**DIRECT IMPACTS**) along the path. On the two link path 'a'-'x', 'x'-'b', the **IMPACT** is the product (L ax * L xb). Thus, if results from work in node 'a' are 25% important to obtaining objectives in node 'x', and results from work in node 'x' are 25% important to obtaining objectives in node 'b', then the IMPACT of node 'a' on node 'b' along the two link path 'a'-'x', 'x'-'b' is 6%. Other functions to represent **IMPACT** along the multi-link path could be defined, but the product of link values appears to be simplest and easiest intuitively to relate to reality.

The **TOTAL IMPACT** of node 'a' on node 'b' is defined as the sum of the **IMPACTS** along every path connecting node 'a' to node 'b' and is the main figure of merit used in the present study. The computational problem for obtaining **TOTAL IMPACT** of node 'a' on node 'b', then, is to trace each path from node 'a' to node 'b', compute the link value products along each path to obtain the **IMPACT** of 'a' on 'b' along the path, and sum the **IMPACTS** over all the paths connecting node 'a' to node 'b'. To eliminate double counting, and to insure that the **IMPACT** of node 'a' on node 'b' decreases as more links are added to the particular path connecting node 'a' to node 'b', the values of all the links coming into node 'b' should not exceed unity.

<u>Normalizing Link Values</u>

This condition is incorporated into the computational process by using a normalized value for each

link value in place of the value provided by the data source; i. e., L' ij = L ij * (1 - L jj)/SUM (L ij) where L ij is the data source link value, L' ij is the normalized link value, L jj represents the fraction of the objectives within node 'j' that can be achieved without input of results from any other nodes in the network, and the sum is taken over all the links coming into node 'j'. The equations without further constraints allow loops to exist in the network. For example, a three link path between node 'a' (Math) and node 'b' (Lasers) could be node 'a' to node 'x' (Physics), node 'x' to node 'a', and node 'a' to node 'b'. While this would be viewed as double counting if it were to occur at one point in time, it is perfectly valid when these steps among nodes occur at different times. Thus, the **IMPACT** of node 'a' on node 'b' has to be interpreted as a cumulative impact over time and is a function of the length of the path from node 'a' to node 'b'. An exact solution for the **IMPACT** would therefore require link values for every step in time from the present to the computational time horizon. Further, each of these link values could not be obtained independently, but would require knowledge of the link values connecting all the nodes at the previous time step, since progress in any one node is assumed to depend on previous progress in all of research and technology. To keep the computational and data generation problem manageable, an approximate solution is obtained by treating the link values as constants rather than functions of time, and interpreting and providing the link values as time-averaged quantities. Without knowledge of the variation of the link values with time, a credible estimation of the error resulting from the constant link value assumption cannot be made.

PILOT STUDY OVERVIEW

Taxonomy Used

It was the author's intent to identify the pathways through which research programs could impact technology areas and eventually naval and other application or mission areas. In parallel, some quantification of the impact of these programs was desired. A complete study would have required hundreds of nodes, many experts or other sources of the raw link value input data, and large amounts of data handling and entry. As a first step, to test the feasibility of the overall method, a small-scale pilot study was performed. Research and technology levels were included in the computational network; missions were not included. The final research taxonomy selected for the study was identical to the categorization which the Office of Naval Research used for research management purposes at the time of the study. The final technology taxonomy selected for the study was similar to functional element breakdowns used in the past by Navy exploratory development programs for management purposes. These two taxonomies had the virtue of being fairly comprehensive in their coverage, at least as far as the Navy is concerned, and there were in-house experts available to provide preliminary link value data for each of the subcategories in these taxonomies. Of necessity, the taxonomy elements used were very broad. Each research taxonomy element (e.g., Mechanics) contained a number of different research programs (e.g., Solid Mechanics, Fluid Mechanics, Energy Conversion), which themselves could have been divided into subprograms.

Data Acquisition

The data was obtained by personal interview. Each in-house expert was provided with a list of the 42

research and technology nodes, and was asked to estimate the importance of results produced from all the other nodes on his particular node of expertise. The expert was asked to provide a number which served as a measure of impact based on the following scoring scale: Crucial(10); Very Important(8); Important(6); Moderately Important(4); Slightly Important(2); Negligible(0). Definitional uncertainties were minimized due to the presence of the interviewer.

Because the approach is based on subjective judgement, there are limitations to the validity of the data, especially with the small numbers of experts per node that were employed. There was no attempt made to normalize the responses, and an impact that one expert labeled Important could have been labeled Moderately Important by another expert. There was no attempt to gauge the degree of expertise of each respondent relative to his field of expertise, and the numerical ratings supplied, therefore, carry different degrees of validity. Because of the broad discipline coverage of each node, the expertise of any respondent relative to the breadth of the discipline was quite limited. Use of a small number of experts per node did not provide a good statistical representation of how each technical community would have perceived impact on its discipline.

Because of the rapid convergence of the link fractional value multiplication process, it was found that timely and accurate results could be obtained with networks whose longest paths were three links in length. Including a fourth link made only a very few percent difference in the results.

Lessons Learned from Pilot Study

The results from the pilot study are described in detail in Kostoff [1994i]. The lessons learned from the pilot study will now be described. The pilot study was limited by a number of factors, especially the broad coverage of each node. To expand the scope and capabilities of the study methodology to the point where study results could support credibly the priorit ization of research areas and produce a more evidentiary basis for establishing program balance, the following steps would be required at a minimum.

1) First, the research and technology nodes need to be subdivided to improve resolution.

2) The second major improvement required over the pilot study is the addition of missions nodes to the network.

3) The third improvement is that research, technology, and missions taxonomies need to be orthogonalized better, so that overlaps among nodes and resultant skewing of the results are minimized.

4) Fourth, the number and range of experts per node need to be expanded to provide more node representative than the one or two experts per node provided in the pilot study.

5) The fifth improvement is that the written material supplied to the respondents needs to be sharpened, especially in the absence of an interviewer.

Operational Value of Present Approach

The final issue in this section addresses the operational value of the present approach. When the pilot study was proposed, the type and significance of results finally obtained were never expected. As the study proceeded, much information about the interlocking nature of research and technology was obtained in addition to that provided on the questionnaires. Thus, much of the study's value derived from the performance of the study, and additional study benefits would be expected from a refined study.

From another perspective, a refined study could serve as a total program assessment. It could identify gaps, duplications, promising research areas, and funding priorities for the total program taken as a whole. The typical technical assessment performed today focuses on a technology or research area, and defines required research to allow attainment of technology and mission objectives. However, in the zero-sum game environment of finite resource constraints, money to fund the required research identified by the assessment has to be taken away from proposed or existing research in some other area. Unless the total impact of unfunding this other research can be identified, it is not clear whether the overall research program would benefit by funding that research identified by the technology assessment. In fact, it is evident that unless all technology and research are assessed simultaneously, funding reallocations based on one or two specific technology assessments could be highly suboptimal and misleading and could affect the overall research program adversely. A refined study could serve as a total research and technology assessment, performed at the project level, and may perhaps be the only sensible way to perform a technical assessment.

APPENDIX 9-B.

NETWORK MODELING FOR ROADMAPS [Zurcher and Kostoff, 1997]

Introduction

One of the motivations for research assessment and evaluation studies is to gain a better understanding of the potential myriad impacts of the research, and then use this understanding to help accelerate the transition of the research to useful technology. Accelerating the conversion of science to technology has three essential elements:

1) Information about the science must exist and be readily available to potential users;

2) The need for the converted science (technology) must exist;

3) One or more entrepreneurs who recognize the need, who understand the relationship between the need and the science, and who are willing to obtain the necessary resources and accept the risks inherent in further development of the science, must be available to champion its further development.

Large databases, which describe ongoing and completed research, are commercially available (e.g., journal paper abstracts, federal project and program narratives). With global competition for markets, the need for new technology has never been greater, and many compendia of projected technology requirements are available (National Academy of Science/Engineering Studies, Agency Requirements Documents, etc.).

However, availability of research and requirements information is not sufficient to motivate potential entrepreneurs to invest time and other resources in the high risk research conversion process. Investors must be convinced that the considerable front-end risk of science conversion is more than justified by the potential payoff. Placement of the science conversion step into the larger pathway from research to high-payoff applications is a key component for eliciting investor interest. While relatively large resources have supported the development of the research databases, and substantial study efforts and market surveys have contributed to the volumes of existing requirements, relatively few efforts have focused on fusing together requirements with research systematically.

There are fundamental reasons why little progress has been made on methodologies to identify the characteristics of these linkages. The pathways between research and eventual applications are many, are not necessarily linear, and require significant amounts of data [Kostoff, 1994i; previous section on network modeling]. Substantial time and effort are required to portray these links as accurately as possible, and substantial thought is necessary to articulate and portray this massive amount of data in a form comprehensible to potential investors. Recently, desktop high speed computers with large storage capabilities, intelligent algorithms for manipulating data, and other

tools have become available to allow these research-capabilities pathways (roadmaps) to be constructed and portrayed efficiently and effectively, and to be used as a basis for more detailed analysis.

The main value of these decision aids, or roadmaps, in the science conversion process is to promote, at all phases of the roadmap development process, champion/ investor interest in developing the research further. In planning the roadmap, thought has to be given to all its structural elements, including the extent of the development required, any trade-offs or opportunities lost, and potential costs and payoffs. In building the roadmap, experts in the different levels of development and payoff become involved, and the risks, potential costs and benefits are clarified further. When the completed roadmap is distributed to interested parties, decisions to pursue the science conversion can be made with greater understanding of the larger development context. For a more comprehensive discussion of roadmaps, see Science and Technology Roadmaps [Kostoff, 2001m].

Retrospective studies of successful innovation have shown that at least one champion is required to insure continuity and persistence toward the final goal [Kostoff, 1997c]. Other studies have shown that two champions are preferable, one from the technology-push side and the other from the requirements-pull side [Rubenstein, 1997]. In realit y, there are at least three major parameters which govern the role and impact of champions on the science conversion process. The first is numbers: the more champions, the more likely is the conversion process support. The second is intensity: the more intense the interest and persistence of the champion(s), the more likely is the research to proceed. The third is influence: the greater the influence of the champion(s), the more likely are the chances that the research conversion will be pursued.

Having potential champions involved in the planning, developing, and distribution of the roadmap improves the likelihood of numbers, intensity, and influence of champions being increased if analysis of the roadmap shows downstream potential for substantial payoff. If roadmap analysis does not show convincing evidence of payoff of the research toward the objectives, either due to intrinsic lack of potential payoff or to unawareness of payoff of those constructing the roadmaps, then the research may not proceed further. If the roadmap analysis shows high potential payoff, but with extremely high front-end risk and costs, then the type of champion interest may be limited to government for the initial risk-lowering development phases.

This section overviews the algorithmic component and analyt ic potential of the Graphical Modeling System (GMS), a computer-based process for generating and analyzing roadmaps which link research to technology and eventually to capabilities/requirements. This process has been under development for the past decade [Zurcher, 1997], and its algorithmic component is based on a directed graph/ network model of research/technology/capabilities/ requirements. It uses the latest relational database/ hypertext technology to identify the potential pathways which link research to higher development categories and specific requirements/ targets of interest.

In the past, many methods have been developed to select or evaluate R&D projects [Fahrni, 1990; Cooley, 1986; Jackson, 1983; also see references in previous sect ion on Network Modeling]. These methods typically use simple checklists, scoring, cost/benefit analysis, mathematical programming

or decision trees to determine future value from a current investment. Other methods describe the value of R&D projects by attempting to measure the effectiveness of transfers of technology [Spann, 1995] without explicitly taking into account customer requirements. Some algorithms link research programs to end uses/ capabilities/ requirements [Thomas, 1996; Barker, 1995]. This last method 1) creates a context within which technology projects exist, 2) requires a flexible technology assessment methodology since requirements change and emerging technologies will modify current plans, and 3) demands continual dialog between customers and developers. As shown in the previous section on network modeling, in the classical matrix approach [Dean, 1972], impacts flow monotonically upward in the development chain (research --> technology --> capabilities --> requirements/end targets), and in the network/ directed graph approach [Kostoff, 1994i], impacts are allowed to flow upward, downward, or laterally in the development chain (e.g., research --> technology --> research --> research --> technology --> capabilities). GMS is able to show the node-link relationships of both the matrix and network approaches (where a research or technology project, or a capability, is treated as a node in a network, and the impact of one project [node] on another project [node] is portrayed as a quantified link in the network).

In addition, GMS adds a crucial new capability, termed Multiple Perspectives (MP). In GMS, the nodes (projects/ capabilities/ requirements) are treated as multi-valued (multi-attributed) quantities, and are allowed to exist in many different research-requirement pathways simultaneously. This MP capability provides a more accurate depiction of the multi-application nature of most research and technology. The user of GMS is now able to highlight only the specific node-link subnetworks of interest (the desired research-requirement pathways) without being overwhelmed by the massive data which constitutes the larger network.

For example, the MP capability enables the user to select research-requirements pathways to view (e.g., "top-down" requirements perspectives, or "bottom-up" science/ technology perspectives rather than viewing all, potentially complicating, nodes and links, or having a static display that can not change). Researchers can 1) observe the larger context in which their work is being performed, or 2) identify new applications targets for their research, and make informed decisions on how to proceed to maximize payoff for multiple applications. Also, it allows the user and other interested parties to identify the research and technology projects which presently serve as obstacles to reaching desired applications targets in a timely manner.

Methodology

The roadmap, or graphical model, overviewed here is a selected set of requirements, links and R&D projects that describes the state of technology development and potential transfer in a coherent area. It could be composed of a single requirement for a system linked to corresponding R&D projects, or it could encompass multiple requirements linked to numerous projects. A graphical model visually portrays:

- requirements,
- capabilities,
- R&D projects in different development phases;

- relationships between R&D projects and requirements; and
- integration among related R&D projects.

The GMS depiction of the science conversion process is assembled in a two-stage process: 1) Construction of a graphical model; 2) Analysis of the pathway elements between requirements and R&D projects.

a. Model Construction:

Model construction consists of identifying the projects and requirements (nodes) for the roadmap, then identifying the relationships (links) between the projects and requirements.

Step 1: Identifying Types of Projects and Requirements

R&D projects and requirements are partitioned according to the phase of development of the R&D projects and to the level of specificity of the requirements. While the actual graphical models used employ a half-dozen or more bands for subdividing project and requirement types, for purposes of demonstration simplicity the roadmaps shown in Zurcher [1997] have four levels: research, development, capability, requirements.

Constructing the roadmap framework (i.e., identifying the specific nodes to be used in the roadmap and the placement of those nodes at the appropriate level of development) is perhaps the most challenging step in the roadmap development process. It is somewhat paradoxical in that the appropriate expertise must be employed to develop a roadmap, but the appropriate expertise becomes fully known only after a complete roadmap has been constructed. An iterative roadmap development process is therefore essential. For an organization in which many of the roadmap components are being pursued in-house, such as a large focused government or corporate laboratory, much of the expertise can be assembled in-house. Researchers, developers, marketers and others with relevant knowledge of the overall roadmap theme can be readily convened to develop the framework. At the other extreme, organizations with little expertise in the overall roadmap theme, such as venture capital groups or cash-rich organizations that wish to expand their boundaries, will require external assistance to develop credible roadmaps.

The utility of a roadmap increases as it expands to include potentially relevant R&D performed in all sectors of the technical community. The experts constructing the roadmap can draw upon their personal experience and contacts in identifying other R&D performed in the community, and should utilize computerized resources such as program narrative databases to identify relevant external R&D. The quality and credibility of the roadmap increases as more experts are employed in its construction. While it is preferable to have at least one expert in each node technical area (e.g., if ELECTRO-CHEMISTRY RESEARCH is one node, then at least one expert in this area should be part of the roadmap development team), useful roadmaps can be constructed with fewer contributors of broader expertise.

Experience has shown that major benefits accrue during the iterative process when the experts

are convened to develop the framework. The roadmap serves as an important component of both strategic planning and technological forecasting for the organization, and forces the developers to clarify conceptual strategic targets in order to represent them graphically. Awareness of all the contributors to R&D required and R&D available in other sectors of the technical community is increased, sometimes dramatically. In particular, critical path research can be identified, and support for its accelerated development can be strengthened. The main value at this phase is to the developers themselves; additional value accrues when the completed roadmap is provided to external users.

Step 2: Identifying Links Between Projects and Requirements

Once the full complement of nodes has been identified, the next step is to graphically and quantitatively depict the relationships among the nodes. One node is represented as linked to another node when the results emanating from the first node are assumed to have some impact on the achievement of targets of the second node. This relationship is depicted graphically by a line, or link, connecting the two nodes, and is quantified by assigning a value to the link (e.g., Kostoff, 1994i). It is important that node experts from both ends of the link (the results generator node and the results user node) are involved in assigning the link value. Finally, the inherent hypertext capabilities of GMS allow more descriptive information about each node and node-connecting link to be accessed at the touch of a button. These hypertext capabilities allow the rationale for the selection of each node, and selection of node and link values, to be obtained easily, and thereby provide deeper insight to the potential obstacles and impediments to successful research development and transition.

It is assumed that the experts in the node thematic areas are most qualified to assign values to the links entering and exiting their particular nodes of expertise. Experience has shown that most credible impacts are nearest-neighbor (e.g., basic research node outputs tend to impact applied research nodes; applied research node outputs tend to impact early development nodes). The impact of research on far-neighbor nodes, such as advanced technology projects, tends to occur along pathways consisting of nearest-neighbor steps. Thus, the developed network consists of individual node-link subnetworks, each of which has been assigned node and link values by appropriate experts.

Conceptually, however, the developed network is greater than the sum of its nodes, just as the living human body is greater than the sum of its component cells. The developed network includes the intelligence or inherent logic, as quantified by the link values, which connects the nodes to each other and to the overall mission goals, just as the living human body includes the intelligence which links the cells to each other and to the homeostatic operation of the body. As a result of the expert intelligence applied to quantifying each node value as well as the entering and exiting link values, there are at least two new crucial pieces of information provided by the developed network: 1) The strength of the relationships among the projects/ capabilities/ requirements and the subsequent identification of high obstacle and low obstacle paths; 2) Identification of R&D projects being conducted external to the organization, their importance to successful attainment of the organizations goals, and their potential for leveraging by the

organization. Even when node experts have not been identified or cannot be obtained, valuable information about gaps in expertise availability has been generated. The developed network with its enhanced information content now serves to promote communications among all the participants and provide a stronger basis for credible analysis and decisionmaking.

b. Model Analysis

A variety of analyses can now be performed, limited only by the interests and imagination of the analysts. The quantified network, which contains a comprehensive collection of nodes, can serve as the foundation for detailed economic studies, broad systems studies, and parametric tradeoff studies. The initial utilization of the network should serve to foster internal communications and consensus, in preparation for these more detailed analyses.

Obviously, the breadth of information obtained from the different perspectives will be limited by the contents of the total database. In an ideal world, all existing and proposed R&D programs would be entered in the overall database, and the full impact on technology and capabilities of existing and proposed research programs would be identified. In addition, the total R&D available to address required goals and capabilities would be displayed. Because of all the potential node-link combinations, and the attendant enormous amount of data required (Kostoff, 1994i), constructing this complete database is not feasible at present. However, the central thesis of the present paper is that subsets of the total database embedded in the larger analytical process still have substantial value. The existing GMS has a total R&D database constructed from the different specific mission application perspectives which have been performed, and increases in value for an organization as more perspectives are generated.

The value of graphical models is that they show R&D projects and requirements in context rather than in isolation, they can depict new perspectives rapidly, and they can serve as a focal point for enhanced communications and more detailed total systems analyses. Since the context of graphical models is different for each perspective while still using common elements (projects, capabilities, requirements), comprehending a broad R&D program and associated requirements is very difficult without the ability to sort out these elements and how they relate to one another.

Summary and Conclusions

Transferring technology to customers efficiently through a succession of autonomous development groups requires extraordinary coordination. There are many opportunities for technology transfer to become stalled at any point along the way by disparate priorities among many groups. Depicting potential science conversion in a graphical model discloses to the scientists and investors alike the possible transfer points where obstacles may occur to technology transfer or requirements specification [Geisler, 1995].

The benefits of graphical modeling include:

1) showing R&D projects and requirements in context rather than in isolation,

2) multi-attributed nodes which can portray different research-requirement pathways rapidly,

3) serving as a focal point for enhanced communications and more detailed total systems analyses,

4) promoting champion/investor interest,

5) portraying R&D programs as being strategically planned,

6) portraying leveraging of R&D projects from other organizations,

7) identifying obstacles to rapid and low-cost technology development.

APPENDIX 10

EXPERT NETWORKS [Odeyale and Kostoff, 1997q]

Research Impact Assessment is, at its essence, a diagnostic process with many diagnostic tools. In other fields of endeavor, such as Medicine and Machinery Repair, expert systems are increasingly being used as diagnostic tools or as support to diagnostic processes. Recently, there have been efforts to develop expert system approaches combined with artificial neural networks (expert networks) for use in R&D management, including RIA [Odeyale, 1993; Odeyale and Kostoff, 1994a, 1994b]. These efforts will be summarized in this section. Much of the remainder of this section was contributed by Dr. Charles Odeyale, a true visionary in the application of Expert Networks to the broad area of R&D management.

Overview

To increase the degree to which rationality is used to guide decisions, the authors' efforts have been directed towards a comprehensive R&D management tool, a high-tech Peer Review, through a modified version of a previous Office of Naval Research review process. The product of these efforts is Research-Management Expert Network (**R-MEN**) which is characterized by two complementary tools: Organizational/Professional Development and Expert Network. The latter technology is comprised of an expert system (left side brain) and an artificial neural network (right side brain). Given a set of research, and research management policies and strategies, **R-MEN** learns concepts that hierarchically organize those policies and strategies and use them in classifying/triaging research proposals. A brief and non-technical description of how this knowledge technology would foster continuous "learning", improve value and efficiency, increase productivity, and provide excellent performance measures of activities is presented.

Introduction

There is much concern about improving the health of basic research. The increasing politicization of the support of research has awakened many organizations to the risks and realities of survival. There is a growing sentiment that it is no longer enough that research just be excellent, or generate new information; research must contribute results aimed toward national goals. Research and Development (R&D) administrators and managers need a powerful management tool to enable them to predict, assess and monitor the impact(s) of research results and research management processes at the project, program, organizational, and national levels.

As administrators and managers struggle to establish policies/strategies that balance cost issues with research outcomes, establishing systems to predict, assess and monitor the impact(s) of research results and research management processes should be an important consideration. The authors have discovered that successful outcomes-management systems require five basic components:

- openness-to-change,

- specification process,
- information/ knowledge technology,
- measurement instruments, and continuous learning and
- improvement.

For greater processing power, immediate access to information, and powerful applications that monitor, analyze, and manage, the authors have reported [Odeyale, 1993; Odeyale and Kostoff, 1994a, 1994b] a technology whose functionalities surpass these requirements. This value and efficiency improvement technology, which is a comprehensive computer-based Research Impact Assessment (**RIA**), is characterized by two compound mutually complementary tools: Organizational/ Professional Development (**O/PD**) and Expert Network (**EN**).

The framework of Research-Management Expert Network (R-MEN) was reported by Odeyale and Kostoff in the references cited above. It consists of a knowledge base and a data base. Feeding into the knowledge base are four modules:

- a policy/ strategy impartation module and
- a proposal data acquisition module, both of which receive input from the O/PD process; and
- a research impact calculation module and
- a proposal review module.

The knowledge base then feeds into the data base through five modules:

- a project selection module,
- resources allocation module,
- project evaluation and control module,
- investigator evaluation module, and
- organization evaluation module.

Within the framework of Research-Management Expert Network (**R-MEN**), **O/PD** pertains to the relevance, transferability, and system alignment of the training and development efforts of each and every individual in the organization. Most importantly, these criteria of timely selection, training and development of individuals are taken in conjunction with changes in organizational environments and requirements. Through **O/PD**, attitudinal, behavioral, procedural, policy, and structural barriers are uncovered and "removed" to enable effective performance at all levels. To effectively manage this continuous "learning", improve value and efficiency, increase productivity, and provide excellent performance measures of activities, an information/knowledge technology is needed. All these needs, and more, are met by the **EN,** which is comprised of an expert system (left side brain) and an artificial neural network (right side brain). This integration of information processing techniques avoids the limitations of each technique while capitalizing on their unique benefits. Expert Systems, and Knowledge-Based Systems in general, including artificial neural network, are computer programs that deal with

complex problems ordinarily solved by human experts who are highly skilled, trained, and experienced in the specific area of interest.

The conceptual construct that provides the framework for the **OP/D**-based research management processes is described in three phases as shown in Table 1.

Table 1 PARTICIPATIVE R&D MANAGEMENT PROCESS

PHASE.....PROCESS.....MANAGEMENT.....MANAGERIAL

.....LEVEL.....STYLES

I

Position...a...Pre-Vision.....Sr. Executives (with.....Authoritative

Audit.....R-MEN)/Sr. Scientists

.....b....Strategic.....Sr. Executives (with.....Democratic

.....Vision.....R-MEN)/Sr. Scientists

.....c....Design &.....Sr. Executives (with.....Democratic/

.....Planning.....R-MEN)/Sr. Scientist.....Authoritative

II

R&D .....d....Introduction.....R&D Director.....Authoritative

Process

.....e....Implementation...Sr. Scientists/Bench.....Pace Setting/Coaching

.....Level Investigators

III

Control...f....Evaluation &.....Sr. Executives (with.....Coaching/Affiliative/

.....Control.....R-MEN) Sr. Scientists.....Coercive

The above steps and components are identified to facilitate the development of accurate activity

standards to be used in the tracking, evaluation and control to foster accountability and productive efficiency. The general outline of the processes is in spirit with the reports of Dubnicki and Williams [1991], Englert [1991], and Kostoff [1992a]. The phases are briefly described below (see Odeyale [1993] for detail).

PHASE I

This phase includes the development of the strategic plan, which defines and communicates longer-term research directions, and the development of the operating plan, which specifically identifies the projects that will implement the strategic plan taking into consideration the goals, quantifiable objectives and development of the individual investigator and the organization. Series of processes with interlacing feed-back- and feed-forward-loops in operation during this phase include:

1. Formation of a top-management pre-vision team composing of theorists, technologists and practitioners who must demonstrate interest and commitment to this process and the **RIA** program as a whole. This team must be able to explain the "whys" behind directions or decisions in terms of the employees' and/or the organization's interests. Top management must include in their considerations: a) the uncertainties of innovation and the environments; b) the recognition of technology push (the brilliant idea seeking a field/market) and field/market pull (a field/market need seeking a product), and what the general corporate climate or attitude is on projects based on either; c) the determination of attribute, and formation of attribute tables with the disciplines or sciences which are determined to be absolutely necessary in the support of R&D unique to the organization.

2. Transformation of research, and research management policies and strategies into key terms that are used later in proposal text-body content analysis. Policies and strategies may include the research direction, preferred research technology, goals, objectives, values, etc.

3. Machine learning of the policies and strategies by **R-MEN** whose method of learning is incremental concept formation. The policies and strategies are grouped by research area as they are learned. They become a form of long term memory that remains the same until a change in policy and strategy is recognized and implemented by the management.

4. Collection of contract/grant applications through a Bulletin-Board-Service-like client/server system. From anywhere in the world through a software like "PC ANYWHERE", individual investigators can call in to fill out grant application electronic forms that visually resemble their paper counterparts. In addition, the bottom of the forms and/or the last page contain(s) control buttons for the collection of prediction/assessment related data which are needed for network computing such as benefit, contribution, feasibility, need, impact value, and proposal index value calculations. This same method is used for the collection of proposal review, and evaluation/monitoring related data such as solicitation of quantifiable opinions and objectives from reviewers and individual investigator, respectively. For example, investigator-objectives are projected and quantified for each evaluation period (one year) as follows:

a) No. of Poster Presentations (0.5 point each);

b) No. of Abstract Publications (1 point each);

c) No. of Paper Publications (1.5 points each);

d) No. of Graduate Seminar Lectures (2 points for a "once-a-week-one-semester" lectures);

e) No. of Developments (2 points each);

f. No. of Patent Applications (3 points each).

As an element of vision, the top management may envision or set as objectives for the whole (private or public) organization 300 publications, 450 published abstracts, 200 postal displays at major scientific and/or engineering society meetings, 10 developments, and the assignment of at least three patent rights in a one year period. All objectives must be in-line with those of the organization. After the completion of the forms, with appropriate warnings, access to application forms are denied once the "SEND" button is pressed.

5. The applications are grouped by research area as they are collected. At the end of funding agency published collection period, coded policies and strategies are used in proposal text-body content analysis of each proposal. That is, **R-MEN** will search the text-body of each application for the coded key terms, counting and adding only one instance of each key term. A major concern about the use of this technique is that investigators who know the key terms may write their proposals directly to address the key terms. Ideally, that is what the administration should require, i.e., the alignment of the investigators' goals and objectives with those of the organization. Besides, the investigators must meet their projected quantified objectives if they want their projects funded the next time around. This is outcomes-management, placing greater reliance on standards and guidelines. Furthermore, such resourceful proposal writing will be revealed during feasibility, need, and benefit calculations as described below. Anyway, the result of this content analysis changes (triage) the state of the application to either exclusion or inclusion in further review process.

6. For R&D_Area-Science Relationships (feasibility), Science-Requirement Relationships (need), and Requirement-Value Relationships (benefit), a portion of **R-MEN**'s inference technique uses a modified version of the Multiattribute Utility Technology (**MAUT**) in electronically obtaining the views of experts (from universities, government and industries), respectively, on: a) the potential impact of break-throughs in a research area on disciplines, and specific research subject; b) the contribution of the Science to satisfying operational requirements through suggested research opportunities (proposals); and c) the magnitude of the contribution of a set of proposals to satisfy a set of needs. Refer to Edwards [1980, 1982] for detail on **MAUT**. When a reviewer calls in to contribute his/her opinion to the opinion table, he/she will be asked to: i) review provided list of value disciplines and areas of interest in the terms of their being affected by any research break-through in one of the areas of interest (say blood substitutes); ii) rank order the value disciplines and provided areas of interest to reflect their being affected by research break-through in blood substitutes; and iii) weigh the value disciplines - assign 10 points to the least affected disciplines,

then accordingly assign the relative impact of blood substitutes research break-through on each discipline, (the limit is 100 and as many as 100, 500, etc. experts can "review" a proposal).

7. Before final proposal review and indexing, a mean for hypothesis testing is provided. This nonprimitive function provides relationship Congruency or Entropy values ranging between zero and a system determined value, depending on the data provided. It provides a choice of 99, 95, 90, 75 or 50% confidence level for the calculation of the entropy value. A value of zero means that the newly generated information/knowledge from **MAUT** obtained data adds relatively no useable information/knowledge to the existing one. A break-through research in a project may insignificantly contribute to a limited number of disciplines, i.e., there is no cross-fertilization. Replacing the entry in the cell of interest with a new value and repeating the calculation will generate a new value which may or may not be acceptable. Thus, it assists in the identification of special problems to be addressed before project selection. On the other hand, a value other than zero indicates a level of added useable information/ knowledge to the existing one. A break-through research in a project may significantly contribute to a number of disciplines, i.e., there is cross-fertilization.
8. Impact and index values are calculated for each of the applications using data including investigator's performance record, stated objectives, and desired outcomes. Every application whose "CRITERIA MATCH" field is occupied is included in the organization's R&D portfolio and automatically indexed based impact and index values. If they have not already been entered, the system will ask for available resources and minimum reserve, then, it will start assigning fund to projects starting from the one with the highest index value until the minimum reserve is reached.

PHASE II

This phase represents the necessary education, and management support needed to prepare the staff to participate in such an "Action Research" effort. This phase identifies and utilizes the critical components required to develop an environment that facilitates participative research management activities. A significant activity occurring during this phase is daily verification of individual scheduled training and development. If an individual has no recorded training and/or development within a preset period, the system will generate and send a report through E-mail directly to the office of the director for R&D. The system will be able to look at a training and/or development description(s) and compare it/them with the background of the individual to determine if the training and/or development is/are suitable for that individual. This is one of the ways how **R-MEN** shows concern for human feelings and human needs for support, dignity, and fulfillment in work.

PHASE III

This phase represents a means by which participative methods can be put into operation in developing productivity tracking systems. Significant activities occurring during this phase include project evaluation and control. This entails periodic monitoring of project milestones for applied research, and research objectives for the more basic research. If a project has no recorded fulfillment of a milestone within a preset period, the system will generate and send a report through E-mail directly to the office of the director for R&D.

ANTICIPATED BENEFITS

Frequently in human affairs, past intellectual baggage hinders our ability to forge novel approaches. Therefore, we advocate the use of **R-MEN** concurrently with present research review process. During this period, **R-MEN** is foreseen as a supplement in the form of a guide to data generation, acquisition and processing, and a validity check. Before long, just as the **R-MEN**'s anticipated review period is very significantly (62.5 - 66.67%) less than that required by un-aided review, other **R-MEN** benefits, including those presented below, will standout as well.
With appropriate implementation and maintenance, this knowledge technology, which utilizes demonstrated and proven approaches, methods, procedures and techniques in an innovative and unique way, would:

1. Provide a means for **effective**, policy- and strategy-oriented management through outcomes-management.

2. Improve management quality, reduce operation costs, and increase productivity and public trust.

3. Foster impact evaluation to document Federally funded program and management effectiveness.

4. Provide short-term (three-year) program progress tracking and long-term (ten-year) result(s) impact tracking.

5. Shield administrators, managers, and other policy-makers from the complexity of the mathematics of the inference machine.

6. Permit the evaluation of a range of alternatives.

7. Permit handling large amounts of data.

8. Permit policy-makers to have a better understanding of existing technical attributes of and capabilities for potential projects.

9. Facilitate choice of strategy compatible with agency structure and processes, and with the policy or the nature of decision making for activities scheduling and control.

According to Nonaka [1991], "In an economy where the only certainty is uncertainty, the one sure source of lasting competitive advantage is knowledge. And yet ... few managers grasp the nature of the knowledge-creating company - let alone how to manage it. The reason: They misunderstand what knowledge is and what companies must do to exploit it."

Is the reader up to date in strategic information/knowledge technology application? Is his strategy-structure and/or reward and training systems barriers or opportunities to professional and organizational success? Does the reader know how to integrate information technology with your

research management processes? These are where the authors' **R-MEN** technology comes in.

APPENDIX 11

POTENTIAL USE OF ENTROPY IN RESEARCH EVALUATION [Kostoff, 1997n]

In the assessment of research or research impact, many types of distribution patterns occur. There are:

- funds allocations across technical disciplines,
- funds allocations across performers,
- funds allocations across levels of development,
- papers produced in different disciplines,
- papers co-authored in different disciplines,
- papers published in different types of journals,
- citations by papers in different disciplines,
- citations by people from different types of institutions and different countries,
- patents produced in different technologies,
- patents cited by papers and patents in different disciplines, etc.

While these distributions are sometimes listed or catalogued during an assessment, they are rarely, if ever, subjected to a pattern analysis. Such an analysis would offer a much richer insight to research impacts or management processes than are offered by the standard examination of magnitudes alone. The use of entropy to characterize these distribution patterns offers a potentially substantial improvement in output interpretation of an assessment.

In statistical mechanics, the entropy is related to the number of micro-states (or states of the system at the atomic level) per macro-state (state of the system at the classical thermodynamic level). The statistical interpretation of the second law is that entropy tends toward the most probable state. The system proceeds from a state of order to disorder.

The information theory use of entropy is related to the statistical mechanics definition. If a system consists of N total units, and these units are distributed among m different states with a distribution function n(i), then the entropy s of the system may be written as:

.....i=m

.....s.=.-SUM..p(i)*ln((p(i)) (1).

.....i=1

where SUM represents the summation over all states i, and p(i) is the ratio of n(i) to N.

Thus, for any distribution n(i), equation (1) allows the entropy to be computed. The entropy can be interpreted as a measure of the order, or breadth, of the distribution, and its change can be tracked with time. It can serve as a single figure of merit for analyzing the distribution diversity of any quantity.
Examples of application of the entropy concept to two of the distribution patterns mentioned above follow.

Funds Allocations Across Disciplines or Levels of Development

Quantitative measures of the degree of vertical or lateral integration in an organization or in a group of programs would be useful to management for tracking purposes. It would also be useful for organizational assessments in being able to display the status of vertical or lateral integration. While quantitative measures are incomplete by themselves, and for the lateral or vertical integration measure here do not address the strength of the linkages among the different related disciplines or levels of development, they do provide a starting point for identifying potential problem areas.

Vertical or lateral integration within an organization makes it easier for multiple level of development or discipline funds to be managed jointly and at lower levels in the organization. The degree of multiple level of development or discipline funds management by an organizational unit is one component of vertical or lateral integration.

The quantitative measure proposed here for ascertaining the funds mixing component of vertical or lateral integration is the degree to which different categories of funds are managed jointly and at the lower levels in the organization. From this perspective, one aspect of vertical or lateral integration can be viewed as a process by which management of different level of development or discipline funds by the same unit diffuses into the lower levels of the organization.

The measure could take different mathematical forms. Some desirable limiting conditions include:

1) for a given amount of funds managed by the unit of interest (say, a Technical Manager), the measure should go to zero as all funds are lumped into one level of development or discipline;

2) the measure should go to one as the funds are equally divided among the levels of development or disciplines;

3) the measure should range between zero and one and be smooth in this region.

Many mathematical measures could be defined which have these desirable properties. Since the problem is in essence a funds mixing problem, and since there is a precedent for using entropy as a measure in physical or chemical mixing problems, the entropy definition above will be used as the metric for assessing the vertical or lateral integration funds mixing component.

The following example is for vertical integration, but with some modifications could apply equally well to lateral integration. Assume there are three levels of funds to be integrated: basic research, applied research, and development. Assume further that the unit of analysis is all programs under each Technical Manager in the organization. Then, for each Technical Manager, the entropy metric for his programs is given by the information theory expression for entropy:

.....i=3

.....s.=.-SUM...p(i)*ln((p(i))/kappa

.....i=1

where p(1) is the fraction of the Technical Manager's funds in basic research, p(2) is the fraction in applied research, p(3) is the fraction in development, and kappa is a constant which will produce an entropy s upper limit of unity.

The following table illustrates how the entropy function varies with different amounts of funds in the different levels of development in the Technical Manager's program. Each column represents different distributions of a $1000 total program.

BAS.RES...999.999..999..990..900..800..700..600..500..400..333

APP.RES.....0005....5....5...50..100..150..200..250..300..333

DEVELOP.....0005....5....5...50..100..150..200..250..300..333

ENTROPY.....0...01...06...36...58...75...87...95...99..1.0

As all funds are concentrated into one level of development, the measure goes to zero, and as the funds are divided equally among levels, the measure goes to one.

The first part of the following discussion applies to implementing the measure for tracking total organization performance, and the second part applies to implementing the measure for tracking individual program performance. The measure would be implemented in the following manner for the total organization. The organization's management at all levels would examine all programs and decide how the funds integration should be structured. This is the key step in the process, and requires that the different modes by which vertical integration will be effected be defined and

planned for implementation. There may be technical areas or Technical Managers where the vertical integration would be effected through close coordination and cooperation rather than funds mixing. For example, generic research areas with multiple higher level of development applications would be one candidate.

Once the degree of desired funds mixing has been determined within the context of the overall vertical integration structure, the measure chosen would be computed for each program and Technical Manager. The measure would be computed for the existing degree of funds mixing and for the desired degree of funds mixing (the funds mixing target). Aggregates of the measure for each Technical Manager, Division, Office, etc., and for the total organization would be computed and tracked. The actual measure levels would be tracked against the measure targets, and progress in achieving the targets monitored.

Because entropy does not define a pattern uniquely, supplemental measures would be of benefit. One such approach would be to track actual funds deviation from a desired funds mixing target. The starting point of this approach is to define the different level of development funds targets for each Technical Manager. Then, the square of the difference between the actual funds each Technical Manager has in each level of development at a point in time and the target funds for each level of development for the Manager would be computed and tracked. As time proceeds, this 'residual' should decrease. Aggregates of this 'residual' over Division, Office, total organization would be computed and tracked as proposed above for the entropy measure. This measure could be normalized in the form of a coefficient for easier interpretation, or could remain in the form of funds.

The entropy measure would also be useful for tracking programs over time as they pass through different levels of development. Well run programs would have hills and valleys in the entropy-time plot, with smooth temporal entropy gradients. A typical program would have low entropy when it is entirely in the basic research phase. Its entropy would rise to near unity as the program transitions from basic to applied research, and both types of funds are used to finance the program. The entropy would decrease again as the basic research funds are phased out and the applied research funds become dominant. The entropy would increase as applied research proceeds and development funds are phased in. These cycles would be repeated as the development process proceeds. In the tracking of the temporal entropy plot, if the entropy remains low during different development phases, this means that abrupt transitions to different phases are occurring. This condition is less desirable than the gradual transitions depicted above, and is readily observable from the entropy trajectory. Again, measures supplemental to entropy could be employed in the tracking process to enhance the interpretation of the output. A quantitative tracking approach as described becomes especially useful when management must track tens or hundreds of programs.

Citations by Papers in Different Journals

One of the measures of research program impact is the number of citations of papers produced by the program. The initial part of this Handbook provides references of some citation studies under the bibliometrics category of the quantitative methods section. While the number of citing papers is

very important, information about the citing papers can be extremely valuable. What is the distribution of citing papers among different technical disciplines; among different journals; among different institutions; among different countries? How can the impact of the program papers on the citing papers be quantified relative to the above and other characteristics of the citing papers? The following application of the entropy concept provides a starting point for the quantification, but it will be shown that additional measures are necessary for further insight into the impact.

Assume that a paper has received 1000 citations by journal papers. Assume also that the citing papers can be categorized by journal quality (level 1, level 2, level 3), where each journal quality category is denoted by i. Then the entropy of the distribution is the same as that given above:

.....i=3

.....s.=.-SUM...p(i)*ln((p(i))/kappa

.....i=1

where p(1) is the fraction of citing papers in journal of level 1 quality, p(2) is the fraction in level 2, p(3) is the fraction in level 3, and kappa is a constant which will produce an entropy s upper limit of unity.

The following table illustrates how the entropy function varies with different numbers of citing papers in the different journal types.

LEVEL.1....998..990..900..800..700..600..500..400..333

LEVEL.2.....1....5...50..100..150..200..250..300..333

LEVEL.3.....1....5...50..100..150..200..250..300..333

ENTROPY.....01...06...36...58...75...87...95...99..1.0

As all citing papers are concentrated into one journal type, the entropy measure goes to zero, and as the citing papers are divided equally among journal types, the measure goes to one. However, the table illustrates the limitations of using the entropy measure alone. If the paper had received 2000 citations distributed among the journal types in the same ratio, the entropy measure would have been the same. Clearly the total impact would not be reflected in the entropy measure as used here. This effect could be overcome by using the analogy with entropy in classical thermodynamic systems. The entropy measure above could be defined as an entropy per unit, and then multiplied by the total number of units in the system to get total entropy. However, the measure would now be substantially greater than unity in the full disorder limit, could be subject to more misinterpretation, and the measure would lose its utility.

To measure impact of the original paper on the citing papers, other measures will be employed in

addition to the entropy function. These other measures are the moments Mj of the citing paper distribution function n(i). The jth moment Mj of the distribution function n(i) is defined as:

.....i=m

Mj.=.SUM..(i^j)*n(i)

.....i=1

where n(i) is the number of citing papers in journal type i.

To show why using the moments of the distribution function is useful, and to aid in the interpretation of what follows, an analogue of the citing process to a nuclear interaction process is provided. For example, if a high energy proton interacts with a natural uranium target, neutrons will be released from the uranium by spallation, evaporation, and fast fission [Kostoff, 1979]. These released neutrons will have a wide range of velocities, which can be characterized by a velocity distribution function. The released neutrons can also interact with other targets and have additional neutron multiplication effects, depending on the energy of the incoming neutron and the composition of the target. With the use of kinetic theory (collisionless for large mean free path neutrons), moments of the released neutron velocity distribution function can be used to obtain macro-state information about the released neutron stream.

The citing process has some analogues to the neutron production process described above. The original published paper is analogous to the high energy proton. The technical community that reads the published paper is analogous to the natural uranium target. The citing papers produced by the technical community are analogous to the neutrons produced. The quality of the journals in which the citing papers are published is analogous to the velocities of the different neutrons.

The zeroth moment of the citing paper distribution function is:

.....i=m

M0.=.SUM..n(i)

.....i=1

In analogy to kinetic theory, where the zeroth moment of the particle velocity distribution is the mass density, the zeroth moment of the citing paper distribution shown above is the number of citing papers, or the citing paper mass.

The first moment of the distribution function is:

.....i=m

M1.=.SUM..i*n(i)

.....i=1

In analogy to kinetic theory, where the first moment of the particle velocity distribution is the momentum (mass*velocity) of the particle stream, the first moment of the citing paper distribution is the citing paper impact.

The second moment of the distribution function is:

.....i=m

M2.=.SUM..(i^2)*n(i)

.....i=1

In analogy to kinetic theory, where the second moment of the particle velocity distribution is the energy (mass*velocity^2) of the particle stream, the second moment of the citing paper distribution is the citing paper energy.

The third moment of the distribution function is:

.....i=m

M3.=.SUM..(i^3)*n(i)

.....i=1

In analogy to kinetic theory, where the third moment of the particle velocity distribution is the flux of particle energy (mass*velocity^3), the third moment of the citing paper distribution is the citing paper energy flux.

Thus, sole use of the zeroth moment of the citing paper journal type distribution provides a very gross measure of the impact (the number of citing papers) but offers little information about the quality of the impact. In this particular example, information about the types of user audience is at least as important as numbers of users. Is the author of the original paper reaching the intended audience? Use of the entropy of the citing paper journal type distribution shows the diversity of the user audience.

Use of the first moment allows the importance assigned to the different journal types to be factored in the analysis. To compute the first moment, journal type i has to be assigned a numerical value which reflects its importance. In analogy to kinetic theory, this numerical value is the effective "velocity" of journal type i. With use of this effective velocity, computation of the first moment

yields the momentum, or total citing paper impact. In analogy to kinetic theory, the ratio of the first moment to the zeroth moment is the citing paper "average velocity", or average impact/citing paper.

Use of the second moment accentuates the difference in importance of the various journals. For distributions which have similar values of total impact, use of the "energy" will identify which of those distributions rely on "velocity" more than "mass" for their impact. For distributions which have similar values of total impact and energy, and where more differentiation is required, third or higher moments can be employed. The following example illustrates this point. In this example, two citing paper journal distributions, A and B, were compared for a domain of six journals of different quality. The distributions were selected such that the entropy and zeroth, first, and second moments were equal. The computational results follow.

.{...1.....2.....3.....4.....5.....6...}--NUMBER.OF.JOURNAL

..n(3)..n(4)..n(5)..n(6)..n(7)..n(8)..s.....M0....M1.....M2.....M3

A.200...100...200...100...300...100...95..1000..5500..33100..212500

B..92...269...218...112....86...223...95..1000..5500..33100..214815

The first row represents the six journals. The first six columns of the second row represent the citing paper distribution function for the six journals. The number in parentheses is the value of quality (effective velocity) assigned to each of the six journals. Thus, the entry in the first column of the second row, n(3), is interpreted as the number of citing papers in journal 1, where journal 1 has a quality value of 3. Continuing on the second row, s is the entropy of the citing paper journal distribution, M0 is the zeroth moment of this distribution, M1 is the first moment, M2 is the second moment, and M3 is the third moment. Rows three and four are the values of these columns for cases A and B.

All of the figures of merit are the same for the two cases except the third moment M3. While two cases with so many equal figures of merit would be an extremely rare occurrence, the example does show the discriminatory capability of the moment approach. In this case, use of even higher moments would provide more separation between the numerical results, and allow more insight for the interpretation of the results.

To track the figures of merit through time, and extract useful information, analogies can be made with aerodynamics trajectory analysis. An aerodynamic vehicle's state can be tracked through space and time to generate its trajectory (position in space and time). The first time derivative of its trajectory is its velocity, the second derivative is the acceleration, and the third derivative is the agility (ability to move inertial forces rapidly). Thus, the entropy and the moments in the above example could be plotted as a function of time, and their derivatives obtained. Valuable information could be obtained from the derivatives to see how the impact of an organization's output is changing over time, and how rapidly shifts are occurring, especially in response to new management initiatives.

In summary, the distribution patterns which occur in research assessments contain much useful information. Present techniques extract relatively little of this information in practice. Use of concepts from thermodynamics and other fields such as entropy, momentum, and energy can improve the information extraction process, and aid in the interpretation of the results through physical analogies.

APPENDIX 12

INFRASTRUCTURE OF S&T METRICS LITERATURE

This final section is addressed to readers who may want information about the S&T metrics literature beyond what the bibliography can provide. This section contains the most prolific authors of S&T metrics papers, journals containing the most S&T metrics papers, the institutions publishing the most S&T metrics papers, the most cited first authors of S&T metrics documents, the most cited journals containing S&T metrics papers, and the most cited S&T metrics documents.

To generate this information, a query was constructed iteratively, and used to retrieve documents from the Science Citation Index for the period 1990-2005. The query used, in addition to all articles in the journal Scientometrics, was:

citation analysis OR bibliometric* OR scientometric* OR research productivity OR scientific productivity OR citation impact OR publication productivity OR citation pattern* OR citation rate* OR citation count* OR (impact factor* AND (journal* OR publish*)) OR citation impact* OR citation data OR scholarly productivity OR total citations OR immediacy index OR citation frequency OR co-authorship links OR science indicator* OR citation frequencies OR database tomography OR scholarly activity OR bibliographic citations OR bibliographic coupling OR citation measures OR citation distribution* OR citation network* OR citation-based indicator* OR high-impact journal* OR self-citation rate* OR self-cited rate* OR citation indicator* OR Lotka's Law OR Bradford's Law OR Bradford Distribution OR number of citations OR citations per paper OR citations per article OR science metric* OR (metric* AND (peer review* OR cost benefit OR rate of return OR citation* OR patent* OR impact factor*)) OR (production function AND productivity AND (research OR science OR technology)) OR co-word OR co-citation OR co-classification OR co-nomination OR (citations AND (science OR indicator* OR indicator* OR productivity)) OR frequency of citation* OR numbers of articles OR numbers of publications OR numbers of papers

Use of this query resulted in retrieval of 4780 records covering the fifteen year period. The author's TextDicer software was used to provide the following bibliometric results.

12-A Most Prolific Authors

AUTHOR	#PAPERS
GLANZEL--W	60
SCHUBERT--A	51
ROUSSEAU--R	50
GARFIELD--E	47
VAN RAAN--AFJ	46
BRAUN--T	42
EGGHE--L	42
MOED--HF	39

THELWALL--M	39
KOSTOFF--RN	33
LEYDESDORFF--L	33
LEWISON--G	28
GUPTA--BM	25
CRONIN--B	24
VINKLER--P	24
BONITZ--M	23
GARG--KC	23
KRETSCHMER--H	20
PERSSON--O	19
BORDONS--M	18
GOMEZ--I	18
TIJSSEN--RJW	18
INGWERSEN--P	17
SMALL--H	17
VAN LEEUWEN--TN	17
WILSON--CS	17
ARUNACHALAM--S	16
BURRELL--QL	16
MCCAIN--KW	16
COURTIAL--JP	14
HARTER--SP	14
LUWEL--M	14
NEDERHOF--AJ	14
WORMELL--I	14
ZITT--M	14
MEYER--M	13
NARIN--F	13
OPPENHEIM--C	13
WHITE--HD	13
BRAHLER--E	12
FERNANDEZ--MT	12
LANG--SB	12

12-B Journals Containing Most Papers

Journal	#ofPapers
SCIENTOMETRICS	1401
JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY	185
JOURNAL OF INFORMATION SCIENCE	69
RESEARCH POLICY	66
JOURNAL OF DOCUMENTATION	60
INFORMATION PROCESSING & MANAGEMENT	54
SCIENTIST	39

BULLETIN OF THE MEDICAL LIBRARY ASSOCIATION	36
MEDICINA CLINICA	35
ACADEMIC MEDICINE	32
COLLEGE & RESEARCH LIBRARIES	31
RESEARCH EVALUATION	30
LIBRARY & INFORMATION SCIENCE RESEARCH	29
JOURNAL OF SOCIAL WORK EDUCATION	29
CURRENT CONTENTS	24
CURRENT SCIENCE	22
BRITISH MEDICAL JOURNAL	21
RESEARCH IN HIGHER EDUCATION	20
JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION	19
LIBRARY QUARTERLY	19
NATURE	19
PROCEEDINGS OF THE ASIS ANNUAL MEETING	18
LIBRI	17
LIBRARY TRENDS	17
ASLIB PROCEEDINGS	16
INTERNATIONAL FORUM ON INFORMATION AND DOCUMENTATION	16
JOURNAL OF THE MEDICAL LIBRARY ASSOCIATION	15
HIGHER EDUCATION	14
SCIENCE	14
WEB OF KNOWLEDGE - A FESTSCHRIFT IN HONOR OF EUGENE GARFIELD	14
ASIST MONOGRAPH SERIES	14
OMEGA-INTERNATIONAL JOURNAL OF MANAGEMENT SCIENCE	14
SOCIAL STUDIES OF SCIENCE	13
SERIALS LIBRARIAN	13
COUNSELING PSYCHOLOGIST	13
PUBLICATIONS OF THE ASTRONOMICAL SOCIETY OF THE PACIFIC	13
CANADIAN JOURNAL OF INFORMATION AND LIBRARY SCIENCE-REVUE CANADIENNE DES SCIENCES DE L INFORMATION E	13
MANAGEMENT SCIENCE	12
FERROELECTRICS	12
LIBRARY RESOURCES & TECHNICAL SERVICES	11
CROATIAN MEDICAL JOURNAL	11
INTERCIENCIA	11
ANNALS OF EMERGENCY MEDICINE	10
JOURNAL OF ANALYTICAL CHEMISTRY	10
PSYCHOLOGICAL REPORTS	10
TECHNOLOGICAL FORECASTING AND SOCIAL CHANGE	10
ACCIDENT ANALYSIS AND PREVENTION	10
STRATEGIC MANAGEMENT JOURNAL	10
JOURNAL OF CRIMINAL JUSTICE	10
PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	10

12-C Institutions Producing Most Papers
(Frequencies Approximate)

INSTITUTION	#PAPERS
LEIDEN UNIV	70
NATL INST SCI TECHNOL & DEV STUDIES	57
INDIANA UNIV	46
WOLVERHAMPTON UNIV	41
HARVARD UNIV	39
CSIC	38
UNIV ILLINOIS	36
HUNGARIAN ACAD SCI	36
DREXEL UNIV	35
UNIV CALIF LOS ANGELES	33
UNIV N CAROLINA	32
UNIV INSTELLING ANTWERP	31
ROYAL SCH LIB & INFORMAT SCI	30
LIMBURGS UNIV CTR	29
UNIV TORONTO	29
UNIV MARYLAND	28
KATHOLIEKE UNIV LEUVEN	27
UNIV GRANADA	27
UNIV MICHIGAN	26
KHBO	25
UNIV MISSOURI	24
OFF NAVAL RES	24
UNIV NEW S WALES	23
UNIV VALENCIA	22
UNIV MINNESOTA	22
PENN STATE UNIV	22
INST SCI INFORMAT	21
UNIV TEXAS	21
UNIV WESTERN ONTARIO	20
UNIV SUSSEX	19
UNIV PITTSBURGH	19
CITY UNIV LONDON	19
UNIV PENN	18
BOSTON UNIV	16
UNIV NEBRASKA	16
RUSSIAN ACAD SCI	16
GEORGIA INST TECHNOL	16
UNIV AMSTERDAM	16
HEBREW UNIV JERUSALEM	15
CORNELL UNIV	15

JOHNS HOPKINS UNIV	15
CHI RES INC	15
UNIV CALIF BERKELEY	15
UNIV FLORIDA	15
UMEA UNIV	14
MCMASTER UNIV	14
UNIV UTAH	13
LIB HUNGARIAN ACAD SCI	13
UNIV GENOA	13
BEN GURION UNIV NEGEV	13
UNIV CALIF SAN FRANCISCO	13
FREE UNIV BERLIN	13
UNIV OKLAHOMA	13
AUSTRALIAN NATL UNIV	13
NANYANG TECHNOL UNIV	12
MICHIGAN STATE UNIV	12
UNIV ZAGREB	12
OBSERV SCI & TECH	12
UNIV ALABAMA	12
KARNATAK UNIV	12
UIA	12
NORTHWESTERN UNIV	12
UNIV WASHINGTON	12
UNIV ALBERTA	11
CNRS	11
UNIV SHEFFIELD	11
RENSSELAER POLYTECH INST	11
OHIO STATE UNIV	11
UNIV COLORADO	11
RUTGERS STATE UNIV	11
UNIV CALIF IRVINE	11
UNIV CHICAGO	11
MUSEUM NATL HIST NAT	11
INRA	11
GEORGIA STATE UNIV	10
UNIV NACL AUTONOMA MEXICO	10
LONG ISL UNIV	10
UNIV OXFORD	10
UNIV GEORGIA	10
IOWA STATE UNIV	10
UNIV LEIPZIG	10
UNIV ANTWERP	10
UNIV ARIZONA	10
UNIV EXTREMADURA	10
RES ASSOC SCI COMMUN & INFORMAT EV	10
CASE WESTERN RESERVE UNIV	10

UNIV CINCINNATI	10

12-D Most Cited First Authors

FIRST AUTHOR	#CITES
GARFIELD E	2153
NARIN F	751
PRICE DJD	604
SMALL H	604
EGGHE L	577
MOED HF	509
BRAUN T	500
CRONIN B	474
SCHUBERT A	462
LEYDESDORFF L	441
WHITE HD	424
THELWALL M	414
SEGLEN PO	397
KOSTOFF RN	381
GLANZEL W	363
MERTON RK	358
ROUSSEAU R	354
MCCAIN KW	324
VANRAAN AFJ	321
GRILICHES Z	291
MACROBERTS MH	266
CALLON M	265
VINKLER P	259
*I SCI INF	248
COLE JR	236
COLE S	234
NEDERHOF AJ	221
BROOKES BC	207
ZUCKERMAN H	205
ARUNACHALAM S	194
SIMONTON DK	193
MORAVCSIK MJ	190
MARTIN BR	184
LONG JS	182
LEWISON G	176
LUUKKONEN T	168
CRANE D	167
BURRELL QL	164
LOTKA AJ	160
HARTER SP	158

INGWERSEN P	155
ALLISON PD	153
NALIMOV VV	151
TIJSSEN RJW	151
GRUPP H	150
LINDSEY D	144
FRAME JD	144
BRADFORD SC	143
BORGMAN CL	143
MANSFIELD E	142
LINE MB	140
PERITZ BC	138
FOX MF	137
JAFFE AB	137
CARPENTER MP	135
PAVITT K	127
COZZENS SE	126
PAO ML	125
SALTON G	124
ABT HA	124
*OECD	124
LATOUR B	123
KATZ JS	123
SMALL HG	122
DIAMOND AM	122
PINERO JML	120
KUHN TS	120
PETERS HPF	118
HICKS D	117
HARGENS LL	116
GRIFFITH BC	112
BONITZ M	111
LAWRENCE S	110
CHUBIN DE	110
NELSON RR	108
SWANSON DR	107
BROOKS TA	107
BRAAM RR	107
IRVINE J	106
COURTIAL JP	104
NOYONS ECM	103
RICE RE	103
HAMILTON DP	103
BARILAN J	101
BEAVER DD	101
OPPENHEIM C	100

CULNAN MJ	100

12-E Most Cited Journals

JOURNAL	#CITES
SCIENTOMETRICS	7317
J AM SOC INFORM SCI	3740
SCIENCE	1896
J DOC	1497
RES POLICY	1474
NATURE	1233
SOC STUD SCI	1017
J INFORM SCI	1012
STRAHLENTHER ONKOL	884
JAMA-J AM MED ASSOC	768
BRIT MED J	748
AM SOCIOL REV	697
AM PSYCHOL	638
AM ECON REV	619
LANCET	600
INFORM PROCESS MANAG	599
NEW ENGL J MED	581
MED CLIN-BARCELONA	474
COLL RES LIBR	460
ASTROPHYS J 1	448
ASTRON ASTROPHYS	388
B MED LIBR ASSOC	329
MANAGE SCI	304
J AM SOC INF SCI TEC	303
LIBR TRENDS	300
ANN INTERN MED	290
RES HIGH EDUC	287
AM J SOCIOL	285
CURR CONTENTS	283
J ECON LIT	279
REV ECON STAT	278
LIBR INFORM SCI RES	269
ANNU REV INFORM SCI	261
HDB QUANTITATIVE STU	258
LITTLE SCI BIG SCI	255
PSICOTHEMA	252
MON NOT R ASTRON SOC	232
DIAGNOSTICA	230
REV INT PSICOLOGIA C	227
ASTROPHYS J 2	224

PSYCHOTHER PSYCH MED	223
STRATEGIC MANAGE J	221
COUNS PSYCHOL	221
AM DOC	220
J HIGH EDUC	220
CITATION INDEXING	214
ACAD MED	214
INFORMATION PROCESSI	214
RES EVALUAT	212
SCI PUBL POLICY	209
ACAD MANAGE J	204
J POLITICAL EC	203
SCHOLARLY COMMUNICAT	202
SOCIOL EDUC	202
ECONOMETRICA	201
Q J ECON	199
P NATL ACAD SCI USA	189
J PERS SOC PSYCHOL	189
ESSAYS INFORMATION S	187
COMMUN ACM	187
ADMIN SCI QUART	185
LIBR QUART	183
COLLECTION MANAGEMEN	180
TOXICON	178
SCI STUD	171
J SOC WORK EDUC	169
ATMOS ENVIRON	166
SALUD MENT	164
SCI TECHNOL	161
J WASHINGTON ACADEMY	151
J POLIT ECON	151
ANN THORAC SURG	150

12-F Most Cited Documents

PAPER	#CITES
GARFIELD E, 1979, CITATION INDEXING	200
GARFIELD E, 1972, SCIENCE, V178, P471	185
LOTKA AJ, 1926, J WASHINGTON ACADEMY, V16, P317	146
SEGLEN PO, 1997, BRIT MED J, V314, P498	126
PRICE DJD, 1965, SCIENCE, V149, P510	124
SMALL H, 1973, J AM SOC INFORM SCI, V24, P265	121
PRICE DJD, 1963, LITTLE SCI BIG SCI	108
MACROBERTS MH, 1989, J AM SOC INFORM SCI, V40, P342	104
SCHUBERT A, 1989, SCIENTOMETRICS, V16, P3	103

CRONIN B, 1984, CITATION PROCESS	95
COLE JR, 1973, SOCIAL STRATIFICATIO	87
GARFIELD E, 1996, BRIT MED J, V313, P411	86
MERTON RK, 1968, SCIENCE, V159, P56	81
NARIN F, 1976, EVALUATIVE BIBLIOMET	78
PRICE DJD, 1976, J AM SOC INFORM SCI, V27, P292	77
SMALL H, 1974, SCI STUD, V4, P17	77
WHITE HD, 1989, ANNU REV INFORM SCI, V24, P119	74
BRADFORD SC, 1934, ENGINEERING-LONDON, V137, P85	74
MARTIN BR, 1983, RES POLICY, V12, P61	71
GARFIELD E, 1955, SCIENCE, V122, P108	70
WHITE HD, 1981, J AM SOC INFORM SCI, V32, P163	69
CALLON M, 1986, MAPPING DYNAMICS SCI	69
SMITH LC, 1981, LIBR TRENDS, V30, P83	64
EGGHE L, 1990, INTRO INFORMETRICS Q	63
INGWERSEN P, 1998, J DOC, V54, P236	62
MOED HF, 1985, RES POLICY, V14, P131	61
MAY RM, 1997, SCIENCE, V275, P793	60
SEGLEN PO, 1992, J AM SOC INFORM SCI, V43, P628	60
KING J, 1987, J INFORM SCI, V13, P261	60
KUHN TS, 1970, STRUCTURE SCI REVOLU	59
SCHUBERT A, 1986, SCIENTOMETRICS, V9, P281	58
MORAVCSIK MJ, 1975, SOC STUD SCI, V5, P86	57
NARIN F, 1997, RES POLICY, V26, P317	56
MCCAIN KW, 1990, J AM SOC INFORM SCI, V41, P433	55
OPTHOF T, 1997, CARDIOVASC RES, V33, P1	55
GRILICHES Z, 1990, J ECON LIT, V28, P1661	55
WHITE HD, 1998, J AM SOC INFORM SCI, V49, P327	54
SMALL HG, 1978, SOC STUD SCI, V8, P327	54
PRICE DJD, 1970, COMMUNICATION SCI EN, P3	51
GIBBONS M, 1994, NEW PRODUCTION KNOWL	51
KESSLER MM, 1963, AM DOC, V14, P10	49
FOX MF, 1983, SOC STUD SCI, V13, P285	48
GILBERT GN, 1977, SOC STUD SCI, V7, P113	48
MOED HF, 1995, SCIENTOMETRICS, V33, P381	47
SCHUBERT A, 1990, SCIENTOMETRICS, V19, P3	47
LIEBOWITZ SJ, 1984, J ECON LIT, V22, P77	47
HAMILTON DP, 1990, SCIENCE, V250, P1331	47
SMALL H, 1985, J INFORM SCI, V11, P147	46
ALMIND TC, 1997, J DOC, V53, P404	46
HAMILTON DP, 1991, SCIENCE, V251, P25	45
JAFFE AB, 1993, Q J ECON, V108, P577	45
MOED HF, 1995, J AM SOC INFORM SCI, V46, P461	44
LUUKKONEN T, 1992, SCI TECHNOL, V17, P101	44
MOED HF, 1996, NATURE, V381, P186	43
CRANE D, 1972, INVISIBLE COLLEGES	43

SMALL H, 1985, SCIENTOMETRICS, V8, P321	43
ALLISON PD, 1974, AM SOCIOL REV, V39, P596	43
PRICE DJD, 1966, AM PSYCHOL, V21, P1011	42
BRAUN T, 1985, SCIENTOMETRIC INDICA	42
EGGHE L, 1990, INTRO INFORMETRICS	42
TODOROV R, 1988, J INFORM SCI, V14, P47	42
VANRAAN AFJ, 1996, SCIENTOMETRICS, V36, P397	42
CALLON M, 1983, SOC SCI INFORM, V22, P191	41
COZZENS SE, 1989, SCIENTOMETRICS, V15, P437	41
CHUBIN DE, 1975, SOC STUD SCI, V5, P423	41
EDGE D, 1979, HIST SCI, V17, P102	41
THELWALL M, 2001, J AM SOC INF SCI TEC, V52, P1157	40
SMITH AG, 1999, J DOC, V55, P577	40
HANSSON S, 1995, LANCET, V346, P906	40
GROSS PLK, 1927, SCIENCE, V66, P385	40

12-G. Seminal Documents

This group includes most cited documents, but adds documents that may not have received large numbers of citations in absolute terms (due to very old or very young documents), but had substantially more citations than their temporal contemporaries. The documents are arranged in chronological order.

AUTHOR	YEAR	DOCUMENT	VOL	PAGE	#CITES
COLE FJ	1917	SCI PROGR	V11	P578	22
LOTKA AJ	1926	J WASHINGTON ACADEMY	V16	P317	146
GROSS PLK	1927	SCIENCE	V66	P385	40
BRADFORD SC	1934	ENGINEERING-LONDON	V137	P85	74
MERTON RK	1942	J LEGAL POLITICAL SO	V1	P115	11
BUSH V	1945	ATLANTIC MONTHLY	V176	P101	9
BRADFORD SC	1948	DOCUMENTATION			28
ZIPF GK	1949	HUMAN BEHAV PRINCIPL			40
LEHMAN HC	1953	AGE ACHIEVEMENT			18
GARFIELD E	1955	SCIENCE	V122	P108	70
SIMON HA	1955	BIOMETRIKA	V42	P425	32
BURTON RE	1960	AM DOC	V11	P18	29
PRICE DJD	1961	SCI BABYLON			22
KUHN TS	1962	STRUCTURE SCI REVOLU			38
PRICE DJD	1963	LITTLE SCI BIG SCI			127
KESSLER MM	1963	AM DOC	V14	P10	49
GARFIELD E	1964	USE CITATION DATA WR			22

PRICE DJD	1965	SCIENCE	V149	P510	124
PRICE DJD	1966	AM PSYCHOL	V21	P1011	42
COLE S	1967	AM SOCIOL REV	V32	P377	32
MERTON RK	1968	SCIENCE	V159	P56	81
KUHN TS	1970	STRUCTURE SCI REVOLU			59
PRICE DJD	1970	COMMUNICATION SCI EN	P3		51
GARFIELD E	1972	SCIENCE	V178	P471	185
CRANE D	1972	INVISIBLE COLLEGES			43
SMALL H	1973	J AM SOC INFORM SCI	V24	P265	121
COLE JR	1973	SOCIAL STRATIFICATIO			87
MERTON RK	1973	SOCIOLOGY SCI			54
SMALL H	1974	SCI STUD	V4	P17	77
ALLISON PD	1974	AM SOCIOL REV	V39	P596	43
MORAVCSIK MJ	1975	SOC STUD SCI	V5	P86	57
CHUBIN DE	1975	SOC STUD SCI	V5	P423	41
NARIN F	1976	EVALUATIVE BIBLIOMET			78
PRICE DJD	1976	J AM SOC INFORM SCI	V27	P292	77
GILBERT GN	1977	SOC STUD SCI	V7	P113	48
SMALL HG	1978	SOC STUD SCI	V8	P327	54
GARFIELD E	1979	CITATION INDEXING			200
EDGE D	1979	HIST SCI	V17	P102	41
WHITE HD	1981	J AM SOC INFORM SCI	V32	P163	69
SMITH LC	1981	LIBR TRENDS	V30	P83	64
MARTIN BR	1983	RES POLICY	V12	P61	71
FOX MF	1983	SOC STUD SCI	V13	P285	48
CALLON M	1983	SOC SCI INFORM	V22	P191	41
CRONIN B	1984	CITATION PROCESS ROL			95
LIEBOWITZ SJ	1984	J ECON LIT	V22	P77	47
MOED HF	1985	RES POLICY	V14	P131	61
SMALL H	1985	J INFORM SCI	V11	P147	46
SMALL H	1985	SCIENTOMETRICS	V8	P321	43
BRAUN T	1985	SCIENTOMETRIC INDICA			42
CALLON M	1986	MAPPING DYNAMICS SCI			69
SCHUBERT A	1986	SCIENTOMETRICS	V9	P281	58
KING J	1987	J INFORM SCI	V13	P261	60
TODOROV R	1988	J INFORM SCI	V14	P47	42
MACROBERTS	1989	J AM SOC INFORM SCI	V40	P342	104

MH					
SCHUBERT A	1989	SCIENTOMETRICS	V16	P3	103
WHITE HD	1989	ANNU REV INFORM SCI	V24	P119	74
COZZENS SE	1989	SCIENTOMETRICS	V15	P437	41
EGGHE L	1990	INTRO INFORMETRICS Q			63
MCCAIN KW	1990	J AM SOC INFORM SCI	V41	P433	55
GRILICHES Z	1990	J ECON LIT	V28	P1661	55
SCHUBERT A	1990	SCIENTOMETRICS	V19	P3	47
HAMILTON DP	1990	SCIENCE	V250	P1331	47
EGGHE L	1990	INTRO INFORMETRICS			42
HAMILTON DP	1991	SCIENCE	V251	P25	45
SEGLEN PO	1992	J AM SOC INFORM SCI	V43	P628	60
LUUKKONEN T	1992	SCI TECHNOL	V17	P101	44
JAFFE AB	1993	Q J ECON	V108	P577	45
GIBBONS M	1994	NEW PRODUCTION KNOWL			51
MOED HF	1995	SCIENTOMETRICS	V33	P381	47
MOED HF	1995	J AM SOC INFORM SCI	V46	P461	44
HANSSON S	1995	LANCET	V346	P906	40
GARFIELD E	1996	BRIT MED J	V313	P411	86
MOED HF	1996	NATURE	V381	P186	43
VANRAAN AFJ	1996	SCIENTOMETRICS	V36	P397	42
SEGLEN PO	1997	BRIT MED J	V314	P498	126
MAY RM	1997	SCIENCE	V275	P793	60
NARIN F	1997	RES POLICY	V26	P317	56
OPTHOF T	1997	CARDIOVASC RES	V33	P1	55
ALMIND TC	1997	J DOC	V53	P404	46
INGWERSEN P	1998	J DOC	V54	P236	62
WHITE HD	1998	J AM SOC INFORM SCI	V49	P327	54
SMITH AG	1999	J DOC	V55	P577	40
THELWALL M	2001	J AM SOC INF SCI TEC	V52	P1157	40
WILKINSON D	2003	J INFORM SCI	V29	P59	21
VAUGHAN L	2003	J AM SOC INF SCI TEC	V54	P29	19