# The North Atlantic Right Whale Consortium Database: A Guide for Users and Contributors

# Version 6

## Robert D. Kenney

North Atlantic Right Whale Consortium
Reference Document 2019-02

March 2010 (original version)
October 2011 (revised)
December 2016 (revised)
February 2018 (revised)
August 2019 (revised)
November 2019

University of Rhode Island
Graduate School of Oceanography
Narragansett, Rhode Island

# TABLE OF CONTENTS

**[THIS PAGE INTENTIONALLY LEFT BLANK]**

[I've always wanted to do that once.]

[But I refuse to put in another page with a list of acronyms and abbreviations.]

# 1. PREFACE

The North Atlantic right whale (*Eubalaena glacialis*) and North Pacific right whale (*E. japonica*) are the two most endangered whale species in the world, although it is presently impossible to conclude with any confidence whether one or the other is in the most precarious situation (Clapham *et al*., 1999; Kenney, 2018). Recent sightings of North Pacific right whales are much rarer, but they have occurred mainly offshore in the Bering Sea (Clapham *et al*., 2004; Shelden *et al*., 2005; Wade *et al.,* 2006), where there are few people to notice. North Atlantic right whales frequent such locations as Cape Cod Bay, the Bay of Fundy, and the nearshore waters off Jacksonville, Florida (Winn *et al*. 1986; Kenney *et al*., 2001; Kraus and Rolland, 2007a), where they are easily accessible for study, but also more susceptible to human impacts (Knowlton and Kraus, 2001; Johnson, 2005; Johnson *et al*., 2005, 2007; Kraus *et al*., 2005, 2016b; Knowlton and Brown, 2007; Kraus and Rolland, 2007b; Moore *et al*., 2007; Parks and Clark, 2007; Knowlton *et al*., 2012, 2016; van der Hoop *et al*., 2013, 2015; Hayes *et al*., 2017).

North Atlantic right whales were the first targets of commercial whaling, beginning along the shores of Europe about a thousand years ago, and by the 18th Century the species was greatly reduced and commercially extinct—no longer economically viable as a target for whalers (Aguilar, 1986; Reeves and Smith, 2006; Reeves *et al*., 2007; Laist 2017). Until the early 20th century, their recovery was likely prevented by continued killing by whalers, most of whom were actually hunting sperm whales. Right whaling was finally stopped by international treaties beginning in the 1930s (Hain, 1975; Laist 2017) except for illegal whaling by the Soviets in the North Pacific and Southern Ocean (Clapham and Ivashchenko, 2009; Ivashchenko *et al*., 2007, 2011), but human impacts are still slowing recovery. The difference is that now we kill them accidentally—by running them over with large ships and drowning them in commercial fishing gear (Knowlton and Kraus, 2001; Johnson, 2005; Kraus *et al*., 2005; Johnson *et al.*, 2007; Knowlton and Brown, 2007; Knowlton *et al*., 2012, 2016; van der Hoop *et al*., 2013, 2015; Hayes *et al*., 2017).

As part of a collaborative research effort that began in the 1980s, a centralized archive of survey and sighting data was created at the University of Rhode Island

(Kenney, 2001, 2015). Every organization and agency conducting surveys for right whales, from Florida to Atlantic Canada, submits their data for inclusion. We also try to obtain any other marine mammal or sea turtle survey data for the region. Every data file is run through a strict quality-control process in order to insure that the archive is as error-free as possible. The database can then be useful for looking at spatial and temporal patterns of occurrence of right whales or many other species of whales, dolphins, porpoises, seals, sea turtles, or specific large fishes, or to synthesize more analytical datasets for use in ecological studies or models. While the database is not publicly accessible, data are made available on request to all qualified researchers, students, agencies, organizations, and institutions for a wide variety of research programs (visit the North Atlantic Right Whale Consortium home page at http://www.narwc.org for a copy of the most recent data-sharing protocols).

This user's guide to the North Atlantic Right Whale Consortium database is intended to serve as a resource for both contributors to the database and users of the data. It provides background on the development of the database, information about collecting and submitting data, and details on the data included. A digital version (PDF) of the most recent update of the document can be downloaded from the Consortium website at https://www.narwc.org/sightings-database.html (click on the blue bar that says "SIGHTINGS DATABASE USERS GUIDE"). Copies of earlier NARWC Reference Documents (Kenney, 2002, 2003) and PDF scans of even older documents (*e.g.*, Owen and Kenney, 1983; Kenney and Winn, 1986) are currently available only by sending me an email request (rkenney@uri.edu), at least until I figure out how to create a folder in Google Docs that can be accessed by anyone.

For the last revision, I decided to scrap the ALL-CAPS formats for the code listings that have persisted since CETAP, if only just to save a little space. This current revision is setting a record for the shortest time between versions (three months), but there was a real need to further clarify the definition of the new aerial survey format added during the last round.

## 2. DATABASE HISTORY

In the fall of 1986, the University of Rhode Island's Graduate School of Oceanography (URI, GSO), New England Aquarium (NEAQ), Center for Coastal Studies in Provincetown (CCS), Woods Hole Oceanographic Institution (WHOI), and Marineland of Florida (MLF) began a cooperative study of right whales in the western North Atlantic. The Principal Investigators involved in the project at that time were Howard E. Winn[*] and Robert D. Kenney at GSO, John H. Prescott* and Scott D. Kraus at NEAQ, Charles A. "Stormy" Mayo at CCS, William A. Watkins* and Karen E. Moore at WHOI, and David K. Caldwell*and Melba C. Caldwell* at MLF. The study was funded through a separate line item in the federal budget for the U.S. Department of Commerce, and came through the Northeast Fisheries Science Center (NEFSC), National Marine Fisheries Service (NMFS), National Oceanic and Atmospheric Administration (NOAA). NMFS is the federal agency with primary responsibility for conservation and management of marine species, including whales and other marine mammals. Lobbying by the conservation community, including Greenpeace and the Connecticut Cetacean Society (now Cetacean Society International), was pivotal in obtaining Congressional support for the program, assisted by testimony by GSO professor Winn and NEAQ president Prescott. This set of cooperating institutions was informally called the "North Atlantic Right Whale Consortium" for many years. The NARWC was established as a more formal organization in 1998.

From the very beginning, a critical component of our project has been to maintain a centralized data archive (Kenney and Winn, 1986; Kenney, 2001, 2015). 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, GSO had the most experience in handling large marine mammal databases. From late 1978 through early 1982 GSO had conducted the Cetacean and Turtle Assessment Program (CETAP, 1982). CETAP was a large project, funded by the Bureau of Land Management (BLM), U.S. Dept. of the Interior, designed to characterize the distribution, abundance, diversity, and seasonality of all whales, dolphins, porpoises,

---

[*] all now deceased

and sea turtles in U.S. continental shelf waters from North Carolina to Maine. The underlying rationale was environmental assessment relative to oil and gas development, both on Georges Bank and in the mid-Atlantic. CETAP resulted in a substantial database that was archived at URI. Because of concern that the data would be too difficult to be useful to future researchers, GSO Dean John Knauss funded a 1-semester research assistantship for Ralph E. Owen, which resulted in a detailed report documenting the database (Owen and Kenney, 1983). NEFSC had also recognized the value of the CETAP database, and had funded a contract to GSO to reformat the database into a form that would be usable on their computer system, transfer a full copy on magnetic tape, and submit a report with the data containing complete data documentation (Kenney and Winn, 1986). Because of this existing data archive and database management expertise, URI/GSO was selected to manage the NARWC database. 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 the CETAP.

There have been dramatic technological changes in data processing since CETAP and the beginning of the NARWC study. All of the CETAP survey and sighting data were computerized via standard 80-column IBM punch-cards. The software selected for all database management processes was an early version of SAS (originally an acronym for "Statistical Analysis System"). At that time, SAS only ran on the large IBM mainframe computer on the URI Kingston campus, and the database was archived on multiple 10-inch, 9-track magnetic tapes. By the time the NARWC project got underway in 1986, personal computers had become commonplace, but SAS still only ran on a mainframe system. The solution was for all the collaborators to use easily available PC software to computerize their data, which would then be transferred into SAS at GSO. We created a set of interactive programs in dBASE-3+ designed to assist collaborators in entering survey data from hand-written field logs into standardized computer datasets for further processing at GSO.

At the present, most survey data are automatically logged in the field using laptop computers linked to GPS receivers. Contributors submit data files in a variety of software formats, including dBASE, MS-Excel, comma-delimited ASCII text, and MS-Access,

although the data formats are standardized regardless of the software used. SAS is still the primary database management software—SAS for Windows, version 9.4 (SAS Institute, Inc., Cary, NC), although work is underway to convert the database to a format that is more user-friendly and accessible. And the entire database is archived on a single desk-top personal computer at GSO, with multiple backup copies on- and off-site.

The NARWC database today has reached the 3.0-gigabyte mark, and includes over 7.5 million records[*]. It includes 51,280 sightings of right whales, and over 425,000 sightings in all. The earliest right whale sighting is from 1762 (extracted from an article in an old Cape Cod newspaper), and the latest are from 2018. Most of the sightings are between Florida and Nova Scotia, and are concentrated in a few habitats, with occasional sightings in the Gulf of St. Lawrence (increasing in frequency after about 2013), the Gulf of Mexico, and Newfoundland, and others as far away as Greenland, Iceland, Ireland, Norway, Spain, the Azores, Madeira, the Canaries, and the Mediterranean (Fig. 1). There is also one sighting from a photoID catalog record (no doubt a southern right whale) in the Gulf of Guinea off western Africa, and we finally created a separate species code for that record, so it disappeared from Fig. 1. The U.S. and Canadian governments have made use of the data to formally designate three Critical Habitats in U.S. waters (subsequently expanded, with the two in the feeding grounds combined into one larger one) and two Critical Habitats (formerly Conservation Areas) in Canadian waters. In addition, the database includes nearly 375,000 sightings of 112 other species or broader categories of whales, dolphins, porpoises, seals, sea turtles, sharks, rays, other large fishes, or other species on rare occasions (*e.g.*, manatee, alligator, polar bear, jellyfish), with regular additions of more species as needed. Finally, there are also over 203,000 sightings of human activities, including large ships, fishing vessels, fishing gear, recreational vessels, and debris or pollution; as well as over 56,000 sightings of birds, the newest addition to the database.

There is also a second critical component of the NARWC program. During nearly every survey project, the field observers try to photograph every right whale they encounter. Right whales have rough, thick patches on their heads, called callosities,

---

[*] All of the summary totals here are as of 1 October 2018, when an updated version of the master database was created as part of the annual progress reporting process.

which are covered by small, light-colored crustaceans known as "whale-lice." The callosity patterns are unique to individuals, like human fingerprints, and can be used to identify each whale (Payne *et al*., 1983; Kraus *et al.*, 1986). Individual whales can also be identified genetically by "fingerprinting" their DNA collected from skin biopsies, sloughed skin, or fecal samples (Frasier *et al*., 2007).
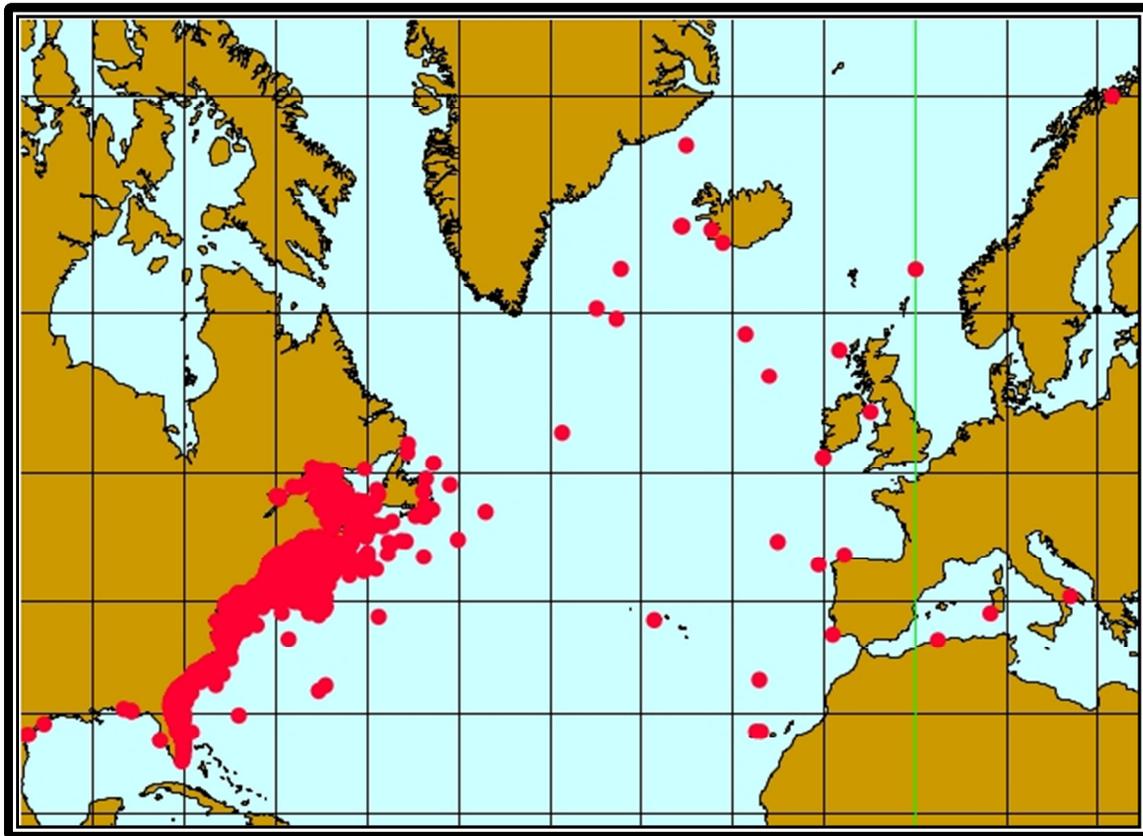


***Figure 1***. *All records of North Atlantic right whales contained in the NARWC database databases, 1762–2019 (n=54,447).*

NEAQ maintains the catalog of photographic and genetic IDs in a database that records each time each whale is identified (Hamilton and Martin, 1999; Hamilton *et al.*, 2007; NEAQ, 2018). The most recent addition to the catalog was tagging data—from both radio tags and satellite tags—with a single location per day added for each tagged whale. The catalog database is extremely valuable for tracking population demographic

parameters—birth rate, inter-birth interval, death rate, survival rate, age at maturity, sex ratio, minimum population abundance, etc.—and how they change over time (Knowlton *et al.*, 1994; Caswell *et al.*, 1999; Kraus *et al.*, 2001, 2007; Pace *et al.*, 2017). A summary catalog database, including basic information about each whale (catalog number, name, sex, reproductive status, mother's ID, and years of birth, death, and last sighting), a composite drawing of identifying features, and selected photographs, is now accessible on-line (http://rwcatalog.neaq.org; NEAQ, 2018). The on-line catalog is linked live to the master catalog database at NEAQ, so data updates are simultaneously made to both. A copy of the catalog database periodically is transferred to GSO, where it gets cross-referenced to the sighting database and file, event, and sighting numbers from the latter are added to the catalog records.

At the same time that the wheels were turning to get the NARWC program started, NMFS contracted URI to provide a copy of the CETAP database on tape, converted to a format of their design. Kenney and Winn (1986) is the final report on that contract, with extensive documentation for the data.

This manual's objective is to ensure that any data submitted are in standardized formats, and that the minimum necessary data variables are included for any particular type of data. It is not to tell anyone how to conduct their surveys or what their objectives should be beyond that minimum. Each program will have its own objectives as determined by their managers and their funders.

# 3. DATA TYPES AND FORMATS

There are three basic types of data that are included in the NARWC database, although the formats, protocols, and sub-types are continuing to evolve. They differ in many aspects—the objective of the field surveys that generated the data, the methods used in the survey, the nature of the data collected, the data formats, and the appropriate uses of the data afterwards (Kenney, 2001, 2015). Whether one is contributing data or using data, there is a need to fully understand the nature and limitations of the data types.

## 3.1. Line-transect Surveys

Line-transect surveys, frequently referred to as "dedicated" surveys, are primarily designed to estimate densities and abundances of any species present within the surveyed area. These surveys are conducted under very rigorous criteria, including use of one or a very few standard sampling platforms, highly trained and experienced observers, survey designs that represent statistically random samples of an area, accurate measurement or estimation of the distances of all sightings from the transect line, and restricted environmental conditions. Detailed information on the design, execution, and data analysis of line-transects surveys is beyond the scope of this document; Buckland *et al.* (2001) would be considered the standard reference to begin with, and Kenney and Shoop (2012) provides an easy-to-understand summary of the methodology. Line-transect surveys can be conducted from either aircraft or ships, however at the present time (1) there are no shipboard line-transect survey data in the NARWC database and (2) the necessary data structures to handle shipboard line-transect survey have never been defined.

CETAP aerial line-transect surveys were flown from 1979 through 1981, including the main surveys throughout the study area, a few designed replicate surveys in two of the blocks, and other surveys focused on oil & gas lease sale areas or on right whale habitats. Other aerial line-transect surveys included in the database are:

- URI right whale surveys off Georgia and Florida in 1987;
- URI right whale surveys off Nova Scotia in 1987;

- NEAQ surveys off the southeastern U.S. in 1989–1992, funded by the Minerals Management Service;

- URI surveys in the Great South Channel region in 1984–1992, with the 1986, 1988, and 1989 surveys comprising part of the South Channel Ocean Productivity Experiment (SCOPEX: Kenney and Wishner, 1995);

- A survey program funded by the Massachusetts Clean Energy Center (and also the Bureau of Ocean Energy Management after the first year) was conducted jointly by NEAQ, CCS, and URI from 2011 to 2015. These North Atlantic Large Pelagic Survey Collaborative (NLPSC) surveys concentrated on the Wind Energy Areas (WEAs) south of Martha's Vineyard and Rhode Island for spatial planning related to offshore renewable energy development. The NLPSC data were incorporated into the NARWC database once the project ended and the final reporting was completed (Kraus *et al*., 2016a). The NLPSC surveys included continuous vertical photography using a fixed camera in the belly of the aircraft (Taylor et al., 2014), which required some modifications of data structures.

- An extension of the aerial surveys of the Mass.-R.I. WEAs, again funded by MassCEC and flown by NEAQ, that commenced in early 2017 and is still continuing at this time. Those surveys also included continuous vertical photographic sampling. A subset of those surveys (called "condensed" surveys) were conducted using a restricted protocol the entailed recording only large whale sightings, necessitating definition of an additional survey type (see FILEID, Section 8.42).

There are some data in the NARWC database that originated from line-transect surveys conducted by others outside of the Consortium, primarily the National Marine Fisheries Service. Those data have been archived in POP survey format (see 3.2 below).

The data from aerial line-transect surveys in the database include detailed information about all sightings (date, time, location, species, number, behavioral data, distance), the track of the aircraft (date, time, position, altitude, heading), environmental conditions (sea state, visibility, cloud cover, sun glare, weather, sea surface temperature), track type (survey line, transit, cross-leg, circling), and watch state (on/off). Density

estimates derived from the surveys are not part of the archived database. Previous estimates from the CETAP, URI, and NEAQ surveys are archived only as hard-copy output in both published and unpublished forms, and as digital data in some cases. NARWC has no direct access to density or abundance estimates from NMFS surveys. It is possible to re-compute densities, using the same or different methodologies as used originally, from the archived data for the CETAP, URI, NEAQ, and NLPSC line-transect surveys. That is not possible from the NMFS survey data, since all of the necessary data fields were not included in the database when the data were added.

## 3.2. Platforms-of-Opportunity Surveys

A significant component of CETAP was the platforms-of-opportunity program (POP). The CETAP POP effort involved a corps of trained observers up and down the coast who were placed aboard a wide variety of aerial and shipboard platforms working in the study area. The platforms utilized included Coast Guard cutters, Coast Guard offshore law-enforcement patrol (OLP) and aerial radio-thermography (ART) aircraft, NOAA vessels, institutional research vessels, foreign research vessels, ferries, commercial fishing vessels, training vessels, and others.

During the CETAP POP surveys, the track of the vessel or aircraft was entirely determined by the primary mission of the platform. The observer(s) simply stood watches and maintained a log—recording the location of the platform periodically and all relevant environmental and sighting information as required. The result is a dataset with relatively complete records of aircraft or ship track and environmental conditions, allowing for subsequent reconstruction of the track and quantification of survey effort. Nevertheless, the surveys were not sufficiently standardized, nor were they designed as statistically random samples, so they can not be used for density estimation.

During the NARWC program, the same POP data format has been used for a variety of directed surveys (*e.g*., aerial and shipboard surveys targeted at right whales) where line-transect density estimation was not an objective. Even though many of these surveys have been directed, systematic efforts following prescribed tracklines, they are formatted as POP survey data and not line-transect survey data (*i.e.*, no sighting

distances, no differentiation of track lines from transits, cross-legs, and circling, etc.) and cannot be reliably used for generating density estimates.

The platform trackline data can be used, however, to quantify survey effort (for that purpose, line-transect and POP survey data are combined). Sighting frequencies can then be scaled by survey effort, resulting in an index we have termed "sightings per unit effort" or "SPUE." This methodology does not allow estimation of the abundance of animals within a given area. It does, however, provide an index of relative abundance that can be quantitatively compared between geographical areas and time periods, or statistically analyzed versus available environmental parameters. As with the density estimates from the line-transect surveys, the effort and SPUE data are not part of the NARWC database. Our practice is to generate new effort and SPUE datasets for each data request, in consultation with the requester, to be sure that the provided data are the optimal available for their purposes.

## 3.3. Intermediate-format Aerial Surveys

An additional, somewhat more rigorous, survey-data format was created in 2015, in collaboration with the aerial survey team leader at CCS. They were interested in recording more detail in their data without needing to switch to a full-scale dedicated line-transect survey format. They were primarily interested in being able to differentiate among designed transect lines, transits, cross-legs, and circling, as well as to record their transect numbers. The format created was intermediate between line-transect and POP surveys, using the same suite of LEGTYPE and LEGSTAGE codes as the former. Since the ability to generate density estimates was not a program objective, they would not be restricted to recording sightings from the trackline so as to keep them "on-effort," but could continue the practice of breaking from the track and recording the sighting location(s) as the closest approach during circling. That data format is available for use by other aerial survey teams, but to date no other team has used it.

### 3.4. Opportunistic Sightings

Opportunistic sighting records include sighting data only, and sometimes limited environmental information. During CETAP there was a differentiation made between opportunistic sightings that were collected during the study and "historical" sightings that had been collected previously and were incorporated into the CETAP database, however there is no real difference in the nature of the data other than the timing. Since there is no survey trackline information associated with opportunistic sighting records, effort and SPUE indices can not be computed. It is therefore impossible to determine whether or not there may be spatiotemporal biases, caused by the pattern of survey effort, in the sighting distributions derived from opportunistic sightings or from all sightings combined.

Some opportunistic sightings were derived from formal, systematic sighting programs, while others came from less-organized sources. Some of the historical information included during CETAP came from a previous attempt by the U.S. Coast Guard to record sightings of marine species from Coast Guard ships at sea. Another example is the Anti-Submarine Warfare Environmental Prediction Service (ASWEPS), a U.S. Navy aerial sampling program to collect sea-surface temperature data and other oceanographic observations—prior to the existence of Earth-orbiting satellites with remote-sensing capability. Observers aboard some ASWEPS platforms during the mid-1960s recorded marine mammal and turtle sightings. At times, there have been opportunistic sighting datasets generated during line-transect or other directed surveys. Data from NEAQ surveys in the Bay of Fundy from 1980 through 1985, prior to the Consortium, were never computerized in the appropriate format. For CCS in some years, absence of available funding precluded creation of anything more than basic sighting tables. Sometimes contributors have declined to submit full survey data, and at times errors or gaps in data (*e.g.*, missing one or more fields from the minimum needed for effort reconstruction) forced treating a dataset from a particular survey as opportunistic sightings only.

There is one other source of opportunistic sightings that is unique to right whales. After each time that the NEAQ photoID catalog is cross-referenced with the database at URI, there remains some number of catalog records that do not match any sighting in the

database. After eliminating any that resulted from surveys where the full survey data are likely to be contributed to the database at some future time, the rest of the unmatched catalog records are extracted and added to the database as a set of opportunistic right whale sightings.

# 4. COLLECTING, RECORDING, AND ENTERING DATA

*To err is human; to really foul things up it takes a computer.*

In the beginning, there was only paper and pencil. And the paper was formless, and often did the pencil create error and chaos. Then Gates said "Let there be personal computers, with expensive electronics, and programming therefor." And he called the electronics "hardware" and the programming "software," and he separated the hardware from the software. Then the hardware and the software did multiply, and the number of the errors and the depth of the chaos became limitless. The engineers of the hardware and of the software did cast blame upon each other for the errors and chaos, and both did heap scorn upon the users. And Gates saw that it was good, and he said "We should be shipping the upgrade early next spring."

The process of recording survey data in the field and getting those data into a useful computer dataset has evolved tremendously since the CETAP study began in October of 1978. As an example, the following describes the steps in the process for a CETAP line-transect aerial survey (knowing something about how the process evolved helps to understand why some things are the way they are). The survey crew in the Beechcraft AT-11 consisted of six individuals—a pilot, a navigator, and four observers. Two observers at a time were on watch in the forward observation bubble, and one was off-duty in the rear of the plane (though usually watching out a rear window on the side with the best viewing conditions at the time). The fourth observer sat in the co-pilot's seat with a clipboard in his or her lap and served as the data recorder. The navigator sat in the back, where there was a duplicate set of instruments, including clock, LORAN-C receiver, compass, and altimeter, as well as a strip-chart recorder for the radiometer that sensed sea-surface temperature.

- The navigator recorded time and latitude/longitude on a paper log at 5-minute intervals, whenever the pilot reported hitting a waypoint at the start or end of line, and whenever an observer reported a sighting or other occurrence. At times, the navigator would be unable to keep up, *e.g*., in the mid-Atlantic in summer when there were often numerous sightings in rapid succession.

- Periodically, the navigator pressed a button on the radiometer recorder to put an index mark on the strip chart, and wrote the time on the chart.

- The data recorder kept a clipboard with survey logs, on which were recorded heading, altitude, track types, leg numbers, weather conditions, etc., as well as all sighting information. Sightings in rapid succession caused as many or more problems for the data recorder as for the navigator.

- The observers maintained photo logs for the hand-held cameras—one sheet per roll of film and one clipboard of sheets per camera. There were also two fixed cameras mounted in the belly of the aircraft (often with different film types) and a third one mounted inside over the navigator's shoulder aiming at the navigation instrument panel. All three were fired simultaneously by pressing a button in the forward bubble. The photo of the nav panel provided the time, position, altitude, and heading for each pair of vertical photos, making a detailed photo log unnecessary; the data recorder merely noted that vertical photos were taken.

- At the end of a survey, the data records included the navigator's log, the survey & sighting log, the radiometer strip chart, and the photo logs, as well as all the photographs after the film eventually was developed, which could take weeks in the case of bulk rolls in the vertical cameras and some of the hand-helds.

- At some time later, two people would sit down to transcribe the data, with all of the logs and photos in hand. The transcribers could be the observers who had collected the data, other observers, or research assistants—both graduate and undergraduate students. The process involved writing numbers in little boxes on 8.5 x 14-inch IBM coding forms, using a detailed transcriber's manual (included as an appendix to Owen and Kenney, 1983) for instructions and lists of codes. Each form was limited to 80 columns, therefore there were five different forms, or "record types." With the exception of Record Type I, all of them were standard 24-row x 80-column grids, with 18 rows for data entry and the top 6 used for variable headings.

  - Record Type I: Header. This form had only a single data-entry row, assigning FILEID, PLATFORM, DDSOURCE, and IDSOURCE (see section 8 for

details on variables) for an entire dataset (one aerial survey day, one to many days of shipboard survey, or one to many opportunistic sightings).

- Record Type II: Location/Environment. On this form was transcribed all of the information related to the survey track and environmental conditions. The first two fields were FILEID and EVENTNO. Event numbers were assigned sequentially. Other fields were *CIRCLFLG*[*], YEAR, MONTH, DAY, TIME, LATDEG, LATMIN, LONGDEG, LONGMIN, LEGTYPE, LEGSTAGE, LEGNO, LEGGOOD, ALT, VISIBLTY, CLOUD, BEAUFORT, *WINDDIR*, DEPTH, SURFTEMP, *WATCOLOR*, *ATEMP*, HUMANACT, DEBRIS, *GLAREAMT*, *GLARELOC*, PORTOBS, and STAROBS. In addition, original data could be recorded in a variety of units, so there were several other columns with unit flags for TIME (time zone), ALT (ft or m), DEPTH (fathoms, ft, or m), SURFTEMP (C or F), and ATEMP (C or F).

- Record Type III: Data I. Basic sighting data were recorded on this form. The first two fields were FILEID and EVENTNO, so that each sighting could be matched to a specific line on Record Type II. The third field was SIGHTNO, which was assigned sequentially within a survey, with one exception. The funding agency (BLM) allowed recording of non-target sightings (seals, sharks, ocean sunfish, etc.) as long as no contract resources were used. To make those sightings easy to delete before any analysis, they were assigned a SIGHTNO of 999. Other columns were for entering *SPECCHAR*[‡], IDREL, NUMBER, CONFIDNC, STRIP, PHOTOS, *FLUKES*, *LINKAGE*, ANHEAD, *RELBAR*, *SIDIST*, *DATAMETH*, BEHAV1–BEHAV15, *REPEAT*, SIGHTOBS, and *TRACKNUM*.

- Record Type IV: Data II. This form was for recording additional, entirely optional, sighting details. The first three fields were FILEID, EVENTNO, and

---

[*] Italicized names represent variables that are no longer included in the NARWC database. In most cases they were dropped completely. For GLAREAMT (glare amount) and GLARELOC (glare location), when the NARWC database was initiated, we decided that coding sun glare separately on the two sides of the survey track (*i.e.*, GLAREL and GLARER) was a much more effective design.

[‡] CETAP used two-letter codes for species (*e.g.*, humpback whale = BG), which were converted into two-digit numeric codes (humpback whale = 11) in the database. Neither system was intuitive or possible to remember. When the NARWC system began, we shifted to four-letter codes (SPECCODE) based on the common name (humpback whale = HUWH).

SIGHTNO—so that each line was related to a specific sighting on Record Type III. The other fields were: *DECOMP*, *TAGTYPE*, *TAGLOC*, *TAGMAT*, *TAGNUM*, *TMSOURCE*, *CARALEN*, *CARAWID*, *CARAWT*, *MAMID*, *NUMADULT*, *NUMMALE*, *NUMFEMAL*, *NUMSUBAD*, *NUMIMMAT*, NUMCALF, *DISTSHOR*, and *DIVETIME*. Record Type 4 was not allowed to be used for any sightings with SIGHTNO = 999.

- Record Type V: Data III. The final form was also to enter additional sighting details, and was also prohibited for "999" sightings. It was added after the project was well underway, primarily because of particular interests of the aerial observer crew, with many of the data relating to the exact location of the sighting instead of the location of the platform at the time of the sighting. Again, the first three variables were FILEID, EVENTNO, and SIGHTNO. The other variables were: *ALATDEG*, *ALATMIN*, *ALONGDEG*, *ALONGMIN*, *ADEPTH*, *AWTEMP*, *GROUPS*, *SIZEGRP*, and *STRUCTURE*.

- The entire stack of coding forms with the transcribed data was submitted for "key-punching." Each line of each form was converted into a separate, individual punch-card.

- The deck of punch-cards was then fed through a card-reader, which converted the holes in the cards to alphanumeric characters in a computer file. The data were then converted into a SAS dataset, at which time all variables were converted to standard units (*e.g.*, Eastern Standard Time, altitudes and depths in meters, temperatures in Celsius).

- There was a Quality-Control step, in which a printed hard copy of the data file was visually scanned for errors against the transcribed coding forms and the original field logs.

- Finally, the survey file was added to the database, which was archived on magnetic tapes in the Academic Computer Center on the URI Kingston Campus.

For POP surveys the process was essentially the same, although slightly simplified. A POP observer, most often working solo, kept a single log for both navigation information and all other survey and sighting data. Beyond that, the steps in

18

the data-entry process were the same. It is relatively easy to see all of the places where errors could be introduced into the final data. In addition, the SAS software only ran on the large IBM mainframe computer on the URI Kingston Campus. All of the SAS jobs for data entry, database management, or data analysis were accomplished from the Bay Campus by either:

- key-punching the program instructions onto a deck of IBM cards and sending those through the card reader to the mainframe, or

- writing and saving a program file on a PR1ME-750 minicomputer, which could then be submitted to the mainframe. The minicomputer was located in the GSO computer center in the Pell Library basement, and was linked by a dedicated remote line to the mainframe. It could be accessed from terminals in the computer center, or by remote terminal via telephone using dial-up acoustic modems. Printed output at first had to be picked up in Kingston, but later hardware modifications allowed the output to be directed to a printer at GSO.

When the Consortium research program was being designed and initiated in 1986, GSO was selected to house and manage the database because of the resident experience remaining from CETAP. By that time, the system had become somewhat more user-friendly. Personal computers were becoming commonplace. In addition to their growing stand-alone functions, they were replacing simple terminals for access to the mainframe and minicomputer. SAS programs could be written, modified, and saved on a PC, but SAS still only ran on the mainframe system. So a SAS program to enter a dataset would be submitted from the PC to the mainframe, using the GSO minicomputer as the remote access point to the mainframe.

Data were still being recorded in the field manually on paper forms, so there remained the necessity of entering data into computerized files. There were two options. One was for all of the NARWC collaborators to submit their logs to GSO, where all of the data would be computerized. The other was for all of the collaborators to computerize their own data and submit computer files to GSO. The second option was more attractive, however a significant difficulty was that none of the other collaborators had access to either a mainframe computer or the SAS software. The solution was to develop a way for

collaborators to computerize their own data using easily available PC software. Those datasets would then be converted into SAS datasets at GSO. We created a set of interactive programs in dBASE-3+, designed to assist collaborators in entering survey data from hand-written field logs into standardized computer datasets. The dBASE programs automatically coded and formatted some variables, presented menus of code selections for others where the number of options was short, directed users to printed lists of codes where the lists were longer (*e.g.*, species, behaviors), forced entry (*i.e.*, did not allow missing data) for some fields, and contained internal loops in the code that checked some variables for values within acceptable ranges.

The first computer systems to automatically log data in the field for any surveys associated with NARWC were developed by Timothy L. Flynn[*] and his staff from Aero-Marine Surveys, Inc. (AMSI), the main aerial survey contractor for CETAP, URI, SCOPEX, and several other projects. He used Hewlett-Packard HP-85 computer systems and custom programming developed in-house at AMSI. Connected to the computer were the outputs of the aircraft's LORAN-C navigation system and the radiometer (the latter through a digital voltmeter to convert the analog output to a numeric value, which the computer then converted to a calibrated temperature value). He had two different programs, depending on which aircraft was being used. In the AT-11 there was enough room so that one of the two off-duty observers could sit at the computer keyboard and serve as the primary data recorder (the other still kept a written log on a clipboard as back-up). In that case, the program allowed recording of full survey data. After pressing a key to mark an event, the program asked about the type of event (*e.g.*, start or end a line, change the environmental conditions, or enter a sighting), and then stepped through a series of fields and waited for the recorder to respond. This system was in use as early as the Southeast Turtle Surveys (SETS) in 1982–1984. The other AMSI aircraft used for surveys, a Cessna 337 Skymaster, was much smaller and carried only four crewmembers. The data recorder (using a clipboard and paper logs) sat in the copilot's seat, the two observers sat in the rear seats, and the computer was installed behind the observers. Although the observers could access the computer keyboard (very awkwardly) to start and stop the program and occasionally hit a key to mark a specific event, it was not

---

[*] now deceased.

20

possible to continuously enter data and maintain an effective watch at the same time. The computer ran continuously, recording time, latitude/longitude, and sea temperature at frequent intervals. The recorder wrote down everything, including times and positions, for all events, but did not need to record routine periodic locations. This system was used as early as 1985 for URI Great South Channel surveys. In both the AT-11 and Skymaster, the computer had an internal printer, which printed each event as it was logged on a roll of 4.25-inch thermal paper, which served as back-up in the event of a computer crash.

At the present time, probably no survey data are being recorded entirely by hand on paper. Navigation by LORAN-C has been replaced by the Global Positioning System (GPS). It seems that a variety of recording methods are in use, and it is not always obvious from the data themselves how a particular survey dataset was recorded. Most survey data are automatically logged in the field using a laptop or tablet computer linked to a GPS receiver. It is a concern that there is no hard-copy back-up in the event of a computer or GPS malfunction, or that can be referred to for questions that may arise afterwards. For example, when a data analysis several years after a survey shows a false killer whale sighting in the northern Gulf of Maine, without written field logs to check—how can we be sure it was really a false killer whale, rather than somebody typing "FKWH" instead of "FIWH"? Many contributors are using a data-logging program called LOGGER that was developed by the International Fund for Animal Welfare, and others are using a newer system called Mysticetus. Some surveys appear to use a GPS system with its own data-logging capability, with either a manual log, a separate computer system, or a voice recorder to record everything other than routine location data. At least one survey appeared to use separate GPS units for two observers on opposite sides, and then merged the outputs afterwards by time (which can be problematic when the times were rounded to whole minutes).

Regardless of how the data are recorded in the field, there are certain minimum requirements. These will differ by platform type and by survey type. Additional details, data requirements, comments, and cautions about particular variables can be found in section 8.

## 4.1. Aerial line-transect surveys

Since line-transect surveys are designed for use in calculating statistically rigorous density and abundance estimates, the data-recording requirements are equally rigorous. Only sightings made by the dedicated observers during standardized, pre-defined census tracks can be used in the density estimates, therefore it is necessary to clearly differentiate those sightings in the data. Different track types must be identified, as must sightings made by anyone other than the defined, dedicated observers. Note that all of this discussion applies equally to shipboard and aerial line-transect surveys, even though the NARWC database is not presently configured to include the former (the necessary codes have never been created). The following discussion is an example of data from the hypothetical survey illustrated in Figure 2.
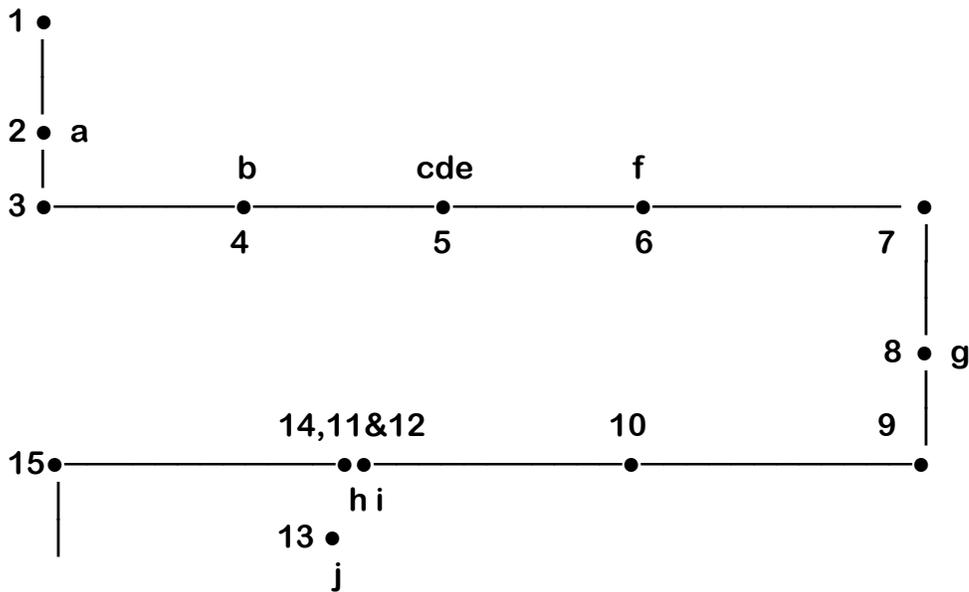


*Figure 2. Representation of a segment of a hypothetical aerial line-transect survey to illustrate the sequence of data records, as detailed in the text following. The black dots and numbers represent events, while the letters represent sightings.*

| EVENTNO | SIGHTNO | DISCUSSION |
|---|---|---|
| 1 | – | Begin a transit. For line-transect surveys, information about observers on watch is recorded only for census legs. Consequently, in data analyses it is presumed that if events are recorded during transits, then somebody is on watch (modified for the NLPSC surveys in 2011; see 8.64 LEGTYPE). LEGTYPE = 1; LEGSTAGE = – (missing). |
| 2 | 1 | Sighting of group **a** made during the transit. Since the sighting was not during a census leg, it makes no difference whether it was by an on-duty observer, off-duty observer, or airplane crewmember. LEGTYPE = 1, LEGSTAGE = –. |
| 3 | – | Begin the first survey line. LEGTYPE = 2, LEGSTAGE = 1. Note that it is not necessary to duplicate the time and position more than once in the dataset to simultaneously mark the end of the transit and the beginning of the census leg, although doing so does not cause a problem (although they <u>must</u> have unique event numbers in the proper logical sequence). |
| 4 | 2 | On-effort sighting of group **b**. For all on-effort sightings, it is necessary to measure the right-angle distance between the trackline and the group. During CETAP this was accomplished by waiting until the sighting was directly abeam of the aircraft, and then classifying the distance into an interval or strip, using marks placed on the glass of the bubble in the AT-11 or the wing struts of the Skymaster (Kenney and Scott, 1981). It could also be measured directly by, *e.g*., clinometer or laser range-finder. Another option is to measure radial distance and bearing immediately and to calculate right-angle distance later. One's own data format can be set up to handle whichever method is being used, since the distance data for most surveys typically would not |

be incorporated into the NARWC database. LEGTYPE = 2; LEGSTAGE = 2.

| | | |
|---|---|---|
| 5 | 3 | On-effort sighting of group **c**. LEGTYPE = 2, LEGSTAGE = 2. |
| 5 | 4 | On-effort sighting of group **d**, at the same location as groups **c** and **e**. They may or may not be associated or in a mixed group. There is no limit to the number of sightings that can be assigned to a single event, and the sightings do not need to be of different species. It is possible for each observer to see an individual or group of the same species on each side of the track at the same time, or even for one observer to see two clearly different animals or groups on the same side but at different distances. With manual log-keeping, it is sometimes necessary to record sightings that occur close together in time at the same event. In the days of data entry using punch-cards, there would be only one Record Type II and any number of associated Record Type IIIs. In the database, and in submitted computer data, there are complete records for each sighting, with duplicate event numbers but different sighting numbers. All of the survey and environmental data MUST match exactly in all the records with the same event number (the only differences should be in the sightings), or else the error-checking software will flag them as incorrect. LEGTYPE = 2; LEGSTAGE = 2. |
| 5 | 5 | On-effort sighting of group **e**, at the same location as groups **c** and **d**. LEGTYPE = 2; LEGSTAGE = 2. |
| 6 | 6 | Sighting of group f during a census track, but the sighting was made by the pilot, so it can not be included in a density estimate, therefore it is assigned a different LEGSTAGE. Standard practice during line-transect surveys is to instruct |

everyone other than the on-duty observers to say nothing about a sighting until it passes behind the wing and has clearly been missed by the on-duty observers. In fact, it is good practice for crewmembers to not do anything that might cue an observer (like nudging the co-pilot and pointing out the window, even without saying anything). LEGTYPE = 2; LEGSTAGE = 6.

| | | |
|---|---|---|
| 7 | – | End the survey line and start turning on the cross-leg. LEGTYPE = 2; LEGSTAGE = 5. |
| 8 | 7 | Sighting of group **g** during the cross-leg, regardless of who made the sighting. LEGTYPE = 3; LEGSTAGE = –. |
| 9 | – | Begin the second survey line. NOTE: There should not be sightings recorded at same event as beginning or ending a line. If they are so close together in time as to require using the same location, they should still be at separate events with different event numbers. It is unlikely that they occurred at the same instant, so you should still be able to determine which to put first in the order of the data records. LEGTYPE = 2; LEGSTAGE = 1. |
| 10 | – | Routine event along a trackline, or perhaps a forced event to change the sea state or other environmental parameter. Of course, with computer-logging of GPS data, there would be <u>many</u> more of these events. |
| 11 | 8 | On-effort sighting of group **h** (with group **i**). The aircraft broke from the track to circle the sightings immediately after the observers marked them. The sighting event must come before the break-off event in order for the density estimation to work correctly. NOTE: each line-transect survey should have a predetermined definition of "on-effort sighting" to use in these situations. The CETAP protocol was to count with the original group any other individuals of the same |

species seen while circling that could reasonably be called "associated" with them. A second un-associated group of the same species seen during circling was also included in the count. Third and subsequent groups, or groups of other species that were not originally seen from the trackline, were recorded as new, off-effort sightings. An exception to the last was when a mixed-species group was sighted from the track; then all species in the group were defined to be on effort. LEGTYPE = 2; LEGSTAGE = 2.

| | | |
|---|---|---|
| 11 | 9 | On-effort sighting of group **i** (with group **h**). LEGTYPE = 2; LEGSTAGE = 2. |
| 12 | – | Break off from the census line to circle the sightings. LEGTYPE = 2; LEGSTAGE = 3. |
| 13 | 10 | Off-effort sighting of group **j** while circling. In this example this is the only event recorded during the circling time, which is typical of manually recorded survey data, sometimes even when the circling went on for 30 minutes or an hour while taking photos of a group of right whales. With computer-logged data there are many routine locations during the circling, so that a plot of the survey track shows a pattern of mostly straight lines, periodically interrupted by little piles of spaghetti. LEGTYPE = 4; LEGSTAGE = –. NOTE: it can be difficult to keep track of other animals in the vicinity while circling for an extended period. Surprisingly, it's actually easier with manual data-logging. When the observer says "Take a mark; mother-calf fin whale," it can be easy for the data recorder to check the log to see that it was the same pair marked 10 minutes prior. With computer data, it may be necessary to plot sightings afterwards to remove duplicates. |

| 14 | – | Resume the census track. This should occur as close as possible to where the break-off point was. LEGTYPE = 2; LEGSTAGE = 4. NOTE: as with LEGSTAGEs 1 and 5, there should not be sightings at LEGSTAGE 3 and 4 events. |
| 15 | – | End the second survey line. |

## 4.2. POP aerial and shipboard surveys

In contrast with line-transect surveys, in POP survey data there is no distinction between different types of flight or cruise tracks—even if there are predefined track lines, they are treated exactly the same as transits, cross-legs, circling, or any other deviation from the track. That makes it much simpler to record the data, but also creates problems if the organization conducting the survey expects to be able to easily pick out the tracklines from everything else after the data have been incorporated into the NARWC database. Each data contributor therefore has to make their own decision about using the line-transect or POP format. The intermediate POP aerial survey format developed in 2015 for CCS (Section 3.3) does allow for differentiating between survey leg types and could be used by other contributors should they desire.

There are not many differences between POP aerial survey data and POP shipboard survey data. They are easily differentiated by the LEGTYPE codes (as well as by the first character of the FILEID). For shipboard surveys LEGTYPE is either 5 (vessel underway) or 6 (vessel not underway). I have never really been convinced of any utility for maintaining that distinction, *e.g.*, which are the appropriate codes for a vessel drifting with a 2-knot current versus one holding an exact position using computer-controlled thrusters? I see no problem with using "5" all the time. For typical POP aerial surveys, the value <u>is</u> constant, but differs between surveys that are recording all sightings and surveys that are recording only some subset of species because of personnel limitations (see the LEGTYPE discussion in Section 8.64).

There is still a need to carefully record watches using the begin-watch (1), on-watch (2), and end-watch (5) LEGSTAGE codes to maximize the data's utility for later analyses of effort and SPUE. For POP surveys, the definitions of on-effort and off-effort

change from what they were in line-transect surveys. On-effort for POP surveys is the equivalent of on-watch. When conducting a SPUE analysis and combining the two data types, line-transect sightings that were off-effort for the purpose of estimating density become on-effort for calculating SPUE. There is no longer any need in POP data to identify breaking off from a track to investigate a sighting or going back to the track afterwards, or to flag sightings not made by an on-duty observer. If the same example trackline shown in Figure 1 represented a POP survey that had done exactly the same things, the ways the data would differ from the description above include:

- LEGTYPE would be the same for every event—5 if it was a shipboard survey and 7 or 9 if it was an aerial survey (see LEGTYPE in Section 8.64 for details).
- LEGSTAGE would be 1 (begin watch) at event 1, 5 (end watch) at event 15, and 2 (continue watch) at every other event.
- Event 12 would not be there at all.
- For the intermediate format (CCS) POP aerial surveys, the LEGTYPE and LEGSTAGE codes would be exactly the same as for the line-transect surveys. Where there might be differences would come with sightings **h** and **i,** which could be recorded after the break-off point rather than before, just like sighting **j**, and a LEGSTAGE of 6 for sighting 6 would not be necessary. Measuring the right-angle distance to sightings would no longer be required, and the objective would change to marking the best possible location for the animal or group.

There are other minor differences between aerial and shipboard survey data. Altitude has no meaning for a vessel, so it is a missing value. (NOTE: The database is designed to incorporate everything into a single dataset rather than separate datasets with their own formats for shipboard effort, aerial effort, survey environmental parameters, shipboard sightings, aerial sightings, etc., as has been done by some data contributors. Consequently, every data record includes every data field, even though many are missing values. One of the real benefits of using SAS software is that missing values are handled much better than, *e.g*., by dBASE or Excel.) Past practice has been not to record sun glare during shipboard surveys, although it easily could be done.

# 5. SUBMITTING DATA

The downside of the greatly increased power that computers have provided for data handling is that computers are even more obsessive-compulsive about data formats than are database managers. Things need to be done precisely according to the protocols and instructions, or the system breaks down.

One of the upsides now is that software formats have become more and more flexible. At one time, reading a dBASE file with another company's spreadsheet software would have been inconceivable. Today things have gotten somewhat easier (although Microsoft did one of their typical annoying things and removed the functionality from Excel to export or import *dbf* files between the MS-Office 2003 version and MS-Office 2007). As in the beginning, the database is archived in SAS, and dBASE is used as an intermediate. But data can be submitted in Excel, Access, or text files. The key is to be careful about following the formats shown here. In addition, don't mix formats between files in a single data submission. It's also a good thing not to change formats from one year to the next (at least without notice). If something worked last year, stick with it.

Among the most important considerations in a submitted dataset are the variable names used. For a spreadsheet program (*e.g.*, Excel), the most important thing is the column location in a dataset, but for a true database management program like dBASE or SAS, the location of a field is unimportant, because both programs work by the names of the variables. That means the variable names (the labels at the top of the Excel columns) must be exact. Computers are idiot savants—they think that VISIBLTY and VISIBILTY and VISIBLITY and VISIBILITY are four different variables. The spelling must be exact, even if the name is technically misspelled. You might think that it would be easier to use much longer and more descriptive names, like Access or Excel would allow. But SAS limits variable names to 8 characters and dBASE limits them to 10. Neither system allows spaces or parentheses or any special characters (except an underscore) in a variable name. One thing that isn't important is capitalization—dBASE converts names to all capitals, and SAS doesn't even care.

Here is an example of what can go wrong. A dataset submitted in Excel is first saved as a dBASE dbf file (now using an old computer running Windows XP and MS-

Office 2003). The column headers in Excel become the variable names in dBASE. Working in dBASE, the initial step is to create an empty data structure in a predefined format, which includes all of the variable names. Then the data are appended to the empty structure from the other file. dBASE doesn't care that the variables are in different columns; it just looks for the column with the same name. If the structure is expecting to find location fields called LAT and LONG because that's what this contributor called them the last three years, but now they're called LATITUDE and LONGITUDE or LATDD and LONGDD, the result is an entire survey with no location data.

Finally, we need to point out that Excel has some really annoying quirks. Submitting YEAR, MONTH, and DAY combined in a single date-formatted field is fine. dBASE also supports date formats and has functions that can easily create the three separate fields. But Excel's time format does not translate directly into dBASE, and will either give you the date again, or the time as a fraction of 24 hours, or something even more incomprehensible (the date <u>and</u> time as the number of days and fractional parts of days since midnight on 1 January 1900). And it insists on AM and PM instead of a 24-hour clock. TIME **must** be submitted strictly as 6-digit numbers in 24-hr format without any colons. And that cannot be done by just defining a custom cell format, because that only changes what you see; the underlying data are still the same.

For numeric data, Excel defaults to "general" format rather than "numeric." General fields are truncated to only the integer part when Excel does the "save as" thing to a dBASE file, stripping off all the decimals (although Microsoft removed that function beginning with Office 2007, so I can keep doing it only so long as my old Windows XP computer with Office 2003 keeps running). So if you've carefully logged 8 hours of an aerial survey to a ridiculous 10 digits of precision in the latitude and longitude, all of the data are reduced to only the degrees. It's easy to spot "general" data in Excel, because the decimal points don't all line up and the decimal parts of numbers never end in zero. All you need to do is force the cell format to numeric with 5 decimal places (1 is sufficient for HEADING, VISIBLTY, and SURFTEMP) before submitting the data, and all will be hunky-dory. Another Excel quirk is that if there is a space before a column header (which will be entirely invisible), the dBASE variable name will have an underscore where the space was, hence be a new name that won't work at all.

30

A format that more and more contributors have been using, and that is more forgiving in some ways than Excel spreadsheets is comma-delimited text (.CSV files). They can be opened in Excel, Notepad, or other word-processing software, and can be directly imported into dBASE. The issue with "general" vs. "numeric" data goes away. Since the field names at the tops of the columns are not really variable names when imported into dBASE, it makes no difference how bad a speller you may be. You don't really even need variable names in the column headers (they are <u>not</u> there when one exports a CSV from dBASE). What is critically important, however, is the order of the columns. They must be consistently in the same columns and the same order, with no other columns in there. If, for example, you submit one file in a set of 30 for a given season where you added a new field somewhere in the middle, when read into dBASE everything from that field and to its right in the dataset will be read into the wrong variables. It is OK to stick in your own unique variables in one or more of your datasets; just put them on the right side, beyond the last column that is part of the standard submission.

# 6. DATA QUALITY-CONTROL AND ARCHIVAL

Every file added to the NARWC database is first run through an intensive quality-control process, involving testing for errors, making corrections, and communicating with the contributor when questions arise and to provide feedback for improvement. The SAS error-checking routines and subroutines have been designed and modified over the years to ensure as much as possible that the data reliably and accurately represent the surveys and sightings. The SAS macros are updated and improved on an on-going basis. As new error types are discovered, the error-checking routines have evolved to keep up:

- All coded variables are checked to make sure that only allowable values of codes are used.

- All continuous numeric variables are checked to make sure they fall within reasonable ranges. Hour and minute parts of times cannot exceed 59; minutes of latitude or longitude cannot exceed 59.99; month cannot be higher than 12; day cannot be higher than 31; year cannot be higher than the current year; aircraft or ship headings cannot exceed 359. There are filters for allowable ranges of latitude and longitude that sometimes need to be adjusted (*i.e.*, for some NMFS surveys that extended to the eastern Scotian Shelf), and other filters with built-in flexibility (*i.e.*, the thresholds that flag possible erroneously high values for numbers of individuals at a sighting are quite different for sea turtles, large whales, dolphins, or cow-nosed rays).

- There are steps that compare values between successive records, or between fields. The year value embedded in the FILEID has to match the YEAR field in every record, and YEAR, MONTH, and DAY from each record in a survey dataset need to match those in the previous record (causing occasional incorrect error flags in multi-day cruise datasets and for the first record in a file, where the previous values don't exist). TIMEs at successive events must occur in proper sequence. The distance covered, elapsed time, and speed between all successive pairs of events are computed, and excessively high or low speeds are flagged (the low-speed test was added most recently and is still being tweaked). This has proven over time to be the most reliable way of detecting errors in recording or

transcribing time and position data during surveys (and now for detecting GPS output errors). Another step compares successive altitude values to compute the rate of climb or descent and check the value against the aircraft specifications, because there have been numerous problems with automated data-logging of altitude values captured from GPS outputs.

- There are steps that insure that the logical sequence of events is followed. Events must increment in sequence, and there cannot be duplicate event numbers except under the restricted circumstances of multiple sightings at the same time and position. There can never be duplicate sighting numbers. A survey crew cannot begin watch if they are already on watch, end watch if they are already off watch, or be on watch if they have not begun one.

- There is one nagging error type that is not possible to detect using analytical methods within SAS. Location errors in opportunistic sighting data cannot be detected by anomalous survey speeds, since by definition there are no associated survey data. Some errors can only be detected by mapping the sightings and looking for ones that plot on land (the clearly obvious errors) or far from other sightings (the less obvious errors that may actually not be errors at all).

The database is archived as a single SAS dataset on the desktop computer in the GSO marine mammal lab. Once or twice a year the database is updated by added all of the new files that have been completed through Q/C processing, at the same time making any corrections to errors noted since the previous update. At least one back-up copy is created on the same computer, and other back-ups are made at the same time (on CD [compressed], DVD, back-up hard drive, flash drive, copied to other computers, etc.). Copies are stored in the GSO lab and off site. About once a year, usually with the fall progress report, a copy is sent to NMFS in Woods Hole.

# 7. DATA SHARING

The following is quoted from the NARWC data-sharing protocol, which is posted on the Consortium web page (https://www.narwc.org/accessing-narwc-data.html):

"Since the materials in these various datasets come from numerous independent individuals and institutions, these datasets are not strictly proprietary. Rather, they represent a scientific resource, and access to the data for scientific, educational, conservation and management purposes is encouraged. Contributors to the datasets are given first access to the data in recognition of their contribution, and contributors have full and unrestricted access to use of their own data. The level of an individual or organization's contribution to the data may weight their rights to, and use of, the full database. After that, proposals for data access from scientists, managers, students or other individuals with a bona fide purpose will be reviewed by the Consortium Board members. Given the great effort required to collect the available data, the Consortium and the curators of the data have an obligation to protect the rights of contributors by placing certain restrictions or conditions upon access to, and use of, the materials within it."

The process of requesting use of NARWC data is actually quite simple. A prospective user submits a brief proposal to the Consortium stating the purpose which the data are to be used for, following the instructions in the protocol document. The proposal goes to the NARWC Board (through Heather Pettis, New England Aquarium, hpettis@neaq.org). The proposal should clearly state whether the work to be undertaken is for publication (peer-reviewed journal article, thesis/dissertation, book, magazine article, conference presentation, etc.) or for management purposes (environmental impact statement or other NEPA document, Take Reduction Team, NMFS research planning, etc.). Publication proposals are reviewed by three Board members; management proposals are not reviewed but are simply passed on to the relevant data curator.

Proposals that clearly state exactly what data are needed (species—remembering that there are **many** more other than right whales, geographic scope, time span, data fields) are much easier to respond to without multiple rounds of questions-and-answers. Approval may often include the requirement for adding one or more Consortium scientists to the authorship of the proposed publication—either representing the organization(s) responsible for collecting the bulk of the data used or because of the amount of analytical effort involved in providing particular data (*e.g.*, SPUE datasets). Users also agree to provide copies of resulting documents for NARWC files.

Proposals may be rejected for a couple of reasons (see the protocol for more detail). One is duplication of effort, since the agreement with a user specifies that we won't provide the data to someone else for another identical project for a reasonable time period. Another is that the proposal does not clearly define what is to be done. Finally, the agreement that is signed by all users specifies that any NARWC data provided are not to be shared with third parties. Consequently, we will reject proposals that include, *e.g.*, including the data in a publicly accessible on-line database.

# 8. DATABASE VARIABLES AND CODES

This section lists all of the variables presently contained in the North Atlantic Right Whale Consortium database, plus all of the variables that were formerly included in the CETAP database before it was streamlined to create the original NARWC database. The latter are being included here both for completeness and to illustrate database solutions to issues that may arise again for the Consortium or an individual collaborator. For each variable, there is a detailed description of the variable, comments on its history, if applicable, comments on the correct usage of the variable, cautions about problems that have been known to occur, and complete listings of codes for all coded variables. Variables that have renamed since their original creation have been listed here under their current name, with cross-references for older names. Also included in the list are a number of variables that are not actually included in any version of the database, but for which SAS macros exist to create them on demand for inclusion in output datasets (*i.e*., "synthetic variables"). Additional macros are relatively easy to create when the need arises to respond to a data request, and then they are generally saved for future use. The code listings for current variables also include obsolete or discontinued values that **cannot** be used in data submissions, but still exist in some older data.

Table 1 is an alphabetical list of all variables, including variable type (numeric or character), format, and label. For a numeric variable, the format **N.** indicates integer values and **N.n** indicates fractional values, where **N** is the field width (number of bytes), counting the decimal point for fractional values, and **n** is the number of decimal places. By SAS convention, character format is designated by a leading **$** and a number showing the field width. For NARWC variables, the format is the currently used one, which may differ from the original format. For CETAP variables, the formats may differ slightly between different versions of the database. SAS limits variable names to eight characters, but allows for the assignment of longer (up to 40 characters) descriptive labels than can be used with or in place of the variable name. Following the table are the complete variable descriptions, with the full listings of codes. Given past experience, the time between a new version of this users' guide being finalized and the first changes to one or more variables that make something in the following pages obsolete or incomplete is

more likely to be measured in days or weeks than in years. It is likely that this guide would be revised to reflect any such changes only at scattered intervals, the very short interval between the last two revisons notwithstanding. Changes in the interim have been and will be continue to be recorded in three places:

- Hand-written on pages of a data-management instruction manual maintained in a 3-ring binder near the computer at GSO, continuously as changes are made. That manual was originally designed in the event that someone ever needed to take over database management. Consequently, many of the sections of step-by-step instructions have become obsolete as computer hardware and software have changed. Only the code listings are presently being kept up to date, and occasionally reprinted and inserted into the binder.

- Similarly hand-written on a hard-copy of this users guide, also kept handy near the GSO database management computer. I attempt to do this simultaneously with the previously noted annotations in the instructions, but do not always manage to keep up.

- Digitally within the Word document for this users guide, working in the "track changes" mode so all changes can be immediately recognized. Those changes are usually made more slowly than the written ones in the paper copies, and finalized updates to the more formal NARWC Reference Document version even more infrequently. We have decided not to continue the prior practice of distributing shorter NARWC Reference Documents for changes to only one or a few variables. A copy of the most recent working version of the user's guide in Word can be provided on request, and is posted on the Consortium website at: https://www.narwc.org/sightings-database— click on the blue bar that says "SIGHTINGS DATABASE USERS GUIDE").

**Table 1**. Master list of all NARWC and CETAP database variables. See pages following the table for complete descriptions and details. Variables in **bold print** are those that currently exist in the NARWC database. Variables in ***bold italics*** are those that currently can be created from the NARWC database via a SAS macro.

| Variable | Type | Format | SAS Label* |
|----------|------|--------|-----------|
| ADEPTH | Numeric | 4. | DEPTH AT ANIMAL POSITION (M) |
| ALATDEG | Numeric | 2. | ANIMAL DEGREES LATITUDE |
| ALATMIN | N | 2. | ANIMAL MINUTES LATITUDE |
| ALATSEC | N | 2. | ANIMAL SECONDS LATITUDE |
| ALONDEG | N | 2. | ANIMAL DEGREES LONGITUDE |
| ALONMIN | N | 2. | ANIMAL MINUTES LONGITUDE |
| ALONSEC | N | 2. | ANIMAL SECONDS LONGITUDE |
| **ALT** | **Numeric** | **4.** | **Aircraft Altitude (m)** |
| **ANHEAD** | **Numeric** | **2.** | **Sighting Compass Heading Code** |
| ATEMP | Numeric | 4.1 | AIR TEMPERATURE (C) |
| AWTEMP | Numeric | 3. | WATER TEMPERATURE AT ANIMAL (C) |
| **BEAUFORT** | **Numeric** | **1.** | **Beaufort Sea State** |
| **BEHAV1** | **Numeric** | **2.** | **Behavior Code 1** |
| **BEHAV2** | **Numeric** | **2.** | **Behavior Code 2** |
| **BEHAV3** | **Numeric** | **2.** | **Behavior Code 3** |
| **BEHAV4** | **Numeric** | **2.** | **Behavior Code 4** |
| **BEHAV5** | **Numeric** | **2.** | **Behavior Code 5** |
| **BEHAV6** | **Numeric** | **2.** | **Behavior Code 6** |
| **BEHAV7** | **Numeric** | **2.** | **Behavior Code 7** |
| **BEHAV8** | **Numeric** | **2.** | **Behavior Code 8** |
| **BEHAV9** | **Numeric** | **2.** | **Behavior Code 9** |
| **BEHAV10** | **Numeric** | **2.** | **Behavior Code 10** |
| **BEHAV11** | **Numeric** | **2.** | **Behavior Code 11** |
| **BEHAV12** | **Numeric** | **2.** | **Behavior Code 12** |

**Table 1**. (continued)

| Variable | Type | Format | SAS Label* |
|---|---|---|---|
| **BEHAV13** | **Numeric** | **2.** | **Behavior Code 13** |
| **BEHAV14** | **Numeric** | **2.** | **Behavior Code 14** |
| **BEHAV15** | **Numeric** | **2.** | **Behavior Code 15** |
| **BLOCK** | **Character** | **$2.** | **Aerial Survey Block** |
| *CALF* | *Character* | *$3.* | |
| CALFHAT | N | 4. | NUMBER OF CALVES/HATCHLINGS |
| *CANADA* | *C* | *$1.* | |
| CARALEN | N | 3. | CARAPACE LENGTH (CM) |
| CARAWID | N | 3. | CARAPACE LENGTH (CM) |
| CARAWT | N | 6. | TURTLE WEIGHT (G) |
| CCOVER | N | 2. | CLOUD COVER (OKTAS) |
| CETSPPCD | C | $4. | CETAP SPECIES CODE |
| CIRCLFLG | N | 2. | CIRCLING TIME ACCURATE? |
| **CLOUD** | **Numeric** | **2.** | **Cloud Cover Code** |
| **CONFIDNC** | **Numeric** | **2.** | **Count Precision Code** |
| DATAMETH | N | 2. | METHOD OF DATA COLLECTION |
| DATED | N | 2. | DAY |
| DATEM | N | 2. | MONTH |
| DATEY | N | 2. | YEAR |
| **DAY** | **Numeric** | **2.** | **Calendar Day** |
| **DDSOURCE** | **Character** | **$3.** | **Direct Data Source** |
| *DEAD* | *C* | *$3.* | |
| DEBRIS | N | 2. | DEBRIS IN WATER |
| DECOMP | N | 2. | DECOMPOSITION |
| **DEPTH** | **Numeric** | **5.** | **Depth at Sighting (m)** |
| DISTSHOR | N | 2. | DISTANCE |
| DIVEMIN | N | 2. | DIVE TIME MINUTES |

**Table 1**. (continued)

___

| Variable | Type | Format | SAS Label* |
|---|---|---|---|
| DIVESEC | N | 2. | DIVE TIME SECONDS |
| DIVETIME | N | 4. | DIVE TIME |
| **EVENTNO** | **Numeric** | **5.** | **Event Number** |
| *FEED* | *C* | *$3.* | |
| **FILEID** | **Character** | **$8.** | **File Identification Number** |
| FLUKES | N | 2. | HUMPBACK FLUKE PHOTOS? |
| *GEAR* | *C* | *$3.* | |
| GLAREAMT | N | 2. | AMOUNT OF GLARE |
| **GLAREL** | **Numeric** | **1.** | **Left Glare Code** |
| GLARELOC | N | 2. | LOCATION OF GLARE |
| **GLARER** | **Numeric** | **1.** | **Right Glare Code** |
| GROUPS | N | 2. | NUMBER OF GROUPS |
| **HEADING** | **Numeric** | **3.** | **Platform Heading (true)** |
| HUMANACT | N | 2. | HUMAN ACTIVITIES |
| *HURT* | *C* | *$3.* | |
| *ID* | *C* | *$4.* | |
| **IDREL** | **Numeric** | **1.** | **Species ID Reliability Code** |
| **IDSOURCE** | **Character** | **$3.** | **Indirect Data Source** |
| *JDATE* | *N* | *3.* | |
| *JELL* | *C* | *$3.* | |
| **LATDEG** | **Numeric** | **2.** | **Latitude Degrees** |
| **LATMIN** | **Numeric** | **5.2** | **Latitude Minutes** |
| LATSEC | N | 2. | SECONDS LATITUDE |
| **LEGGOOD** | **Numeric** | **1.** | **Line-Transect Survey Line Made Good?** |
| **LEGNO** | **Numeric** | **3.** | **Survey Line Number** |
| **LEGSTAGE** | **Numeric** | **1.** | **Survey Watch Stage Code** |
| **LEGTYPE** | **Numeric** | **1.** | **Survey Line Type Code** |

**Table 1**. (continued)

| Variable | Type | Format | SAS Label* |
|---|---|---|---|
| LINKAGE | NUM | 2. | LINKAGE |
| **LONGDEG** | **Numeric** | **2.** | **Longitude Degrees** |
| **LONGMIN** | **Numeric** | **5.2** | **Longitude Minutes** |
| LONGSEC | N | 2. | SECONDS LONGITUDE |
| MAMID | N | 4 | CETACEAN IDENTIFICATION NUMBER |
| *MILL* | *C* | *$3.* | |
| **MONTH** | **Numeric** | **2.** | **Calendar Month** |
| NUMADULT | N | 4. | NUMBER OF ADULTS |
| **NUMBER** | **Numeric** | **5.** | **Number of Animals Sighted** |
| **NUMCALF** | **Numeric** | **3.** | **Number of Calves Sighted** |
| NUMFEMAL | N | 4. | NUMBER OF FEMALES |
| NUMIMMAT | N | 4. | NUMBER OF IMMATURES |
| NUMMALE | N | 4. | NUMBER OF MALES |
| NUMSUBAD | N | 4. | NUMBER OF SUBADULTS |
| **OLDVIZ** | **Numeric** | **1.** | **Old Visibility Code** |
| **PHOTOS** | **Numeric** | **1.** | **Photos Available Code** |
| **PLATFORM** | **Numeric** | **3.** | **Survey Platform ID Code** |
| *POOP* | *C* | *$3.* | |
| **PORTOBS** | **Character** | **$2.** | **Port Observer Code** |
| RELBAR | N | 3. | BEARING |
| REPEAT | N | 3. | REPEATED SIGHTING |
| *SAG* | *C* | *$3.* | |
| *SEASON* | *C* | *$6.* | |
| SIDIST | N | 2. | RADIAL DISTANCE (KM) |
| **SIGHTNO** | **Numeric** | **5.** | **Sighting Number** |
| **SIGHTOBS** | **Character** | **$2.** | **Sighting Observer Code** |
| SIZEGRP | N | 2. | GROUP SIZE |

42

**Table 1**. (continued)

_____

| Variable | Type | Format | SAS Label* |
|---|---|---|---|
| SPECCHAR | C | $2. | SPECIES CHARACTER CODE |
| **SPECCODE** | **Character** | **$4.** | **Species Code** |
| *SPECNAME* | *C* | *$61.* | |
| SPECNUM | N | 2. | SPECIES NUMERIC CODE |
| **STAROBS** | **Character** | **$2.** | **Starboard Observer Code** |
| **STRATUM** | **Character** | **$1.** | **Aerial Survey Stratum** |
| **STRIP** | **Numeric** | **2.** | **Right-Angle Distance Code** |
| *STRK* | *C* | *$3.* | |
| STRUCTUR | N | 2 | GROUP STRUCTURE |
| **SURFTEMP** | **Numeric** | **5.1** | **Sea Surface Temperature (C)** |
| **S_LAT** | **Numeric** | **7.4** | **Sighting Latitude** |
| **S_LONG** | **Numeric** | **7.4** | **Sighting Longitude** |
| **S_TIME** | **Numeric** | **6.** | **Sighting Time** |
| TAGLOC | N | 2. | TAG LOCATION |
| TAGMAT | N | 2. | TAG MATERIAL |
| TAGNUM | C | $10. | TAG NUMBER |
| TAGTYPE | N | 2. | TAG TYPE |
| **TAXCODE** | **Numeric** | **1.** | **Taxonomic Category Code** |
| **TIME** | **Numeric** | **6.** | **Time (hhmmss, EST)** |
| TIMEHR | N | 2. | TIME HOUR (EST) |
| TIMEMIN | N | 2. | TIME MINUTES (EST) |
| TMSOURCE | N | 2. | SOURCE OF TURTLE MEASUREMENT |
| TRACKNUM | N | 3. | NUMBER OF ANIMALS SEEN FROM TRACK |
| *TYPE* | *C* | *$6.* | |
| **VISIBLTY** | **Numeric** | **4.1** | **Visibility (nmi)** |
| *WAKE* | *C* | *$3.* | |

**Table 1**. (continued)

| Variable | Type | Format | SAS Label* |
|---|---|---|---|
| WATCOLOR | N | 2. | WATER COLOR |
| WEATHER | N | 2. | WEATHER CONDITIONS |
| *WHLR* | *C* | *$3.* | |
| WINDDIR | N | 2. | WIND DIRECTION |
| WTEMP | N | 3. | WATER TEMPERATURE (C) |
| **WX** | **Character** | **$1.** | **Weather Condition Code** |
| **YEAR** | **Numeric** | **4.** | **Calendar Year** |

*The variables that are created by SAS macros at present do not have associated labels, but they very easily could by adding only one line of code to a given macro.

## 8.1. ADEPTH

ADEPTH is the water depth (in meters) at the exact position of a sighting. This is a discontinued CETAP variable, which was used only for line-transect aerial surveys. See ALATDEG for more details. Depth could be submitted and entered in either meters, feet, or fathoms, so there was a unit flag variable in the data as originally entered, however the units were converted to meters before the data were added to the database.

## 8.2. ALATDEG

ALATDEG is the degrees of latitude at the exact position of a sighting, which was defined as the position of one individual or the center of a group. This is a discontinued CETAP variable, which was used only for line-transect aerial surveys. The suite of variables for the sighting location (all on the Record Type V card for data entry) was added well after the surveys were started in response to suggestions from the aerial observer team, who had a set of their own research interests. It is important to remember that the latitude and longitude data recorded as sighting locations generally represent the

44

location at the platform at the time of the sighting, and not the actual location of the sighting itself. The same is true of water depth and sea-surface temperature data. See also LATDEG, LATMIN, LONGDEG, and LONGMIN.

It became necessary to begin using similar data fields for the Mass CEC/NLPSC aerial surveys for trigonometric computations of right-angle distances. Rather than six separate variables, we decided to add only two, in decimal degrees (see S_LAT and S_LONG).

## 8.3. ALATMIN

ALATMIN is the minutes of latitude at the exact position of a sighting. This is a discontinued CETAP variable, which was used only for line-transect aerial surveys. See ALATDEG for more details.

## 8.4. ALATSEC

ALATSEC is the seconds of latitude at the exact position of a sighting. This is a discontinued CETAP variable, which was used only for line-transect aerial surveys. See ALATDEG for more details. The original data entered were in degrees, minutes, and tenths of minutes, however in at least some versions of the CETAP database the tenths of minutes were converted to seconds, and the minutes from fractional to integer values.

## 8.5. ALONDEG

ALONDEG is the degrees of longitude at the exact position of a sighting. This is a discontinued CETAP variable, which was used only for line-transect aerial surveys. See ALATDEG for more details.

**8.6. ALONMIN**

ALONMIN is the minutes of longitude at the exact position of a sighting. This is a discontinued CETAP variable, which was used only for line-transect aerial surveys. See ALATDEG for more details.

**8.7. ALONSEC**

ALONSEC is the seconds of longitude at the exact position of a sighting. This is a discontinued CETAP variable, which was used only for line-transect aerial surveys. See ALATDEG for more details; also see ALATSEC.

**8.8. ALT**

ALT is the aircraft altitude in meters. ALT is required for every record for aerial survey data, and is not allowed for shipboard survey or opportunistic sighting data. CETAP data entry allowed altitudes in either feet or meters, therefore included a unit flag in the entered data. The altitudes were then standardized to meters during data processing. The NARWC data protocol still allows submission of altitudes in feet or meters (nearly all submissions are in feet, which is the standard unit used in aviation, although GPS-derived altitudes can be switched by changing settings in the receiver). There is a check-off on the data submission cover sheet to say which units were used. There is also a feet-to-meters conversion step in the SAS code that reads a new dataset into SAS that can be switched off simply by putting an asterisk at the beginning (designating it as a comment line rather than program line).

Altitude data have been a problem at times. Although altitude is a factor in filtering effort for SPUE analyses, precise altitude information is important only for calibrating vertical aerial photos for use in photogrammetry. It is also important that standard altitude be maintained for right-angle distance measurement during line-transect surveys, but the precision is less critical. With manually logged data at, *e.g.*, 5-minute intervals, it is adequate to note changes in the log, *e.g.*, "descended to 300 ft for photos."

However, with computer-logged data at 10-second intervals, a drop from 1000 feet to 300 feet in 10 seconds is a descent rate of 4200 ft/min, which is well in excess of the capabilities of typical survey aircraft. Nevertheless, there is little that can be done with such data.

The output from a GPS unit can be used to log altitude data more precisely, however that has proven to be problematic. First of all, altitude error in a GPS receiver is typically 1.5 times the horizontal error, *e.g.*, if the expected accuracy of the location of a particular GPS is ±15 m 95% of the time, then the accuracy of the altitude output is ±23 m (74 ft) 95% of the time. This doesn't factor in the difference between actual altitude and modeled altitude computed within the GPS, which varies by location. It also assumes an optimal antenna configuration, *i.e.*, an unobstructed view of the entire sky.

After recognizing some extremely variable altitude data in submitted survey datasets, I added a step to the error-checking software that computed the rate of climb (with descent negative) from altitudes at successive events and elapsed time between the events. The error threshold was set at 75% of the maximum rate of climb for a particular aircraft (thresholds can always be adjusted if necessary based on experience developed with datasets over time). It appears that errors are most likely during tight turns and circling, suggesting that antenna configuration may be implicated. Problems with location data from GPS also are most likely during circling; interestingly, problems with the altitude data often happen first, before the latitude/longitude data go down the tubes.

Sometimes data are clearly unreliable and must be corrected, either individual outliers or sequences of bad data. Depending on how long they are, series of bad numbers can be replaced by averages of the preceding and following events, or simply by the intended survey altitude. Regular noting of intentional altitude changes in comments in computer data can be very helpful in this regard.

## 8.9. ANHEAD

ANHEAD is a two-digit code for the heading of a sighting, using a 16-point compass rose. ANHEAD is optional for sightings in any of the data types, and not allowed for non-sighting records. It was originally designed for sightings of an animal or

group or group of animals, but it was immediately applicable to heading of a recorded vessel when we began recording human activities as sightings rather than using HUMANACT (see that, also SPECCODE). At that time, the two last codes applicable only to vessels were added to the list.

Some datasets have been submitted with these headings in degrees rather than being coded. These must be converted during SAS data entry, and there is program code to do so. The listing below shows the degree equivalents for each code in brackets.

| | | | | |
|---|---|---|---|---|
| 00 = N | [349 – 011] | | 11 = WSW | [237 – 258] |
| 01 = NNE | [012 – 033] | | 12 = W | [259 – 281] |
| 02 = NE | [034 – 056] | | 13 = WNW | [282 – 303] |
| 03 = ENE | [057 – 078] | | 14 = NW | [304 – 326] |
| 04 = E | [079 – 101] | | 15 = NNW | [327 – 348] |
| 05 = ESE | [102 – 123] | | 16 = circling | |
| 06 = SE | [124 – 146]` | | 17 = various courses | |
| 07 = SSE | [147 – 168] | | 21 = stationary, but not anchored | |
| 08 = S | [169 – 191] | | (vessels only) | |
| 09 = SSW | [192 – 213] | | 22 = anchored (vessels only) | |
| 10 = SW | [214 – 236] | | | |

## 8.10. ATEMP

ATEMP is the air temperature measured at the survey platform, in degrees Celsius. This is a discontinued CETAP variable.

## 8.11. AWTEMP

AWTEMP is the sea-surface temperature measured at the exact sighting location, in degrees Celsius. This is a discontinued CETAP variable. See ALATDEG for more details.

## 8.12. BEAUFORT

BEAUFORT is sea state estimated on the Beaufort scale, which is actually a wind force scale. BEAUFORT is required for all on-watch records for both aerial and shipboard surveys, and optional for off-watch records during surveys and opportunistic sightings. The important factor is the presence of white-caps and their strong effect on sightability. Some surveys in the past have used a sea state scale that is different from the Beaufort scale. At least two other scales exist; values can be roughly converted, although it can be difficult or impossible to know which scale has been used. Table 2 (duplicated from NARWC Ref. Doc. 2003-01) shows the expected wind speeds and wave heights associated with each level on the Beaufort scale. Also included is descriptive information from Appendix R ("Beaufort Scale with Corresponding Sea State Codes") of *American Practical Navigator* (U.S. Naval Hydrographic Office Pub. No. 9), as well as additional descriptive information and tips on estimating Beaufort from an aerial survey platform.

During CETAP, BEAUFORT was a two-digit field, allowing any Beaufort number up to the maximum of 12. There was also a defined code of "99," meaning "few or small waves, but large swells interfering with visibility and making survey difficult or impossible." Since few or no surveys would be conducted at high sea states, the NARWC standard is that "7" means 7 or greater and "9" replaces the CETAP "99" code, reducing the field width to 1 digit.

Some survey data have been submitted with BEAUFORT to one decimal place of precision (*e.g.*, 1.3, 2.4, 3.5), which in my opinion is artificially precise and of dubious reliability. The defined standard for the database has always been that sea states are in whole numbers, so all data submitted like that have been and will continue to be rounded to integers (*i.e.*, 3.4 becomes 3 and 3.5 becomes 4). Submitting BEAUFORT data with tenths but no explicit decimal (*e.g.*, 13, 24, 35), unless accompanied by an explanation in the cover letter will result either in an error message during quality control (values beyond the allowable range) or in forcing all such data to Beaufort 7 (the maximum allowable value), depending on the exact data format submitted.

**Table 2.** Beaufort sea state levels with associated wind speeds (knots), wave heights (feet), and descriptive details.

| Beaufort | Wind | Waves | Description |
|---|---|---|---|
| 00 | 0–1 | 0 | *Calm*—Sea smooth and mirror like. From the air, the surface looks like a mirror, and glare from the sun is reduced to a very small area, maybe even only a reflection of the sun's disk. |
| 01 | 1–3 | 1/4 | *Light air*—Scale-like ripples without waves or white-caps. The surface looks scaly; sun glare extends less than half-way to the horizon. |
| 02 | 4–6 | 1/2 | *Light breeze*—Small, short wavelets; crests have a glassy appearance; occasional white-caps. From the air, white-caps look like points, with never more than one–three in view at once. |
| 03 | 7–10 | 2 | *Gentle breeze*—Large wavelets; some crests begin to break; foam of glassy appearance; scattered white-caps. From the air, white-caps still appear small and point-like. There may be many in view at one time, but they generally can be seen only within a half-mile to a mile. They tend to disappear quickly and do not persist. |
| 04 | 11–16 | 4 | *Moderate breeze*—Small waves, becoming longer; fairly frequent white-caps. From the air, the white-caps become elongate rather than point-like and persist as the wave moves away. White-caps are now visible beyond one or two miles away. |

**Table 2.** (continued)

_____

| Beaufort | Wind | Waves | Description |
|---|---|---|---|
| 05 | 17-21 | 6 | *Fresh breeze*—Moderate waves, taking a more pronounced long form; many and longer white-caps; there may be some spray. From the air, white-caps begin to look more like breakers, with foam patches persisting long after the wave breaks. White-caps are visible nearly to the horizon. |
| 06 | 22–27 | 10 | *Strong breeze*—Large waves begin to form; white-caps are more extensive everywhere; spray is more frequent and pronounced. Huge rolling breakers can be seen from the air, with persistent foam patches behind the waves now forming longer lines parallel to the wind. |
| 07 | 28–33 | 14 | *Near gale*—Sea heaps up and white foam from breaking waves begins to be blown in streaks along the direction of the wind ahead of the waves (spindrift). |
| 08 | 34–40 | 18 | *Gale*—Moderately high waves of greater length; edges of crests breaking into spindrift; foam is blown in well-marked streaks along the direction of the wind. |
| 09 | 41–47 | 23 | *Strong gale*—High waves; dense streaks of foam along the direction of the wind; crests of waves begin to topple, tumble, and roll over; spray may reduce visibility. |

**Table 2.** (continued)

_____

| Beaufort | Wind | Waves | Description |
|----------|------|-------|-------------|
| 10 | 48–55 | 29 | *Storm*—Very high waves with long overhanging crests; resulting foam in great patches is blown in dense white streaks along the direction of the wind; entire surface of the sea is white in appearance; tumbling of the sea becomes heavy and shocklike; visibility is reduced. |
| 11 | 56–63 | 37 | *Violent storm*—Exceptionally high waves may obscure small and medium-sized ships; sea is completely covered with long white patches of foam lying along the direction of the wind; everywhere the edges of wave crests are blown into froth; visibility is reduced. |
| 12–17 | 64–118 | 45+ | *Hurricane*—The air is filled with foam and spray; sea completely white with driving spray; visibility very much reduced. |

_____

## 8.13. BEHAV1 – BEHAV15

There are fifteen separate BEHAVn fields, which are two-digit codes for behaviors observed at particular sightings. BEHAVn is optional for sightings, and not allowed for non-sighting events. Some codes were used in the 1979 CETAP data and then dropped as too subjective and inadequately defined. These are not available for use as new codes (they are in bold italics and the former meaning is given in parentheses), but are included here since they do exist in the 1979 data, and those codes are precluded from being assigned to different behaviors. One of those (64, formerly defined as

"associated with sharks," has now been re-admitted to the fold as indicating shark scavenging on an identified carcass). The behavior codes have evolved over the years, with new codes added in response to needs and interests, and some codes falling into disuse. When we switched to recording human activities as sightings, a couple of codes were added for vessel "behaviors," and two others (73 and 74) are being held for the same purpose at some future time. And I just added two new codes for birds (56 and 57, anticipating more data in the future).

There is a certain inherent degree of difficulty in dealing with the behavioral information in the database, given that there are 15 identical fields and 90 possible codes. That was the impetus behind development of many of the SAS macros described earlier, which produce the synthesized behavior variables listed in Table 1 without SAS labels. For example, someone might request all sightings of one or more species of whales, and wants to be able to identify sightings of mother-calf pairs. There are three different behavior codes (40, 41, 42) that indicate a mother-calf pair; as does a non-zero value for NUMCALF. Rather than dump all 16 variables in response to the request, I can simply run the appropriate macro as part of the SAS job creating the output dataset, which searches all 16 fields for the values indicated and outputs one new variable—CALF—with values of either "YES" or "NO."

There is a great deal of variability in how complete or not the behavioral information might be from a given survey or in a given dataset. Some survey teams have been meticulous about recording behavior, some recorded only limited subsets of behaviors or behavior types, some recorded behavior somewhat haphazardly (perhaps depending on personnel), and some have not recorded any behavioral data at all. This needs to be considered in any analysis using the behavior data, and it should never be assumed that an absence of a behavior recorded at a given sighting means that behavior was not observed.

| | |
|---|---|
| 00 = dead, in water | 04 = stranded alive & rescued |
| 01 = dead, stranded | 05 = visible injury |
| 02 = dead, in fishing gear | 06 = fast swimming (>10 knots) |
| 03 = killed by whalers | 07 = moderate swimming (1-10 knots) |

08 = slow swimming (<1 knot)

09 = obvious speed change

10 = apparently influenced by platform

11 = porpoising

12 = riding vessel bow wave

13 = breach (whales)

14 = aerobatics (dolphins)

15 = swimming upside down/ rolling

16 = swimming on side

17 = swimming at surface

18 = swimming below surface

19 = flippering

20 = lobtailing, tail slashing (whales); tail-slapping (dolphins)

21 = spyhopping

22 = motionless at surface

23 = dive, flukes not raised (whales)

24 = dive, flukes raised (whales)

25 = blow, mist visible

26 = blow, mist not visible

***27 = unavailable (respiration intervals recorded)***

28 = dive intervals recorded

***29 = unavailable (synchronous diving)***

30 = swimming in wake of vessel

34 = swimming steadily in one direction, "traveling"

35 = circular movement

36 = obvious change of direction

37 = defecation

38 = close (<1/2 mi) to fishing gear

40 = mother with young

41 = apparent calving

42 = apparent nursing

43 = penis observed

44 = body contact, not belly-to-belly

45 = riding whale bow wave

***46 = unavailable (synchronous swimming)***

***47 = unavailable (synchronous breathing)***

***48 = unavailable (shallow dive)***

50 = associated with seaweed

51 = associated with other cetaceans

52 = associated with pinnipeds

53 = associated with birds

54 = apparent feeding

55 = feeding on fishery catch or by-catch

56 = flying (birds)

57 = sitting on the water (birds)

58 = bubbles observed

59 = associated with small fish

60 = associated with large fish

61 = associated with squid

62 = associated with jellyfish

63 = associated with visible zooplankton

64 = shark scavenging (formerly associated with sharks)

65 = distinct sub-groups

***66 = unavailable (apparent cooperative behavior)***

67 = belly-to-belly contact

| 68 = motionless below surface | 82 = apparent oil avoidance |
| 69 = diving (turtles) | 83 = apparent oil attraction |
| 70 = on beach, nesting or other (turtles) | 84 = in contact with oil |
| 71 = fishing/trawling (fishing vessels) | 85 = apparently not influenced by oil |
| 72 = hauling/setting gear (fishing vessels) | 86 = change in group heading |
| | 87 = change in group structure |
| 73 = hold for future use (vessels) | 88 = biopsy darted |
| 74 = hold for future use (vessels) | 89 = tagged (any type) |
| 75 = hauled out on ice (seals) | 90 = surface-active group (right whales) |
| 76 = hauled out on beach (seals) | 91 = thrashing, violent behavior |
| 77 = hauled out on rocks (seals) | 92 = tangled in fishing gear |
| 78 = milling | 93 = abnormal behavior |
| 79 = associated with physical feature | 94 = uncodeable behavior |
| 80 = audible sounds produced | 97 = mud on animal |
| 81 = underwater sounds recorded | 98 = struck by vessel |

## 8.14. BLOCK

BLOCK identifies a predefined survey block used for line-transect aerial surveys. During CETAP it was a 1-character field, which was entered into the first space for the survey line number (LEGNO) on Record Type II. The only values used were letters A–S and numbers 0–7. There were nine dedicated aerial blocks in Year 1, designated A through I. In Year 2 (1980), the sampling design was stratified by water depth (see STRATUM). Because of the more complicated bathymetry in the Gulf of Maine and Georges Bank, four of the blocks were cut into two: A → J & K; B → L & M, C → N & O; and D → P & Q. In year three, two additional slope-water blocks, R and S, were added. The numbers designated special right-whale and endangered-species aerial surveys in both 1980 and 1981 in the Great South Channel and off Nova Scotia. The block numbers were duplicated between the surveys; users must refer to the figures in Kenney and Winn (1986) to ensure they are using the correct blocks.

The variable was expanded to two characters for the NARWC program to allow for more flexibility in identifying survey areas.

| | |
|---|---|
| A–S = | CETAP dedicated aerial surveys |
| W = | 1985 Great South Channel right whale surveys |
| 0–7 = | CETAP 1980 right whale surveys (duplicate numbers used for Nova Scotia and Great South Channel/Georges Bank) |
| 1–2 = | CETAP 1981 endangered species surveys |
| GS = | Great South Channel right whale surveys |
| S1–S0 = | Southeast Survey blocks 1–10 (SE Turtle Survey blocks extended and stratified for mammal surveys) |
| MC = | Muskeget Channel sub-block, 2011–15 Massachusetts Clean Energy Center (NLPSC) project |
| MN = | NOREIZ (Northeast Offshore Renewable Energy Innovation Zone) sub-block, 2011–15 Massachusetts Clean Energy Center (NLPSC) project |
| MV = | Martha's Vineyard Block, 2011–15 Massachusetts Clean Energy Center (NLPSC) project |
| M2 = | General survey block, 2017+ Massachusetts Clean Energy Center WEA aerial surveys |
| ME = | Eastern block of condensed surveys, 2017+ Massachusetts Clean Energy Center WEA aerial surveys |
| MW = | Western block of condensed surveys, 2017+ Massachusetts Clean Energy Center WEA aerial surveys |
| NB = | Browns/Baccaro Bank block on Scotian Shelf |
| NE = | Emerald Bank block on Scotian Shelf |
| NO = | Offshore block on Scotian Shelf |
| SE = | MMS Southeast surveys, 1989–92, Cape Hatteras, NC to Savannah, GA |
| GA = | MMS Southeast surveys, 1989–92, Savannah, GA TO Jacksonville Beach, FL; E-W lines |
| FL = | MMS Southeast surveys, 1989–92, Jacksonville Beach to Miami, FL; lines parallel to beach |

OB =    UNC-Wilmington Shallow-Water Training Range surveys, 1998–99, Onslow
        Bay

WI =    UNCW SWTR Surveys, 1998–99, Wallops Island


## 8.15. CALF

CALF is first (alphabetically) of the "synthetic" variables. These variables do not
exist in either the CETAP or NARWC database, but instead are created on demand from
the NARWC database. It is created by a SAS macro, which is nothing more than a
separate file of SAS code that can be read into any other SAS program using a single
"%INCLUDE" step. In this case, the macro looks through all fifteen BEHAVn fields for
any code value that would indicate that at least one calf was observed at a particular
sighting, and also for values of NUMCALF greater than zero. CALF can take the value
of "YES" or "NO." In some cases, the variable for output purposes has been reduced to
just the first letter. When the output file is a dBASE file (_____.dbf), the variable format
can also be defined as a "logical" variable.


## 8.16. CALFHAT

CALFHAT is the original CETAP variable designating the number of cetacean
calves or turtle hatchlings observed at a sighting. The variable name was changed to
NUMCALF for the NARWC database.


## 8.17. CANADA

CANADA is a synthetic variable that flags any event as to whether it is in waters
under Canadian jurisdiction. It was created to more easily and accurately respond to
requests for only sightings in Canada. The output value is either "Y" or "N," although in
most cases the variable would merely be used to subset a dataset and not actually be
included in the output data. The macro works by approximating the Hague Line by two
simple linear equations and testing which side of those lines any sighting falls on.

Therefore the sub-setting is not exact in the vicinity of the line, but it will divide sightings beyond the EEZ by the offshore extension of the line.

**8.18. CARALEN**

CARALEN is the carapace length (in cm) of a sea turtle, expected most often to be a stranding. CETAP included a separate turtle research group, headed by C. Robert Shoop from URI, which resulted in a database design that included a lot of turtle-specific. There were separate defined data types for historical and opportunistic turtle stranding reports (see FILEID). These were used very rarely (1 and 8 times, respectively), therefore the turtle measurement fields were used even less often, and were not included in the first copy of the CETAP database transferred to NMFS (Kenney and Winn, 1986). (see also CARAWID, CARAWT, TAGLOC, TAGMAT, TAGNUM, TAGTYPE, and TMSOURCE)

**8.19. CARAWID**

CARAWID is the carapace width (in cm) of a sea turtle, most often a stranding (see CARALEN).

**8.20. CARAWT**

CARAWT is the weight (in grams) of a stranded sea turtle (see CARALEN).

**8.21. CCOVER**

CCOVER was the original CETAP variable for cloud cover. It was nominally measured in "oktas," or eighths of sky coverage, however the transcriber's manual (see Kenney and Owen, 1983) specified that, *e.g.*, a value of 5 meant ">4–5 oktas" (">50.0–62.5%). The value could then range from 0 to 8. There were also four additional values

defined (not surprisingly, because they are not exclusive from values 0–8, 10 was used rarely, while 11 and 12 were never used):

9 = sky obscured, or cloud amount cannot be estimated

10 = shadows caused by scattered cloud cover

11 = no shadows

12 = overcast.

Since it is highly unlikely that observers could quantify cloud cover that precisely, and because the variable was never useful enough to justify the effort needed by the field observers, the cloud cover variable was greatly simplified for the NARWC database (see CLOUD).

## 8.22. CETSPPCD

CETSPPCD is exactly the same variable as SPECCODE under a different name (although with fewer available species at the time). In the process of transferring a copy of the CETAP database to NMFS (Kenney and Winn, 1986) and designing the NARWC database, it was clear that we needed a better system than the two-letter or two-number codes used by CETAP. We settled on four-letter codes based on common names as simple and, importantly, relatively easy to remember. Apparently, SPECCODE was used for a different code system in other NMFS databases, so they insisted on a different name and decided on CETSPPCD (CETAP species code, even though CETAP never used them). See also SPECCODE, SPECCHAR, SPECNUM, and SPECNAME.

## 8.23. CIRCLFLG

CIRCLFLG was a simple flag (1 = yes; 2 = no) denoting whether the time spent circling a given sighting to verify species identification and get accurate counts could be accurately computed by subtracting the time at the LEGSTAGE 3 event from the following LEGSTAGE 4 event. It applied only to line-transect aerial surveys, and was required to be filled in for every on-census sighting (LEGTYPE = 2; LEGSTAGE = 2)

that was immediately followed by breaking off from the track to circle (LEGTYPE = 2; LEGSTAGE = 3).

## 8.24. CLOUD

CLOUD is the NARWC cloud cover code. CLOUD is required for all on-watch records for shipboard and aerial survey data, and optional for off-watch survey records and opportunistic sightings. In the original CETAP data, cloud cover was recorded in "oktas" or eighths (see CCOVER), and other survey programs have used tenths of sky coverage or even more precise percentages. When the Consortium database was begun, a simpler system was adopted. Based on our experience, observers had great difficulty in consistently estimating cloud cover on any precise scale (especially when looking out an aircraft window and not seeing the entire sky at once). In addition, cloud cover was not one of the more important environmental parameters that were used to define acceptable survey conditions. We decided to use a simpler, descriptive classification based on that used in aviation weather reporting—clear, scattered, broken, and overcast. Because the values 0–8 were preoccupied by the CETAP okta values and 9 and 10 had also been used, we settled on two-digit codes (11, 12, 13, and 14). In early 2004, the older data were converted from oktas to the descriptive systems (0–1 = clear, 2–4 or 10 = scattered, 5–7 = broken, and 8 = overcast), enabling simplification of this code to one digit. A "0" value for "not recorded" was also added to accommodate contributors who do not allow missing values in their data, since the usual "9" value has another meaning, but will still be converted to and archived in SAS as a missing value.

1 = clear, <10% cloud cover
2 = scattered, 10-50% cloud cover
3 = broken, 50-90% cloud cover
4 = overcast, >90% cloud cover
9 = sky obscured, or cloud amount cannot be estimated (night, fog)
0 = not recorded

60

## 8.25. CONFIDNC

CONFIDNC is a two-digit code for the estimated precision associated with the number of animals counted at a particular sighting. CONFIDNC is required for all sighting records in all data types, and is not allowed for non-sighting records. This is one of the two variables which have been most often misused in contributed data and caused the most exasperation (see also IDREL). Many datasets have been submitted with the variable simply filled in with "00" down the line. I have had survey team leaders tell me they had been trained that IDREL should always be 3 and CONFIDNC should always be 00 without ever knowing what they meant. It is relatively obvious when sightings of tens to hundreds (or even thousands) of animals are reported as exact counts, and even more so when "00" has been entered into every line of a file, including non-sighting events. After years of dealing with dubious CONFIDNC data, for many surveys the SAS code in the data entry program had a line that changes CONFIDNC for all sightings of more than 5 animals from "00" to "10." Files submitted with missing values for CONFIDNC are also changed to "10."

Some datasets have been submitted with minimum and maximum estimates of group size, or minimum/maximum/best. These have generally been converted to NUMBER and CONFIDNC as closely as possible, given the broad gaps in the CONFIDNC levels.

This is one of many variables that are defined as numeric fields in SAS but as character fields in dBASE. One of the weaknesses of dBASE is in its handling of missing values. Missing values for numeric variables are treated as zeroes, even though there is not an explicit 0 in the field. In most cases, we want to differentiate zero from missing, which SAS is entirely capable of doing. In those cases, the defined value here and in other code lists may have leading zeroes, simply for data entry purposes and because NMFS did not like empty spaces when we were designing the database. Don't worry if you are unable to enter "00" for CONFIDNC (or "005" for HEADING) if you are using, *e.g.*, Excel; 0 and 00 are exactly equivalent.

00 = +/- 0                    01 = +/- 1

| 02 = +/- 2 | 09 = +, "at least" for group |
| 03 = +/- 5 | counts |
| 04 = +/- 10 | 10 = no estimate of confidence |
| 05 = +/- 25 | level, number estimated |
| 06 = +/- 50 | 11 = number of animals |
| 07 = +/- 100 | unknown |
| 08 = +/- 1000 | |

## 8.26. DATAMETH

DATAMETH is a discontinued CETAP variable that describes how the animals in a sighting were counted.

1 = visual count
2 = visual estimate
3 = photographic count
4 = acoustic count

## 8.27. DATED

DATED is the obsolete CETAP variable name for day of the month (see DAY).

## 8.28. DATEM

DATEM is the obsolete CETAP variable name for month (see MONTH).

## 8.29. DATEY

DATEY is the obsolete CETAP variable name for year, including only the last two digits and assuming that the first two are "19" (see YEAR).

## 8.30. DAY

DAY is the calendar day of the month (1–31). DAY is required for all records in shipboard and aerial (POP and line-transect) survey data, and optional (although strongly encouraged) in opportunistic sightings. Data can be submitted in dBASE, Excel, or Access formats with complete dates in a defined date-format field (preferably called "DATE"). Separate MONTH, DAY, and YEAR fields can be easily extracted during the translation to SAS for archival (the extraction is done within dBASE). Dates could probably be submitted in text files in any format that could be recognized by dBASE, Excel, or Access (*e.g.*, August 19, 2008; 8/19/08; 08/19/2008; 08-Aug-2001).

## 8.31. DDSOURCE

DDSOURCE (direct data source) is an identifier for the original data collector or contributor. During CETAP, it was mainly used to identify the project PI who was directly responsible for the data, so it was dominated by individuals' initials. The primary meaning has changed in the NARWC database to identify the organization that conducted a survey or created a dataset. In the list of codes that follow, those shown in italics are obsolete CETAP codes that are no longer in routine use. In fact, because usage of DDSOURCE and IDSOURCE also seemed to be quite inconsistent in the CETAP data, most of them have been converted in the NARWC data to "CET."

DDSOURCE is assigned after a dataset has arrived at GSO, so is "invisible" to data contributors. But it is something that some data users may be interested in obtaining in a dataset.

ASW = Associated Scientists at Woods Hole

*CAM = Charles A. Mayo*

CCS = Center for Coastal Studies

*CET = CETAP*

*CHS = Cape Hatteras National Park Service*

CMA = Clearwater Marine Aquarium Research Institute (formerly Sea-to-Shore Alliance)

*CRS = C. Robert Shoop*

CSA = Continental Shelf Associates

CWI = Canadian Whale Institute

DAL = Dalhousie University

DFO = Dept. of Fisheries & Oceans, Canada

ECE = East Coast Ecosystems

FFJ = FACSFACJAX (Navy Jacksonville)

FLA = Florida state agencies

FWW = Fundy whale-watchers

GEO = Georgia state agencies

GMI = Geo-Marine, Inc.

GMW = Grand Manan Whale & Seabird Research Station

*GWO = Guy W. Oliver, III*

*HEH = Herbert E. Hayes*

*HEW = Howard E. Winn*

IFA = International Fund for Animal Welfare

*JGM = James G. Mead*

LIT = Literature

MAQ = Mystic Aquarium

MBO = Manomet Bird Observatory

MWR = Massachusetts Water Resources Authority

NEA = New England Aquarium

NLP = Northeast Large Pelagics Survey Collaborative

NNE = NMFS, Northeast

NSE = NMFS, Southeast

NYS = New York State agencies

*OGB = Ollie G. Brazier*

OPB = Offshore Petroleum Board (Nova Scotia)

OTH = Other (miscellaneous)

*RAR = Richard A. Rowlett*

RFM = Riverhead Foundation for Marine Research & Preservation

*RXP = Robert Prescott*

SSA = Sea to Shore Alliance (changed to Clearwater Marine Aquarium Research Institute)

*TXA = Thomas A. Azarovitz*

UCG = U.S. Coast Guard

UNC = Univ. of North Carolina–Wilmington

URI = University of Rhode Island

VAM = Virginia Aquarium & Marine Science Center

WCN = Whale Center of New England

WDC = Whale & Dolphin Conservation Society

WHO = Woods Hole Oceanographic Inst.

WLT = Wildlife Trust (name changed to EcoHealth Alliance)

**8.32. DEAD**

DEAD is a synthetic variable with a value of "YES" or "NO" that identifies whether the animal observed in a particular sighting was dead (stranded, floating, killed by whalers, etc.—see BEHAVn).

**8.33. DEBRIS**

DEBRIS is an abandoned code for debris or pollution of various types observed in the water. DEBRIS was optional for all records and all data types. Following the example of the NMFS surveys, the database was modified to allow such observations (including a much broader variety of types) as sightings. A large number of new codes was added to SPECCODE. This provided much more detail, accuracy, and flexibility. In late 2003, use of the DEBRIS field was discontinued, and observations of debris and pollution were only allowed to be submitted using the sighting format. In early 2004, all previously archived with DEBRIS codes were converted into the newer style, using the SPECCODE values shown following the items in the list below and sighting numbers of 999.

| | |
|---|---|
| 0 = Sargassum weed line | [DE-S] |
| 1 = Flotsam (natural origin) | [DE-F] |
| 2 = Jetsam (human origin) | [DE-J] |
| 3 = Oil slick | [DE-O] |
| 4 = Oil sheen | [DE-O] |
| 5 = Oil patches | [DE-O] |
| 6 = Other | [DE-U] |
| 7 = Combinations of above | [DE-U] |

**8.34. DECOMP**

DECOMP is a discontinued CETAP variable describing the stage of decomposition of a stranded sea turtle.

1 = shell and body intact, little or no foul odor or bloating

2 = shell falling apart (on leatherback, skin gone from top shell, shell totally
exposed), body bloated and discolored, strong foul odor

3 = skeleton

## 8.35. DEPTH

DEPTH is the water depth at a sighting location, in meters. DEPTH was formerly required for all sighting records for all data types. However, the depth data were often of questionable reliability. Most were interpolated from depths printed on navigation charts during data reduction, with a high probability of imprecision and error. There was also no easy way to detect depth errors during quality-control review. During one study looking at sperm whale sightings in continental shelf waters south of New England, all sperm whale sightings were classified by DEPTH into two categories: < 200 m and > 200 m. The sightings were plotted on a map with the 200-m isobath shown, using two different symbols for the two categories. There were substantial numbers of both symbols plotted on the wrong side of the line, clearly demonstrating the unreliability of the depth data. Since it is now possible to use GIS methods to associate sighting locations and water depth for analyses, with much more accuracy and reliability, depths are no longer required. Depth data may still be submitted, in meters, feet, or fathoms (units should be noted at the place provided on the cover sheet, and the method for obtaining the data noted in the comment section).

## 8.36. DISTSHOR

DISTSHOR is the distance from shore of a sighting, classified into intervals and coded. It was almost never used and not especially worthwhile. The CETAP transcription manual said this: "This distance is most applicable for nearshore sightings and/or when latitude/longitude positions are crude approximations only. This field will generally be used in relation to nearshore sightings of turtles." This was apparently another field

included at the request of Dr. Shoop's turtle group, who regularly flew beach surveys in single-engine aircraft without LORAN-C capability.

| | | | | | |
|---|---|---|---|---|---|
| 0 = | 0 m | 0 n.mi. | 4 = | >1000–2000 m | >0.5–1.35 n.mi. |
| 1 = | >0–100 m | 0–0.05 n.mi. | 5 = | >2000–4000 m | >1.35–2.7 n.mi. |
| 2 = | >100–500 m | >0.05–0.3 n.mi. | 6 = | >4000 m | >2.7 n.mi. |
| 3 = | >500–1000 m | >0.3–0.5 n.mi. | | | |

## 8.37. DIVEMIN

See DIVETIME.

## 8.38. DIVESEC

See DIVETIME.

## 8.39. DIVETIME

DIVETIME was intended to be the duration of a dive measured at a particular sighting, in minutes and seconds, with DIVEMIN and DIVESEC entered into the data form as the first two digits and last two digits, respectively. The intention was that multiple dive times could be recorded for one sighting by filling in multiple Record Type IIIs. It never worked out very well, and was rarely used. The transcriber's directions specified that: (1) measured times should be entered into both fields, zero-filled, with DIVEMIN explicitly "00" for times of less than a minute (*e.g.*, 3:05 = "03," "05;" 50 seconds = "00," "50"); (2) estimated dive times of longer than a minute should be entered only in DIVEMIN, with DIVESEC left blank (*e.g.*, "03," "   "); (3) estimated dive times of less than a minute should be entered only in DIVESEC, with DIVEMIN left blank (*e.g.*, "   ," "50"). The result seemed to be a lot of confusion (either in data entry or in the creation of DIVETIME from the other two fields, or both), so that DIVETIME sometimes had minutes in the seconds columns. Overall, this is a good example of which

types of data should not be included in a sightings database, but instead should be maintained in a separate database (which can still be linked by FILEID and SIGHTNO).

**8.40. EVENTNO**

EVENTNO is the event number, which is simply a sequentially assigned record number. EVENTNO is required for all records in all data types. In an opportunistic sighting file, all "events" are sightings. In survey data, "events" include sightings, watch changes, changes in environmental conditions, turns, etc., or simply periodic fixes for later reconstruction of the survey track. At first glance it might seem that recording frequent vessel or aircraft positions at times when nothing is happening or changing is a waste of time, however that assumption is incorrect. The length of survey track covered, with or without sightings, is used to quantify effort, therefore it is extremely important to record periodic locations even when there are no sightings. The farther apart the locations are, the less likely it is that a straight line connecting them is an accurate reconstruction of the survey track. This can be very important when someone later wishes to analyze SPUE at a very high resolution (some data users have requested and received SPUE data at a 1x1-minute grid). When recording data by hand, there should be an effort to record positions approximately every 5 minutes for aerial surveys and every 15 minutes for shipboard surveys. For computer data-loggers, 30-second and 5-minute intervals, respectively, are reasonable intervals; even closer intervals may be desirable. Extremely close intervals in aerial survey data (e.g., closer than 5 seconds) have proven to be a problem. It seems that not all GPS/computer data-logger systems are capable of updating that quickly.

EVENTNO <u>must</u> increase sequentially within a file; many analyses sort data on EVENTNO and SIGHTNO one or more times during processing. Skipped numbers are allowed, however duplicate numbers are not, except in one circumstance. Recognizing that multiple species, human activities, and/or kinds of debris/pollution can occur at the same place, there may be multiple records with the same event number, but each with a different sighting number. Whenever there are multiple sightings with the same event number, all of the non-sighting variables <u>must</u> be identical across all records. A common

error here occurs when something is missed during initial data entry and a record is inserted at some later time with an event number that duplicates one already existing. When inserting missed events into a data file, it is usually necessary to correct the event numbers from that point forward in the file. To make corrections and insertions easier during CETAP, all of the event numbers assigned during transcription were multiplied by 10 when the dataset was entered into the database.

## 8.41. FEED

FEED is a synthetic variable with a value of "YES" or "NO" that identifies whether feeding behavior by one or more of the animals in a group was observed to be feeding.

## 8.42. FILEID

FILEID is the file identification number that identifies a particular dataset—one day of aerial survey, one shipboard survey (single day or multiple days), or one collection of opportunistic sightings. As such, it must be unique to that file. FILEID is generally assigned to a dataset after arrival at GSO, so there is less concern about possible duplication. The format presently in use is one character longer than that used during CETAP. The field is eight characters in width—one letter in the first position, followed by six 6 numbers, then an optional final letter.

FILEID has evolved in usage since CETAP. In the CETAP data, the first character (a letter) indicates data type. The second character (a number) could have one of three different meanings. For data types A, D, E, and G, it designates survey number. For the line-transect aerial surveys (A) in 1979, the year was partitioned into eight 45-day windows (surveys 1–8), with the design calling for one completed survey in each block during each window. In 1980 and 1981, the number of windows was cut to four and sampling was both stratified and intensified. (Note that the final survey days of a given sampling year may have occurred at the beginning of the next calendar year.) For data types F, L, M, T, M, and I, the second character designates data sub-type (see below). For

the remaining data types, it simply differentiates multiple files of the same type beginning on the same date. For example, two shipboard surveys that both began on 1 October 1979 would have been assigned FILEIDs of P179274 and P279274. The third and fourth characters are the last two digits of the year, and characters 5–7 are the Julian date of the first day of data included in the file.

A = dedicated aerial (line-transect) survey

D = dedicated aerial replicate survey in block D

E = dedicated aerial endangered species survey

F = POP aerial survey

  F0 = miscellaneous

  F1 = USCG radiothermography

  F2 = USCG fisheries patrol

  F3 = turtle survey (Shoop)

  F4 = coastal survey (Mead)

  F5 = USCG harbor patrol

G = dedicated aerial replicate survey in block G

H = historical data (cetaceans)

I = air/ship interactive survey

  I1 = ship

  I2 = air (never used)

J = historical data (turtle sightings)

K = historical data (turtle strandings)

L = dedicated aerial lease sale survey

M = POP aerial survey during 1979 "Minimum Right Whale Count"

  M1 = AT-11

  M2 = Skymaster (Aero-Marine)

  M3 = Skymaster (Katona)

  M4 = Islander

  M5 = Cessna 206

  M6 = Cessna 150

O = opportunistic data (cetaceans)

P = POP shipboard survey

Q = miscellaneous right whale sightings (post-CETAP, but pre-
   Consortium)

R = dedicated aerial right whale survey

      R1 = AT-11

      R2 = Skymaster

T = miscellaneous aerial survey

      T1 = "hot spot" (POP format)

      T2 = training (dedicated format)

      T3 = oil spill response (dedicated format)

      T4 = oil spill response (pop format)

U = opportunistic data (turtle sightings)

W = "dedicated" shipboard right whale survey (still in POP format, not
   line-transect)

Y = opportunistic data (turtle strandings)


In the NARWC data, the first character still designates data type, but the number of options was greatly reduced. As a work-around for the "Y2K" problem (*i.e.*, in order to avoid duplicate FILEIDs for data in the 20th and 21st Centuries), all FILEIDs for data beginning in year 2000 use a lower case letter for the first character. A different solution will be needed in 2100, but that will be someone else's problem. A simple option would be to expand the field by two characters and use all four digits of the year; converting all previous FILEIDs to that format would be a simple coding task in SAS.

The second character originally was used as during CETAP to differentiate files with the same date. Keeping track and avoiding duplication began to become a problem given the proliferation of concurrent surveys being conducted (in theory, there could be as many as eight or nine aerial surveys flown simultaneously at certain times of year, and data contributed by organizations using whale-watch vessels may have multiple vessels each making 1–3 trips per day). The practice became to assign different numbers for each data contributor, still changing the number for multiple flights or boat trips in one day.

Even that proved problematic; the issue was resolved in 2005 by expanding the field from seven to eight characters. The eighth character is an optional letter ("a," "b," etc.) to differentiate multiple surveys on the same day by a given contributor. Whether upper-case or lower-case letters are used makes little difference, as long as it is consistent within a survey file. As the suite of contributors changed, the numbers assigned to them also changed at times. The list below identifies both the current usage and former usage.

The third and fourth characters still represent the last two digits of the year. Similarly, the last three numbers are still the Julian date of the first day of data. An exception to the last is for a few regular collections of opportunistic data where the date is not important; for these the last three digits identify the data source (501 = NMFS or Smithsonian marine mammal stranding data; 901/902/903 = CCS opportunistic sightings; 911 = sightings extracted from the right whale catalog; 921 = NEFSC Sightings Advisory System opportunistic records; 951 = FACSFACJAX records; 961 = MWRA records; 971 = CRESLI).

There was an additional format created in March 2009, when the complete shipboard survey database from the Manomet Bird Observatory surveys during the 1980's was obtained, reformatted, and added to the NARWC database. In those surveys, MBO observers conducted POP-style watches on board NOAA fisheries and oceanographic research cruises. The FILEID format created was P9yynnM, where "yynn" was the NOAA cruise identifier number. This format keeps the last two digits of the year in the same place in FILEID, and at the same time allowed very easy subsetting of the data via the final character.

Another format was created in April 2015 for locations of tagged right whales extracted from the photoID catalog during the cross-referencing process (DDSOURCE = NEA and IDSOURCE = TAG). The FILEID format is O9yyEGNO, where "yy" is the last two digits of the year (and the first character will be upper or lower case depending on whether it was before or after 1 Jan 2000), and EGNO is the catalog number of the tagged whale.

A/a = dedicated aerial survey

b = dedicated aerial survey, but focused on North Atlantic right whales
    and other large whales, and only recording other species

opportunistically (defined specifically for the "condensed" surveys in the MassCEC/NEAQ aerial surveys in the Mass.-R.I. WEAs that began in 2017)

c0 = aerial survey, intermediate (relaxed line-transect) format (CCS), beginning in April 2015. If somebody else decides to use the same format, the second character can vary with contributor, as with the following.

F/f = POP aerial survey

F0 = Center for Coastal Studies

F1 = Florida Wildlife Research Inst. (EWS-south)

F2 = Wildlife Trust/Ecohealth Alliance/Sea to Shore Alliance/Clearwater Marine Aquarium Research Institute (Georgia, EWS-north, Georgia EWS)

F3 = Wildlife Trust/Ecohealth Alliance/Sea to Shore Alliance (South Carolina) [was East Coast Ecosystems]

F4 = NMFS Northeast (Sighting Advisory System/North Atlantic Right Whale Sightings Surveys)

F5 = AMAPPS [was NMFS Northeast—Protected Species Branch]

F6 = New England Aquarium/Florida Wildlife Research Inst. (EWS-central; Florida EWS)

F7 = Univ. of North Carolina-Wilmington [was Southeast offshore]

F8 = Riverhead Foundation

F9 = other

H/h = historical data (not expected to be used at all)

O/o = opportunistic data

oyy901, 902, ... = CCS opportunistic

o1yy911 = NEAQ catalog opportunistic

o1yy921 = NEFWS SAS opportunistic

o1yy951 = FACXFACJAX opportunistic

o1yy961 = MWRA opportunistic

o1yy971 = CRESLI opportunistic

P/p = POP shipboard survey

P0 = Center for Coastal Studies

P1 = New England Aquarium (Fundy)

P2 = New England Aquarium (Roseway)

P3 = New England Aquarium (GSC) [was East Coast Ecosystems]

p1, p2, p3 = New England Aquarium/Canadian Whale Institute (In
later years NEAQ ran multiple vessels at the same time, CWI
took over some surveys, and some surveys expanded to the
eastern Scotian Shelf and Gulf of St. Lawrence. It became
easier to just assign the subtypes by vessel within a given
year; they can be further sorted out by DDSOURCE and
PLATFORM.)

P4 = other Bay of Fundy (whalewatchers) [was East Coast
Ecosystems]

P5 = NMFS Northeast [was East Coast Ecosystems]

P6 = NMFS Southeast

P7 = Whale Center of New England

P8 = Gulf of St. Lawrence (whalewatchers)

P9 = other

## 8.43. FLUKES

FLUKES is an obsolete, discontinued CETAP variable denoting whether fluke
photographs of a given humpback sighting were available in CETAP files, with 1 = no
and 2 = yes. While this variable may be useful for a single program, it is of limited utility
when there are multiple contributors to the database who each maintain their own files of
photographs. (See also PHOTOS).

**8.44. GEAR**

GEAR is a synthetic variable with a value of "YES" or "NO" that identifies whether one or more of the animals in a sighting were observed to be entangled in fishing gear.

**8.45. GLAREAMT**

GLAREAMT is a discontinued CETAP variable that described the amount of sun glare on the water's surface that was interfering with the observer's ability to spot animals. It was used in conjunction with GLARELOC (see below). It was difficult to use these fields for several reasons. The order of the codes was illogical. There was no "slight" value, forcing an observer to decide between "none" and "moderate." Finally, the GLARELOC field was too inflexible, since it did not allow for different values on the two sides of the track, which was the usual condition encountered. When the NARWC database was being designed, GLAREAMT and GLARELOC were discarded in favor of separate fields for the left and right sides of the track (see GLAREL and GLARER).

1 = moderate

2 = none

3 = severe

**8.46. GLAREL**

GLAREL describes the amount of sun glare affecting observer visibility on the left side of the trackline. GLAREL and GLARER are required (technically, they are optional, but <u>strongly</u> encouraged) during on-watch records for aerial surveys, optional during off-watch aerial survey records, not allowed for opportunistic sightings, and theoretically optional for shipboard surveys, although they have never been used there to date. Since glare is a function of heading, it will change rapidly during turns and circling, which is the reason that missing values are allowed.

|  |  |
|---|---|
| 0 = none | 2 = moderate |
| 1 = slight | 3 = severe |

## 8.47. GLARELOC

GLARELOC is a discontinued CETAP variable that described the location of sun glare on the water's surface that was interfering with the observer's ability to spot animals. It was used in conjunction with GLAREAMT (see above)

.

|  |  |
|---|---|
| 1 = both port and starboard | 3 = starboard |
| 2 = port | 4 = unknown (not reported) |

## 8.48. GLARER

GLARER describes the amount of sun glare affecting observer visibility on the right side of the trackline. (See GLAREL).

## 8.49. GROUPS

GROUPS is a discontinued CETAP variable that described the number of separate sub-groups into which the animals in a sighting were divided. If a sighting was not divided into distinct sub-groups then a value of 01 was to be assigned. If there were too many sub-groups to count accurately or the number was >99, then it was to be left blank. For that to be useful, it would presume that the variable was required for all sightings, however it was input on Record Type V, which was optional. That makes it impossible to determine whether a blank means lots of groups or simply a missing value.

## 8.50. HEADING

HEADING is the heading of the survey aircraft or vessel, in degrees true. HEADING is strongly encouraged for all records in both aerial and shipboard survey

76

data, and not allowed for opportunistic sightings. Allowable values are 000-359. For headings recorded in degrees magnetic, there is a space on the cover sheet to fill in the compass variation for the region of survey, so that the conversion can be made when the data are added to the database. For those using computer data-loggers that are capturing GPS output directly, the "course made good" output from the GPS can be input directly into HEADING.

Since heading can change rapidly during turns and requiring accurate records of that would be an unreasonable burden on a data recorder using manual data-logging, missing values are allowed. However, the interactive dBASE data-entry programs provided in the early years of the Consortium research to data contributors were intentionally designed to force users to enter a value into every required field, as well as some that were not absolutely required. In the case of HEADING, the user was instructed to enter "999" for changing or variable headings. Over the years, that often became interpreted as a code with a particular meaning (*i.e.*, off-track or circling), to the point where some actually would replace perfectly good heading data captured from the GPS with long strings of 999s. But it was never more than a dummy variable to force users to enter something. All "999" headings were converted to missing values when the data were added to the database.

## 8.51. HUMANACT

HUMANACT is a two-digit code describing human activities (vessels, fishing gear, etc.). HUMANACT was optional for all records in all data types. As was the case with DEBRIS, it was very inflexible and not always clear. In addition, it was possible to enter only one HUMANACT code (and one DEBRIS code) at any event. Treating observations of this type as "sightings" allows the incorporation of much more detail and accuracy. For example, it is possible to enter the actual numbers of vessels, there are more "species" options, and there can be any number of sightings of different species, human activities, and debris/pollution types at a single event. In late 2003, use of HUMANACT was discontinued and observations of human activities had to be submitted as sightings. In early 2004, all previously archived data were converted into the newer

style, using using SIGHTNO=999 and the SPECCODE values shown following the items in the list below (italicized codes were created strictly to match existing HUMANACT codes, and are not available for current use). See also SPECCODE.

00 = commercial fishing fleet   (FV-U)

01 = supersonic aircraft   (*AC-S*)

02 = subsonic aircraft   (*AC-J*)

03 = turboprop aircraft   (*AC-T*)

04 = propeller aircraft   (*AC-P*)

05 = helicopter   (HELO)

06 = diver(s)   (*DIVE*)

07 = swimmer(s)   (*SWIM*)

08 = multiple activities   (*MULT*)

09 = several commercial fishing vessels
     (FV-U)

10 = oil drillship   (OI-L)

11 = fixed fishing gear   (FG-U)

12 = oil rig   (OI-D)

13 = sonar in use   (*SONR*)

14 = buoy tending   (CG-B)

15 = tending oil rig   (OI-D)

16 = research activity   (RV-L)

17 = explosive discharge   (EXPL)

18 = cable/pipe laying   (*CABL*)

19 = drilling   (OI-D)

20 = dredging   (DR-W)

21 = garbage dumping   (*DU-G*)

22 = toxic waste dumping   (*DU-T*)

23 = oil seepage   (DE-O)

24 = commercial longline fishing   (FV-
     L)

25 = commercial net fishing   (FV-U)

26 = single commercial fishing vessel
     (FV-U)

27 = sport fishing vessel(s)   (SPFV)

28 = single large merchant vessel   (MV-
     L)

29 = multiple large merchant vessels
     (MV-L)

30 = single small merchant vessel
     (MV-S)

31 = multiple small merchant vessels
     (MV-S)

32 = single sailing vessel   (SV-U)

33 = multiple sailing vessels   (SV-U)

34 = single recreational motorboat
     (RECV)

35 = multiple recreational motorboats
     (RECV)

36 = light boat traffic   (*BT-L*)

37 = heavy boat traffic   (*BT-H*)

38 = submarine   (NV-U)

39 = unidentified vessel(s)   (UNVE)

40 = moderate boat traffic   (*BT-M*)

41 = cruise ship   (crsh)

42 = ferry   (FE-U)

43 = coast guard ship   (CG-C)

44 = military vessel   (NV-L)

45 = whale-watch vessel   (WHAL)        48 = seismic/air guns in use   (OI-S)

46 = small military vessel   (NV-S)     49 = Personal watercraft ("jet ski")

47 = aquaculture facilities   (AQUA)          (JETS)

## 8.52. HURT

HURT is a synthetic variable with a value of "YES" or "NO" that identifies whether one or more of the animals in a sighting were observed with a visible injury.

## 8.53. ID

ID is a synthetic variable created from IDREL (see below) to provide clearer information in response to data requests. Rather than providing coded values that then need to be explained, users often are provided with text values of "possible," "probable," "sure," and "not recorded."

## 8.54. IDREL

IDREL is a one-digit code for the observer's judgment about the reliability of the stated identification of the species observed. IDREL is required for all sightings in all data types. This is one of the most frequently misunderstood, misused, and/or ignored variables (see CONFIDNC), and poses a much larger problem in terms of the scientific reliability of the data.

The value of IDREL that is assigned should apply to the species identification that is used, not to the sighting generally. As an example, consider a sighting of a distant school of dolphins during a shipboard survey. The ship is never able to approach close enough to make a good identification, and the sighting is finally recorded as "unidentified dolphins." The observers may have been "unsure" about which species of dolphin they had seen, but they were "sure" the animals were dolphins (and that they could not identify the dolphins). In that case, the proper value to use for IDREL is 3, not 1. A sighting of "unidentified dolphins" with IDREL = 1 would mean that the observers were

unsure if whatever they had seen was dolphins at all, and such a sighting record would likely be deleted as of minimal value.

The choice of what combination of SPECCODE and IDREL to record for a particular sighting where the exact identification is questionable can be problematic. Is it better, for example, to record a sighting as definitely unidentified dolphins, probably unidentified *Stenella* sp., or possibly striped dolphins? All three could be correct. My recommendation would be to select the most precise ID that has at least a "probable" reliability, which would be "unidentified *Stenella*" in this example. For the majority of analyses conducted from the database on individual species, only the definite and probable sightings are used.

More important is the clear lack of critical consideration given to species identification during some surveys, whether due to poor observer training, over-confidence, or some other factor(s). The task of field observers is to report exactly what they see, not what they believe they should have seen—observations rather than interpretations. Observers should carefully look at every animal or group sighted and record the species seen using only the ID cues actually observed. Do not identify a group of dolphins as bottlenose dolphins because that's the only species you expect to see. Record them as bottlenose dolphins if you got a really good look and saw the appropriate ID cues. Record them as positively identified if you are fully convinced of the ID (from the observed cues, not from any expectations). If you are only relatively sure, but not positive, record the species ID as "probable." If you are really unsure, go with "possible," or even with "unidentified dolphin" (with a "definite" reliability).

A data file submitted with every or almost every sighting recorded as positively identified is likely to be looked at more closely as potentially questionable during quality-control review. (See pages 22–27 in Kenney, 2002 for a more detailed discussion and a comparison of several different surveys.) The usual process with such data is to downgrade the IDREL values for all sightings by one level (*e.g.*, sure to probable), or more likely to downgrade all sightings except for really obvious and easily identified species (*e.g.*, right and humpback whales),

1 = unsure / possible

2 = probable

3 = definite / sure

9 = unknown/not recorded (also should be used for all sightings of vessels, fishing
    gear, human activities, pollution, debris, etc.)

## 8.55. IDSOURCE

IDSOURCE is the indirect data source. During CETAP, it was mainly used to identify the original data collector; as with DDSOURCE, it was dominated by individuals' initials (and often was the same as DDSOURCE). The primary meaning has changed in the NARWC database to identify (1) the funding source of a survey program, (2) the overall project that a survey was a part of, (3) the organization that provided the data to the Consortium, or (4) the original data collector. In the list of codes that follow, those shown in italics are obsolete CETAP codes that are no longer in routine use. Some do not currently exist in the NARWC database, either because they were never used or because individual's initials were converted to an organizational abbreviation to minimize duplication. A few were post-CETAP—used on opportunistic right whale sighting datasets but not currently in use.

As with DDSOURCE, IDSOURCE is assigned after a dataset has arrived at GSO, so is "invisible" to data contributors. But it is something that some data users may be interested in obtaining in a dataset.

AMN = American Museum of Natural
    History

AMP = AMAPPS (Atlantic Marine
    Assessement Program for
    Protected Species)

BWS = Blandford Whaling Station

*CAM = Charles A. Mayo*

CAT = Right Whale PhotoID Catalog

CCS = Provincetown Center for Coastal
    Studies

*CET = CETAP Flight Team*

*CHS = Cape Hatteras National Park
    Service*

COE = U.S. Army Corps of Engineers

CRE = Coastal Research & Education
    Society of Long Island

*CRS = C. Robert Shoop*

CWI = Canadian Whale Institute

DFO = Dept. of Fisheries & Oceans, Canada

*EDM = Edward M. Mitchell*

EWS = Early Warning Survey (Southeast)

GMI = Geo-Marine, Inc.

GSC = Great South Channel Survey

*GWO = Guy W. Oliver, III*

*HEH = Herbert E. Hayes*

*HEW = Howard E. Winn*

*JGM = James G. Mead*

*JPR = J. Perran Ross*

MAQ = Mystic Aquarium

MBO = Manomet Bird Observatory

MCE = Massachusetts Clean Energy Center

*MRW = Minimum Right Whale Count*

NEA = New England Aquarium

NNE = NMFS, Northeast

NSE = NMFS, Southeast

*OGB = Ollie G. Brazier*

OTH = Other (miscellaneous)

PIR = Programme Intégré de Recherches sur les Oiseaux Pélagiques

*POP = POP Observers*

PSB = Protected Species Branch and North Atlantic Right Whale Sightings Surveys (NARWSS, NMFS, Northeast)

*RAR = Richard A. Rowlett*

RWC = Right Whale Consortium

*RXP = Robert Prescott*

SAS = Sightings Advisory System (NMFS, Northeast)

SCO = SCOPEX

SEA = Sea Education Association

*SKK = Steven K. Katona*

SRW = MMS Southeast Right Whale Project

*SSS = Samuel S. Sadove*

SWF = Sea Watch Foundation (U.K.)

TAG = Tagged whale locations (from NARWC catalog)

*TXA = Thomas A. Azarovitz*

UCG = U.S. Coast Guard

URI = University of Rhode Island

USN = U.S. Navy

*WAW = William A. Watkins*

WHO = Woods Hole Oceanographic Inst.

## 8.56. JDATE

JDATE is a synthetic variable created from MONTH and DAY (with YEAR factored in). It is the Julian date (sometimes referred to as :yearday")—the day of the year numbered from 001 on 1 January to 365 on 31 December (366 in leap years).

## 8.57. JELL

JELL is a synthetic variable with a value of "YES" or "NO" that identifies whether a sighting was recorded as "associated with jellyfish" in one of the BEHAVn codes.

## 8.58. LATDEG

LATDEG is the degrees of latitude. LATDEG is required for all records in all data types (See LATMIN). The usual convention is for north latitudes to be positive numbers and south latitudes to be negative numbers, but since there are no data from south of the equator in the database, there is no issue.

## 8.59. LATMIN

LATMIN is the minutes of latitude. LATMIN is required for all records in all data types. The standard for many years was for all data to be submitted in separate degrees and minute fields, with minutes to the nearest tenth. Given computer data-loggers and GPS accuracy, additional precision can be reasonably expected to be reliable, and data recording of latitudes and longitudes in decimal degrees has become more frequent. Data submissions in either degrees and minutes or decimal degrees will be acceptable, as they can be quickly inter-converted (the conversion is equally simple in both directions, and data users often prefer decimal degree output for GIS mapping). It is not necessary to go through the effort to convert one format to another prior to submission, nor to submit data containing both formats simultaneously. For data submissions in decimal degrees, at least

three decimal places are required (equivalent to minutes with one decimal place), with four or five desirable (provides the second decimal precision for minutes). If your original data have high precision, please do not round or truncate the data that are submitted to fewer than two decimal places for minutes or four decimal places for decimal degrees. Positions in decimal degrees to 8 or 10 decimal places are just silly, GPS units may display that many decimal places, but expecting precision to the millimeter or better is not quite realistic.

We fully expect in the near future to convert the entire database to latitudes and longitudes in decimal degrees, since most submissions come that way, most data-sharing outputs go out that way, and it will likely be simpler for any updated version of the data management system.

## 8.60. LATSEC

LATSEC is the seconds of latitude, which was apparently how the data were archived during CETAP (with LATMIN reduced to only the integer part). It is no longer in use (hundredths of minutes are more precise than seconds).

## 8.61. LEGGOOD

LEGGOOD is a one-digit code indicating whether a line-transect aerial survey line was "made good," with 1 = "no" and 2 = "yes." The CETAP definition of "made good" was that at least two-thirds of the track length was completed in acceptable survey conditions. LEGGOOD applies to an entire track, and should not be used to identify segments of good survey effort. Its purpose was to define whether a survey block had been completed during a particular survey window. If the first two lines were not made good, then the survey was aborted and the aircraft either returned to base or started a different block, depending on conditions. Keeping track of this was necessary only because of BLM contractual requirements to make good one survey of each block during each survey window, *i.e.*, it was essentially a "check-box." It was determined to be unnecessary for the NLPSC aerial surveys.

**8.62. LEGNO**

LEGNO is the aerial line-transect survey track number. Generally, tracks within a given survey block are pre-defined and numbered sequentially from 1. A survey would fly a random subset of the lines, either drawn completely randomly (CETAP surveys) or spaced systematically with a random start line (SCOPEX and other Great South Channel surveys, NLPSC surveys. WEA surveys).

**8.63. LEGSTAGE**

LEGSTAGE is a one-digit code for the stage of watch during a survey. LEGSTAGE is required for aerial and shipboard survey data (although not for all records), and is not allowed for opportunistic sighting data.

For POP surveys, LEGSTAGE is recorded independently of LEGTYPE (see below), with values for "begin watch," "continue watch," and "end watch." Since beginning and ending a watch period are instantaneous occurrences, they can only occur singly. And they must occur in logically consistent order—"1" can not follow "2," "2" can not follow "5" or a missing value, and "5" can not follow a missing value. **There were three additional values defined during CETAP that were rarely used and of limited value, so they were dropped from active use for the NARWC database (they are italicized in the code listing below and should <u>not</u> be used in new data).**

Both LEGSTAGE and LEGTYPE are much more complex for line-transect survey data, with LEGSTAGE only recorded during census tracks (LEGTYPE = 2), with one exception for off-transect sightings in vertical camera images. Based on experience, the presumption in analyses using quantified effort for SPUE calculations was that aerial line-transect observers maintained at least an informal watch on all cross-legs and transits when they were recording regular positions. This is less likely to be the case with computer data-loggers that run constantly, therefore a new method for recording on and off watch during cross-legs and transits was created for the Mass CEC/NLPSC surveys (see LEGTYPE). Those surveys also incorporated automated vertical aerial photography (mainly for sea turtles), requiring an additional code to differentiate visual sightings from

photographic sightings. Finally, sightings should not occur at events with LEGSTAGEs of "1" or "5" (or "3" or "4" for line-transect surveys), although this has often been allowed to slide through in the past. If you must log a sighting at the same location as a start- or end-watch event (*e.g.*, a bottlenose dolphin group in the surf just as you fly across the beach to start a survey), include duplicate events (but with different event numbers) so that the sighting clearly shows as occurring while on watch, or off-watch, whichever is correct.

For computer-logged datasets, it is not necessary to fill in the large number of "2" values between the begin- and end-watch events. There is a dBASE macro that will do that automatically, as long as the "1" and "5" values have been entered in the appropriate records.

Dedicated aerial survey:  recorded only during defined census lines
(LEGTYPE=2), except for 7
1 = begin line
2 = continue line
3 = break off line to circle
4 = resume line
5 = end line
6 = sighting by anyone other than an on-duty observer
7 = sighting detected in a vertical photograph

POP ship and aerial surveys: recorded independently of LEGTYPE
0 = off watch (optional if more efficient for data recording,
will be changed to blanks in archived data)
1 = begin watch period
2 = Continue watch period
5 = end watch period
NOTE: the following three should not be used in new datasets.
6 = *sighting by other crewmember during a watch period*
7 = *sighting by observer during an off-watch period*

*8 = sighting by other crewmember during an off-watch period*

## 8.64. LEGTYPE

LEGTYPE indicates the line type during line-transect or POP surveys. LEGTYPE is required for all records in all shipboard and aerial survey data, and is not allowed for opportunistic sightings. It is a very simple code for POP surveys, but becomes more complex for line-transect surveys or the newer intermediate format (relaxed line-transect) surveys. For shipboard POP surveys, the codes denote either "vessel underway" or "vessel not underway," (*i.e.*, drifting). For POP aerial surveys, there was only a single value available until 2003. After a Skymaster crash that resulted in the deaths of the pilot and three observers, NMFS instituted a two-pilot requirement for all aircraft contracted with NMFS funding. That eliminated the seat for a dedicated data recorder, and the observers could not maintain a full sighting log and keep their attention on their primary duties. Southeast surveys at that time switched to recording only large whales and ships. A new code was defined so those data could be differentiated (*e.g.*, they could be used in SPUE analysis for right whales, but not bottlenose dolphins or loggerhead turtles).

0 = line-transect (including relaxed) aerial, off-watch during transit, cross-leg, or circling

1 = line-transect (including relaxed) aerial, transit

2 = line-transect (including relaxed) aerial, survey line

3 = line-transect (including relaxed) aerial, cross-leg

4 = line-transect (including relaxed) aerial, other (circling)

5 = POP ship, underway

6 = POP ship, not underway

7 = POP aerial

9 = POP aerial, but with restricted data-recording

**8.65. LINKAGE**

LINKAGE (there were actually six separate linkage fields eventually created, LINKAGE1–LINKAGE6) is a discontinued CETAP variable. It was intended to identify sightings that were clearly associated with one another. Association was defined as something beyond mere occurrence at the same location, including, *e.g.*, dolphins bow-riding on large whales or mixed-species herds of delphinids. All linked sightings at a particular event were to be given the same linkage number, beginning with "001" within a given dataset and incrementing in sequence. In practice, the variable was confusing and was often "clarified" and re-defined (leading to a number of very amusing memos, with and without accompanying cartoons). It was very likely used inconsistently in the data, frequently ignored, and not completely reliable.

**8.66. LONGDEG**

LONGDEG is the degrees of longitude. LONGDEG is required for all records in all data types (See LATMIN for extended discussion and detail). The usual convention (as defined in ArcGIS software, but not in all mapping software) is for east longitudes to be positive numbers and west longitudes to be negative numbers. Since the CETAP and NARWC survey effort has all been in the western North Atlantic, all longitudes had been assumed to be west and were positive values. There was actually one catalog record with an east longitude—the sighting of "Porter" at the North Cape of Norway (Fig. 1). When it was transferred to the database as an opportunistic sighting, the result was that LONGDEG was a missing value. Later, additional opportunistic/historical right whale sightings in the Mediterranean were added. Since there were so few records, it was decided that modifying the entire database to accommodate the ArcGIS standard for east vs. west longitudes would not be worth the effort right now. Therefore, in the NARWC database, east longitudes are negative. (NOTE: in those cases, both LONGDEG and LONGMIN must be negative, or converting the longitude to decimal degrees will create erroneous values.)

88

**8.67. LONGMIN**

LONGDEG is the minutes of longitude, to two decimal places. LONGDEG is required for all records in all data types (See LATMIN for extended discussion and detail).

**8.68. LONGSEC**

LONGSEC is the seconds of longitude—only used in the CETAP data. (See LATSEC, also LATMIN for extended discussion and detail).

**8.69. MAMID**

MAMID is an obsolete CETAP variable, which was a three-digit numeric code number assigned to identified individuals. Numbers 001–799 were to be used for known animals (only 024–053 were ever assigned; all were humpbacks except for 047, which denoted a fin whale with a streamer or spaghetti tag). Numbers 800–899 were to be used for "special classification." Only one value was ever assigned: 800 = albino.

**8.70. MILL**

MILL is another synthetic behavior variable with a value of "YES" or "NO." It identifies whether "milling" or "circling" behavior a sighting was recorded for a sightings, based on both BEHAVn codes and ANHEAD.

**8.71. MONTH**

MONTH is the calendar month (01–12). Full date information is required for all records for all aerial and shipboard survey types. For opportunistic sighting records, full date information is nominally required, although the requirement can be a little flexible, especially when using sighting or other data from historical sources. Just MONTH and

YEAR, with DAY missing, still provides valuable information. One step lower is to use coded values in MONTH to represent seasons (13 = winter; 14 = spring; 15 = summer; 16 = fall). The lowest level allowed is for YEAR alone, with both MONTH and DAY missing.

## 8.72. NUMADULT

NUMADULT is the number of adults counted out of the total number of individuals observed at a sighting. It was one of six variables in the CETAP database for numbers of different age or sex classes (see below). Of the six, only NUMCALF (renamed from CALFHAT) was retained in the NARWC database.

## 8.73. NUMBER

NUMBER is the number of animals (or vessels, etc.) counted at a sighting. NUMBER is required for all sightings for all data types, and not allowed for non-sighting records. If the number of animals is not known (or for many pollution/human activity sightings where a number is neither logical nor practical), the field may be left blank, however in those cases the value for CONFIDNC must be "11." For field efforts where counts are collected in a high/low/best format, it would be "best" that would be put in here.

## 8.74. NUMCALF

NUMCALF (renamed from CALFHAT in the CETAP database) is the number of calves counted at a sighting. NUMCALF is optional for animal sightings for all data types, and not allowed for non-sighting records. This should be used only when the number of calves in a group can be counted precisely. Observers should not attempt to judge juveniles by relative body sizes, but should limit themselves to obvious calves, generally less than half the length of the accompanying adult. When calves are present, but an accurate count is not possible, this can be recorded using the BEHAVn fields (40 =

"mother with young"). It is not necessary to enter "0" into this field when no calves are observed. In addition, NUMCALF should be less than NUMBER; values that are equal to or greater than NUMBER will trigger an error message during Q/C processing (although sightings of unaccompanied whale calves with NUMBER = NUMCALF = 1 do occur and will be recognized as acceptable; in those cases it is very helpful for there to be a note along the lines of "calf alone" in the comments).

## 8.75. NUMFEMAL

NUMFEMAL is an obsolete CETAP variable for the number of females counted at a sighting (see NUMADULT).

## 8.76. NUMIMMAT

NUMIMMAT is an obsolete CETAP variable for the number of immatures counted at a sighting (see NUMADULT).

## 8.77. NUMMALE

NUMMALE is an obsolete CETAP variable for the number of males counted at a sighting (see NUMADULT).

## 8.78. NUMSUBAD

NUMSUBAD is an obsolete CETAP variable for the number of sub-adults counted at a sighting (see NUMADULT).

## 8.79. OLDVIZ

OLDVIZ is the discontinued CETAP visibility and weather code. It was originally named WEATHER, and the values were simply "clear," "fog," "haze," "rain," and

"snow." It was clear from the outset, however, that the intention was to use this field to describe visibility, with the threshold value for a "good" survey being visibility of at least 2 nautical miles. (E.g., visibility of 4 miles in haze was to be coded as "clear" rather than "haze.") The NARWC variable was changed to VISIBLTY to better reflect its purpose.

It was long recognized that combining visibility and descriptive weather conditions into a single field limited its usefulness. Beginning with the 2003–04 winter surveys, VISIBLTY was required to be submitted as the actual visibility (in n.mi.). In 2004, all of the older archived data were converted to the new format. Rather than arbitrarily assign visibility distances—OLDVIZ was created, the existing VISIBLTY codes transferred, and then VISIBLTY was changed to "–1." OLDVIZ is missing for all new records with VISIBLTY in actual distances.

1 = clear visibility for at least 2 nautical miles
2 = visibility less than 2 miles, fog
3 = <2 miles, haze
4 = <2 miles, rain
5 = <2 miles, snow

## 8.80. PHOTOS

PHOTOS was originally a simple yes/no during CETAP, indicating whether photographs of a given sighting existed in CETAP files. In the NARWC data, the field was been expanded to cover different types of photographic records, and to simply indicate whether photographs were taken of a sighting. PHOTOS is required for all sightings for all data types, and not allowed for non-sighting records. As with FLUKES, this field is of limited usefulness, since knowing about the potential presence of photos in a wide variety of institutional or personal collections has little value for the Consortium and for external data users.

PHOTOS would be more useful as a flag to identify any right whale sighting where there is at least one corresponding record (and possibly many) in the identification

92

catalog. Making that change would require some creative programming, and probably a significant time commitment, so it is not likely to be done any time soon.

> 1 = no
>
> 2 = yes, slides or prints (including digital)
>
> 3 = yes, cine
>
> 4 = yes, video (including digital)
>
> 5 = yes, more than one type

[3, 4, 5 seem to be generally ignored and just all pooled as 2]

## 8.81. PLATFORM

PLATFORM is a three-digit code for survey/sighting platform. PLATFORM is required for every record in all data types, however it is assigned during the process of entering a dataset into SAS at GSO, and is therefore invisible to data contributors. It is probably the variable with the longest list of possible codes, which continues to grow.

The ranges of codes set aside for particular types of platforms were created for CETAP. In some cases, we ran out of options within one range, but another range had far more unused values that needed. Therefore some ranges have been modified from their original definitions. A few codes have been re-defined entirely, so that the current code for a particular platform may be different from the one originally defined during CETAP.

**(020-059: U.S. Coast Guard Vessels)**

020 = Alert

021 = Vigilant

022 = Unimak

023 = Vigorous

024 = Ingham

025 = Active

026 = Tamaroa

027 = Taney

028 = Chilula

029 = Decisive

030 = Cherokee

031 = Duane

032 = Sherman

033 = Reliance

034 = Jefferson Island

035 = Juniper

036 = Hammerhead

037 = Northland

038 = Grand Isle

039 = Campbell

040 = Shearwater

041 = Cochito

042 = Dependable

043 = Tybee

044 = Spencer

045 = Hollyhock

046 = Dallas

047 = Tahoma

048 = Willow

049 = Seneca

050 = Chinook

051 = Mignon

052 = Escanaba

053 = Legare

054 = Finback

055 = Tigershark

056 = Albacore

057 = Diligence

058 = USCG Small Vessel

059 = unid./other USCG
     Cutter/Vessel


**(060-074: NOAA Vessels)**

060 = Albatross IV

061 = George B. Kelez

062 = Mt. Mitchell

063 = Delaware

064 = Advance

065 = Oregon II

066 = Nancy Foster

067 = Gordon Gunter

068 = Henry Bigelow

069 = Gloria Michelle

074 = misc./mixed/unknown
     NOAA vessels


**(075-089: Foreign Research Vessels)**

075 = Wieczno

076 = Anton Dohrn

077 = Argus

078 = Alliot

079 = Belagorsk

080 = Evrika


**(090-149: Institutional & Other**
**Research Vessels)**

090 = Henlopen

091 = Oceanus

092 = Annandale

093 = Endeavor

094 = Alert (EPA charter)

095 = Challenge

096 = Beluga

097 = Jere A. Chase

098 = Trident

099 = Nereid

100 = Halos

101 = Lauri Lee

102 = John M. Kingsbury (SML)

103 = Argo Maine

104 = Seward Johnson

105 = Indigo

106 = Lady Hammond

107 = Shearwater (CCS)

108 = Gannet (CCS)

109 = Hannah T (CCS/WHOI)

110 = Dixie (CCS)

111 = Pelican

112 = John B Heiser (SML)

113 = Silver (WCNE)

114 = Galatea (NEA)

115 = Tioga (WHOI)

116 = Shelagh

117 = Callisto (NEA)

118 = Connecticut (UCONN)

119 = Shackleton (CCS)

120 = Ibis (CCS)

121 = MysticetE (WCNE)

122 = Auk (SBNMS)

123 = Hugh Sharp (UDel)

124 = Junet (NEA)

125 = Marindin (UMass-Boston)

126 = misc. NEAQ charter

127 = misc./unknown WHOI

128 = Ocean Explorer

129 = Ocean Researcher

130 = USNS Bartlett

131 = Marlin

132 = Abel J

133 = Navaho

134 = Esperanza (Green-Peace)

135 = Aquamonitor (MWRA/

    Batelle)

136 = Selkie (RHIB)

137 = Easterly

139 = misc./other R/V

**(150-169: Sailing Research Vessels)**

150 = Regina Maris

151 = Westward

152 = Sirius

153 = Corwith Cramer

154 = Balaena (Hal Whitehead)

155 = Sedna IV

156 = Rosita

157 = Spirit of Massachusetts

**(170-249: Whale/Bird Watch Vessels)**

170 = Miss Ocean City

171 = Dolphin Fleet

173 = Viking Queen

174 = Viking Starship

175 = Super Squirrel Ii

176 = Voyager

177 = Finback

178 = Cetacean Quest

179 = Frances Fleet

180 = Miss Cape Ann

181 = Miss Gloucester

182 = Capt. John Fleet

183 = Elsie Manota/Grand Manan
      Whalewatch

184 = Norwood Boat Tours (N.S.)

185 = Quoddy Link (N.B.)

186 = Sea Watch Tours (N.B.)

187 = Portuguese Princess

188 = Seafarer (NEA)

189 = Granite State

190 = Friendship

191 = Bar Harbor/Allied Whale

192 = Tales Of The Sea

193 = Mingan Island Cetacean
      Study

199 = misc./other whale-watch
      vessel

**(250-264: CETAP Charter)**

250 = Stone Horse

251 = Walter E. Phipps

252 = Three Of A Kind

253 = When And If

254 = Flying Sorceress

255 = Tioga

**(265-289=Canadian Vessels)**

265 = Canadian whaling vessel,
      Blandford station

266 = Canadian Coast Guard

267 = D.F.O vessels (incl. charters)

280 = misc./unknown Canadian
      vessel

**(290-310: Commercial Fishing
      Vessels)**

290 = Christina M.

298 = NOAA fishery observer

299 = misc./other fishing vessel

**(311-374: Seismic, Geophysical, &
      Ocean-Drilling Vessels)**

311 = Maurice Ewing

312 = Western Monarch

313 = Veritas Viking

314 = Akademik Shatskiy

315 = Geco Diamond

316 = Geco Triton

317 = Geco Topaz

318 = Geco Sigma

319 = Geco Rho

320 = Western Orient

321 = Western Aleutian

322 = Polar Duke

323 = Zod 2

324 = unknown seismic vessel

325 = Fugro Explorer

326 = Dina Polaris

**(375-424: Charter Fishing Vessels)**

375 = Miss Ocean City

376 = Flying Sorceress

377 = Yankee Captain

378 = Sea Doll Meditation

379 = Lucky Seven

380 = Divecom III

381 = Sneak Attack

382 = Reuby

383 = Sea Holly

384 = Ezy Duz It

385 = Rominic

386 = Gulf of St. Lawrence snow
crab boat

399 = misc. fishing vessels/party
boats

**(425-474: Passenger Ferries)**

425 = Block Island

426 = Marine Evangeline

427 = Caribe

428 = Blue Nose

429 = The Cat

444 = misc./other ferry

**(475-525: Tugs And Work Boats)**

475 = Alaskan Seahorse

**(526-550: BLM Charter)**

526 = Sub Sig

527 = Edgerton

528 = Oceanus

529 = Eastward

530 = Elizabeth

531 = Atlantic Twin

**(551-599: Miscellaneous Vessels)**

551 = Walter E. Phipps

552 = Sunbeam

553 = Tioga

554 = Bagatelle

555 = State Of Maine

556 = Valhalla

557 = D'Sonoqua

558 = Shearwater (Nova Scotia)

559 = Todd & Cherida

560 = Song of the Whale

561 = Orion

562 = Cloud 9

563 = Captain's Lady

564 = Dominion Victory

570 = inflatable deployed from
other vessel

571 = misc. small boat/inflatable

572 = Campobello Whale Rescue
BOAT

575 = WHOI miscellaneous
(Watkins)

576 = Massachusetts
Environmental Police/Marine
Patrol

577 = U.S. Navy

578 = Maine Marine Patrol

580 = private yachts (sail/power)

581 = misc. recreational vessel

582 = dredge

583 = merchant vessel

584 = pilot boat

585 = tugboat

586 = misc. Coast Guard
    oppor/hist

587 = cruise ship

588 = misc. foreign R/V oppor/hist

589 = misc. institutional R/V
    oppor/hist

590 = misc. harbor master/police
    vessel

591 = misc. whale/bird watch
    oppor/hist

593 = misc. commercial F/V
    oppor/hist

594 = misc. charter F/V oppor/hist

595 = misc. merchant vessel
    opport.

596 = misc. naval vessel opport.

597 = misc. yacht oppor/hist.

599 = misc. other oppor/hist

**(600-619: Helicopters)**

600 = United Helicopters

601 = USCG ship-based

602 = Massachusetts
    Environmental Police

603 = USCG helicopter

604 = DFO helicopter

605 = misc. helicopter

**(620-625: Private Aircraft)**

620 = private drone

625 = private aircraft

**(626-639: Dedicated Aircraft)**

626 = Skymaster (Aero-Marine)

627 = AT-11 (Aero-Marine)

628 = Islander (New England
    Airways)

629 = Skymaster (Katona charter)

630 = Cessna 150 (Katona charter)

631 = Cessna 206 (CETAP recon
    charter)

632 = Piper Aztec (Aero-Marine)

633 = NOAA Twin Otter

634 = NOAA (charter) Wigeon

635 = NOAA (charter) Goose

636 = NOAA Shrike

637 = APEM high-res photo
    survey aircraft (Partenavia)

**(640-644: U.S. Coast Guard Aircraft)**

640 = Fisheries Patrol-Cape Cod

641 = Fisheries Patrol-Cape
    Hatteras

642 = Harbor Patrol

643 = Radiothermography (ART)

**(645-699: Miscellaneous Aircraft)**

645 = New England Airways

646 = Southeast misc.

647 = Bay of Fundy/Browns Bank
    misc.

648 = Cape Cod Bay misc.

649 = other Skymaster

650 = fishspotter

651 = Partenavia

652 = misc. low-wing, twin engine

653 = other Twin Otter

654 = Aircam (ASWH)

655 = misc. Canadian aircraft –
    DFO, Transport Canada,
    A&P, etc.

656 = research drone

661 = Metlife airship

662 = Fuji airship

663 = Airship Shamu

664 = Virgin Lightship

665 = U.S. Navy Airship

669 = misc. other Airship

691 = misc. Coast Guard
    oppor/hist

695 = Navy Anti-Submarine
    Warfare Environmental
    Prediction Services
    (ASWEPS)

699 = misc./other oppor/hist

**(700-999: Other)**

700 = shore station

701 = Whale Alert app

702 = DFO Whale Alert system

800 = misc./mixed platforms
    oppor/hist

801 = mixed platforms, photoid
    catalog

802 = tagged animal locations
    from catalog

803 = mixed platforms, stranding
    data

900 = unknown

## 8.82. POOP

POOP is another synthetic behavior variable with a value of "YES" or "NO." It identifies whether defecation was observed at a sighting. In the case of right whale records taken from the catalog, it could mean that feces were collected.

**8.83. PORTOBS**

PORTOBS is a code to identify the observer on the left side. PORTOBS, STAROBS, and SIGHTOBS were used during CETAP only for the line-transect aerial surveys and for the POP aerial surveys by the sea turtle group aboard Coast Guard aircraft. The objective was to use the information, if necessary in data analyses, to quantify inter-observer differences. The codes started out as one-digit numbers and quickly expanded into letters. Early in the NARWC studies, the field was expanded to two characters and additional observers' initials were used. The usefulness of the variable rapidly declined to nil. When last quantified, 136 different codes had been utilized, with most of them unidentifiable and no way to know when multiple observers may have had the same initials. Standard practice now is to delete all three of these fields from submitted data.

Keeping track of individual observers is only useful for line-transect surveys. In the event that such surveys are undertaken as part of NARWC research in the future, this issue could be revisited.

**8.84. RELBAR**

RELBAR is a discontinued CETAP variable denoting the relative bearing from the observer to a sighting, primarily designed for shipboard observers. Units were degrees, with the vessel heading = 000.

**8.85. REPEAT**

REPEAT is a discontinued CETAP variable used to flag repeated sightings of the same animal or group during a particular survey. The first group sighted more than once during a survey was assigned REPEAT=001 for all sightings. The second group seen multiple times was assigned REPEAT=002 for all sightings, and so on. The variable was added to the database in 1980, and apparently caused so much confusion that its use was discontinued in June 1981.

**8.86. SAG**

SAG is a synthetic behavior variable, with a value of "YES" or "NO," that identifies sightings of right whales involved in surface-active group behavior (*i.e.*, socializing or apparent courtship).

**8.87. SEASON**

SEASON is a synthetic variable denoting season, created from MONTH or MONTH and DAY. Values are the season names—winter, spring, summer, or fall (or autumn if requested), and can be abbreviated for users interested in minimizing output dataset size. Currently three different SAS macros exist with different definitions of seasons—one using calendar conventions (spring = 21 March – 20 June) and the other two using whole-month approximations (spring = April – June or spring = March – May). SAS code for other definitions could be created on request.

**8.88. SIDIST**

SIDIST is a discontinued CETAP variable denoting distance from the survey platform to a sighting. It was intended for use primarily by shipboard observers. Even though the SAS variable label indicated that distance was in kilometers, this was actually a coded variable (with completely ridiculous precision in the definitions).

| | | | |
|---|---|---|---|
| 1 = | 0 – 1/16 n.mi. | 0 – 380 ft | 0 – 115.8 m |
| 2 = | >1/16 – 1/8 n.mi. | >380 – 760 ft | >115.8 – 231.7 m |
| 3 = | >1/8 – 1/4 n.mi. | >760 – 1,520 ft | >231.7 – 463.3 m |
| 4 = | >1/4 – 1/2 n.mi. | >1,520 – 3,040 ft | >463.3 – 926.6 m |
| 5 = | >1/2 – 1 n.mi. | >3,040 – 6,080 ft | >926.6 – 1,853.2 m |
| 6 = | >1 – 3 n.mi. | >6,080 – 18,240 ft | >1853.2 – 5559.6 m |
| 7 = | >3 n.mi. | >18,240 ft | >5559.6 m |

### 8.89. SIGHTNO

SIGHTNO is the sighting number (see EVENTNO). SIGHTNO is required for all sighting records for all data types, and is not allowed for non-sighting records. Datasets submitted in dBase format will show SIGHTNO = 0 for all non-sighting records, but this is a recognized dBase quirk, and they will be converted to true missing values during the process of conversion to SAS. Sighting numbers are generally sequential from 1 within a file, and duplicate numbers within a file are not allowed (but see next paragraph). SIGHTNOs do not have to run continuously within a dataset. Some computer data-logging programs (*e.g.*, LOGGER) assign a sequential "sighting" number for every event when the data-recorder presses the key to force a record—including sightings, but also beginnings and ends of lines and watches, weather changes, altitude changes, etc. Those SIGHTNOs are deleted during data processing by searching for records where SPECCODE is missing but SIGHTNO > 0. The result is that there are gaps in the sequence, which is perfectly acceptable. On the other hand, duplicate SIGHTNOs are a recurring problem in submitted datasets. Sometimes sightings don't get recorded in the computer data at the time (usually because of things happening in quick succession), and are added in afterwards. When this happens, you need to be very careful that the added sighting does not have a duplicate SIGHTNO, even if it means manually changing all the subsequent records in that dataset. That is not even necessary unless your OCD is worse than mine. Go to the end of the file, find the last SIGHTNO used, and just assign something higher. Unlike EVENTNOs, SIGHTNOs do not need to be in order within a dataset.

During CETAP, sightings of non-target species (seals, sharks, sunfish, etc.) were assigned sighting numbers of 999 to facilitate removal prior to any analysis, since BLM was adamant that no resources were to be expended on those species. There are certainly duplicate 999 sightings in many survey files. In addition, when all of the archived data records with HUMANACT and/or DEBRIS codes were converted to sighting format, those record were similarly assigned SIGHTNO = 999.

102

**8.90. SIGHTOBS**

SIGHTOBS is a code identifying the observer making a sighting (see PORTOBS).

**8.91. SIZEGRP**

SIZEGRP is an obsolete CETAP variable coding the modal size of sub-groups within a particular sighting (see also GROUPS and STRUCTURE).

| | |
|---|---|
| missing = only one group | 4 = 5 – 10 animals per sub-group |
| 0 = no modal sub-group size | 5 = 11 – 20 animals per sub-group |
| 1 = 1 – 2 animals per sub-group | 6 = 21 – 100 animals per sub-group |
| 2 = 3 animals per sub-group | 7 = >100 animals per sub-group |
| 3 = 4 animals per sub-group | |

**8.92. SPECCHAR**

SPECCHAR is the original two-letter species code used in the CETAP data (see SPECCODE).

**8.93. SPECCODE**

SPECCODE is a four-letter code for the species sighted. SPECCODE is required for all sightings in all data types, and must be blank for all non-sighting records. SPECCODEs are essentially abbreviations of common names to make them easy to remember (following the long-established practice of field ornithologists). There are a few oddities forced by the need to avoid duplicates. If any contributor is still using the old dBASE file structures, this variable may be called CETSPPCD. Data contributors should never create new codes on their own, as this will create error messages during

quality-control review, but should make a request to the database manager if they believe that one or more new codes is necessary.

The following list is intended to be a master list of all species codes that have ever been used for both CETAP and NARWC data. Each line includes, in order, SPECCODE, SPECNAME, SPECCHAR, and SPECNUM, with missing values as underscores. The list is in four parts. The first includes all of the currently available codes for animal sightings (other than birds), sorted in order of TAXCODE (i.e., major categories). That is followed by the codes used during CETAP but not currently in use, with SPECCODE missing and the definition from the transcriber's manual substituted for SPECNAME. Any of those codes that were actually used in the CETAP (some were never used, and some only once or twice) would have been "translated" to one of the NARWC codes when the NARWC database was first created in 1986. The third part includes the codes for recording vessels, fishing gear, other human activities, and debris/pollution as sightings (TAXCODE=0). In that section, lines in italics are the codes created only to accommodate some of the CETAP HUMANACT codes **but that cannot be used now.** The fourth and final part includes the codes added in 2009 to accommodate bird sightings (TAXCODE=8), and subsequently updated to reflect modern species names and codes currently in use by the U.S. Fish & Wildlife Service.

**Active NARWC Codes for Animal Sightings (except birds):**

_____

**TAXCODE = 1: large whales**

| SPECCODE | SPECNAME | SPECCHAR | SPECNUM |
|---|---|---|---|
| BLWH | Blue Whale (*Balaenoptera musculus*) | CE | 17 |
| BOWH | Bowhead Whale (*Balaena mysticetus*) | — | — |
| BRWH | Bryde's Whale (*Balaenoptera brydei*) | CH | 20 |
| FIWH | Fin Whale (*Balaenoptera physalus*) | CF | 18 |
| GRWH | Gray Whale (*Eschrichtius robustus*) | — | — |
| HUWH | Humpback Whale (*Megaptera novaeangliae*) | BG | 11 |
| RIWH | North Atlantic Right Whale (*Eubalaena glacialis*) | BF | 10 |
| SEWH | Sei Whale (*Balaenoptera borealis*) | CG | 19 |
| SPWH | Sperm Whale (*Physeter macrocephalus*) | BE | 9 |

| | | | |
|---|---|---|---|
| SRWH | Southern Right Whale (*Eubalaena australis*) | — | — |
| UNBA | Unidentified *Balaenoptera* | CC | 15 |
| UNBS | Unidentified Bryde's or Sei Whale | — | — |
| UNFS | Unidentified Fin or Sei Whale | — | — |
| UNLW | Unidentified Large Whale | AB | 2 |
| UNRO | Unidentified Rorqual (Balaenopteridae) | CA | 13 |
| UNWH | Unidentified Whale (Size Unknown) | AA | 1 |

### TAXCODE = 2: medium whales

| | | | |
|---|---|---|---|
| BEWH | Beaked Whale (*Mesoplodon* sp.) | EC | 35 |
| BLBW | Blainville's Beaked Whale (*Mesoplodon densirostris*) | — | — |
| GEBW | Gervais' Beaked Whale (*Mesoplodon europaeus*) | — | — |
| GOBW | Cuvier's Beaked Whale (*Ziphius cavirostris*) | ED | 36 |
| KIWH | Killer Whale (*Orcinus orca*) | DG | 28 |
| MIWH | Minke Whale (*Balaenoptera acutorostrata*) | CI | 21 |
| NBWH | Northern Bottlenose Whale (*Hyperoodon ampullatus*) | EE | 37 |
| SOBW | Sowerby's Beaked Whale (*Mesoplodon bidens*) | — | — |
| TRBW | True's Beaked Whale (*Mesoplodon mirus*) | — | — |
| UNBW | Unidentified Beaked Whale | EA | 33 |
| UNMW | Unidentified Medium Whale | AC | 3 |

### TAXCODE = 3: small cetaceans

| | | | |
|---|---|---|---|
| ASDO | Atlantic Spotted Dolphin (*Stenella frontalis*) | — | — |
| BELU | Beluga (*Delphinapterus leucas*) | FA | 39 |
| BODO | Bottlenose Dolphin (*Tursiops truncatus*) | HE | 54 |
| CLDO | Clymene Dolphin (*Stenella clymene*) | KC | 66 |
| DSWH | Dwarf Sperm Whale (*Kogia sima*) | — | — |
| FKWH | False Killer Whale (*Pseudorca crassidens*) | DF | 27 |
| FRDO | Fraser's Dolphin (*Lagenodelphis hosei*) | HF | 55 |
| GRAM | Risso's Dolphin (*Grampus griseus*) | FD | 42 |
| HAPO | Harbor Porpoise (*Phocoena phocoena*) | FC | 41 |

| LFPW | Long-finned Pilot Whale (*Globicephala melas*) | DJ | 31 |
| MHWH | Melon-Headed Whale (*Peponocephala electra*) | DH | 29 |
| OBDO | Offshore Bottlenose Dolphin (*Tursiops truncatus*) | — | — |
| PIWH | Pilot Whale (*Globicephala* sp.) | DE | 26 |
| PSDO | Pan-Tropical Spotted Dolphin (*Stenella attenuata*) | — | — |
| PSWH | Pygmy Sperm Whale (*Kogia breviceps*) | — | — |
| PYKW | Pygmy Killer Whale (*Feresa attenuata*) | DI | 30 |
| RTDO | Rough-Toothed Dolphin (*Steno bredanensis*) | GG | 49 |
| SADO | Common Dolphin (*Delphinus delphis*) | JG | 63 |
| SFPW | Short-finned Pilot Whale (*Globicephala macrorhynchus*) | DK | 32 |
| SNDO | Spinner Dolphin (*Stenella longirostris*) | KB | 65 |
| SPDO | Spotted Dolphin (*Stenella attenuata/frontalis*) | JE | 61 |
| STDO | Striped Dolphin (*Stenella coeruleoalba*) | JF | 62 |
| UNBD | Unidentified Beaked Dolphin (*Delphinus*, *Stenella*, *Tursiops*) | | |
| UNBF | Unidentified Blackfish | DA | 22 |
| UNCW | Unidentified Common or White-sided Dolphin | GB | 44 |
| UNDO | Unidentified Dolphin/Porpoise | AD | 4 |
| UNGD | Spotted or Bottlenose Dolphin | GE | 47 |
| UNKO | Pygmy or Dwarf Sperm Whale (*Kogia* sp.) | FB | 40 |
| UNLD | Unidentified *Lagenorhynchus* | HB | 51 |
| UNSB | Unidentified Small Blackfish (*Feresa* or *Peponocephala*) | DD | 25 |
| UNST | Unidentified *Stenella* | JC | 59 |
| WBDO | White-Beaked Dolphin (*Lagenorhynchus albirostris*) | HD | 53 |
| WSDO | Atlantic White-Sided Dolphin (*Lagenorhynchus acutus*) | HC | 52 |

**TAXCODE = 4: other marine mammals**

| BESE | Bearded Seal (*Erignathus barbatus*) | — | — |
| GRSE | Gray Seal (*Halichoerus grypus*) | — | — |
| HASE | Harbor Seal (*Phoca vitulina*) | — | — |
| HGSE | Harp or Gray Seal | — | — |
| HOSE | Hooded Seal (*Cystophora cristata*) | — | — |

| | | | |
|---|---|---|---|
| HPSE | Harp Seal (*Pagophilus groenlandicus*) | — | — |
| MANA | Manatee (*Trichechus manatus*) | — | — |
| PINN | Unidentified Pinniped | — | — |
| POBE | Polar Bear (*Ursus maritimus*) | — | — |
| RISE | Ringed Seal (*Pusa hispida*) | — | — |
| UNSE | Unidentified Seal | XA | 79 |
| WALR | Walrus (*Odobenus rosmarus*) | — | — |

## TAXCODE = 5: sea turtles

| | | | |
|---|---|---|---|
| GRTU | Green Turtle (*Chelonia mydas*) | VD | 76 |
| HATU | Hawksbill Turtle (*Eretmochelys imbricata*) | VF | 78 |
| LETU | Leatherback Turtle (*Dermochelys coriacea*) | UB | 71 |
| LOTU | Loggerhead Turtle (*Caretta caretta*) | UC | 72 |
| ORTU | Olive Ridley Sea Turtle (*Lepidochelys olivacea*) | — | — |
| RITU | Kemp's Ridley Turtle (*Lepidochelys kempii*) | VE | 77 |
| UNTU | Unidentified Turtle | TA | 67 |

## TAXCODE = 6: sharks

| | | | |
|---|---|---|---|
| ANSH | Angel Shark (*Squatina dumerili*) | — | — |
| BASH | Basking Shark (*Cetorhinus maximus*) | YK | 90 |
| BLSH | Blue Shark (*Prionace glauca*) | YL | 91 |
| DUSH | Dusky Shark (*Carcharhinus obscurus*) | — | — |
| GHSH | Great Hammerhead Shark (*Sphyrna mokarran*) | — | — |
| HHSH | Hammerhead Shark (*Sphyrna* sp.) | YJ | 89 |
| LMSH | Long-finned Mako Shark (*Isurus paucus*) | — | — |
| MKSH | Mako Shark (*Isurus* sp.) | — | — |
| SDOG | Spiny Dogfish (*Squalus acanthias*) | — | — |
| SMSH | Short-finned Mako Shark (*Isurus oxyrinchus*) | — | — |
| THSH | Thresher Shark (*Alopias* sp.) | — | — |
| TISH | Tiger Shark (*Galeocerdo cuvier*) | — | — |
| UNSH | Unidentified/Other Shark | YA | 80 |

| | | | |
|---|---|---|---|
| WHSH | Whale Shark (*Rhincodon typus*) | — | — |
| WTSH | White Shark (*Carcharodon carcharias*) | — | — |

**TAXCODE = 7: other fish**

| | | | |
|---|---|---|---|
| BFTU | Bluefin Tuna (*Thunnus thynnus*) | — | — |
| BLFI | Bluefish (*Pomatomus saltatrix*) | — | — |
| CDRA | Chilean Devil Ray (*Mobula tarapacana*) | | |
| CNRA | Cow-Nosed Ray (*Rhinoptera bonasus*) | — | — |
| FLFI | Flying Fish (Exocoetidae) | — | — |
| MAHI | Mahi-mahi/Dolphin-fish (*Coryphaena hippurus*) | — | — |
| MARA | Manta Ray (*Mobula birostris*) | — | — |
| MOBU | Mobulid ray, not identified to species (*Mobula* sp.) | | |
| OCSU | Ocean Sunfish/Sharp-tailed Mola (*Mola mola/lanceolata*) | YC | 82 |
| OTBI | Other Billfish | YI | 88 |
| SCFI | Fish School | YF | 85 |
| SCRA | Schools of Rays | YD | 83 |
| SWFI | Swordfish (*Xiphius gladius*) | YH | 87 |
| TUNS | Unidentified Tuna | YG | 86 |
| UNFI | Unidentified/Other Fish | YB | 81 |
| UNRA | Unidentified/Other Ray | YE | 84 |
| WHMA | White Marlin (*Tetrapterus albidus*) | — | — |
| YFTU | Yellowfin Tuna (*Thunnus albacares*) | — | — |

**TAXCODE = 9: other**

| | | | |
|---|---|---|---|
| AMAL | American Alligator (*Alligator mississippiensis*) | — | — |
| JELL | Jellyfish | — | — |
| LMJE | Lion's-Mane Jellyfish (*Cyanea capillata*) | — | — |
| PMOW | Portuguese Man of War (*Physalia physalus*) | — | — |
| UNCE | Unidentified Cetacean | — | — |
| UNID | Unidentified Animal | — | — |
| UNMM | Unidentified Marine Mammal | — | — |

108

**Obsolete CETAP Codes:**

_____

| | | | |
|---|---|---|---|
| ____ | Unidentified large whale, not right whale | BA | 5 |
| ____ | Unidentified large whale, not sperm whale | BB | 6 |
| ____ | Unidentified large whale, not right or sperm whale | BC | 7 |
| ____ | Sperm or humpback whale | BD | 8 |
| ____ | Sperm, humpback, or right whale | BH | 12 |
| ____ | Unidentified rorqual, not minke whale | CB | 14 |
| ____ | Blue, fin, sei, or Bryde's whale | CD | 16 |
| ____ | Unidentified large blackfish | DB | 23 |
| ____ | Pilot or false killer whale | DC | 24 |
| ____ | *Mesoplodon* or *Ziphius* | EB | 34 |
| ____ | Sperm or beaked whale | EF | 38 |
| ____ | Unidentified beaked dolphin | GA | 43 |
| ____ | *Stenella* or *Tursiops* | GC | 45 |
| ____ | Rough-toothed of bottlenose dolphin | GD | 46 |
| ____ | Unidentified dolphin/porpoise, not *Grampus* | GF | 48 |
| ____ | Unidentified short-beaked dolphin | HA | 50 |
| ____ | Unidentified long-beaked dolphin | IA | 56 |
| ____ | *Stenella* or *Delphinus* | JA | 57 |
| ____ | Striped or bottlenose dolphin | JB | 58 |
| ____ | Unidentified *Stenella*, not spinner or Clymene dolphin | JD | 60 |
| ____ | Spinner or Clymene dolphin | KA | 64 |
| ____ | Unidentified large turtle | TB | 68 |
| ____ | Unidentified small turtle | TC | 69 |
| ____ | Unidentified large turtle, not leatherback | UA | 70 |
| ____ | Unidentified small turtle, not hawksbill or Kemp's ridley | VA | 73 |
| ____ | Unidentified small turtle, not hawksbill | VB | 74 |
| ____ | Unidentified small turtle, not Kemp's ridley | VC | 75 |

## Human Activity and Debris/Pollution Codes:

_____

| | | | |
|---|---|---|---|
| *AC-J[1]* | *Subsonic aircraft* | — | — |
| *AC-P* | *Propeller aircraft* | — | — |
| *AC-S* | *Supersonic aircraft* | — | — |
| *AC-T* | *Turboprop aircraft* | — | — |
| *BT-H* | *Heavy boat traffic* | — | — |
| *BT-L* | *Light boat traffic* | — | — |
| *BT-M* | *Moderate boat traffic* | — | — |
| *CABL* | *Cable/pipe laying* | — | — |
| CG-B | Coast Guard, buoy tender | — | — |
| CG-C | Coast Guard, cutter | — | — |
| CG-S | Coast Guard, security patrol boat | — | — |
| CG-U | Coast Guard, utility boat | — | — |
| CRSH | Cruise ship | — | — |
| DE-B | Debris/pollution, balloon(s) | — | — |
| DE-F | Debris/pollution, flotsam (natural origin) | — | — |
| DE-G | Debris/pollution, loose/"ghost" fishing gear | — | — |
| DE-J | Debris/pollution, jetsam (human origin) | — | — |
| DE-O | Debris/pollution, oil slick or sheen | — | — |
| DE-P | Debris/pollution, plastic | — | — |
| DE-R | Debris/pollution, rope/line | — | — |
| DE-S | Debris/pollution, Sargassum/seaweed patches or lines | — | — |
| DE-U | Debris/pollution, unspecified/other | — | — |
| DE-W | Debris/pollution, wood | — | — |
| *DIVE* | *Diver(s)* | — | — |
| DR-D | Dredge, dumping | — | — |
| DR-T | Dredge, in transit | — | — |
| DR-U | Dredge, unspecified | — | — |
| DR-W | Dredge, in operation | — | — |

_____

[1] **Do not use any of the italicized codes**

| | | | |
|---|---|---|---|
| *DU-G[1]* | *Garbage dumping* | __ | __ |
| *DU-T* | *Toxic waste dumping* | __ | __ |
| EXPL | Explosive discharge | __ | __ |
| FE-H | Ferry, high-speed | __ | __ |
| FE-S | Ferry, standard type | __ | __ |
| FE-U | Ferry, unspecified | __ | __ |
| FG-A | Fixed fishing gear, aquaculture pens or facilities | __ | __ |
| FG-C | Fixed fishing gear, crab pot | __ | __ |
| FG-D | Fixed fishing gear, drift net | __ | __ |
| FG-G | Fixed fishing gear, sink gill net | __ | __ |
| FG-I | Fixed fishing gear, inshore lobster | __ | __ |
| FG-L | Fixed fishing gear, pelagic long line | __ | __ |
| FG-O | Fixed fishing gear, offshore lobster | __ | __ |
| FG-U | Fixed fishing gear, unspecified type | __ | __ |
| FRNT | Front/visible water-mass boundary | __ | __ |
| FV-C | Fishing vessel, lobster/crab/other pot/trap fishery | __ | __ |
| FV-D | Fishing vessel, drift-netter | __ | __ |
| FV-F | Fishing vessel, unspecified/unknown fixed-gear | __ | __ |
| FV-G | Fishing vessel, gill-netter | __ | __ |
| FV-H | Fishing vessel, party ("head") boat | __ | __ |
| FV-L | Fishing vessel, long-liner | __ | __ |
| FV-P | Fishing vessel, purse-seiner | __ | __ |
| FV-S | Fishing vessel, shrimper | __ | __ |
| FV-T | Fishing vessel, trawler/dragger | __ | __ |
| FV-U | Fishing vessel, unspecified type | __ | __ |
| FV-W | Fishing vessel, whale-watching | __ | __ |
| FV-Z | Fishing vessel, factory trawler | __ | __ |
| HELO | Helicopter | __ | __ |
| JETS | Personal watercraft ("jet-ski") | __ | __ |
| KAYK | Kayak, canoe, rowboat | __ | __ |

[1] **Do not use any of the italicized codes**

| | | | |
|---|---|---|---|
| LE-V | Law enforcement/Security patrol vessel (including USCG) | — | — |
| METT | Meteorological tower | — | — |
| *MULT[1]* | *Multiple activities* | — | — |
| MV-B | Tug and barge | — | — |
| MV-C | Container ship | — | — |
| MV-L | Merchant vessel, large | — | — |
| MV-O | Tanker | — | — |
| MV-S | Merchant vessel, small | — | — |
| MV-T | Tug alone | — | — |
| MV-U | Merchant vessel, unspecified | — | — |
| MY-L | Motor yacht, large | — | — |
| MY-S | Motor yacht, small (but larger than RECV) | — | — |
| NV-L | Naval vessel, large | — | — |
| NV-S | Naval vessel, small | — | — |
| NV-U | Submarine | — | — |
| OI-D | Drilling rig | — | — |
| OI-L | Drill ship | — | — |
| OI-P | Production platform | — | — |
| OI-S | Seismic exploration ("air-guns") | — | — |
| OW-B | Oceanographic, weather, or other data buoy | — | — |
| PIBO | Pilot boat | — | — |
| RECV | Recreational vessel, motorboat | — | — |
| RV-G | Research vessel, geophysical/seismic profiling (with the airgun arrays deployed, otherwise use RV-L) | — | — |
| RV-L | Research vessel, large | — | — |
| RV-S | Research vessel, small | — | — |
| RV-U | Research vessel, unspecified | — | — |
| RV-W | Wave glider, AUV, or other autonomous sampler | — | — |
| SONO | Sonobuoy deployed | — | — |
| *SONR* | *Sonar in use* | — | — |

---

[1] **Do not use any of the italicized codes**

112

| | | | |
|---|---|---|---|
| SPFV | Sport-fishing vessel | — | — |
| SV-L | Sailing vessel, large (>50 feet) | — | — |
| SV-S | Sailing vessel, small (<50 feet) | — | — |
| SV-U | Sailing vessel, unspecified | — | — |
| *SWIM*[1] | *Swimmer(s)* | — | — |
| UNVE | Unknown vessel | — | — |
| WHAL | Whale-watching vessel | — | — |

**Bird Codes:** (updated to currently accepted codes and names; entries with question marks are codes used in the Manomet data that have still not been entirely figured out)

_____

| | | | |
|---|---|---|---|
| ABDU | American Black Duck (*Anas rubripes*) | — | — |
| ARTE | Arctic Tern (*Sterna paradisaea*) | — | — |
| ATBR | Atlantic Brant (*Branta bernicla*) | — | — |
| ATPU | Atlantic (Common) Puffin (*Fratercula arctica*) | — | — |
| AUSH | Audubon's Shearwater (*Puffinus lherminieri*) | — | — |
| BBPL | Black-bellied Plover (*Pluvialis squatarola*) | — | — |
| BCPE | Black-capped Petrel (*Pterodroma hasitata*) | — | — |
| BHGU | Common Black-headed Gull (*Larus ridibundus*) | — | — |
| BLGU | Black Guillemot (*Cepphus grylle*) | — | — |
| BLKI | Black-legged Kittiwake (*Rissa tridactyla*) | — | — |
| BLSC | Black Scoter (*Melanitta nigra*) | — | — |
| BLTE | Black Tern (*Chlidonias niger*) | — | — |
| BOGU | Bonaparte's Gull (*Larus philadelphia*) | — | — |
| BRNO | Brown Noddy (*Anous minutus*) | — | — |
| BRPE | Brown Pelican (*Pelecanus occidentalis*) | — | — |
| BRTE | Bridled Tern *(Sterna anaethetus)* | — | — |
| BSTP | Band-rumped Storm Petrel (*Oceanodroma castro*) | — | — |
| BUFF | Bufflehead (*Bucephala albeola*) | — | — |
| CAGO | Canada Goose (*Branta canadensis*) | — | — |

[1] **Do not use any of the italicized codes**

| | | | |
|---|---|---|---|
| CATE | Caspian Tern (*Sterna caspica*) | __ | __ |
| COEI | Common Eider (*Somateria mollissima*) | __ | __ |
| COLO | Common Loon (*Gavia immer*) | __ | __ |
| COMU | Common Murre (*Uria aalge*) | __ | __ |
| COSH | Cory's Shearwater (*Calonectris diomedea*) | __ | __ |
| COTE | Common Tern (*Sterna hirundo*) | __ | __ |
| DCCO | Double-crested Cormorant (*Phalacrocorax auritus*) | __ | __ |
| DOVE | Dovekie (*Alle alle*) | __ | __ |
| DOWI | Dowitcher sp. (*Limnodromus* sp.) | __ | __ |
| FOTE | Forster's Tern (*Sterna forsteri*) | __ | __ |
| GBBG | Great Black-backed Gull (*Larus marinus*) | __ | __ |
| GLGU | Glaucous Gull (*Larus hyperboreus*) | __ | __ |
| GRCO | Great Cormorant (*Phalacrocorax carbo*) | __ | __ |
| GRSC | Greater Scaup (*Athya marila*)? | __ | __ |
| GRSH | Great Shearwater (*Puffinus gravis*) | __ | __ |
| GRSK | Great Skua (*Catharacta skua*) | __ | __ |
| HERG | Herring Gull (*Larus argentatus*) | __ | __ |
| HOGR | Horned Grebe (*Podiceps auritus*) | __ | __ |
| ICGU | Iceland Gull (*Larus glaucoides*) | __ | __ |
| LAGU | Laughing Gull (*Larus atricilla*) | __ | __ |
| LBBG | Lesser Black-backed Gull (*Larus fuscus*) | __ | __ |
| LESP | Leach's Storm-Petrel (*Oceanodroma leucorhoa*) | __ | __ |
| LETE | Least Tern (*Sterna antillarum*) | __ | __ |
| LIGU | Little Gull (*Larus minutus*) | __ | __ |
| LTDU | Long-tailed Duck (*Clangula hyemalis*) | __ | __ |
| LTJA | Long-tailed Jaeger (*Stercoraria longicauda*) | __ | __ |
| MABO | Masked Booby (*Sula dactylata*) | __ | __ |
| MAGW | Marbled Godwit (*Limosa fedoa*) | __ | __ |
| MALL | Mallard (*Anas platyrhynchus*) | __ | __ |
| MASH | Manx Shearwater (*Puffinus puffinus*) | __ | __ |
| NOFU | Northern Fulmar (*Fulmarus glacialis*) | __ | __ |

| | | | |
|---|---|---|---|
| NOGA | Northern Gannet (*Sula bassanus*) | — | — |
| PAJA | Parasitic Jaeger (*Stercoraria parasiticus*) | — | — |
| PEEP | Unidentified Shorebird | — | — |
| POJA | Pomarine Jaeger (*Stercoraria pomarinus*) | — | — |
| RAZO | Razorbill (*Alca torda*) | — | — |
| RBGU | Ring-billed Gull (*Larus delawarensis*) | — | — |
| RBME | Red-breasted Merganser (*Mergus serrator*) | — | — |
| RBTR | Red-billed Tropicbird (*Phaethon aethurus*) | — | — |
| REKN | Red Knot (*Calidris canutus*) | — | — |
| REPH | Red Phalarope (*Phalaropus fulicaria*) | — | — |
| RNPH | Red-necked Phalarope (*Phalaropus lobatus*) | — | — |
| ROST | Roseate Tern (*Sterna dougalli*) | — | — |
| ROYT | Royal Tern (*Sterna maxima*) | — | — |
| RTLO | Red-throated Loon (*Gavia stellata*) | — | — |
| RUTS | Ruddy Turnstone (*Arenaria interpres*) | — | — |
| RWBL | Red-winged Blackbird (*Agelaius phoeniceus*) | — | — |
| SAGU | Sabine's Gull (*Xema sabini*) | — | — |
| SATE | Sandwich Tern (*Sterna sandvicensis*) | — | — |
| SOSH | Sooty Shearwater (*Puffinus griseus*) | — | — |
| SOTE | Sooty Tern (*Sterna fuscata*) | — | — |
| SPPL | Semipalmated Plover (*Charadrius semipalmatus*) | — | — |
| SPSK | South Polar Skua (*Catharacta maccormicki*) | — | — |
| SUSC | Surf Scoter (*Melanitta perspicillata*) | — | — |
| TBMU | Thick-billed Murre (*Uria lomvia*) | — | — |
| UNAL | Unidentified Alcid? | — | — |
| UNBI | Unidentified Bird | — | — |
| UNCO | Unidentified Cormorant (*Phalacrocorax* sp.) | — | — |
| UNCT | Unidentified ? (crested tern?) | — | — |
| UNDU | Unidentified Duck | — | — |
| UNGO | Unidentified Goose | — | — |
| UNGU | Unidentified Gull | — | — |

| | | | |
|---|---|---|---|
| UNJA | Unidentified Jaeger (*Stercoraria* sp.) | — | — |
| UNLA | Unidentified Large Alcid? | — | — |
| UNLG | Unidentified Larus Gull (*Larus* sp.) | — | — |
| UNLO | Unidentified Loon (*Gavia* sp.) | — | — |
| UNLS | Unidentified Large Shearwater? | — | — |
| UNME | Unidentified Merganser | — | — |
| UNMU | Unidentified Murre (*Uria* sp.) | — | — |
| UNPH | Unidentified Phalarope (*Phalaropus* sp.) | — | — |
| UNPL | Unidentified Plover | — | — |
| UNSA | Unidentified Small Alcid? | — | — |
| UNSC | Unidentified Scoter/Scaup? | — | — |
| UNSK | Unidentified Skua (*Catharacta* sp.) | — | — |
| UNSP | Unidentified Storm-Petrel | — | — |
| UNSS | Unidentified Small Shearwater? | — | — |
| UNSU | Unidentified ? | — | — |
| UNSW | Unidentified Shearwater | — | — |
| UNTE | Unidentified Tern (*Sterna* sp.) | — | — |
| WFSP | White-faced Storm-Petrel (*Pelagodroma marina*) | — | — |
| WHIM | Whimbrel (*Numenius phaeola*) | — | — |
| WISP | Wilson's Storm-Petrel (*Oceanites oceanicus*) | — | — |
| WTTR | White-tailed Tropicbird (*Phaethon lepturus*) | — | — |
| WWSC | White-winged Scoter (*Melanitta fusca*) | — | — |

## 8.94. SPECNAME

SPECNAME is a synthetic variable created by a SAS macro from SPECCODE, either for inclusion in output datasets directly, or for creation of a SPECCODE look-up table to be provided with an output dataset (see SPECCODE). SPECNAME includes both the common name and the scientific name, if applicable (as included in the SPECCODE lists above). In the early years of the NARWC, different macros were used to assign separate COMNAME and SCINAME fields for all copies of the database

116

transferred to NMFS. At NMFS request (they had a strong aversion to blank values), SCINAME for unidentified sightings was "NOT APPLICABLE," and both COMNAME and SCINAME for all non-sightings were "NO SIGHTING." Those macros still exist, but do not include any SPECCODEs assigned after their use was discontinued.

## 8.95. SPECNUM

SPECNUM is a two-digit numeric species code, created from SPECCHAR, used in the archived version of the CETAP database (see SPECCODE).

## 8.96. STAROBS

STAROBS is a code identifying the observer on the right side of the plane during an aerial survey (see PORTOBS).

## 8.97. STRATUM

STRATUM is a one-character code to identify the depth stratum (or other stratum/subset) of a stratified or subdivided line-transect aerial survey block (see BLOCK). Beginning in 1980 (year 2 of the project), the CETAP aerial surveys were stratified by depth, and sampling intensity was also stratified—based on cetacean densities from the first-year surveys. The defined stratum boundaries were the 20- and 50-fathom isobaths. At the end of the study, the 1979 survey data were post-stratified for calculating the final 3-year average density and abundance estimates. Scotian Shelf right whale survey blocks designed in 1987 were stratified, but not by depth, into two halves that could be made good independently. The 1989-1992 MMS-funded Florida surveys incorporated four lines approximately parallel to the shoreline. The "inshore lines" were 1/2 and 4 miles offshore; the "offshore" lines were 8 and 12 miles off. After the first year of the 2011–15 NLPSC surveys the survey area was expanded into the Rhode Island portion of the Wind Energy Area, so a stratum code was added to differentiate from the original and the added portion of the study area.

X = 0-20 FATHOMS

Y = 20-50 FATHOMS

Z = >50 FATHOMS

0 = NON-STRATIFIED AERIAL SURVEY BLOCK

A,B = SCOTIAN SHELF BLOCK HALVES

I = FLORIDA, INSHORE

O = FLORIDA, OFFSHORE

M = NLPSC, YEAR 2+, MARTHA'S VINEYARD (SAME AS
    YEAR 1 SURVEYS)

R = NLPSC, YEAR 2+, RHODE ISLAND (ADDED)

## 8.98. STRIP

STRIP is a two-digit code identifying the right-angle distance interval of a given sighting from the trackline for a line-transect aerial survey. STRIP is required for on-effort animal sightings during line-transect aerial surveys (*i.e.*, LEGTYPE=2, LEGSTAGE=2), optional for human activity/debris sightings during those same legs, and not allowed for any other sighting or non-sighting events. Some aerial survey teams have found it useful to records STRIPs for sightings during transits, cross-legs, and circling to maintain both consistency and proficiency, however those will not be maintained in the archived database. Sightings were classified into intervals using calibrated markings on the rear edge of the forward observation bubble in the AT-11 or on the wing struts of the Skymaster (see Kenney and Scott, 1981 for details). Originally, the closest interval was out to 1/4 n.mi., but it was subsequently split at 1/8 n.mi. The intervals beyond 1 n.mi. differed between the platforms, with only one classification for the AT-11 and a split at 2 n.mi. for the Skymaster. In the codes, odd numbers indicate the left (port) side of the track and even numbers indicate the right (starboard) side. Code 0 only applies to the AT-11, because of restricted downward visibility in the Skymaster. That restricted downward visibility also means that distances for the Skymaster in the CETAP data (not the NLPSC or WEA data) are actually measured from about 1/8 mile to either side of the survey line.

0 = directly on the trackline

1,2 = 0–1/4 n.mi.

3,4 = 0–1/8

5,6 = 1/8–1/4

7,8 = 1/4–1/2

9,10 = 1/2–3/4

11,12 = 3/4–1

13,14 = >1 (AT-11)

13,14 = 1– 2 (Skymaster)

15,16 = >2 (Skymaster)

A new set of values was defined for the NLPSC/Mass CEC surveys that began in October 2011, which were also flown using a Skymaster.

1,2 = <1/8 n.mi.

3,4 = 1/8 – 1/4

5,6 = 1/4 – 1/2

7,8 = 1/2 – 1

9,10 = 1 – 2

11,12 = 2 – 4

13,14 = > 4

## 8.99. STRK

STRK is a synthetic behavior variable, with a value of "YES" or "NO," that identifies sightings of animals that had been struck by a vessel.

## 8.100. STRUCTUR

STRUCTUR is a discontinued CETAP code describing the geometric relationships of the animals in a defined group of cetaceans. It was generally called "herd type" by the CETAP aerial observer crew, and it was somewhat subjective.

0 = diffuse or intermediate structure

1 = clustered or bunched

2 = multiple sub-group clusters

3 = vee formation

4 = triangle formation

5 = lateral formation (in a rank, flank to flank)

6 = linear formation (head to tail)

7 = oval formation

8 = four-sided formation (diamond, square, rectangle)

9 = diagonal or echelon formation

## 8.101. SURFTEMP

SURFTEMP is the sea surface temperature (SST) in degrees C at an event. SURFTEMP is optional for all records. The variable was originally called WTEMP during CETAP. SSTs could be measured remotely by radiometer during an aerial survey, or in situ during a shipboard survey by a bucket thermometer or by a more automated system. SURFTEMP should represent the SST actually measured at the location specified by the latitude/longitude coordinates for that event, therefore practices such as, *e.g.*, entering the SST from a weather buoy as reported over NOAA Weather Radio in event 1 for an aerial survey are not acceptable. Temperatures are archived in Celsius, but may be submitted in either Celsius or Fahrenheit (note the units on the cover sheet).

SSTs measured from airborne radiometers have been erratic in some datasets, with values that are anomalously high and excessively variable. As yet, no satisfactory

solution for detecting and removing bad or suspicious data has been developed. Data users are cautioned about the potential for unreliable SST data.

## 8.102. S_LAT

S_LAT is the latitude (in decimal degrees) of the exact position of a sighting recorded during a line-transect aerial survey, much like the ALATDEG, ALATMIN, and ALATSEC fields during CETAP. It was added for the 2011-12 Northeast Large Pelagics Survey Consortium (MassCEC) aerial surveys, with the intention of calculating right-angle sighting distance from the trackline and exact positions—both when the strip number could not be recorded and as a double-check of the accuracy of estimating strips.

## 8.103. S_LONG

S_LONG is the longitude (in decimal degrees) of the exact position of a sighting recorded during a line-transect aerial survey, much like the ALONDEG, ALONMIN, and ALONSEC fields during CETAP. Although the general convention is to represent West longitudes as negative values for GIS mapping purposes, they are positive values in the Consortium database to be consistent with LONGDEG and LONGMIN. Trigonometric functions in SAS require them to be converted from degrees to radians, anyway.

## 8.104. S_TIME

S_TIME is the time of an exact-location sighting recorded during a line-transect survey. It was added after the first year of the NLPSC surveys when questions arose about the accuracy of using exact locations to verify accuracy of STRIP estimates, when a relatively long time may have elapsed between the initial sighting from the track-line and the subsequent sighting during circling.

**8.105. TAGLOC**

TAGLOC is a discontinued CETAP code for the location of a tag observed on a marine mammal or sea turtle. See also the other three associated TAG____ variables.

| | |
|---|---|
| 1 = dorsal fin | 5 = left front |
| 2 = body | 6 = right front |
| 3 = fluke | 7 = left rear |
| 4 = head | 8 = right rear |

**8.106. TAGMAT**

TAGMAT is a discontinued CETAP code for the material comprising a tag observed on a marine mammal or sea turtle. See also the other three associated TAG____ variables.

| | |
|---|---|
| 1 = plastic | 3 = other |
| 2 = metal | |

**8.107. TAGNUM**

TAGNUM is a discontinued CETAP field used for recording the number of a tag recovered from a stranded sea turtle, with up to ten characters (letters or numbers). See also the other three associated TAG____ variables.

**8.108. TAGTYPE**

TAGTYPE is a discontinued CETAP code for the type of tag observed on a marine mammal or sea turtle. See also the other three associated TAG____ variables.

| | |
|---|---|
| 0 = unmarked | 2 = disk tag |
| 1 = spaghetti tag | 3 = Discovery tag |

| 4 = brand | 7 = radio tag |
| 5 = roto tag | 8 = natural mark |
| 6 = metal crimp | 9 = streamer tag |

## 8.109. TAXCODE

TAXCODE is a one-digit code to identify the general species category of a sighting. TAXCODE is required for all sightings, however it is assigned by a SAS macro during data entry at GSO, and is therefore invisible to data contributors. During CETAP, the code was limited to 1 = large cetacean, 2 = small cetacean, 3 = turtle, and 4 = other, however its use was expanded in the NARWC database. When we shifted from recording vessels, gear, other human activities, and debris/pollution using HUMANACT and DEBRIS to recording them as sightings, an additional value for TAXCODE was also created.

1 = large cetacean (includes "unidentified whale")
2 = medium cetacean (minke, beaked, and killer whales)
3 = small cetacean
4 = other marine mammal (seals, manatee, polar bear)
5 = sea turtle
6 = shark
7 = other fish
8 = bird
9 = other/unknown
0 = vessel, gear, human activity, debris/pollution

## 8.110. TAXTYPE

TAXTYPE is a synthetic variable, simply an expansion of TAXCODE to spell out the taxonomic categories to make it easier for data users.

**8.111. TIME**

TIME is the clock time of an event, in EST, hhmmss, in 24-hour format. TIME is required for all records in all aerial and shipboard survey data, and optional (although desirable and strongly encouraged) for opportunistic sighting records. TIME was formerly a four-digit field (hhmm), but was expanded to six digits to accommodate the greater precision possible with computer data-loggers. Four-digit times may still be submitted; these will have "00" appended before archival. Do not, however, submit a set of data files that mixes four-digit and six-digit times, which is a guarantee of problems.

Unlike the date format, dBASE does not support a defined time format like Excel or Access, and times submitted in that format will not be usable after conversion. The Excel/Access time format includes an embedded date. The actual value stored in an Excel date- or time-formatted field is the number of days since midnight on 1 January 1900, with the time in the fractional part of the number. Excel times are translated into dBase as only the date components, which are always 01/00/1900. Times should be submitted simply as six-digit numbers, without intervening colons or any AM/PM labels.

Although essentially arbitrary, the standard that was defined during CETAP and that has been maintained ever since is to archive all data in Eastern Standard Time. There is a space on the cover sheet to note the time zone used to record the data. A common error has been to enter, *e.g.*, "EST" on the cover sheet, when the data were actually recorded during the summer (Eastern Daylight Time/EDT). If a set of survey files submitted together were collected using local time overlapping one of the "spring ahead–fall back" dates, it would be helpful to note the date of the time change on the cover sheet. If you are using time from your GPS output in GMT, you don't have to be concerned with Daylight Saving Time. In the long run, I expect that the database times will all be converted to GMT and all contributors will be required to forget time zones and submit everything in GMT.

Another problem that has come up with TIME in submitted data seems to arise when times from multiple sources are merged into a single dataset. Examples would be using a GPS to log routine locations while using a computer to record sightings, or separate observers maintaining separate written logs, each using their own watch for

times. If the two clocks are not synchronized, the result after merging the two datasets and sorting by time can be that the survey's path ends up as a strange zig-zag back and forth along the track. If the two datasets are different by an hour (*e.g.*, when the time zones don't match), the track can be truly bizarre. If you do plan to use multiple time sources during a survey, make sure they are synchronized exactly at the start of each day. If you are going to stick with only one—use the GPS time; it is extremely precise.

## 8.112. TIMEHR

TIMEHR and TIMEMIN are subsets of TIME with only the hours and minutes parts, respectively. They were included as separate variables in the CETAP data but are not in the NARWC database. They can easily and quickly be created whenever necessary for a computation (*e.g.*, elapsed time between events).

## 8.113. TIMEMIN

See TIMEHR.

## 8.114. TMSOURCE

TMSOURCE is an obsolete CETAP code for the method used to obtain turtle measurements (see CARALEN, CARAWID, CARAWT).

      1 = direct measurement (lab or field)
      2 = indirect measurement (*e.g.*, photographic)
      3 = estimate

## 8.115. TRACKNUM

TRACKNUM is a discontinued CETAP variable for the number of animals originally sighted from the trackline. It was used only for on-effort sightings during line-transect aerial surveys. The field was added to the database in 1980 in an attempt to better

assess how many additional animals were sighted relative to the duration of circling. For sightings that were circled, both NUMBER and TRACKNUM were filled in. By definition, TRACKNUM ≤ NUMBER. For sightings that were not circled, only TRACKNUM was filled in and NUMBER was left blank. When the CETAP database was converted to start the NARWC database, all of the missing NUMBERs were filled with the appropriate TRACKNUM values.

**8.116. TYPE**

TYPE is a synthetic variable used to identify the general type of data for interested users, rather then forcing them to decipher the different FILEID categories. The possible output values are "aerial," "shipbd," and "opport."

**8.117. VISIBLTY**

VISIBLTY is the estimated clear visibility, in nautical miles, during a survey. VISIBLTY is required for all on-watch records in aerial and shipboard survey data, and is optional for all other records (although encouraged for all records in survey data). It was formerly a one-digit code indicating whether or not the visibility was at least 2 nautical miles (the standard for acceptable survey conditions that was defined during CETAP) and, if not, the weather conditions causing the reduced visibility. It was not particularly useful for anything beyond the effort threshold, so in 2004 it was modified to use the actual visibility distance, estimated as precisely as possible. The general descriptive weather conditions were then separated into a new field (see WX). For all data records created under the old system, the codes were moved into a newly defined field called OLDVIZ and VISIBLTY was set at -1 as an identifying flag. Data may be submitted in any units; submissions in other than nautical miles should be clearly noted in the "comments" block on cover sheet. If the visibility conditions differ significantly on the opposite sides of the airplane or ship track, use the average.

One of the intentions with this variable was to allow recording visibility with more precision. An alternative model using codes for intervals (which would be finer for

shorter visibility distances) was rejected in favor of using visibilities estimated down to the closest 0.1 n.mi. precision. It hasn't quite worked out that way yet.

As originally established (Kenney, 2003), the maximum allowable value was to be 5.0 n.mi., since that was well beyond the 2 n.mi. effort threshold. The value used to flag records where OLDVIZ was to be used would have been 9. The NMFS aerial survey crew objected, and wanted to record their visibility estimates out to 20-30+ n.mi. That made using 9 for the OLDVIZ flag impossible, but making the field one character wider enabled using -1.

The name VISIBLTY was not created because of spelling difficulties—SAS has an eight-character limit for variable names. When submitting data in any database software format, all variable names must be properly misspelled according to the defined standard misspelling, or the variable(s) in question will be treated as missing during data entry.

Because VISIBLTY is used along with BEAUFORT to define "acceptable" survey effort, it should be treated as important information. It is disconcerting to see day after day of aerial survey data for a particular area with the visibility never varying from 5.0 n.mi. Part of the problem could be the originally defined maximum, but that can't be the case when the visibility stays constant even though the weather varies between clear, gray, haze, showers, snow flurries, etc.

## 8.118. WAKE

WAKE is a synthetic behavior variable, with a value of "YES" or "NO," that identifies sightings of animals observed to be swimming in the wake of a vessel.

## 8.119. WATCOLOR

WATCOLOR is a discontinued CETAP code for the water color observed at a particular event. It was rarely ever used, and was discontinued after 1979.

| | |
|---|---|
| 1 = green | 3 = gray |
| 2 = blue | 4 = red |

|         |                |                 |              |
|---------|----------------|-----------------|--------------|
| 5 = chalky |             | 7 = luminescent |              |
| 6 = brown  |             |                 |              |

## 8.120. WEATHER

WEATHER is the original CETAP variable name for the field that first changed to VISIBLTY and then to OLDVIZ.

## 8.121. WHLR

WHLR is a synthetic variable, with a value of "YES" or "NO," that identifies records of animals that were killed by whalers. The majority of these were records of the Blandford (Nova Scotia) whaling station obtained from the Canadian Dept. of Fisheries and Oceans, however some are historical records extracted from old newspaper articles of from Smithsonian files.

## 8.122. WINDDIR

WINDDIR is a discontinued CETAP code for the wind direction at an event, using a 16-point compass rose.

| | | | |
|---|---|---|---|
| 00 = N | [349 – 011] | 10 = SW | [214 – 236] |
| 01 = NNE | [012 – 033] | 11 = WSW | [237 – 258] |
| 02 = NE | [034 – 056] | 12 = W | [259 – 281] |
| 03 = ENE | [057 – 078] | 13 = WNW | [282 – 303] |
| 04 = E | [079 – 101] | 14 = NW | [304 – 326] |
| 05 = ESE | [102 – 123] | 15 = NNW | [327 – 348] |
| 06 = SE | [124 – 146] | 16 = variable | |
| 07 = SSE | [147 – 168] | 98 = not available | |
| 08 = S | [169 – 191] | 99 = no wind | |
| 09 = SSW | [192 – 213] | | |

### 8.123. WTEMP

WTEMP is the original CETAP variable name for SURFTEMP.

### 8.123. WX

WX is a code for descriptive weather conditions, added in 2004 when the new VISIBLTY code was created. WX is optional, although strongly encouraged for all types of survey data.

B = both rain (or other precipitation) and fog

C = clear

D = drizzle

F = fog

G = gray (heavy overcast and dark, but no precipitation)

H = haze

L = light rain, intermittent showers

P = patchy fog

R = rain

S = snow

T = thunderstorms, squalls

X = not recorded

### 8.124. YEAR

YEAR is the calendar year. YEAR is required for all records for all data types (even an old historical sighting record is of little or no value without at least the year). YEAR was formerly a two-digit field, but was expanded to four digits to accommodate the change from 1999 to 2000. (See DAY).

# 9. DIGITIZED BACK-UP DATA

On many occasions it has been necessary, not to mention extremely helpful, to dig into the old paper files stored in the lab at GSO to check on something (Was that really a false killer whale? Who reported that white-beaked dolphin sighting offshore of North Carolina in 1979? Were any behaviors noted in the logs but not entered into the data for all those basking shark sightings in 1980 and 1981?) Originals (and copies) of all of the data files from CETAP were stored in bankers' file boxes in the trailer next to the lab, and they all had to be piled up on tables in the lab after the dean decided the campus would be more attractive without a bunch of old trailers. Newer data files filled up six file cabinets. All of that took up a lot of space, and it was obvious that none of it would survive if I retired completely.

Beginning during the summer of 2016, we began the process of scanning into PDF all of those old paper files. I hired an undergraduate work-study student at first, while there was some funding available, and then began working on it myself as time permitted. At first, I was very careful to make the student scan different parts of the field data into separate PDFs (using "DocumentType FILEID.pdf" as a filename convention) and then store those in an individual folder with the FILEID for the title. For example, the first set of scans for the first CETAP dedicated aerial survey on 16 January 1979 is in a folder called "A179016." That folder contains eight separate documents (see the discussion on page 16 about the different documents that constituted a complete aerial survey file)—

CoverSht A179016.pdf
FieldNotes A179016.pdf
NavLog A179016.pdf
PhotoLog A179016.pdf
Report A179016-A178043.pdf
SurveyLog A179016.pdf
TempLog A179016.pdf
Transcription A179016.pdf

There are other document types for other data files. When I started doing the scanning myself, I decided this was all way too much trouble, so I would just stack all of the forms together (after first removing all of the dang staples and cursing myself for stapling things together before filing them) and scan them into a single document with just the FILEID for a file name. So any particular set of scans would no longer be in separate folders.

The scanner that I've been using (Fujitsu ScanSnap iX500, which I heartily endorse) is capable of duplex scanning (i.e., both sides of a page simultaneously) and of skipping blank pages when "duplex" is selected, but many pieces of data files were printed on recycled paper, so the blank backs of pages were not really blank. It is also capable of automatically rotating pages to the proper orientation, but is not infallible at that task—so many are sideways or upside-down. I've made no attempt at fixing either of those issues. If you happen to be looking up something in a particular PDF (and using the full Acrobat software instead of just the Reader), feel free to rotate pages or remove obvious junk pages while you're in there.

All of those scanned files are backed up onto an external hard drive in a folder called SUPPORTING DATA. The good news is that the job is complete, so every paper log is now backed up in PDF format. The folder structure at the present time looks like the following (not including all of the individual FILEID folders:

**SUPPORTING DATA**
   CETAP
         BOX 1-CETAP dedicated aerial surveys, 1979, part 1
         BOX 2-CETAP dedicated aerial surveys, 1979, part 2
         BOX 3-ETAP dedicated aerial surveys, 1980, part 1
         BOX 4-CETAP dedicated aerial surveys, 1980, part 2
         BOX 5-CETAP dedicated aerial surveys, 1981
         BOX 6-CETAP POP aerial surveys, 1978-1981
         BOX 7-CETAP POP ship surveys, part 1 (2 Nov 1978-5 May 1979)
         BOX 8-CETAP POP ship surveys, part 3 (6 May-12 Sep 1979)
         BOX 9-CETAP POP ship surveys, part 3 (15 Sep 1979-15 Dec 1980)
         BOX 10-CETAP historical data, part 1 (1963-1975)

132

BOX 11-CETAP hist, part 2 (1975-1978, Mead), Misc (J,K,U,Y), opportunistic

DIGITAL FILES
    AMAPPS
    ASWH aerial surveys
    CCS
    Miscellaneous data
    NEAQ & CWI
    NEFSC
    NLPSC aerial surveys
    SEUS EWS
    WEA (MassCEC) aerial surveys

DRAWER 1.1-historical - Jun 1989
    1984 & before
    1985
    1986
    1987
    1988
    1989

DRAWER 1.2-Aug 1989 - Oct 1991
    1989
    1990
    1991

DRAWER 1.3-Dec 1991 - Jan 16 1994
    1991
    1992
    1993
    1994

DRAWER 1.4-Jan 17 1994 - Feb 8 1995

    1994

    1995


DRAWER 2.1-Feb 9 1995 - Apr 1996

    1995

    1996


DRAWER 2.2-Jul 1996 - Apr 1997

    1996

    1997


DRAWER 2.3-Jun 1997 - Mar 1998

    1997

    1998


DRAWER 2.4-Apr 1998 - Apr 1999

    1998

    1999


DRAWER 3.1-May 1999 - May 2000

    1999

    2000


DRAWER 3.2-Jun 2000 - Feb 2001

    2000

    2001


DRAWER 3.3-Mar 2001 - Feb 2002

    2001

    2002

DRAWER 3.4-Feb 2002 - Dec 2005

    2002

    2003

    2004

    2005

DRAWER 4.1-Jan 2006 - Sep 2011

    2006

    2006

    2008

    2009

    2010

    2011

The box numbers refer to the eleven file boxes remaining after consolidating all of the CETAP files and discarding any duplicates; they should all have box numbers and contents summaries written on the outsides. DIGITAL FILES are all of those more recent data that were submitted entirely electronically without any paper. Those are essentially just copied from folders in a Documents folder called DATABASE on the GSO lab computer. Within each one there may be sub-folders by dataset subset and/or year, and eventually folders by FILEID (sometimes). DRAWER N.n refers to the file cabinet number (N) and the drawer within that cabinet (n). With the actual file drawer, the files are generally in chronological order and grouped into hanging folders by surveys (e.g., successive folders within a drawer might contain the January EWS-South surveys, the January EWS-Central surveys, the January EWS-North surveys, the January South Carolina surveys, and the January Cape Cod Bay surveys, then the same series for February, etc. Within any folder of the backup data, however, things will be sorted just by the filenames.

I have no real expectation that the actual paper files will survive beyond the time I depart from the campus. As of the current time, I am the only person remaining in this building, and it looks like I will need to move offices before the end of this year. The

building recently tested over the state limits for radon exposure (probably related to no occupants, limiting air flow in and out), and there won't be any spending for mitigation. In addition, the plans for campus renovations call for demolishing this building, which is expected to happen within the next few years (it also costs money to tear a building down). When that time does come, I would expect all that paper to go into the recycling.

# 10. REFERENCES

Aguilar, A. 1986. A review of old Basque whaling and its effect on the right whales (*Eubalaena glacialis*) of the North Atlantic. *Reports of the International Whaling Commission, Special Issue* 10: 191–199.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. *Introduction to Distance Sampling: Estimating Abundance of Biological Populations.* Oxford University Press, New York, NY.

Caswell, H., M. Fujiwara, and S. Brault. 1999. Declining survival probability threatens the North Atlantic right whale. *Proceedings of the National Academy of Sciences of the USA* 96: 3308–3313.

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.

Clapham, P. J., and Y. V. Ivashchenko. 2009. A whale of a deception. *Marine Fisheries Review* 71(1): 44–52.

Clapham, P. J., S. B. Young, and R. L. Brownell, Jr. 1999. Baleen whales: conservation issues and the status of the most endangered populations. *Mammal Review* 29: 35–60.

Clapham, P. J., C. Good, S. E. Quinn, R. R. Reeves, J. E. Scarff, and R. L. Brownell, Jr. 2004. Distribution of North Pacific right whales (*Eubalaena japonica*) as shown by 19th and 20th century whaling catch and sighting records. *Journal of Cetacean Research and Management* 6: 1–6.

Frasier, T. R., B. A. McLeod, R. M. Gillett, M. W. Brown, and B. N. White. 2007. Right whales past and present as revealed by their genes. Pp. 200–231 *in* S. D. Kraus and R. M. Rolland, eds. *The Urban Whale: North Atlantic Right Whales at the Crossroads*. Harvard University Press, Cambridge, MA.

Hain, J. H. W. 1975. The international regulation of whaling. *Marine Affairs Journal* 3: 28–48.

Hamilton, P. K. and S. M. Martin. 1999. *A Catalog of Identified Right Whales from the Western North Atlantic: 1935–1997.* Edgerton Research Laboratory, New England Aquarium, Boston, MA.

Hamilton, P. K., A. R. Knowlton, and M. K. Marx. 2007. Right whales tell their own stories: The photo-identification catalog. Pp. 75–104 *in* S. D. Kraus and R. M. Rolland, eds. *The Urban Whale: North Atlantic Right Whales at the Crossroads*. Harvard University Press, Cambridge, MA.

Hayes, S. A., E. Josephson, K. Maze-Foley, and P. E. Rosel, eds. 2017. *US Atlantic and Gulf of Mexico Marine Mammal Stock Assessments – 2016.* NOAA Technical Memorandum NMFS-NE-241. National Marine Fisheries Service, Woods Hole, MA. 282 pp. (updated annually; available online from the NMFS Protected Resources website at https://www.nmfs.noaa.gov/pr/sars/ or whatever they've just changed the address to this year)

Ivashchenko, Y. V., P. J. Clapham, and R. L. Brownell, Jr., eds. 2007. *Scientific Reports of Soviet Whaling Expeditions in the North Pacific: 1955–1978.* NOAA Technical Memorandum NMFS-AFSC-175. National Marine Fisheries Service, Seattle, WA. 36 pp.

Ivashchenko, Y. V., P. J. Clapham, and R. L. Brownell, Jr. 2011. Soviet illegal whaling: the devil and the details. *Marine Fisheries Review* 73(3): 1–19.

Johnson, A. J., G. S. Salvador, J. F. Kenney, J. Robbins, S. D. Kraus, S. C. Landry, and P. J. Clapham. 2005. Fishing gear involved in entanglements of right and humpback whales. *Marine Mammal Science* 21: 635–645.

Johnson, A. J., S. D. Kraus, J. F. Kenney, and C. A. Mayo. 2007. The entangled lives of right whales and fishermen: Can they coexist? Pp. 380–408 *in* S. D. Kraus and R. M. Rolland, eds. *The Urban Whale: North Atlantic Right Whales at the Crossroads*. Harvard University Press, Cambridge, MA.

Johnson, T. 2005. *Entanglements: The Intertwined Fates of Whales and Fishermen*. University Press of Florida, Gainesville, FL.

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

Kenney, R. D. 2002. *Quality-control Issues for Data Submissions to the North Atlantic Right Whale Consortium Database.* North Atlantic Right Whale Consortium Reference Document 2002-02. University of Rhode Island, Graduate School of Oceanography, Narragansett, RI.

Kenney, R. D. 2003. *Modifications and Clarifications of Data Formats for the North Atlantic Right Whale Consortium Database.* North Atlantic Right Whale Consortium Reference Document 2003-01. University of Rhode Island, Graduate School of Oceanography, Narragansett, RI.

Kenney, R. D. 2015. Status update: the NARWC survey and sightings database. *Right Whale News* 23(1): 15–21.

Kenney, R. D. 2018. Right whales. *Eubalaena glacialis, E. japonica,* and *E. australis*. Pp. 817–822 *in:* B. Würsig, J. G. M. Thewissen, amd K. M. Kovacs, eds. *Encyclopedia of Marine Mammals*, third edition. Elsevier/Academic Press, San Diego, CA.

Kenney, R. D., and G. P. Scott. 1981. Calibration of the Beechcraft AT-11 forward observation bubble for population estimation purposes. Pp. III.1–III.11 *in*: Cetacean and Turtle Assessment Program, University of Rhode Island (CETAP). *A Characterization of Marine Mammals and Turtles in the Mid- and North-Atlantic Areas of the U.S. Outer Continental Shelf, Annual Report for 1979*. Bureau of Land Management, Washington, DC.

Kenney, R. D., and C. R. Shoop. 2012. Aerial surveys for marine turtles. Pp. 264–271 *in:* R. W. McDiarmid, M. S. Foster, C. Guyer, J. W. Gibbons, and N. Chernoff, eds. *Reptile Biodiversity: Standard Methods for Inventory and Monitoring*. University of California Press, Berkeley, CA.

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.

Kenney, R. D., and K. F. Wishner. 1995. The South Channel Ocean Productivity EXperiment. *Continental Shelf Research* 15**:** 373–384.

Kenney, R. D., C. A. Mayo, and H. E. Winn. 2001. Migration and foraging strategies at varying spatial scales in western North Atlantic right whales: A review of

hypotheses. *Journal of Cetacean Research and Management, Special Issue* 2: 251–260.

Knowlton, A. R., and M. W. Brown. 2007. Running the gauntlet: Right whales and vessel strikes. Pp. 409–435 *in* S. D. Kraus and R. M. Rolland, eds. *The Urban Whale: North Atlantic Right Whales at the Crossroads*. Harvard University Press, Cambridge, MA.

Knowlton, A. R., and S. D. Kraus. 2001. Mortality and serious injury of northern right whales (*Eubalaena glacialis)* in the western North Atlantic Ocean. *Journal of Cetacean Research and Management, Special Issue* 2: 193–208.

Knowlton, A. R., S. D. Kraus, and R. D. Kenney. 1994. Reproduction in North Atlantic right whales. *Canadian Journal of Zoology* 72: 1297–1305.

Knowlton, A. R., P. K. Hamilton, M. K. Marx, H. M. Pettis, and S. D. Kraus. 2012. Monitoring North Atlantic right whale *Eubalaena glacialis* entanglement rates: a 30 year retrospective*. Marine Ecology Progress Series* 466: 293–302.

Knowlton, A. R., J. Robbins, S. Landry, H. A. McKenna, S. D. Kraus, and T. B. Werner. 2016. Effects of fishing rope strength on the severity of large whale entanglements. *Conservation Biology* 30: 318–328.

Kraus, S. D., and R. M. Rolland. 2007a. Right whales in the urban ocean. Pp. 1–38 *in* S. D. Kraus and R. M. Rolland, eds. *The Urban Whale: North Atlantic Right Whales at the Crossroads*. Harvard University Press, Cambridge, MA.

Kraus, S. D., and R. M. Rolland. 2007b. The urban whale syndrome. Pp. 488–513 *in* S. D. Kraus and R. M. Rolland, eds. *The Urban Whale: North Atlantic Right Whales at the Crossroads*. Harvard University Press, Cambridge, MA.

Kraus, S. D., K. E. Moore, C. A. Price, M. J. Crone, W. A. Watkins, H. E. Winn, and J. H. Prescott. 1986. The use of photographs to identify individual North Atlantic right whales (*Eubalaena glacialis*). *Report of the International Whaling Commission, Special Issue* 10: 145–151.

Kraus, S. D., P. K. Hamilton, R. D. Kenney, A. R. Knowlton, and C. K. Slay. 2001. Reproductive parameters of the North Atlantic right whale. *Journal of Cetacean Research and Management, Special Issue* 2: 231–236.

Kraus, S. D., M. W. Brown, H. Caswell, C. W. Clark, M. Fujiwara, P. K. Hamilton, R. D. Kenney, A. R. Knowlton, S. Landry, C. A. Mayo, W. A. McLellan, M. J. Moore, D. P. Nowacek, D. A. Pabst, A. J. Read, and R. M. Rolland. 2005. North Atlantic right whales in crisis. *Science* 309: 561–562.

Kraus, S. D., R. M. Pace III, and T. R. Frasier. 2007. High investment, low return: the strange case of reproduction in *Eubalaena glacialis*. Pp. 172–199 *in* S. D. Kraus and R. M. Rolland. eds. *The Urban Whale: North Atlantic Right Whales at the Crossroads*. Harvard University Press, Cambridge, MA.

Kraus, S. D., S. Leiter, K. Stone, B. Wikgren, C. Mayo, P. Hughes, R. D. Kenney, C. W. Clark, A. N. Rice, B. Estabrook, and J. Tielens. 2016a. *Northeast Large Pelagic Survey Collaborative Aerial and Acoustic Surveys for Large Whales and Sea Turtles*. OCS Study BOEM 2016-054. US Department of the Interior, Bureau of Ocean Energy Management, Sterling, VA.

Kraus, S. D., R. D. Kenney, C. A. Mayo, W. A. McLellan, M. J. Moore, and D. P. Nowacek. 2016b. Recent scientific publications cast doubt on North Atlantic right whale future. *Frontiers in Marine Science* 3: 137.

Laist, D. W. 2017. *North Atlantic Right Whales: From Hunted Leviathan to Conservation Icon*. Johns Hopkins University Press, Baltimore, MD.

Moore, M. J., W. A. McLellan, P.-Y. Daoust, R. K. Bonde, and A. R. Knowlton. 2007. Right whale mortality: a message from the dead to the living. Pp. 358–379 *in* S. D. Kraus and R. M. Rolland, eds. *The Urban Whale: North Atlantic Right Whales at the Crossroads*. Harvard University Press, Cambridge, MA.

NEAQ (New England Aquarium). 2018. The North Atlantic Right Whale Catalog (on-line resource). New England Aquarium, Boston, MA. http://rwcatalog.neaq.org

Owen, R. E., and R. D. Kenney. 1983. *The CETAP Data Base: A Guide for Prospective Users*. Graduate School of Oceanography, University of Rhode Island, Narragansett, RI.

Parks, S. E., and C. W. Clark. 2007. Acoustic communication: social sounds and the potential impacts of noise. Pp. 310–332 *in* S. D. Kraus and R. M. Rolland, eds. *The Urban Whale: North Atlantic Right Whales at the Crossroads*. Harvard University Press, Cambridge, MA.

Payne, R., O. Brazier, E. M. Dorsey, J. S. Perkins, V. J. Rowntree, and A. Titus. 1983. External features in southern right whales (*Eubalaena australis*) and their use in identifying individuals. Pp. 371–445 *in* R. Payne, ed. *Communication and Behavior of Whales.* AAAS Selected Symposium 76. Westview Press, Boulder, CO.

Reeves, R. R., and T. D. Smith. 2006. A taxonomy of world whaling eras. Pp. 82-101 *in* J. A. Estes, D. P. DeMaster, D. F. Doak, T. M. Williams, and R. L. Brownell, Jr., eds. *Whales, Whaling, and Ocean Ecosystems*. University of California Press, Berkeley, CA.

Reeves, R. R., T. D. Smith, and E. A. Josephson. 2007. Near-annihilation of a species: Right whaling in the North Atlantic. Pp. 39–74 *in* S. D. Kraus and R. M. Rolland, eds. *The Urban Whale: North Atlantic Right Whales at the Crossroads*. Harvard University Press, Cambridge, MA.

Shelden, K. E. W., S. E. Moore, J. M. Waite, P. R. Wade, and D. J. Rugh. 2005. Historic and current habitat use by North Pacific right whales *Eubalaena japonica* in the Bering Sea and Gulf of Alaska. *Mammal Review* 35: 129–155.

Taylor, J. K. D., R. D. Kenney, D. J. LeRoi, and S. D. Kraus. 2014. Automated vertical photography for detecting pelagic species in multitaxon aerial surveys. *Marine Technology Society Journal* 48: 36-48.

Van der Hoop, J. M., M. J. Moore, S. G. Barco, T. V. N. Cole, P.-Y. Daoust, A. G. Henry, D. F. McAlpine, W. A. McLellan, T. Wimmer, and A. R. Solow. 2013. Assessment of management to mitigate anthropogenic effects on large whales. *Conservation Biology* 27: 121–133.

Van der Hoop, J. M., A. S. M.Vanderlaan, T. V. N. Cole, A. G. Henry, L. Hall, B. Mase-Guthrie, T. Wimmer, and M. J. Moore. 2015. Vessel strikes to large whales before and after the 2008 Ship Strike Rule. *Conservation Letters* 8: 24–32.

Wade, P. R., M. P. Heide-Jørgensen, K. Shelden, J. Barlow, J. Carretta, J. Durban, J., R. LeDuc, L. Munger, S. Rankin, A. Sauter, and C. Stinchcomb. 2006. Acoustic detection and satellite-tracking leads to discovery of rare concentration of endangered North Pacific right whales. *Biology Letters* 2: 417–419.

Winn, H. E., C. A. Price, and P. W. Sorensen. 1986. The distributional ecology of the right whale *Eubalaena glacialis* in the western North Atlantic. *Reports of the International Whaling Commission, Special Issue* 10: 129–138.