Animal telemetry network data assembly center: Phase 2

Barbara Block & Randy Kochevar
Hopkins Marine Station
Stanford University
120 Oceanview Blvd.
Pacific Grove, Ca
phone: (831) 655-6236 fax: (831) 375-0793 email: bblock@stanford.edu

Award Number: N000141512500

LONG TERM GOALS

In Phase I of the ATN DAC program, programmers and researchers from Stanford University and the NOAA SWFSC Environmental Research Division worked together with IOOS and ONR program managers, to develop a beta version of the IOOS ATN DAC (**http://oceanview.pfeg.noaa.gov/ATN/**). The objective of the Stanford team in Phase was to leverage their prior development for tag data management (e.g. TOPP, GTOPP, GulfTOPP) of animal telemetry data management into a single system (DAC) with an intuitive front end, capable of delivering and visualizing US telemetry data streamed from multiple animal and platform types. The specific aim was to create for the ATN web portal capable of demonstrating the opportunity that exists for utilizing ATN datasets.

For this Phase I goal the Stanford team utilized existing hardware, database software, algorithms for tag quality control, automated Argos downloads, and data visualizations (tracks) in a Google Maps-based user interface, to engage users with U.S. animal telemetry data. The front end was created with simple, color-coded icons, representing five distinct tag platform types: real time satellite tags with oceanographic data, satellite tags with position only, pop-up satellite archival tags, archival tags, acoustic tags, and autonomous buoys with receivers for detection. For each platform type, the user can display additional data types (e.g., animal or glider track, acoustic detections) and metadata (platform type, date and duration of deployment) by simply clicking on the icon. An icon click also presents the user with a variety of additional options, which vary by tag platform. These include: display or download of ocean profile data, display or download location data for track visualization, display or download acoustic detection data, view data in ERDDAP server, etc. Additional education and outreach material is also provided in pop-up windows and displays.

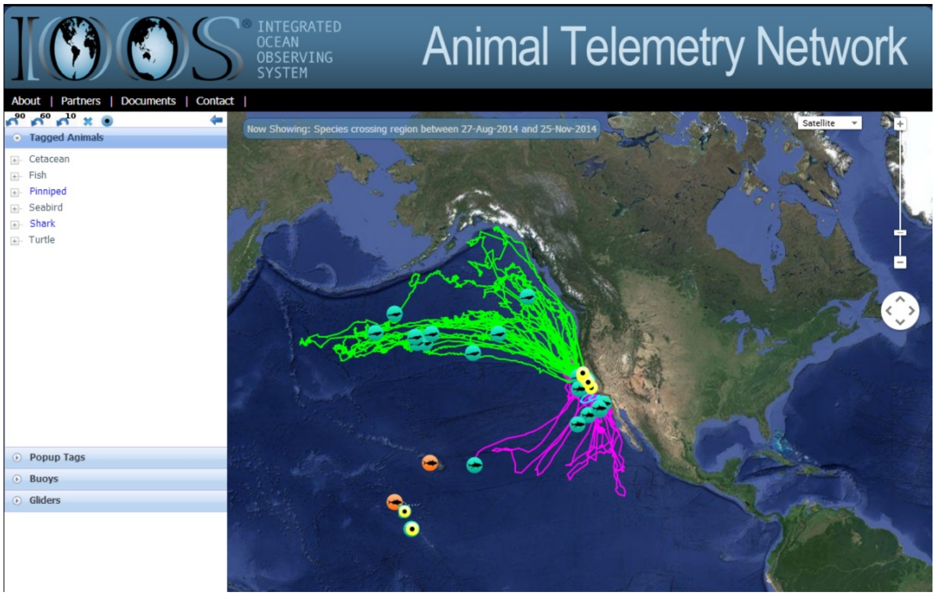

Figure 1. The home screen of the ATN DAC beta with tracks displayed.

The ATN DAC currently through Phase I development provides access to four data streams: 1) A "live" from the animal borne platform data stream that reports automatically from Argos satellites via codes that directly download from the CLS to Stanford servers and then display location and data sets to the DAC in near real time. This data stream is customized for tag platforms built by SMRU (Satellite Relay Data Logger, Fastloc GPS), Wildlife Computers (SPOT and SPLASH) 2) Acoustic data (white sharks) via automated Iridium satellite-linked Vemco receivers mounted on stationary buoys or mobile platforms such as Liquid Robotics Wave Gliders; and 3) Pop-up satellite tags (e.g. Wildlife Computers MK10, MiniPAT) that report throughout the year (e.g. blue marlin from the IGFA Great Marlin Race), and then take approximately 20 days to download as the tag floats at the surface and transmits data to the DAC. The DAC servers collect oceanographic, position and behavioral data archived on the tag and rapidly displays it, and 4) the archival based datasets drawn from the several thousand animal tracking deployments and datasets collected through the prior TOPP, GTOPP programs (Census of Marine Life), as well as Atlantic bluefin tuna tagged through the Tag A Giant program with implantable archival tags and pop up archival tags (already deployed and recovered).

Currently, 48 different species are represented in the ATN DAC (with datasets), with deployment dates ranging from 2000-2014. In addition, the interface provides access to acoustic buoy detection data from 20 receivers located in 13 different sites -in the US and Canada. The beta DAC system also demonstrates the capacity to pull data from remote locations across the globe, with streams coming from US territories (Palmyra-acoustic tagged mantas) and the Chagos Archipelago (reef sharks, silvertips, grey reef sharks). On these live acoustic receiver buoys at the DAC, there are currently 7 species reporting daily which complements the 4-5 species reporting through the satellite ARGOS tags. There is also a placeholder in the data structure for archival glider data. Three glider expeditions are currently planned early in the 2015 and code for reporting of this data stream has already been designed and tested. This placeholder will be used to display the track of the glider and animal acoustic detections.

The default view of the IOOS DAC interface shows the most recent data (users can select from 10, 60 or 90 day displays), and the interface features a pull-out "Data Menu" which allows them to view or hide datasets from all these species and platforms by clicking check-boxes arranged in a hierarchical, nested structure - similar to that used in Google Earth to activate and deactivate various data layers. For example, a user can view all the blue whale tag datasets by clicking a single box; or they can drill down to individual deployment years (e.g., to display all the tags deployed in 2004), or even down to individual tags (tag number "2204001-2306-MatePTT"). At the individual tag level, users can also view, download or access the data through an ERDDAP server, all directly from the Data Menu (i.e., without having to locate that specific tag on the map first.) The ERDDAP server enables users to quickly query, visualize and download data in several different formats. More importantly, once the user has created a query, ERDDAP provides a restive URL for that query which can be incorporated directly into other systems (e.g., websites, Matlab routines, models) which require ongoing access to those data streams.

The ATN DAC development team under 2014 funds continued to add and refine additional datasets available through the ATN DAC interface, which includes adding species not currently represented, continuing to upload data sets from the TOPP and TAG programs which are in highest demand (whales, tunas), and ingesting additional data associated with 2014 deployments and display tracks already in the interface (e.g., depth and temperature profiles). Other ongoing tasks include: production and integration of profile plots for real-time data from oceanographic tags into website interface; production and integration of buoy detection plots into acoustic buoy displays; and production and integration of buoy metadata into the ATN DAC interface. These tasks were completed by the end of 2014. At the completion of Phase I, the demonstration of being capable of serving live and archived data from numerous tag platforms will have been successfully achieved.

OBJECTIVES

There are several key areas of focus the next phase of ATN DAC development. These include: 1) Improvements in infrastructure design and hardware, 2) Integration of external data sets and development of new web-tools which will enable collaborators to submit new data and data streams into the ATN DAC, 3) Development and completion of compliance with IOOS standards, and 4) Ingestion of TOPP/TAG data sets in high demand (uploading all data from key data sets modelers and assessment scientists are requesting- Atlantic and Pacific bluefin tuna). Each of these will be discussed in greater detail below:

APPROACH

In its present form and for the short term goal, the ATN DAC system is built as a quasi-centralized system, with data being ingested into servers at Stanford University and SWFSC. At Stanford the data is processed from providers (CLS, Iridium, users returning archival tags), backed up at several locations, and served to NOAA SWFS and then displayed as the ATN DAC user interface. This telemetry data system was initiated in 2000 at the beginning of the TOPP program with Census of Marine Life investments (Sloan Foundation). It has been maintained with private (Stanford, Packard, Moore) and federal funding (NOAA, GulfTOPP). However, the capacity of the current infrastructure is limited and the servers are aging. In order to make the system to meet the short-term ATN community needs, we need to immediately establish a replicated server system that ensures 100% up time, even in the event of complete failure of a primary server.

In order to meet the short-term ATN community needs, hardware needs to be upgraded (purchased, installed) and made compatible with the existing system. In order to accomplish this, we will need at a minimum $20,476, which will provide sufficient funding to establish three new paired server/backup systems, each with 30 TB of storage space. These machines will be co-located with two systems at Hopkins Marine Station (the primary server/backup, plus a developmental server/backup), and a third system at the Stanford University main campus in Palo Alto, CA. The time required to upgrade, install and configure the new system is estimated to be around six months following receipt of the required hardware. A fourth server, which receives daily updates from Hopkins and which serves the ATN DAC website, is located at SWFSC in Santa Cruz, CA. They are in the process of configuring an additional server to mirror this server, so there will be an on-site backup there as well for website services. This latter pair will be operated and maintained by SWFSC ERD staff.

A significant portion of our data is obtained from direct downloads from Argos CLS, Vemco, and/or Iridium. These download and automated processing steps are achieved routinely with codes that run on the Stanford servers every night, delivering updated datasets via ftp to the visualization engine. Less routine is the post- processing of the data stream, the requirements for each tag platform sets of unique algorithms and corrections. In Phase II, we want to continue to develop these processes to be more fully automated (to the extent possible), so that they can be carried out more efficiently and with greater regularity. We also want to make these transparent (codes will be published on the site where possible) Our focus will be on optimizing routines for data processing, reducing human steps in the process but maintaining the highest possible QC standards.

Integration of External Datasets

Now that the ATN DAC system is being developed and tested using a variety of different tag platforms, we are in a position to begin working with external data providers to integrate their unique data streams from a wider diversity of animals and regions into the system. The primary purpose of the ATN DAC once operational is to 1) integrate data sets from any U.S. Principal Investigator conducting animal telemetry in the ocean, great lakes or river systems in the U.S., and 2. Provide that animal telemetry data publicly to all interested customers using appropriate standards. This process should be straightforward as the TOPP team received data from diverse investigators during the Census of Marine Life program. We envision establishing the same type of routines we ran during TOPP, based on automated data ingest protocols called “predeployment” that enabled external data to be linked to deployment data streams at Argos CLS. For example, we might work with the Bureau of Ocean Energy Management (BOEM) Marine Arctic Ecosystem Study (MARES) project, which in partnership with ONR is deploying satellite tags on marine mammals in the coming year for the Arctic program. We may also work with Sean Hayes and the NMFS/SWFSC to ingest salmon data, and the NMFS/AKFSC to ingest marine mammal data through Alaska Ocean Observing System (AOOS). So, we would collaborate with (AOOS an IOOS RA that will manage MARES and AKFSC data. This collaboration with AOOS would involve 1) developing, reviewing and updating tools developed during the TOPP program which enabled external collaborators to submit deployment metadata into the TOPP data management system, and to 2) working with the RA directly to develop protocols for transferring their tag data into the ATN DAC through automated uploads, that ensure that the process is as simple and straightforward as possible. Finally in process 3) the two systems (IOOS ATN DAC) and the RA for the MARES.SWFSC/AKFSC collaboration will share codes when necessary to improve the capacity for the RA to display data on their own website.

Throughout the TOPP program, our data management team ran specific routines for automatic tag ingests and deployment metadata ingests for 14 different taxa, 7 tag platforms. We organized TOPP into guilds (e.g. marine mammals, turtles, fish, sharks, seabirds), and each guild made a tag ingest metadata standard for their group of animals. In Phase II we will build upon those TOPP efforts, utilizing the deployment IOOS metadata standards and further developing, refining and publishing the TOPP tag data conventions to create automated pages that literally ingest the data from a user group, once they have entered a web-based pre-deployment form. This then allows the capture of all telemetry data registered to the system. Data from the International Great Marlin Race (IGMR), an ongoing collaborative tagging program between Stanford University and the International Game Fish Association (IGFA) which consists of data from more than 150 pop-up satellite tags, will serve as a case study. A web-based portal will be developed for this program to enter pre-deployment and deployment metadata and data ingestion processes will be further tested and refined.We anticipate using similar tools at IOOS ATN DAC. This process will then create an automated data stream of Animal Telemetry data, from multiple animals and tag platforms entering into the DAC from across the country. Stanford routinely handles data from several ocean basins (TOPP, TAG, and Remote Territory programs), so there are few unknown issues associated with this step in the 2015 workplan.

Documentation
A final aspect of Phase 2 will involve developing documentation tools to accompany the interface itself. This documentation will include:

- Documentation on the DAC itself, including servers, database structure, QC routines, Web services, visualization software, etc.
- Instructions for use of the DAC
- Links to relevant resources (e.g., methods papers on data processing)
- Bug reports, user comments and feature requests

The development of these tools will help to foster the growth of the ATN DAC as a community asset, by providing ways of facilitating an ongoing conversation among data providers, data managers, software developers and data consumers.

Ingestion of TOPP/TAG Data
The Tagging of Pacific Pelagics and TAG A Giant represent two of the largest global program for animal archival and satellite telemetry. Both data sets are in high demand with critical requests coming from several communities. NOAA and state governments who are involved with Marine spatial planners, along the west coast of North America, send in two requests per month to map some aspect of the TOPP data set in relationship to power usage, marine protected areas, human usage, or other activities. Having the data freely available within the ATN is, and has been a goal. The datasets are large, with thousands of deployments and a wealth of location, oceanographic and behavioral data. To date the ATN DAC has roughly 20% of the TOPP dataset ingested (See Appendix Table 1). We currently include only a sampling of the Pacific bluefin tuna data and no white shark data from TOPP, as well as a small sample of the Atlantic bluefin tuna data from TAG and the marlin tagging data from the IGMR. And for the vast majority of the animal tagging datasets that are in the system, the ATN DAC currently provides access only to the location data (i.e., date, latitude and longitude). In many cases there is also extensive depth and temperature data in the TOPP database, which has not yet been ingested into the ATN DAC system. We believe focused effort should be made to get the entire data set in. Similarly

stock assessment scientists in the Atlantic and Pacific oceans and RFMOS that manage these important fisheries have stated requests in their executive summary documents of what is needed to properly assess these fisheries. TAG data provides critical information for life history inputs as well as fisheries and natural mortality estimates. By placing both data sets (2000 archival tag deployments) into the ATN DAC, the utility for using the ATN for stock assessment purposes would be validated. Finally increased interest is emerging to utilize animal movement data sets for climate redistribution models. ATN DAC with both TOPP and TAG and newer data from more recent deployments will provide rich attributes to studies examining coastal ecosystems and resolving animal distributions in a changing sea. US IOOS/ONR DAC will be the largest source of multiple animal data once this archive is created. This has always been a legacy goal of Drs. Block, Costa and Kochevar and the entire TOPP/TAG team.

Compliance with IOOS Standards

A key aspect of the ATN that was not implemented in the beta release is to bring the system, and the content on it, into compliance with IOOS controlled vocabulary and metadata conventions and standards. This includes the need to use standard terminology throughout the ATN DAC, both on the front-end browser interface as well as in the data descriptors used in the various databases which are made accessible through ERDDAP.

A second aspect of this task is the requirement to create metadata citations, using ISO or DOI standards, which will allow end users to quickly and easily identify all of the characteristics of the dataset in question, including the species, tag type, PI, post-processing routines, etc. In some cases, as for acoustic tags, there are well-established, thoroughly documented and widely-used metadata conventions already in place. In other cases, we may need to play a lead role in defining those standards, using the work we already did during TOPP to standardize data from a variety of different tagging platforms.

In order to carry out these tasks, we will need to update the database descriptors to ensure that they use appropriate terminology and units throughout. We will also need to assess the current state of metadata that exists in the TOPP database, and develop a process to assimilate the metadata we have into the ATN DAC system. Because there will certainly be gaps in the existing metadata, this process is certain to require some level of human intervention and manual data entry – which could be extensive.

In order to carry out these projects, we will need to provide some limited support for database formatting, etc.; but the bulk of the work will probably involve reviewing and manually updating data and metadata fields in the database.

RESULTS

In this section, we are detailing what the plan for this phase of the project will entail.

Project Tasks and Timeline

A. Develop ATN DAC Infrastructure Updates and Improvements (5 months/24 weeks; led by Stanford database manager)

a. Develop and implement infrastructure improvements and order all components required
b. Assemble, integrate and test one paired server/backup system (primary server) (4 weeks)
c. Assemble, integrate and test second paired server/backup system (development server) (2 weeks)
d. Develop and test mirroring protocols (6 weeks)

e. Assemble and test third paired server/backup system (3 weeks)
f. Relocate third system to Stanford main campus, establish database communications, mirror server and backup server operations among all three systems (4 weeks)

B. Develop capability for Integration of External Datasets (times below based on AOOS/MARES example; may vary depending on needs of data providers and types of data being ingested. To be coordinated by Stanford project manager and carried out by database manager, database programmer and GIS specialist. Final website integration by SWFSC information technology specialist). Other external datasets to be integrated include AKFSC marine mammal program (includes engagement with AOOS and Axiom) and SWFSC acoustic telemetry dataset.

a. Identify available datasets and data operators
b. Develop and establish protocols for delivery of deployment metadata and tag data into ATN DAC (5 weeks)
c. Review current status of deployment metadata forms and applicability to (e.g., AOOS/MARES) program (2 week)
d. Develop and test web-based deployment metadata form(s) for (e.g., AOOS/MARES) project use (6 weeks)
e. Develop and establish process to ingest metadata records and tag data into ATN DAC database (timeline dependent upon fieldwork schedule and volume of data to be ingested)
f. Process data as needed (time required TBD) using algorithms
g. Integrate (e.g., AOOS/MARES) data into ERDDAP server and ATN DAC front-end user interface (2 weeks from final data processing)
h. Develop process to review forms, processes and outcomes with (e.g., AOOS/MARES) project staff to provide feedback for future development of data sharing tools (4 weeks)

C. Improve Data Interoperability (to be coordinated by project manager; all ATN DAC team)

a. Review existing data sharing tools available on other, comparable sites (e.g., OTN, Movebank, etc.) and identify key features and functions needed for ATN DAC
b. Convene meeting with Axiom technologist to explore the use of the Research Workspace (https://workspace.aoos.org/) to enable interoperability with ATN DAC . May include ingestion of AKFSC marine mammal data into DAC through research workspace platform.

D. Documentation (to be coordinated by project manager; all ATN DAC team)

a. Create information architecture for system documentation (4 weeks)
b. Develop system documentation and instructions for use (4 weeks)
c. Identify and develop tools for user feedback, bug reports, etc. (4 weeks)
d. Integrate documentation features into ATN DAC website (4 weeks)

E. Ingestion of TOPP/TAG data sets in high demand

a. Ingest all remaining TOPP datasets, including Pacific bluefin tuna and white shark data as well as depth and temperature data for all species. (3 months of effort).
b. Ingest all TAG Data (3 months of effort)

F. Improve Compliance with IOOS Standards (to be coordinated by project manager; IOOS/CeNCOOS DMAC team and all ATN DAC team)

a. Review extant ATN DAC data and document all data structures (e.g., date, time, latitude, longitude, depth, temperature, etc.) (4 weeks)
b. Identify appropriate standard nomenclature/units (where available) for all structures (1 week)
c. Update database to reflect ISO standard nomenclature/units (2 weeks)
d. Review deployment metadata records for all extant ATN DAC datasets (2 weeks)
e. Define required metadata schema for all current taxa and tag platforms (1 weeks)
f. Create database structures to accommodate all required ATN DAC metadata (2 weeks)
g. Ingest TOPP, TAG and IGMR metadata into updated metadata database (4 weeks)
h. Develop and deploy updated views of deployment metadata in front-end user interface (2 weeks)

Deliverables

a. Updated server infrastructure with three paired server/backup systems; two located at Hopkins Marine Station and one at Stanford University Main Campus in Palo Alto
b. Web-based deployment metadata forms for external data providers (e.g. MARES project researchers, AFSC marine mammals program and/or SWFSC fish program)
c. Integrated external data users' data and metadata, available through front-end user interface map as well as ERDDAP server.
d. Refined metadata views for all TOPP, TAG and IGMR datasets, integrated into database, ERDDAP server and front-end user interface compliant with IOOS DMAC standards and protocols.
e. New website sections providing system documentation and user feedback
f. More data available from TOPP and TAG, including addition of Pacific bluefin tuna, Atlantic bluefin tuna and adult white shark datasets, as well as associated depth and temperature data for all species in ATN DAC.

Technical Points of Contact:

Dr. Barbara Block
Tagging of Ocean Pelagic Predators Program
Stanford University, Department of Biological Sciences
Hopkins Marine Station, Pacific Grove, CA 93950, USA
831-655-6236**,** bblock@stanford.edu

Dr. Michael J. Weise
Marine Mammals & Biological Oceanography Program
Office of Naval Research - Code 32
One Liberty Center - Rm 1068
875 N. Randolph St.
Arlington, VA 22203-1995
703-696-4533, michael.j.weise@navy.mil

Dr. Hassan Moustahfid
US Integrated Ocean Observing Program
National Ocean Service/NOAA
1100 Wayne Ave, Suite 1225, Silver Spring, MD 20910
301-427-2429, hassan.moustahfid@noaa.gov

Lynn DeWitt
NOAA SWFSC Environmental Research Division
110 Shaffer Road
Santa Cruz, CA 95060
831-420-3668, lynn.deWitt@noaa.gov

WORK COMPLETED

Funding was only recently received, so no work beyond planning has occured.

IMPACT/APPLICATIONS

We expect in Phase II to deliver more realtime on line products, and have summary data that may intuitively be digested by the public who might be viewing the data. One type of very common tag is the pop up satellite archival tag. We anticipate having all tags in a uniform output that provides access to time series data in a raw format, as well as sophisticated plots of the data for summarized download. Table 1 shows the tagged animal datasets currently available in ATN DAC.

RELATED PROJECTS

The ATN DAC project builds on a decade of work done in the Tagging of Pacific Predators (TOPP) program, for which the TOPP data management system was first developed. The software for this system was further refined and improved for the GulfTOPP system, used in the NOAA National Resources Damage Assessment process following the *Deepwater Horizon* oil spill, and many features in the ATN DAC were first developed as part of the GulfTOPP effort. We are also using many of the data processing and management tools in collaboration with the International Game Fish Association for our IGFA Great Marlin Race project; the data from which are being incoroporated in their entirety into the ATN DAC. We are also using animal telemetry data in a variety of public education and outreach projects, including NSF-funded Ocean Tracks and Ocean Tracks: College Edition programs, the Concord Consortium's NSF-funded Common Online Data Analysis Platform (CODAP) project, and the Exploratorium's NSF-funded Living Liquid exhibit development program. In all of these cases, discussions are happening about how to either leverage or tie in to the ATN DAC to facilitate greater and more efficient data delivery to educators, students and the general public that uses these resources.

Future phases of this work, currently being proposed, will add more data to the portal and enhanced capabilities for data access, distribution, archiving, analysis and visualization based in part on requirements defined in the first steps of this project and on user feed-back when the initial product is deployed.

REFERENCES

No Citations

Appendix:

Table 1: Tagged animal datasets currently in ATN DAC and TOPP database

Animal	PI	Years	ATN DAC	TOPP DB
Cetaceans	Bruce Mate			
blue whale		2004-2008	32	32
fin whale		2004, 2006	1	1
humpback whale		2003-2005	13	13
sperm whale		2008	4	4
Fishes	Block			
Atlantic bluefin tuna		2007-2008	6	1270
black marlin		2011-2014	30	30
blue marlin		2002-2014	91	91
Pacific bluefin tuna		2002, 2005	9	849
Pacific sailfish		2005-2014	21	21
striped marlin		2003-2014	49	49
swordfish		2004-2009	12	12
white marlin		2013	6	6
yellowfin tuna		2003, 2013	11	609
Pinnipeds	Costa			
California Sea Lion		2003-2008	136	136
Cape Fur Seal		2008	6	6
crabeater seal		2007	11	11
Galapagos sea lion		2005-2006	18	18
Northern elephant seal		2002-2014	331	331
Southern elephant seal		2007-2009	32	32
Weddell seal		2010	22	22
Seabirds	Shaffer			
black-footed albatross		2002-2009	103	103
Laysan albatross		2002-2009	137	137
sooty shearwater		2005	10	10
Sharks				
basking shark	Wilson	2009-2011	3	3
blue shark	Dewar	2002-2014	90	90
common thresher shark	Dewar	2003-2008	23	23
Galapagos shark	Block	2006	17	17
Greenland shark		2004	1	1
grey reef shark	Block	2013	7	7
Juvenile white shark	Jorgensen	2004-2011	32	32
manta ray	Block	2004-2014	15	15
oceanic whitetip shark	Block	2004, 2006	6	6
porbeagle shark		2008-2009	3	3

salmon shark	Block	2002-2012	133	133
silky shark	Block	2004-2013	8	8
silvertip shark	Block	2013-2014	8	8
tiger shark		2010	1	1
whale shark		2010-2011	5	5
white shark	Block	2002-2013	74	74
Humboldt squid	Gilly	2004-2009	32	32
Sea Turtles				
Hawskbill turtle		2010	1	1
leatherback sea turtle	Block	2004-2007	101	101
Loggerhead sea turtle		2005-2007	16	16