Data integration for conservation: Leveraging multiple data types to advance ecological assessments and habitat modeling for marine megavertebrates using OBIS–SEAMAP

Ei Fujioka a,⁎, Connie Y. Kot a, Bryan P. Wallace a,b, Benjamin D. Best a, Jerry Moxley a, Jesse Cleary a, Ben Donnelly a, Patrick N. Halpin a

a Marine Geospatial Ecology Lab, Nicholas School of the Environment, Duke University, P.O. Box 90328, Durham, NC 27707, United States
b Oceanic Society, Washington, DC, United States

Abstract

Spatially explicit conservation efforts to identify, designate, and prioritize protected areas or biologically significant areas require analyses beyond basic species distribution and abundance studies, including assessments of migration patterns, habitat use, and ecological drivers of behavior. With the advent of alternate survey methods and platforms within the marine environment (e.g. satellite telemetry, passive acoustics, photo identification, nesting site monitoring and genetic sampling) in addition to traditional shipboard or aerial visual surveys, researchers have been developing novel analytical and modeling methodologies to fulfill such in-depth ecological assessments. This trend has raised interests and needs not only in filling spatial, temporal and ‘ecological’ gaps but also in the synthesis of these disparate data from multiple methods/platforms. OBIS–SEAMAP, a thematic node of the Ocean Biogeographic Information System (OBIS) specializing on marine megavertebrates, takes a unique approach to data integration into the OBIS–SEAMAP database to enable novel applications of a global biogeographic database. In this paper, we summarize our efforts to accomplish this integration and to develop novel mapping and visualization tools available on the OBIS–SEAMAP web site. We also discuss advantages and implications of an integrated database in advancing ecological assessments and modeling efforts based on preliminary assessments of the OBIS–SEAMAP data and derived products. Finally, we make critical suggestions for the design and function of biogeographic databases to make contributed data more useful for conservation efforts.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Spatial conservation efforts require ecological understanding beyond the distribution or abundance of the target species for more effective conservation strategies and management. This need is particularly clear for the designation of marine protected areas or the delineation and prioritization of biological hotspots. For example, the Convention on Biological Diversity (CBD) has defined seven criteria for the identification of Ecologically or Biologically Significant Areas (EBSAs) in the open ocean and deep seas: 1) Uniqueness or rarity; 2) Special importance for life history of species; 3) Importance for threatened, endangered or declining species and/or habitats; 4) Vulnerability, fragility, sensitivity, slow recovery; 5) Biological productivity; 6) Biological diversity; and 7) Naturalness (CBD, 2009). Identification of EBSAs that meet these criteria relies on spatio-temporal assessments of migration patterns, habitat uses (breeding and foraging), site fidelity, behavioral differences by sex or age class, and other ecological aspects (CBD, 2009; Olsen et al., 2009; Redfern et al., 2006). Such global scale ecological assessments require extensive collections of high-quality biological data, which was the driving purpose for establishing global biogeographic databases such as the Ocean Biogeographic Information System (OBIS; http://www.iobis.org; Ausubel et al., 2010; Costello and Vanden Berghe, 2006; Grassle, 2000). In the case of marine megavertebrates (e.g. marine mammals, seabirds and sea turtles), most data in such databases come from museum specimens or sightings recorded during visual surveys (both line-transect and opportunistic surveys by ship or plane). Visual sightings provide confirmation of species’ presence in surveyed locations and are often used to examine marine biodiversity across taxa, geographic distribution and abundance of species (Mora et al., 2008; Tittensor et al., 2010; Webb et al., 2010). Given the costs and restricted spatial and temporal coverage of visual surveys (Ackleh et al., 2012; Marques et al., 2009), however, these platforms are limited in filling spatio-temporal gaps for marine megavertebrates, many of which are long-distance migratory or elusive, deep-diving species (Gedamke et al., 2001; Olsen et al., 2009). Moreover, visual surveys are not optimal platforms to collect animal characteristics
such as age, sex, body size, or genetic information, which results in ‘eco-
logical gaps’ in biogeographic databases that hinders in-depth assess-
ments of animal ecology.

To fill spatio-temporal or ecological gaps, other technologies and
methodologies have been developing (i.e. Birdlife International, 2010;
Lascelles et al., 2012; Teloni et al., 2008), including animal tracking,
photo-identification, genetic stock delineations and passive acoustic
monitoring (PAM). Data collected through these methods or platforms
(hereafter ‘data types’) are complementary to those from visual surveys
and provide a wealth of additional information on animal behavior and
ecology. Combined with novel analytical methodologies, the data
collected with various data types facilitate a growing number of studies
investigating animal behavior, migration patterns, habitat uses and
population structures for seabirds (Camphuysen et al., 2012; Louzao
et al., 2009; Skov et al., 2007), marine mammals (DeAngelis et al.,
2011; Mate et al., 2011; Pirotta et al., 2011; Weller et al., 2012), and
sea turtles (Godley et al., 2010; Seminoff et al., 2012).

These demands and requirements highlight the need for biogeo-
graphic databases to take a comprehensive approach to filling in
data gaps by incorporating a wide variety of data types in one place
and to make them available for researchers to select the most appro-
priate suite of datasets that meet their analytical or modeling goals.

Table 1
Data sources and citations used in the case studies. The #records column for Case Study 2 show the number of records of sperm whales with the total number of records in a dataset in parentheses. (*) Contributing providers can be seen by going to the species page and opening up the Datasets on Map window. For more details, see http://seamap.env.duke.edu/content/
help/DatasetsOnMap.

<table>
<thead>
<tr>
<th>Data source and link to the OBIS-SEAMAP web site</th>
<th>Data type</th>
<th>#Records</th>
<th>Citations or providers</th>
</tr>
</thead>
</table>
| **Case study 1: Integration of multiple sea turtle data types**
  The State of the World’s Sea Turtles
  http://seamap.env.duke.edu/swot
  130 Datasets for Caretta caretta
  http://seamap.env.duke.edu/species/173830
  58 Datasets for Chelonia mydas
  http://seamap.env.duke.edu/species/173833
  82 Datasets for Dermochelys coriacea
| **Case study 2: Applications of integrated data for modeling work**
  BOEM Sperm Whale Seismic Study: S-Tag Argos Telemetry
  http://seamap.env.duke.edu/dataset/810
  SEFSC GoMex Oceanic 1992
  http://seamap.env.duke.edu/dataset/13
  SEFSC GoMex Oceanic 1993
  http://seamap.env.duke.edu/dataset/17
  SEFSC GoMex Oceanic 1994
  http://seamap.env.duke.edu/dataset/19
  SEFSC GoMex Oceanic 1996
  http://seamap.env.duke.edu/dataset/25
  SEFSC GoMex Oceanic 1997
  http://seamap.env.duke.edu/dataset/27
  SEFSC GoMex Oceanic 1999
  http://seamap.env.duke.edu/dataset/29
  SEFSC GoMex Shelf 1994
  http://seamap.env.duke.edu/dataset/35
  Gomex Sperm Whale Survey 2000
| | Visual | 3 | NOAA Southeast Fisheries Science Center |
| | Visual | 4 | NOAA Southeast Fisheries Science Center |
| | Visual | 6 | NOAA Southeast Fisheries Science Center |
| | Visual | 2 | NOAA Southeast Fisheries Science Center |
| | Visual | 1 | NOAA Southeast Fisheries Science Center |
| | Visual | 2 | NOAA Southeast Fisheries Science Center |
| | Visual | 1 | NOAA Southeast Fisheries Science Center |
| | Visual | 20 | NOAA Southeast Fisheries Science Center |
(Pennisi, 2010; Sala and Knowlton, 2006; Scholes et al., 2012). The Ocean Biogeographic Information System–Spatial Ecological Analysis of Megavertebrate Populations (OBIS–SEAMAP; http://seamap.env.duke.edu; Halpin et al., 2009), a thematic node of OBIS that specializes on marine megavertebrates, is a rare initiative in that it has the ability to store multiple data types including opportunistic and line-transect visual sightings, animal tracking data, photo-identification data (sighting history and fin images), PAM data, seabird colony observations, marine mammal stranding data and sea turtle nesting data. Improved integration of these multiple data types would help to fill data gaps and provide the means for significant improvements in ecological assessments and habitat modeling. More specifically, OBIS–SEAMAP has been addressing the following technical challenges: 1) how to integrate different data types while maintaining the differences among them without losing the vital information associated with each data type; 2) how to map and visualize different types of data in a consistent, informative way; and 3) how to extend the database structure and tools for a particular data type without losing the integrity of other types.

Herein, we first summarize our integration efforts and the improvement of the mapping and visualization tools on the OBIS–SEAMAP web site. Next, we present preliminary assessments of the integrated database to discuss how the integration would contribute to site. Next, we present preliminary assessments of the integrated data-type stacked (visual sightings: orange, telemetry: red, nesting: blue).

2. Materials and methods

2.1. Classification of multiple data types into three structuring levels

As of December 2013, the OBIS–SEAMAP database holds more than 3.1 million visual sighting records from ship/airplane-based surveys and over 710,000 telemetry location records from tagged animals, a majority of which are tagged sea turtle data shared through the Satellite Tracking and Analysis Tool (STAT) program (Coyne and Godley, 2005). The database also holds a large volume of sighting history and fin images from several photo-identification catalogs and an increasing number of PAM data recorded by stationary or towed hydrophones. In addition, over 6500 sea turtle nesting records from over 3000 nesting sites worldwide have been contributed to the State of the World’s Sea Turtles initiative (SWOT; http://www.seaturtlestatus.org), a distinct node of OBIS–SEAMAP (Kot et al., 2014).

To assess the similarities and differences among data types, we reviewed all of the data types in the OBIS–SEAMAP database and sorted out what fundamental information all types possess and how remaining parts of the data could be grouped and linked to the fundamental information. A record in a visual survey dataset represents a species occurrence or encounter at a given location, and at a specific time and date during the survey. Most of the shipboard or aerial line-transect survey datasets submitted to OBIS–SEAMAP hold attributes that are required for the calculation of the distance to the sighted animals from the survey platform and the estimation of species density, such as sighting angles, survey conditions (e.g. Beaufort sea state and glare) and the observer’s positions (Best et al., 2012; Buckland et al., 1993; Thomas et al., 2010). Line-transect datasets are often accompanied with effort data in the form of a time series of ship or airplane locations. Telemetry studies and photo-identification studies both trace individual animals’ movements over time and provide unique identifiers of animals or tagging devices and animal characteristics (e.g. age class, sex). However, telemetry data record the consecutive movement of an animal in a relatively short period of time, whereas photo-identification data reveal animal movement at sporadic times over a longer period. Additional information accompanying telemetry and photo-identification data are also different. Telemetry data often include quality flags (e.g. Argos quality indicator) and parameters to the filtering algorithm applied whereas photo-identification data contain classification of images (e.g. image quality and distinctiveness) and fin features (e.g. top notch or trailing edge). Sea turtle nest monitoring counts nesting females, clutches or crawls at fixed locations (i.e. nesting beaches) at various intervals (e.g. daily, and weekly) for one nesting season. Nesting data also include information describing monitoring efforts (e.g. start and end
dates of monitoring efforts and percent of beaches covered by monitoring).

The similarities of these different data types were condensed to where (latitude and longitude of the location), when (date and time) and what (species identification). This fundamental information is accompanied with a variety of attributes depending on data types, survey objectives and survey methods, which is usually denoted in a tabular format corresponding to individual records. In addition to the fundamental information and tabular attributes, there are relevant resources collected during surveys or derived from the data that would help understand and investigate the data further. For example, it is more informative and better meets project’s objectives to visualize the locations of individual animals captured in a photo-identification along with their fin images. Likewise, model outputs derived from biological data should be visualized along with the input data (Barlow et al., 2009). Supplemental georeferenced data such as inland locations where animals were tagged and released are also regarded as relevant resources. Therefore, these resources need to be integrated to help researchers gain as much insight from their collected data as possible.

In summary, we identified three data structuring levels: level 1) fundamental information required to map and visualize the data as points (i.e. locations, date/time, species identification, group size and data type) and lines (i.e. effort tracklines and travel paths of animals tagged). These attributes are regarded as common or minimum required ones that are needed for all datasets submitted to OBIS–SEAMAP. Animal characteristics such as sex and age class are also principal. Telemetry locations of tagged animals or sighting history of photo-identified animals are often accompanied with these attributes but visual surveys are usually not able to determine such characteristics. Thus, animal characteristics are grouped into level 1 but not regarded as required attributes; level 2) additional attributes unique to a data type (e.g. sighting angles and survey conditions in a line-transect survey) or particular dataset (e.g. in-situ oceanographic measurements) that
are suitable to store in database tables. These attributes are optional and attributes are provided in formats that vary between datasets; and level 3) relevant resources. These resources are also optional and usually provided as files (e.g. fin images from a photo-identification survey or raster of model outputs) that may or may not be suitable to be put in a relational database. Resources can be associated with an entire dataset or an individual observation record in a dataset.

2.2. Implementation of a hierarchical structure for data integration

We developed a hierarchical structure representing the three data levels that maximizes the flexibility of the database structure and data processing, while retaining valuable information from each data type (Fig. 1). At the top level (level 1), the fundamental information was put together in standardized, spatially-enabled tables for point and line data. Though being common attributes, values can be provided in different formats (e.g. single tabular column having ‘2000/10/20’ or three columns having ‘2000’, ‘Oct’ and ‘20’ for an observation date). The OBIS–SEAMAP data manager cleaned and standardized the attribute values (e.g. ‘2000-10-20 10:15:00’ for an observation date and time). Each record in the standardized tables was given a flag indicating the data type. Tabular attributes unique to individual datasets were stored in separate tables at level 2. Some attributes (e.g. glare) are common in datasets of the same data type but coding scheme, recording method or value range can vary (e.g. numeric code or string vocabulary for glare). Instead of standardizing these attributes to fit them in a single database schema, we took a straightforward approach where the attributes were stored in the OBIS–SEAMAP database in their original format. Since attributes at level 2 were not standardized, we did not design the database and online tools to use these attributes for any statistical calculations or comparisons among datasets. The linkage between the standardized tables (level 1) and individual dataset tables (level 2) was established using dataset identification numbers at a dataset level and row identification numbers at a record level. For level 3, we devised a means of linking datasets or records in the standardized point table with relevant resources such as fin images and overlaying relevant geographic data (e.g. model outputs) over the biological data. For individual records, the link information is stored in a column in the standardized table. Relevant resources are usually stored in a hard disk on the OBIS–SEAMAP server. Relevant geographic data are defined as map layers in the OBIS–SEAMAP map engine. Geographic data can be stored in a hard disk in the original format (e.g. ESRI shapefile) or imported into the OBIS–SEAMAP database depending on how they are visualized and assessed.

2.3. Enhancements of mapping and visualization tools

One of the imperative challenges in mapping multiple data types simultaneously was to avoid misinterpretation of data while achieving a consistent data processing and visualization of multiple data types. For example, although the OBIS–SEAMAP web site had been providing gridded summaries of the data representing the number of records and animals per unit cell, it had counted individual locations from telemetry data as unique animals, which would result in an overestimate of the species abundance. For improved interpretation of different data types, we infused the data type as a parameter in the rendering processes and most of the statistical calculations to produce more appropriate results per data type before combining them in visual outputs (e.g. gridded summary maps or time series graphs). With the improved calculations, for example, an individual animal tagged or

Table 2
Seasonal gaps represented by the number of records of visual sightings and telemetry locations for loggerhead, green and leatherback sea turtles. Seasons are defined as: winter — December, January and February; spring — March, April and May; summer — June, July and August; fall — September, October and November.

<table>
<thead>
<tr>
<th>Season</th>
<th>Loggerhead Visual</th>
<th>Loggerhead Telemetry</th>
<th>Green Visual</th>
<th>Green Telemetry</th>
<th>Leatherback Visual</th>
<th>Leatherback Telemetry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter</td>
<td>2111</td>
<td>21,129</td>
<td>81</td>
<td>3971</td>
<td>398</td>
<td>1124</td>
</tr>
<tr>
<td>Spring</td>
<td>4172</td>
<td>22,862</td>
<td>77</td>
<td>2201</td>
<td>344</td>
<td>1783</td>
</tr>
<tr>
<td>Summer</td>
<td>4672</td>
<td>52,374</td>
<td>76</td>
<td>4097</td>
<td>1289</td>
<td>641</td>
</tr>
<tr>
<td>Fall</td>
<td>1592</td>
<td>38,757</td>
<td>32</td>
<td>7021</td>
<td>769</td>
<td>87</td>
</tr>
</tbody>
</table>

Fig. 5. Green turtle (Chelonia mydas) visual sightings (circle), telemetry locations (triangle) and Northwest Atlantic RMU (hatched area). The green and yellow triangle points off the RMU are telemetry locations of two juveniles.
photo-identified is counted as one animal no matter how many telemetry locations or sightings of that animal fall within a unit cell. We also implemented different representations for each data type on a map. Furthermore, we added query options that allow users to filter point data based on data types.

Since the hierarchical structure unrolls from the standardized table to individual dataset tables at the second level, we had to develop a blanket mechanism to abstract the differences among the dataset tables in order to display dataset-specific attributes in a consistent fashion. The blanket mechanism has two phases. First, it determines what attributes are appended to a record or dataset of interest. Second, it extracts values of these attributes from the dataset table using the record or dataset identifier to display online.

To visualize relevant resources associated with a record or dataset, we developed a similar mechanism that locates the resources using the link information in the standardized table and makes tailored views on the map interface depending on the type of the resources. For example, images are displayed with an HTML IMG tag while georeferenced data are rendered through the OBIS–SEAMAP map engine according to the layer definition that specifies symbols, colors, etc.
2.4. Data sources and assessment methods for case studies

To demonstrate the application of a biogeographic database integrating multiple data types, we describe two case studies with visual, qualitative assessments conducted using tools available on the OBIS–SEAMAP web site, summary statistics of records within the OBIS–SEAMAP database, and spatial analyses on the data downloaded from OBIS–SEAMAP using ArcGIS Desktop version 10.1 (ESRI, 2012).

2.4.1. Case study 1: integration of multiple sea turtle data types

For the first case study, we integrated the SWOT nesting data, which had been technically separated from the main OBIS–SEAMAP database, with other sea turtle data types (i.e. visual sightings and telemetry data) in OBIS–SEAMAP. SWOT is a partnership among the Oceanic Society, the IUCN Marine Turtle Specialist Group (MTSG) and the OBIS–SEAMAP initiative along with the network of more than 500 sea turtle experts. Every year, SWOT collects sea turtle nesting data from around the world and disseminates the collected data through the OBIS–SEAMAP web site (http://seamap.env.duke.edu/swot; Kot et al., 2014). Each record in the SWOT database holds the location of the beach, species observed and count, reported as an exact or binned number of nesting females, clutches, or crawls. Additional attributes include beach names, beach length and details of monitoring effort according to the SWOT minimum data standards (MDS; SWOT Scientific Advisory Board, 2011). In addition to nesting data records, the SWOT database holds relevant sea turtle biogeographic data, including species distribution ranges, genetic stocks, mark-recapture studies, and telemetry data for each species (Wallace et al., 2010). By fitting the SWOT nesting data into the three data levels, the location of the beach, species monitored and count fell in level 1. Other attributes, such as beach name, beach length, and monitoring effort details, were grouped in level 2. Biogeographic data were classified as relevant resources at level 3.

Using the integrated sea turtle data (Table 1), we addressed spatio-temporal complementarity of data types for loggerhead (Caretta caretta), green (Chelonia mydas) and leatherback (Dermochelys coriacea) sea turtles by aggregating data per 1-degree grid per data type and comparing the number of cells that had only one data type present with cells that had multiple data types present. We also examined seasonal differences for the number of records OBIS–SEAMAP currently had from visual sightings vs. telemetry data. In addition, we examined how many visual sighting and telemetry points lay outside of the RMUs. While the OBIS–SEAMAP database holds sea turtle visual sighting data as early as 1758, we filtered data from 1996 through 2013 for this case study because both telemetry and visual sighting data were available for those years.

2.4.2. Case study 2: applications of integrated data for modeling work

The second case study is aimed at illustrating applications of a merged dataset to modeling work using an intensive data set from the Sperm Whale Seismic Study (SWSS) supported by the US Minerals Management Service (now Bureau of Ocean Energy Management — BOEM; Jochens et al., 2006, 2008; Table 1). The SWSS conducted line-transect surveys, recorded visual sightings of marine mammals and deployed Argos telemetry tags when possible. In addition, skin samples and fin images were taken for certain individuals. Animal characteristics such as sex, age class, body size and haplotypes were identified by skin samples or visual observations, or estimated with images. In this case study, we associated Argos transmitted locations of tagged sperm whales with animal characteristics and a link to a fin image if the animal was photo-identified. Argos data (e.g. locations and date/time) with animal characteristics were appended to the standardized point table (level 1). Fin images were classified as relevant resources and put in the server hard disk (level 3).

Then, we overlaid the merged telemetry data over a summer predictive habitat layer for sperm whales in the Gulf of Mexico derived from an environmental niche model developed by the Strategic Environmental Research and Development Program (SERDP) project (Best et al., 2007, 2012; Department of the Navy, 2007a, 2007b). The model ingested 49 sperm whale sighting data from 8 shipboard and aerial line-transect surveys registered in OBIS–SEAMAP (Table 1). Using the interactive tool on the OBIS–SEAMAP site that allows for selection of individual animals based on sex, we compared the movements of male and female sperm whales over the predictive habitat layer. Habitat suitability from the predictive habitat layer was sampled and summarized for each animal for all Argos locations from the SWSS dataset.
3. Results

3.1. Mapping and visualization of integrated data

The integration of multiple data types using the hierarchical data structure allows users to apply most of the advanced mapping and visualization tools available on the OBIS–SEAMAP web site (e.g. interactive time-series charts or histograms for oceanographic variables sampled) while mapping multiple data types simultaneously. Moreover, the hierarchical structure and the mechanism to abstract the differences among data types contributes to a compatible look and interface that presents disparate attributes from different datasets and to custom-tailored visualization tools specific to a data type (e.g. displaying fin images for photo-identification data). The mechanism also grants us the capability to incorporate any sorts of geographic data relevant to a particular dataset or species (e.g. species distribution maps, study areas or an extension of the beach covered by nest monitoring efforts) that help comprehend the details of the data.

Since most of the statistical calculations are done per data type, time series graphs are able to display summary statistics of the data mapped (e.g. the number of animals in a designated study area aggregated per year) with bars per data type stacked (Fig. 2). Similarly, points on the
map are given distinct symbols according to the data type (Fig. 2). As the fundamental information of line data is stored in the standardized table, it becomes possible to map telemetry tracklines and survey tracklines from multiple datasets or multiple species on a single map, which provides useful information to clarify which direction an animal migrated to, how often the species of interest was encountered during a transect survey or how well transects were located compared with actual animal movements (Fig. 2). Through these improvements, researchers are able to explore spatio-temporal distributions of diverse sets of data while clearly distinguishing the scientific implications that different data types deliver.

3.2. Case study 1

The enhanced flexibility of OBIS–SEAMAP enables an extension and consolidation of separate applications into a unified framework. Originally launched as a separate database and application, SWOT is now fully integrated into OBIS–SEAMAP, both data-wise and tool-wise. Researchers are able to assess a variety of sea turtle data including visual sighting, telemetry and nesting data along with the SWOT RMUs using exploratory tools tailored for SWOT (e.g. filtering nesting beaches by country or color-coding symbols by colony size) combined with those commonly available on the OBIS–SEAMAP web site (Fig. 3).

For all of the three sea turtle species assessed, 1-degree cells containing either visual sighting or telemetry data only outnumbered those with both visual sighting and telemetry data (Fig. 4). Most of the cells for nesting data do not overlap with other sea turtle data types, since nesting sea turtle data are collected on beaches (land) where shipboard or aerial surveys do not cover and tagged animals do not usually visit (except during the nesting seasons). The seasonal comparison of the number of visual sighting and telemetry data revealed that the telemetry data for loggerheads and greens compensate fewer visual sightings in fall than the other three seasons (Table 2). Across the seasons, whereas there are more leatherback visual sightings in summer and fall than in winter and spring, telemetry data demonstrates the opposite trend with more data in winter and spring than summer and fall.

Out of 18,094 green turtle records (266 visual sightings and 17,828 telemetry locations), 30 visual sightings and 243 telemetry locations lay outside of the green turtle RMUs mainly in the North Atlantic (Fig. 5). A few loggerheads traveled outside the loggerhead RMUs with 42 visual and 650 telemetry records (Fig. 6). There were 7736 records of leatherbacks, among which 61 visual sightings and 88 telemetry locations fell outside of the leatherback RMUs concentrating in the North Atlantic and the eastern side of Scotland (Fig. 7).

3.3. Case study 2

The integration of multiple data types from the SWSS surveys resulted in a single dataset containing telemetry locations of 51 tagged sperm whales (34 females, 6 males and 11 unknown) among which 21 whales were linked to fin images and all 40 sexed animals were associated with haplotypes. The compiled dataset provides researchers with the ability to not only trace individual sperm whale movements over time but also see detailed animal information including sex, age class and a fin image of the whale (Fig. 8).

When overlaid on the SERDP summer habitat suitability map for sperm whales in the Gulf of Mexico, the compiled SWSS telemetry dataset showed that most of the tagged females remained around offshore areas in the Gulf where high habitat suitability was predicted (Fig. 9a). Some females, however, frequented low habitat suitability areas in the northwestern part of the Gulf. While summer visual sightings of sperm whales within the boundary of the model presented an average suitability of 0.1594 (n = 49, SD = 0.054), Argos transmitted locations of three female sperm whales (3700846, 3705648, and 3705726; a full list of individuals is seen in Jochens et al., 2008) loomed in areas with an average suitability around 0.0480 (n = 98, SD = 0.0196) in the summer. This average for the three females is lower than a mean suitability of 0.0969 for all 34 females (n = 767, SD =

![Fig. 10. Three females identified as 3700846, 3705648, and 3705726 stayed in low habitat suitability areas in summer.](image)

### Table 3

Summary statistics of habitat suitability (HS) sampled on transmitted locations of three female sperm whales, all females and all males from the SWSS telemetry dataset and all visual sightings used in the model.

<table>
<thead>
<tr>
<th>Group</th>
<th>Min HS</th>
<th>Mean HS</th>
<th>Max HS</th>
<th>STD</th>
<th>#Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three females</td>
<td>0.0004</td>
<td>0.0480</td>
<td>0.0851</td>
<td>0.0196</td>
<td>58</td>
</tr>
<tr>
<td>All females</td>
<td>0.0004</td>
<td>0.0969</td>
<td>0.3864</td>
<td>0.0653</td>
<td>767</td>
</tr>
<tr>
<td>All males</td>
<td>0.0003</td>
<td>0.1275</td>
<td>0.4353</td>
<td>0.0787</td>
<td>142</td>
</tr>
<tr>
<td>Visual sightings</td>
<td>0.0009</td>
<td>0.1594</td>
<td>0.3088</td>
<td>0.054</td>
<td>49</td>
</tr>
</tbody>
</table>
Fig. 11. Movements of the male sperm whale 2505725 from the SWSS telemetry dataset in summer (left); the fall movement was also overlaid on the fall density layer in the Atlantic (right).
0.0653) though the difference is not statistically significant (Fig. 10 & Table 3).

Male sperm whales tended to have wider ranges than females where an average suitability range (difference between the maximum and minimum suitability values) for males is 0.2566 (SD = 0.0734; Fig. 9b) whereas that for females is 0.1168 (SD = 0.0688). One of the male sperm whales identified as 2505725 (body size: 11 m, haplotype: C) showed distinct migratory patterns crossing between the Gulf of Mexico and US east coast in summer and fall (Fig. 11).

4. Discussion

4.1. Advantages of an integrated biogeographic database

One of the most direct benefits of an integrated biogeographic database with advanced mapping and visualization tools is that it provides complementary, multi-perspective views of biological data for a species of interest. The ease of spatio-temporal exploration of a variety of data will promote qualitative assessments of animal distribution and migration patterns, which expedites data mining and accession of data appropriate to planned analytical work. While OBIS–SEAMAP primarily remains as an exploratory framework or discovery portal, novel approaches that we developed bridge qualitative and quantitative assessments, the latter of which is supported by the data download functionality with multi-faceted criteria (i.e., selection of multi-species, time period, data type, and region). The flexibility of the OBIS–SEAMAP database and application framework also enables reintegration of quantitative analyses or modeling work back into the database and to visualize results with source data for further investigation.

4.2. Case study 1: enhancing conservation efforts with the integrated data

Since sea turtles cover large geographic ranges and present complex life history patterns (Godley et al., 2010; James et al., 2005; Polovina et al., 2004), biological assessments using the combination of telemetry, visual sighting and nesting data are required to fill spatio-temporal and ecological gaps and to generate a more resolute understanding of total range and distribution, habitat use, and migratory patterns of the sea turtles. Thus, the integration of the SWOT and OBIS–SEAMAP databases was a perfect coupling that can greatly facilitate global assessments. The case study presented here verified that the integrated database covered a wider geographic area and supplemented seasonal gaps of each data type.

Further assessments of spatio-temporal gaps in data and strategic planning of visual sighting and tagging surveys will need to take into account survey effort. For example, when constrained to sighting records from visual line-transect surveys with effort data, the sighting rate (the number of records per effort hour) of loggerheads in summer is the second lowest (Fig. 12). This low rate is in contrast with visual sightings (line-transect plus opportunistic) and telemetry data of loggerheads that are most numerous in summer (Table 2). This may indicate that summer line-transect surveys have been conducted in areas where loggerheads visit less frequently in summer or that a low sighting rate in summer reflects loggerhead’s different behavior from other seasons causing them to be more difficult to visually observe during a line-transect survey.

The SWOT initiative is expanding its database to include available satellite telemetry data in an effort to identify “Important Turtle Areas” throughout the world’s oceans. The OBIS–SEAMAP/SWOT database will facilitate the assessments of finer-scale spatial and temporal variation in densities of turtle movements or presence, which in turn allows for improvements of the existing RMUs. The existence of records falling outside of an RMU illustrated in our case study will give researchers an opportunity to review the adequacy of and reexamine the boundaries of current RMUs. First, subadult green turtles visit waters off Cape Cod (Lazell, 1980), which was confirmed by the telemetry data from two juveniles identified as 37189 (McClenan and Read, 2007) and 78066a (data provided by Connie Merigo, New England Aquarium through the STAT program) that traveled outside of the Northwest Atlantic RMU (Fig. 5). Visual sightings of green turtles were also found outside of the RMU though the age class was not known. Second, a recent study documented trans-Atlantic migrations of juvenile loggerheads (Eckert et al., 2008). Although a similar migration was illustrated in our case study, a middle section of the pathway is not contained in the loggerhead RMUs (Fig. 6). More data collection and investigations will be needed to assess if this pathway is regularly followed by many juveniles. Third, although the congregation of leatherbacks at the sea off Nova Scotia, Canada has been studied (James et al., 2005) and is covered by the leatherback northwest Atlantic RMU, the visual and telemetry data indicate that leatherbacks travel further north (Fig. 7). Although RMUs were delineated to be core areas based on data from various data types and not intended to encapsulate all locations of sea turtle data, these points are worth considering when RMU boundaries are revisited in the future. The redefinition of RMUs or “RMUs 2.0” will directly advance international and regional conservation processes such as defining EBSAs or promoting sustainable fishing practices with reduced bycatch.

4.3. Case study 2: implications of integrated data for modeling work

The compiled SWSS dataset was a first attempt in the OBIS–SEAMAP database to merge Argos telemetry data and photo-identification images along with animal characteristics obtained from multiple methodologies, which in turn can be explored all together for a better account of sperm whale distribution. Sperm whales are known to favor waters with depths of 400–1200 m (Watwood et al., 2006) and display sexual segregation with females remaining in small-ranging social groups at tropical and subtropical latitudes while males migrate solitary or in
Fig. 13. Sequences of telemetry locations of individual animals in the OBIS–SEAMAP database, such as those of (Sterna paradisaea; http://seamap.env.duke.edu/dataset/705; Egevang et al., 2010; left) are aggregated by 1-degree cell per species and per day before being uploaded into OBIS (right).
small groups to high latitudes along continental shelf-edges (Griffin, 1999; Rice, 1989; Teloni et al., 2008). These ecological preferences were qualitatively confirmed by the SWSS telemetry data in which the female sperm whales remained around the continental shelf in the Gulf of Mexico and the tagged male sperm whale moved further offshore (Figs. 8 & 9).

When compared with the SWSS telemetry locations, the SERDP model results were consistent with off-shelf areas in the Gulf of Mexico, although a lower habitat suitability prediction in the northwestern part seemed to contradict the numerous telemetry observations there (Figs. 9 & 10). This may be the result of other environmental conditions or inherent variability with their distribution. Moreover, it is generally unreasonable to extrapolate tracks from only a few individuals to capture the distribution of a full population (Aarts et al., 2008). Given that only 49 sperm whale sightings were ingested in the habitat model, a similar issue of adequacy should also be addressed in a modeling work based on visual sightings.

The track of the male sperm whale 2505725 reasonably suggested a migratory corridor between shelf habitats of the Gulf of Mexico and US east coast, separated by the tip of Florida where the lowest suitability was predicted (Fig. 11). Predicted density surfaces based on line-transect surveys favor capturing behaviors with the greatest residence time (e.g. foraging and breeding) and not ephemeral ones (i.e. migration). Hence this migratory route is not captured well in the seasonal average habitat model.

These assessments highlight the need for examining multiple data types in order to fill spatio-temporal, ecological gaps when evaluating areas of importance for different animal behaviors that may also vary by characteristics (e.g. age and sex). This is especially the case for sperm whales that display strong sexual segregation (Pirota et al., 2011), which are also hard to detect visually (Ackleh et al., 2012). Thus, the simultaneous visualization of these various types of data and ability to interactively explore them spatially and temporally yields an extremely useful synoptic view for planning or modeling purposes.

4.4. Necessity of extended data schema for biogeographic databases

There are several challenges to incorporating data from multiple data types into a single biogeographic database. While occurrence data from biogeographic databases indicate broad-scale species distribution or biodiversity (Evans and Hammond, 2004; Sala and Knowlton, 2006), occurrence data alone are generally not suitable for other modeling approaches such as estimates of absolute abundance, density or identification of important areas based on ecological aspects (e.g. migration, feeding and breeding). For example, a density estimate using visual sighting data from line-transect surveys requires additional attributes such as sighting angles, weather conditions and observer’s positions as well as effort data (Barlow et al., 2009; Buckland et al., 1993). Although the OBIS–SEAMAP database holds such attributes and effort data, they are dropped when the OBIS–SEAMAP data are transferred to biogeographic databases that rely on data schema derived from Darwin Core (Darwin Core Task Group, 2009), originally intended to document records of museum specimen (OBIS, 2005; Wieczorek et al., 2012).

Furthermore, reviewing a time series of transmitted positions from tagged animals within a biogeographic database could be misleading when these data are used without unique identifiers of individual animals for abundance estimates or density models. For example, multiple positions from one tagged animal may come across as multiple animals counted once for each position. To alleviate such misrepresentation, sequences of telemetry locations of individual animals in the OBIS–SEAMAP database are aggregated per species, per day and per 1-degree grid before being uploaded to OBIS, so the resulting grids at least illustrate the presence of species at a given location on a given day. This results in a loss of finer scale accuracy and the capability to assess individual animals’ movements (Fig. 13).

To overcome these challenges and make public biogeographic databases better suited to various ecological assessments and conservation efforts, it is crucial to extend and improve their data schemas. There are on-going efforts to extend database schemas. For example, in partnership with the US Integrated Ocean Observing System (IOOS) and other US federal agencies, USGS/OBIS-USA is establishing the Marine BioGeography data standard (MBG), which is an enhanced implementation of Darwin Core designed to hold richer data such as absence, abundance, tracking and sampling details (Fornwall et al., 2012). This effort, which OBIS–SEAMAP is participating in, will extend the utility of the data held in OBIS-USA.

4.5. Conclusion

We strive to improve the integration of multiple data types and to develop novel mapping and visualization tools that allow researchers to discover, explore and assess integrated data with complementary, multi-perspective views. This effort will help in determining reliable conservation strategies when used within a justified, systematic analytical framework (Lascelles et al., 2012). The case studies demonstrate advantages of our approach to leveraging multiple data types to advance ecological assessments and habitat modeling for marine megavertebrates using OBIS–SEAMAP. Qualitative assessments similar to these case studies, exploring data through time, space, taxon, and data type, can be conducted directly using the OBIS–SEAMAP web site and its array of tools. As OBIS–SEAMAP project members are deeply involved in the development of the OBIS and OBIS-USA databases and their portals, lessons learned from our efforts in developing OBIS–SEAMAP can be applied to these databases (Fujikoa et al., 2012). It is equally important to facilitate collaboration in mutual data exchanges among taxon/data type/region specific biogeographic databases, as seen in the SWOT and OBIS–SEAMAP integration, to fill data gaps, and advance analytical methodologies for a better understanding of the biology and ecology of marine megavertebrates.

Acknowledgments

The OBIS–SEAMAP project, launched in 2002 through the support from the Alfred P. Sloan Foundation, is supported by a funding from NOMP/NSF (NSF-OCE-07-39199 & NSF-OCE-11-38046). We thank NOAA NMFS, the US Navy, BOEM and other US federal agencies for their data contribution and collaborative efforts to improve OBIS–SEAMAP. The accumulation of sea turtle telemetry data would not be possible without the support by Michael Coyne, seaturtle.org. For the SWOT initiative and its database, our acknowledgement goes to Rod Mast and Brian Huthchinson at the Oceanic Society and all the nesting data contributors worldwide as well as the organizers of WIDECAST. We also would like to give thanks to the past OBIS–SEAMAP project members and the current ones who were not directly involved in this paper work, including Andy Read and Kim Urian at Duke University, Melissa Soldevilla at NOAA Southeast Fisheries Science Center, and Andrew DiMatteo at the US Navy. Above all, we express the deepest appreciation to OBIS–SEAMAP data providers worldwide.

References


