



OPINION PIECE

Searching for meaning in marine mammal shared data

Giovanni Bearzi^{1,2,*}, Olivier Gimenez³

¹Dolphin Biology and Conservation, via Cellina 5, 33084 Cordenons PN, Italy

²OceanCare, Gerbestrasse 6, 8820 Wädenswil, Switzerland

³CEFE, CNRS, Univ Montpellier, Univ Paul Valéry Montpellier 3, EPHE, IRD, Montpellier, France, Campus CNRS, 1919 route de Mende, 34293 Montpellier Cedex 5, France

ABSTRACT: The sharing of marine mammal data is a worthwhile practice, but there are caveats. Data interpretation may be difficult, sometimes resulting in misleading information or inappropriate formulation of research questions. Here, we point out some of the challenges when dealing with shared marine mammal datasets. We emphasize the importance of collecting, publishing and sharing data in ways that can produce unbiased and meaningful knowledge, ultimately inspiring and directing management action. Finally, we suggest that bridging the gap between data sharing and data reuse will require enhanced spatially referenced online databases as well as direct collaboration between the data analysts and the field researchers who possess relevant place-based expertise.

KEY WORDS: Data sharing · Data reuse · Marine mammals · Mediterranean Sea · Public data archive · Reproducible science · Research network · Online database

INFORMATION, OR JUST DATA?

The sharing of biodiversity and ecological data (Costello et al. 2014, Michener 2015) is a worthwhile practice that should be encouraged. National and regional policies rely increasingly on spatial information and electronic data sharing to inform environmental management. For instance, the EU Directive 'Inspire' — to be fully implemented by 2021 (<http://inspire.ec.europa.eu/>) — aims to create a major infrastructure to facilitate access to geo-referenced data across European boundaries. Unquestionably, shared datasets have been yielding remarkable advances in knowledge, including in the field of marine mammalogy (e.g. Kaschner et al. 2006, 2011, Schipper et al. 2008). There is a substantial difference, however, between data — whether or not they are shared — and information that ultimately brings knowledge (Mazaris 2017). Emphasis should be put on the latter.

Here, we emphasize the value of publishing marine mammal data as timely, thoroughly and honestly as we can, producing information that can advance knowledge as well as support and complement data sharing and reuse (including for validation/replication purposes). Without such a knowledge component, data collection may not bring the desired benefits.

Field researchers, as well as organisations relying on volunteers, sometimes become 'addicted' to data collection, perhaps with the illusion that all the data they store will eventually be employed for some good cause. Unfortunately, data use or reuse cannot be taken for granted, and even when information is shared, important data may end their life as numbers, codes or dots on a map. First, few conservation practitioners seem to use evidence-based knowledge to support their management (Bearzi 2007, Cook et al. 2010). Second, most public data in ecological and

*Corresponding author: giovanni.bearzi@gmail.com

evolutionary research journals (intended to increase accessibility to data underlying scientific publications) seem to be archived in ways that partially or entirely prevent reuse (Roche et al. 2015). Third, shared data may incur the risk of being misinterpreted, due inter alia to inappropriate sharing platforms, scarcity of associated information, poor or unknown data collection protocols and inconsistent methodology. Fourth, distributional databases are spatially biased due to uneven effort of sampling and data storage (Beck et al. 2014). The ultimate goals of data reuse consist of extracting 'true' and valuable information from data, giving meaning to such information and ensuring that our findings are actually used by others—ideally for worthwhile reasons including effective conservation management (Michener 2015, Mazaris 2017).

DIFFERENT METHOD, DIFFERENT DATA

Dots on a map indicating the geographic position of marine mammal sightings look good, but information contents may be poor and sometimes misleading. Here are a few caveats based on our own research experience. Different cetacean species have different likelihoods of being encountered. Some species perform very long dives (Schorr et al. 2014), some surface inconspicuously, some are solitary, whereas others tend to stay in very large groups. A marine mammal sighting may be composed of 1 individual, several hundreds or even thousands (Acevedo-Gutiérrez 2009). The very definition of what is a 'sighting' or a 'group size' can vary greatly depending on factors including methodology and group dynamics (Mann 1999, Whitehead 2004), therefore yielding data that may be remarkably different (for instance when counts of highly gregarious dolphins refer to 'focal groups' instead of large and virtually uncountable actual group sizes). Ten dots indicating dolphin sightings on a map may be 10 repeated sightings of the same individual, or it may be 10 different individuals. It may be 10 sightings of exactly the same group, or 10 sightings of different groups totalling a much higher number of individuals. Even if an online database allows for retrieval of group size and other metadata, one will not know how many actual animals are represented by a given set of dots, let alone their abundance trends and whether population density has remained the same after a given survey, or if the population has increased or vanished altogether. Inclusion of survey effort may not help, as long as information on survey methods is unavail-

able. How was the sea state when the survey was carried out? A few hours of survey on a flat sea may bring more encounters with cetaceans than a week-long campaign challenged by breaking waves (and seasick observers).

Marine mammals perform extensive movements, sometimes encompassing 1000s or even 10 000s of km (Stern 2009). Regardless of whether these movements qualify as migrations, most species travel to obtain resources such as prey and mates, and they respond to variables such as water temperature, resulting in shifts in population density and distribution across seasons and years. When data end up being frozen in a database or on a map, temporal shifts in distribution become less apparent, and one may wrongly assume that density has been permanently high within areas where animals just happened to have passed through or were temporarily stationed. While movement data can aid assessments of animal stocks and boundaries, and assist in ecosystem-based management, the incorporation of these data into place-based conservation strategies remains underused (Hays et al. 2016). When using shared data, researchers must take into account the dynamic nature of marine mammal movements and potential bias associated with temporal shifts in distribution.

THE CHALLENGE OF DATA INTERPRETATION

Having access to some data is better than having access to no data at all. But can the data be put to good use if differences in survey method and effort are not seriously taken into account, and if the analyses are not grounded in the expertise of local researchers? It does take skilled analysts and experienced field researchers to come up with meaningful hypotheses about cause–effect relationships. Data analysts can use shared data to test a number of hypotheses and find correlations. The question is: Will they be able to tell if there is causation without field experience and local knowledge? When understanding of local context is poor, there is a risk of misinterpreting the data and obtaining results that may look compelling, but make no ecological sense. This point is humorously made by a web site of spurious correlations (tylervigen.com/spurious-correlations), which shows, for example, that the number of people who drowned by falling into a pool correlates with films featuring Nicolas Cage. An even higher correlation rate was found between the number of IKEA furniture stores and Nobel laureates (Maurage et al. 2013).

How much useful inference can anyone draw from raw marine mammal data without knowing that an area has been impacted by overfishing or habitat destruction for decades? Can one formulate hypotheses without considering that culling campaigns in historic times have eradicated important cetacean populations, or without taking into account that a rare marine mammal species has become accustomed to mating with a more abundant different species, producing maladaptive hybrids? Such contexts are often essential for data interpretation, but they can hardly be found in online data repositories.

In recent years, our team recorded roughly 1000 marine mammal sightings in the Gulf of Corinth, a 2500 km² semi-enclosed inland system in central Greece (Fig. 1). The Gulf would look like a marine mammal heaven if such a dataset was added to a spatially referenced online database (e.g. OBIS-SEAMAP: seamap.env.duke.edu; Best et al. 2007, Halpin et al. 2009)—especially if compared with the rest of the eastern Mediterranean, where shared records are scarce. On a digital map, the high density of dots might suggest that the Gulf is a hotspot for marine mammals—including striped dolphins *Stenella coeruleoalba*, common dolphins *Delphinus delphis*, common bottlenose dolphins *Tursiops truncatus*,

Risso's dolphins *Grampus griseus* and Mediterranean monk seals *Monachus monachus* (Fig. 1). However, one would hardly appreciate the actual degree of ecological complexity. Over 1300 striped dolphins live in the Gulf of Corinth, and they seem to be doing fine (Bearzi et al. 2016). Common dolphins number only about 20 and are predicted to be declining (their local population has been recently suggested to be Critically Endangered within the Gulf; Bearzi et al. 2016, Santostasi et al. in press). Common dolphins are believed to mate with striped dolphins and produce hybrids (a total of 55 individuals of intermediate pigmentation have been estimated to occur in the area). These intermediate animals, incidentally, would not show in online maps because they do not belong to any of the categorised taxa. Only 1 Risso's dolphin has been recorded: all the sightings on the map would refer to this single individual, which was repeatedly encountered over 8 yr of the study. Monk seal sightings also appear to indicate a single individual that was observed just a few times. Moreover, 3 of the dolphin species (striped, common and Risso's dolphins, plus the *Stenella* × *Delphinus* hybrids) are resident in the Gulf of Corinth and they live together in mixed groups—a rather unusual behaviour that would not be easy to infer from online data repositories. Bottlenose dol-

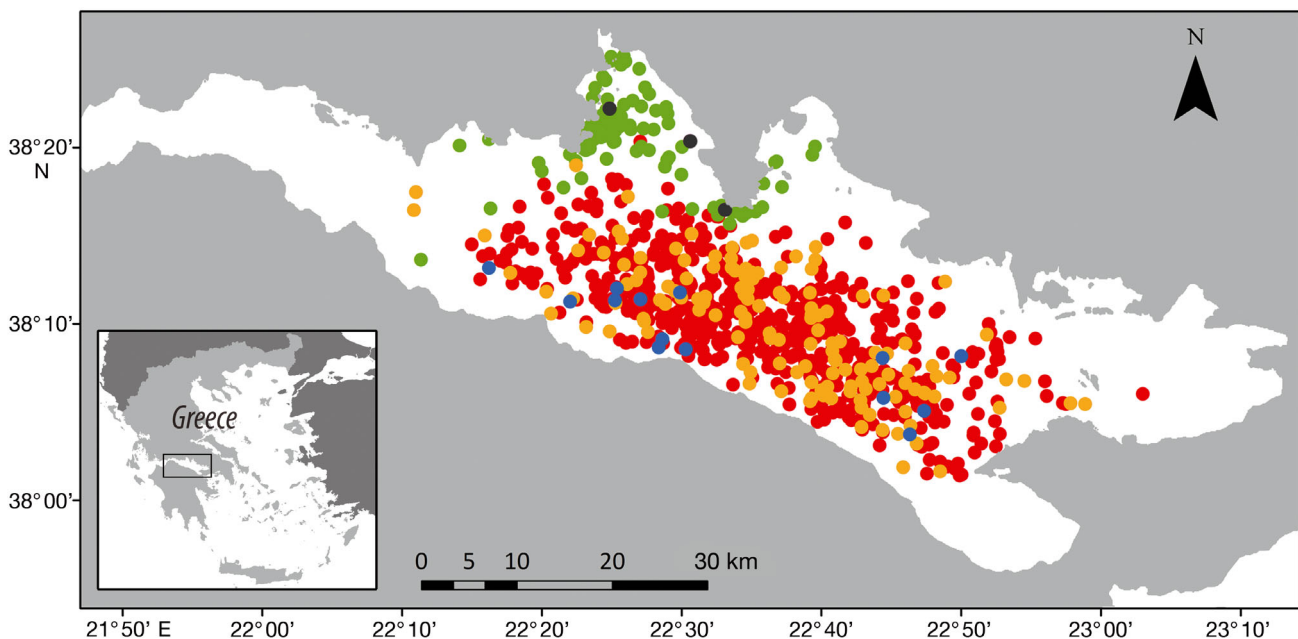


Fig. 1. Marine mammal sightings in the Gulf of Corinth, Greece (2009–2017): striped dolphin (red dots), common dolphin (orange), common bottlenose dolphin (green), Risso's dolphin (blue) and Mediterranean monk seal (black). This distribution map may convey misleading information, considering that estimates of population abundance in the Gulf are 1324, 22, 39, 1 and 1, respectively, and that 3 odontocetes (striped, common and Risso's dolphins) are found in mixed-species groups (Bearzi et al. 2016). Marine mammal distribution maps like this one also depend heavily on variables such as survey effort and sea state

phins, on the other hand, never mix with any of the other species, and they are not resident in the Gulf. They come and go—in and out of the Gulf—and the same individuals have been observed hundreds of km apart, in completely different environments, feeding on completely different prey (Bearzi et al. 2011, 2016).

In the special case of the Gulf of Corinth, a comprehensive overview of present marine mammal knowledge is now available, together with data allowing for replication of population abundance analyses (Bearzi et al. 2016, Santostasi et al. 2016). We believe that such understanding of context is often essential, and ideally it should become available prior to (or together with) the raw data. But even when the scenario has been fairly well elucidated, the problem is that online repositories rarely associate shared data with context. Biologically relevant factors are generally difficult to extract from most of the available data sharing platforms, and in some cases not taking these factors into account may affect the conclusions drawn from the dataset—be it in the context of a local or a regional study. One may retrieve marine mammal data for a given area without even knowing that background information is available for that area, therefore overlooking many of the associated issues and uncertainties.

THE FUTURE: BEST PRACTICES, ENHANCED SHARING TOOLS AND SYMBIOSIS

What can be a way forward? Coherent ‘best practice’ processes exist for the sharing and management of ecological and biodiversity data (e.g. Costello et al. 2013, 2014, Costello & Wieczorek 2014, Michener 2015, Roche et al. 2015), and to provide data contributors with appropriate motivation and reward (Costello et al. 2013). We believe that such processes and ideas should also inspire and direct the sharing of marine mammal data. The available platforms have been making remarkable progress over the years, but they still seem to lag behind in terms of essential requisites that meet the expectations of both the data providers and the data users. More sophisticated, flexible and user-friendly platforms facilitating data interpretation (e.g. relative to effort and methods as well as time and space factors), combined with area-based repositories of information (with narrative and literature describing the local and regional context) would go a long way, ultimately bridging the gap between data sharing and data reuse (Roche et al. 2015).

In addition, a closer collaboration between those who collect and share the data and those who analyse and publish them would be beneficial. A recent editorial in *The New England Journal of Medicine* (Longo & Drazen 2016) portrays the (re)users of shared data as ‘research parasites’, which sounds like a blatant overstatement (Berger et al. 2016, McNutt 2016). Whether dealing with medical or marine mammal datasets, data analysts clearly are not ‘people who had nothing to do with the design and execution of the study but use another group’s data for their own ends, possibly stealing from the research productivity planned by the data gatherers, or even use the data to try to disprove what the original investigators had posited’ (Longo & Drazen 2016, p. 276). However, we would concur with Longo & Drazen (2016) that the analysis and interpretation of shared data is best done symbiotically, ideally by involving the marine mammal scientists who actually collected the data in the field (e.g. as done by Cañadas et al. 2018). Such procedure would not only be fair to the data collectors, but would also allow the analysts to benefit from place-based expertise and understanding of context, and ensure that the relevant scientific questions are asked.

Acknowledgements. Our thanks go to Maddalena Bearzi, Silvia Bonizzoni, Lavinia Eddy and Nina Luisa Santostasi for contributing valuable comments. Silvia Bonizzoni kindly produced the map in Fig. 1. We are grateful to the editor Konstantinos I. Stergiou, to Antonios Mazaris and to an anonymous reviewer, whose comments helped improve our manuscript. Marine mammal research in the Gulf of Corinth, Greece, was supported by OceanCare.

LITERATURE CITED

- Acevedo-Gutiérrez A (2009) Group behavior. In: Perrin WF, Würsig B, Thewissen JGM (eds) *Encyclopedia of marine mammals*, 2nd edn. Academic Press, Amsterdam, p 511–520
- ✦ Bearzi G (2007) Marine conservation on paper. *Conserv Biol* 21:1–3
- ✦ Bearzi G, Bonizzoni S, Gonzalvo J (2011) Mid-distance movements of common bottlenose dolphins in the coastal waters of Greece. *J Ethol* 29:369–374
- Bearzi G, Bonizzoni S, Santostasi NL, Furey NB, Eddy L, Valavanis VD, Gimenez O (2016) Dolphins in a scaled-down Mediterranean: the Gulf of Corinth’s odontocetes. In: Notarbartolo di Sciara G, Podestà M, Curry BE (eds) *Mediterranean marine mammal ecology and conservation*. *Adv Mar Biol* 75. Academic Press, Oxford, p 297–331
- ✦ Beck J, Böller M, Erhardt A, Schwanghart W (2014) Spatial bias in the GBIF database and its effect on modeling species’ geographic distributions. *Ecol Inform* 19:10–15
- ✦ Berger B, Gaasterland T, Lengauer T, Orengo C, Gaeta B, Markel S, Valencia A (2016) ISCB’s initial reaction to *The New England Journal of Medicine* editorial on data shar-

- ing. PLOS Comput Biol 12:e1004816
- ✦ Best BD, Halpin PN, Fujioka E, Read AJ, Qian SS, Hazen LJ, Schick RS (2007) Geospatial web services within a scientific workflow: predicting marine mammal habitats in a dynamic environment. *Ecol Inform* 2:210–223
 - ✦ Cañadas A, Aguilar de Soto N, Aissi M, Arcangeli A and others (2018) The challenge of habitat modelling for threatened low density species using heterogeneous data: the case of Cuvier's beaked whales in the Mediterranean. *Ecol Indic* 85:128–136
 - ✦ Cook CN, Hockings M, Carter RW (2010) Conservation in the dark? The information used to support management decisions. *Front Ecol Environ* 8:181–186
 - ✦ Costello MJ, Wiczorek J (2014) Best practice for biodiversity data management and publication. *Biol Conserv* 173:68–73
 - ✦ Costello MJ, Michener WK, Gahegan M, Zhang ZQ, Bourne PE (2013) Biodiversity data should be published, cited, and peer reviewed. *Trends Ecol Evol* 28:454–461
 - ✦ Costello MJ, Appeltans W, Bailly N, Berendsohn WG and others (2014) Strategies for the sustainability of online open-access biodiversity databases. *Biol Conserv* 173:155–165
 - ✦ Halpin PN, Read AJ, Fujioka E, Best BD and others (2009) OBIS-SEAMAP: the world data center for marine mammal, sea bird, and sea turtle distributions. *Oceanography* 22:104–115
 - ✦ Hays GC, Ferreira LC, Sequeira AMM, Meekan MG and others (2016) Key questions in marine megafauna movement ecology. *Trends Ecol Evol* 31:463–475
 - ✦ Kaschner K, Watson R, Trites AW, Pauly D (2006) Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. *Mar Ecol Prog Ser* 316:285–310
 - ✦ Kaschner K, Tittensor DP, Ready J, Gerrodette T, Worm B (2011) Current and future patterns of global marine mammal biodiversity. *PLOS ONE* 6:e19653
 - ✦ Longo DL, Drazen JM (2016) Data sharing. *N Engl J Med* 374:276–277
 - ✦ Mann J (1999) Behavioral sampling methods for cetaceans: a review and critique. *Mar Mamm Sci* 15:102–122
 - ✦ Maurage P, Heeren A, Pesenti M (2013) Does chocolate consumption really boost Nobel award chances? The peril of over-interpreting correlations in health studies. *J Nutr* 143:931–933
 - ✦ Mazaris AD (2017) Open data and the future of conservation biology. *Ethics Sci Environ Polit* 17:29–35
 - ✦ McNutt M (2016) #IAmAResearchParasite. *Science* 351:1005
 - ✦ Michener WK (2015) Ecological data sharing. *Ecol Inform* 29:33–44
 - ✦ Roche DG, Kruuk LEB, Lanfear R, Binning SA (2015) Public data archiving in ecology and evolution: How well are we doing? *PLOS Biol* 13:e1002295,
 - ✦ Santostasi NL, Bonizzoni S, Bearzi G, Eddy L, Gimenez O (2016) A robust design capture-recapture analysis of abundance, survival and temporary emigration of three odontocete species in the Gulf of Corinth, Greece. *PLOS ONE* 11:e0166650,
 - ✦ Santostasi NL, Bonizzoni S, Gimenez O, Eddy L, Bearzi G (in press) Common dolphins in the Gulf of Corinth are Critically Endangered. *Aquat Conserv*
 - ✦ Schipper J, Chanson JS, Chiozza F, Cox NA and others (2008) The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science* 322:225–230
 - ✦ Schorr GS, Falcone EA, Moretti DJ, Andrews RD (2014) First long-term behavioral records from Cuvier's beaked whales (*Ziphius cavirostris*) reveal record-breaking dives. *PLOS ONE* 9:e92633,
 - ✦ Stern SJ (2009) Migration and movement patterns. In: Perrin WF, Würsig B, Thewissen JGM (eds) *Encyclopedia of marine mammals*, 2nd edn. Academic Press, Amsterdam, p 726–730
 - ✦ Whitehead H (2004) The group strikes back: follow protocols for behavioral research on cetaceans. *Mar Mamm Sci* 20:664–670

Editorial responsibility: Konstantinos Stergiou, Thessaloniki, Greece

Submitted: August 28, 2017; Accepted: November 22, 2017
 Proofs received from author(s): January 26, 2018