RIGHT WHALE NEWS

An independent forum for right whale conservation and recovery, publishing several times each year.

Volume 23 Number 1 ***March 2015***

As described in the article on page 6, the Right Whale Catalog along with the genetics database are central to knowledge and recovery of the North Atlantic right whale. The third component is the long-term sightings and survey-effort database maintained at the Graduate School of Oceanography, University of Rhode Island. As elsewhere, changes are on the horizon. The history, status, and future of this important component are reported below.

Status Update: The NARWC Survey and Sightings Database

Contributed by Robert D. Kenney, URI Graduate School of Oceanography

After a one-year hiatus (FY2014) caused by scarcity and uncertainty in federal funding, in September 2014 I got back into active curation and management of the North Atlantic Right Whale Consortium (NARWC, or simply Consortium) database at the University of Rhode Island, Graduate School of Oceanography (URI-GSO). Adding to the uncertainty for the funding agency (NMFS) was my semi-retirement and, in the not-too-distant future, retirement. The database had been in more or less continuous service from the fall of 1986 through August 2013. This report is intended as an update on the status of the Consortium database at URI-GSO. I will provide a summary description of the database and what information it contains, try to explain how it relates to the photoID catalog curated at New England Aquarium (NEAq), clarify some common misconceptions about the database, and finally describe the near-term changes in store. However, my first task will be to present a little bit of the history of the Consortium—as a refresher for the other old veterans out there and as background for the younger readers.

NARWC History

The Consortium's beginnings actually go back to June 1983. At an IWC workshop on right whales held at the NEAq in Boston (the proceedings and papers are summarized in Brownell *et al.* 1986), Howard Winn (a URI-GSO professor and my Ph.D. advisor) stood up and made one of his regular rants about the inadequacy of federal funding for marine mammal research and recovery activities in the U.S. He suggested that the research community should ask Congress for a dedicated appropriation focused specifically on right whales. Lobbying by the NGO community, including Greenpeace and the Connecticut Cetacean Society (now Cetacean Society International), was pivotal in obtaining Congressional support for an appropriation. Both Howard Winn and NEAq president John Prescott testified at a hearing of the Senate Appropriations Subcommittee on Commerce, Justice, State, and Judiciary. The subcommittee approved a budget for the Dept. of Commerce that included a line item directing $500,000 to right whale research, which was enacted by the Congress.

After some foot-dragging by the agency, NMFS made $381,000 available and requested a single proposal from the research community for a collaborative research program. The Principal Investigators on that proposal were Howard Winn[*] and myself at URI-GSO, John Prescott* and Scott Kraus at NEAq, Stormy Mayo at the Center for Coastal Studies (CCS), Bill Watkins* and Karen Moore at the Woods Hole Oceanographic Institution (WHOI), and Dave Caldwell* and

[*] Howard, John, Bill, Dave, and Melba are all now deceased.

Melba Caldwell* at Marineland of Florida (MLF). The proposal was submitted with URI-GSO as the lead and the other organizations as subcontractors. After the usual negotiations, a contract was awarded to begin in October 1986. From the beginning, we called ourselves the "North Atlantic Right Whale Consortium" although there was no formal organization. The Consortium as we know it today was created in 1998.

From the very beginning, a critical component of the Consortium project was to maintain a centralized data archive to which everyone would contribute and have full access (Kenney and Winn, 1986; Kenney, 2001). We attempted as much as possible to standardize the methods for conducting field surveys and recording data, as well as for managing the resulting computerized information. Of the NARWC collaborators, URI-GSO had the most experience in handling large databases. From late 1978 through early 1982, we had conducted the Cetacean and Turtle Assessment Program (CETAP, 1982). CETAP was a large project designed to characterize the distribution, abundance, and seasonality of all whales, dolphins, porpoises, and sea turtles in U.S. continental shelf waters from North Carolina to Maine. The underlying rationale was environmental assessment relative to proposed oil and gas development. CETAP resulted in a substantial database that was archived at URI. Because of this existing data archive and database management expertise, URI-GSO was selected to manage the NARWC database, and, as the only person left standing from CETAP who was familiar with the data, I became the data manager. The CETAP database became the original core of the NARWC database with many of the data structures, conventions, and protocols following or adapted from those that were originally developed for CETAP.

At the beginning, the NARWC contract included all of the familiar tasks: aerial surveys, shipboard surveys, photoID, data management, data analysis, and publication. After the first few years of the project, NMFS decided to split the subcontracts off into their own separate contracts to the individual organizations, at least, in part, to avoid double-charging of overhead. After that time MLF no longer had any part of the funded research. Over the years, budgets got tighter, priorities changed, and some tasks began to drop out. Beginning in 1993, URI's contracts had been narrowed down to include only database management, including data analyses and publication, except in the last few years as funding shrank even further. For 1986–1993, NMFS funding to URI for all right whale projects, including the subcontracts in the first several years, totaled $1.62M (a seven-year period, at about $230,000 per year). Funding for 1993–2013 for database management only, totaled $1.77M (a 20-year period, at about $89,000 per year).

The Database

The Consortium database currently encompasses four separate databases, and all of the summary statistics below refer to the aggregated whole. The four databases, all of which are in identical formats so they could be easily combined, include:

- The main Consortium database.
- A second similar collection of files put together for a project where I worked with Geo-Marine, Inc. on a series of environmental summaries ("Marine Resources Assessments") for the Navy. This was almost entirely NMFS surveys, mostly but not exclusively in the Southeast. I never got around to asking all the relevant individuals for permission to add the data to the main archive, so for now they have been kept separate.

- All of the survey files from the Northeast Large Pelagics Survey Collaborative (NLPSC)—the aerial surveys that have been conducted in the Massachusetts–Rhode Island Wind Farm Area since the fall of 2011 (see Kenney 2011, 2014). The contract from the Massachusetts Clean Energy Center places restrictions on data release until the completion of the project.
- A collection of mostly stranding data, including large-whale records for the Northeast from the Smithsonian Institution's computer files and more general stranding records from NMFS for Rhode Island, Connecticut, New York, and New Jersey for 1993–2005. These were assembled for a publication project on the marine mammals of New York.

Eventually, I expect that the NMFS survey database and the NLPSC dataset will both be incorporated into the main archive. On the other hand, because the stranding data are available elsewhere where they are updated regularly and because the coverage (geographic and taxonomic) in what I have is incomplete, it does not seem wise to incorporate them.

The NARWC database today comprises about 5 million records. The data are archived, managed, and analyzed using SAS software (SAS Institute, Inc., Cary, NC). At the beginning in 1986, SAS only ran on a full-sized IBM mainframe computer where the data were stored on 9-track magnetic tape reels. Today, SAS runs in the standard personal computer environment (either 32-bit or 64-bit), so the entire operation is housed in a desktop computer in my lab (and backed up in multiple locations).

The database includes both survey and sighting data. The database was designed so that there is a single data structure that fits all data types. For a survey dataset, track (*e.g.*, date, time, location, heading, altitude), environmental (*e.g.*, weather, sea state, visibility), and sighting (*e.g.*, species, number, behavior) parameters are included within the same data records rather than being separated into different files as in some data-logging or archival systems. Several different classes of data are included:

- Line-transect ("dedicated") aerial surveys: These are surveys designed to generate estimates of density and abundance of the species encountered using distance-sampling methods. Survey methodologies are strictly defined to maintain statistical rigor. There are also line-transect shipboard surveys, but we have never conducted any ourselves, so the necessary data structures have never been created.
- POP aerial surveys: During CETAP we established a "Platforms of Opportunity Program" (POP). We had a cadre of skilled observers under contract who could be placed aboard any aircraft flying over the study area, most typically Coast Guard fishery patrols and aerial radio-thermography missions (this was before satellites measuring sea surface temperature existed). The trackline of the aircraft was determined by the primary mission, and the observer kept a continuous log of both track and environmental data and recorded all sightings. Today, the primary missions of most of our aerial surveys include detection of right whales for sighting alert systems and photoID, and the data are recorded in this POP format. Some line-transect surveys by others (*e.g.*, NMFS) are archived in the NARWC database in the POP format, since we consider it outside the scope of our research to be re-doing density estimates from other people's data.
- POP shipboard surveys: This is essentially the same as above, but using vessel platforms rather than aerial.

- Opportunistic and historical sightings: An opportunistic record is simply a sighting without any associated platform or track information. There may or may not be associated environmental data; most often there are not. There is nearly no difference between opportunistic and historical data other than the time factor. During the first year of CETAP, there was an effort to identify and incorporate any pre-existing sighting records as "historical" data. Since that time, older and current opportunistic sightings have not really been distinguished.
- Strandings: In addition to the separate stranding dataset identified above, there are some stranding records included in the primary NARWC database. Stranding records in the original CETAP data were identified by a different first character of the FILEID variable, which differed between mammal and turtle strandings. Currently, a stranding record would be an opportunistic sighting with a behavior code that indicates "dead, stranded" or "dead, floating".

The survey data in the NARWC database go back to the first CETAP surveys in the fall of 1978. The opportunistic data go back further, with three records before 1800, all representing whales killed by whalers. The two oldest right whale records were whales killed off Cape May, New Jersey: two in April 1762 and one in April 1764. Both of those came from old newspaper articles in a folder of miscellaneous historical data from the NEAq. The oldest record in the database is for a humpback killed in Nantucket Harbor on an unknown date in 1608—"when a party of Indians killed a humpback whale which got stranded on a part of Nantucket, called Caton, in the inner harbor" (Allen 1916, quoting from A. H. Clark, in G. B. Goode, *Fisheries and Fishery Industries of the U.S.*, 1887). That record came from the file of mostly strandings obtained from the Smithsonian. Jim Mead, the retired curator of marine mammals there, does not get enough credit for being a pioneer in computerizing marine mammal data. The Smithsonian dataset includes many older records painstakingly extracted from Allen's monograph and many other similar sources.

The combined database presently includes 44,413 right whale records. That total includes 44,412 records of North Atlantic right whales and one southern right whale. One sighting in the Gulf of Guinea off West Africa came into the database from the photoID catalog (see below).

Misconceptions

Maybe it's because the Consortium includes "North Atlantic Right Whale" in its name, but some people mistakenly conclude that the database contains only right whale sightings. In fact, right whales are only the third most frequently sighted species in the database following loggerhead sea turtles and bottlenose dolphins. Right whales represent only 11% of the total number of 392,417 records of marine biota. The total includes 77 species of cetaceans, pinnipeds, other marine mammals, sea turtles, sharks, other fishes, and other marine species. Other broader categories, including two species (*e.g.*, fin or sei whale, pilot whale sp., common or white-sided dolphin), multiple species in a genus (*e.g.*, *Balaenoptera* sp., *Mesoplodon* sp., *Stenella* sp.), and broad general terms (*e.g.*, unidentified large whale, unidentified dolphin, unidentified sea turtle), bring the total number of "species" to 111. In addition, there are 55,946 sightings of 77 species and 22 unidentified categories of birds, mainly from the Manomet

surveys (see below) but a few from whale-watch naturalists. Finally, there are another 170,291 sightings of vessels, fishing gear, human activities, debris, or oceanographic phenomena.

Conversely, there are those who think that the database includes parameters or estimates that it actually does not. Occasionally, I receive requests for density or abundance estimates. While it may be possible to calculate those estimates from the line-transect survey data in the database, the estimates themselves are not included. Likewise sightings-per-unit-effort (SPUE) values are not contained in the database. Those are computed from the aggregate of aerial line-transect, aerial POP, and shipboard POP survey data. The survey area, the target species, and the study objective all impact how SPUE data are worked up. A new SPUE dataset is typically generated for each request; however, at times a data requester is able to utilize a previously generated dataset without incurring additional expense or time delays. One final item frequently requested by those interested in distribution patterns is GIS shapefiles. The data are not stored in a GIS environment so there are no shapefiles, although it is a simple matter to output data files in the appropriate format for input to ArcGIS or any other software.

The biggest misconception is that the NARWC database is so biased toward right whales that it is not useful for analysis for other species. The corollary is that there is little or no survey effort outside of right whale habitats. Both of these statements are simply untrue. It is the case that the current surveys by the Consortium partners are focused in right whale habitats: the winter EWS surveys in the Southeast, the winter-spring aerial Cape Cod Bay surveys by CCS, and the summer-fall Bay of Fundy and Scotian Shelf surveys by NEAq. However, the database does include extensive broad-scale survey effort. In addition to the CETAP surveys, there was a survey program conducted by the Manomet Bird Observatory from 1980 to 1988. Manomet observers went out on NMFS fisheries and oceanographic research cruises and collected POP-type data on seabirds, marine mammals, and sea turtles. There also are past and current more broad-scale surveys by NMFS, including some focused on right whales and others on general stock assessment. The whole point of SPUE analysis is to factor out as much as possible any bias introduced by sampling patterns. Even broken down month by month, SPUE distributions derived from the NARWC data are geographically complete and reliable enough to be used in the co-occurrence model of entanglement risk in the Atlantic Large Whale Take Reduction Plan (http://www.greateratlantic.fisheries.noaa.gov/protected/whaletrp/eis2013/march_2014_draft_vl_model_documentation_appendices.pdf).

The Database vs. the Catalog

One other class of information that is not included in the NARWC database is the right whale photoIDs. Those data, including individual whale identity, age, and sex, are contained in the photoID database (i.e., the "Catalog") curated at NEAq. It is not always easy to keep the differences between the two datasets straight, not least because records in both are referred to as "sightings". A sighting in the database and a sighting in the catalog, however, are not exactly equivalent. As an example, consider an aerial survey flying down a transect line. They sight something and break from the track to investigate. After circling and taking lots of photographs, they record a "sighting" of 23 right whales. Back in the lab they go through all of their photographs, and decide that they actually saw 25 different whales so they amend their data, and eventually submit the data for incorporation into the database. They also submit all of the

photographs to NEAq for inclusion in the Catalog. If they were perfectly correct in their analysis, the catalog will end up with 25 "sightings" of right whales—one per animal. It could be more or less, depending on the quality of the photos, the existence of duplicates of the same whale, or the presence of more individuals than showed up in the preliminary analysis. For that reason, and to keep things straight in my own mind, I always try to restrict the term "sightings" to the database and to refer to records in the Catalog as "records" or "identifications."

The two datasets periodically get cross-referenced, which is one of my least favorite jobs. I generate a chronological listing of all right whale sightings in the database. Then, I obtain an updated copy of the Catalog data from NEAq, sort it chronologically, and filter it for only records not already matched to sightings in the database. Then, I just match up catalog records line by line to database sightings (by matching date, time, location, and source), and manually enter the FILEID, EVENTNO, and SIGHTNO fields from the database into the appropriate record(s) in the Catalog. The process cannot be automated because the data attached to the photoID records do not always exactly match those in the database. Because it takes around two years for the photo-matching at NEAq to be considered relatively complete, and the cross-referencing task happens only once a year at best, data users can never expect the database cross-references within the Catalog to be fully up to date. Once the matching process is complete, any photoID records that do not correspond to sightings in the database and that are not likely to be from surveys where the data are expected to be submitted in the future are extracted, creating a file of new opportunistic right whale sightings to be added to the database.

The Future

A new one-year contract is now in place through August of 2015. Some changes have already taken place, and others are coming. I now have four co-PI's on the project: Dr. Peter August, Charles LaBash, and Christopher Damon from the URI Dept. of Natural Resources and the URI Environmental Data Center (EDC), and Dr. Kathleen Vigness-Raposa from Marine Acoustics, Inc. EDC (www.edc.uri.edu) is the primary provider of spatial data at URI and within the State, and the home of both the state geographic information system (RIGIS; http://www.edc.uri.edu/rigis/) and mapping for NOAA's Large Marine Ecosystems initiative (http://lme.edc.uri.edu/). Dr. August has a background in terrestrial mammals and bats before turning to landscape ecology, but little experience with marine mammals. Dr. Vigness-Raposa's role is to provide the marine mammal expertise. She completed her M.S. in oceanography with Howard Winn on modeling visual vs. acoustic whale surveys before going on to earn a Ph.D. in landscape ecology with Pete August on modeling North Atlantic humpback whale habitats.

The first task in the new contract is to get caught up on the backlog of data that did not get processed during the hiatus year. We also need to stay current with newly generated data submissions, and complete a database-catalog cross-referencing on two years of new catalog records.

The more important task is to begin designing a new database structure in a more modern, useful software environment. The way that the current system has evolved piece-meal over the years has resulted in a system that relies on software that is not particularly available or user-friendly (dBASE, SAS) and that has a large number of relatively "clunky" processes. They work

perfectly well for me, but I could never expect another person to step in and figure out the whole thing. We need to develop new software to perform all the current functions: data entry, conversion of data from a variety of input formats to the defined standard, quality-control testing and corrections, archival, queries via the Consortium data-sharing process, and the analyses often necessary for particular queries (*e.g.*, SPUE analysis). The long-term expectation would be to transfer all database management, quality-control, archival, and query functions to the EDC, although expecting that to be fully complete within this one year may be optimistic. The final, even longer-term goal is to develop a stable source of funding, rather than needing to rely on a series of one-year federal contracts in what is sure to remain an unpredictable budget environment.

Acknowledgements

The U.S. Dept. of the Interior, Bureau of Land Management supported the CETAP study and got the ball rolling. The U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service provided near-continuous funding for database management for 28 years. NMFS also supported some survey projects, as did other agencies including the Minerals Management Service, Marine Mammal Commission, U.S. Army Corps of Engineers, U.S. Navy, state agencies, and others. I also thank all of the data contributors—far too many to list individually.

References

Allen, G.M. 1916. The whalebone whales of New England. Memoirs of the Boston Society of Natural History 8(2):107–322.

Brownell, R.L., Jr., P.B. Best, and J.H. Prescott, eds. 1986. Right Whales: Past and Present Status. Reports of the International Whaling Commission, Special Issue 10. International Whaling Commisson, Cambridge, U.K.

CETAP (Cetacean and Turtle Assessment Program, University of Rhode Island). 1982. A Characterization of Marine Mammals and Turtles in the Mid- and North-Atlantic Areas of the U.S. Outer Continental Shelf, Final Report. Bureau of Land Management, Washington, DC.

Kenney, R.D. 2001. The North Atlantic Right Whale Consortium databases. Maritimes 43(2):3-5.

Kenney, R.D. 2011. New survey to begin south of Martha's Vineyard. *Right Whale News* 19(2): 8-9.

Kenney, R.D. 2014. Mass-RI Wind Area surveys extended. *Right Whale News* 22(2): 5-6.

Kenney, R.D., and H.E. Winn. 1986. Marine Mammal Data Transfer and Documentation. Final Report, Contract no. 40-EANF-501629. National Marine Fisheries Service, Woods Hole, MA.