Data and Metadata Profile

The data which I have chosen for my term project is titled *Shaken, not stirred: blue whales show no acoustic response to earthquake events* and it was authored in 2022 by Dawn R. Barlow, Mateo Estrada Jorge, Holger Klinck, and Leigh G. Torres from Oregon State and Cornell Universities. The data available via Figshare consists of the "processed data … and data analysis code" for acoustic recordings of blue whale 'songs' and 'D-calls' as well as seismic events and activity, both types of data which originated from an area of the coast of Aotearoa New Zealand called South Taranaki Bight (STB). Data of recordings of blue whale calls were sourced from another larger effort to study their patterns, while earthquake data came from a database of seismic activity in the same area. (Barlow 2022) In total there are 17 files associated with these data: 15 files of processed data in .csv files, which is one of the most common file format for spreadsheets and can be opened with a text editor or spreadsheet program like Excel and; two files of data analysis code that appear as HTML files and are written in, I believe, the R programming language and, therefore, would require the user to download the R software in order to run. Access to all of these files is unrestricted and I haven't found any proprietary notice for the analytical code so, assuming a user had the R software installed, they would be free to 'check' the data analysis themselves or likely even adapt it for other data analysis situations. Additionally these data are licensed under the CC BY 4.0 (Creative Common Attribution 4.0 International) license, which allows them to be reproduced and even adapted as long as the correct conditions of crediting/attribution, which include disclosing if any changes have been made, are met.

In terms of metadata, there is a research and data analysis article, under the same name as these data, published by the authors in the *Royal Society Open Science* journal; however, since that is 'outside' metadata, the most informative metadata in Figshare is the Abstract. The same abstract provided for the journal article, it presents the numerical results and data analysis of the processed data, along with the answer to the research question posed by the authors. According to their analysis and observations, while blue whale calls demonstrate fluctuations surrounding earthquakes, they do not differ significantly from the regular fluctuations in calls that have been previously recorded. This implies that blue whale vocalizations and calls are used for more complex communication resembling language, rather than the simplistic function of just 'making noise when something happens." (Barlow 2022) Neither the data and files available in Figshare nor the published article directly reference a specific metadata standard, however, the database used to track seismic data, GeoData, states that they use the ISO 19115:2003/19139 metadata standard, which is also the same standard that the Metadata Standards Catalog states can be used for both marine biology and oceanography. (GNS Science 2019) It is also important to note that there is no specific metadata or any other definitions that illustrate exactly what all the variables are in all of the data files available, which makes it difficult to fully understand what is being measured without any further context.

Overall, I would consider the abstract on Figshare and the journal article from the *Royal Society Open Science* - which I found by searching the Figshare entry title in the UW

Libraries catalog - to be the most important and informative metadata related to these data, but there are still some small pieces of unstructured metadata that can be found in the Figshare entry. Although the HTML documents featuring the R code to be used for data analysis are mostly code, there are some small bits of commentary or clarification in them that can serve as metadata, as well as some parts of the code itself depending on how well the user might be able to read it. The email of the 'primary' author, Dawn Barlow, is listed in the Figshare entry, as well as the universities and departments that each of the authors come from. Additional metadata provided by the Figshare entry includes how many views, downloads, and citations these data have been involved in, as well as the categories and keywords that can be applied to these data. One thing that caught my attention regarding the metadata was that there was only one place - the "Keywords" - where Aotearoa New Zealand, the place that these data come from, is mentioned. Neither the R code, the abstract, the processed data, nor any other part of the Figshare entry mentions it and, in fact, because all four of the authors are listed as being associated with American universities and Oregon State University in particular, I believe that is potentially misleading. Overall, this doesn't seem to be too significant of a problem, because the processed data includes coordinates for the acoustic monitoring devices as well as the locations of earthquakes, both of which are in that STB area of Aotearoa New Zealand, but there's still some potential for confusion.

  The stakeholders for these data that are explicitly named are the four authors as well as their respective universities and departments; the academic and financial supporters named in the Acknowledgements and Funding sections of the corresponding research paper; Aotearoa New Zealand, and the journal involved in publishing the research paper and analysis related to these data, the Royal Society Open Science. The implicit stakeholders of these data could include: marine biologists; geologists and/or seismologists; possibly amateur whale scientists or whale watchers; anyone commercially involved in an industry that might be affected by a change in blue whale patterns and; scientists involved in studying noise pollution. Additionally, the data of the acoustic recordings of blue whale calls that the authors used was sourced from a semi-related larger research study examining the general blue whale population in Aotearoa New Zealand, so anyone involved in that study would also be considered a stakeholder. Finally, because the authors of these data got their seismic data from the GeoData database, anyone working for that database or with the same data as these authors could be considered a loosely related stakeholder.

Repository Profile

  To store the dataset I am working with, I am using the PANGAEA Data Publisher repository which is, in its own words, is "an Open Access library aimed at archiving, publishing and distributing georeferenced data from earth systems research." (Felden, et. al. 2023) This repository is a good fit for my dataset because its scope encompasses a wide range of earth and environmental science topics; agriculture, atmosphere, biological classification, biosphere, chemistry, cryosphere, ecology, fisheries, geophysics, human dimensions, lakes and rivers, land surface, lithosphere, oceans, and paleontology; and it also already includes data for both

earthquakes/seismic activity and blue whales. In particular, their About section emphasizes that they specialize in "georeferenced observational and experimental data" which I am hoping will be particularly suited to my dataset because it deals with coordinates of seismic activity along with a relatively new type of acoustic monitoring device, the MARU unit. While PANGAEA does require potential submitters to register an account with them and encourages them to also connect their ORCID iD, there don't appear to be any other limitations or official restrictions for submissions, so it is otherwise open to submissions from anyone.

In terms of limitations and guidance for submissions, the authors and editors of PANGAEA have created a Wiki Authors Guide, along with several other supplemental materials, to assist submitters in their work. As mentioned previously, this repository prioritizes observational and experimental research data, and the guide states specifically that it "specialize[s] in field observation and experimental data in two-dimensional tabular format with parameters/variables measured provided in columns." (Schumacher[a] 2024) The primarily preferred submission formats are TAB-deliminated TEXT-files (UTF-8) and Excel files, but other accepted formats are documentation (PDF/a, ODF, plain UTF-8 text, MS Office files), images (.tiff, .jpeg, .png), video (MP3, MPEG2, AAC, MPEG-4), audio (MP3, WAVE), seismic (segy), ADCP, non-model large-array oriented scientific data, compressed zip files, and references to proprietary software that can be used to open files. Submission formats that are not accepted include raw data (which PANGAEA defines as "data processing level 0" data that is entirely unprocessed and does not have any metadata), sequence data, model data, software code, device-specific and proprietary tables, and any format that would not be compatible with open-source technology. (Schumacher[b] 2024) Along with all of this more general information, the Authors Guide also provides a step-by-step walkthrough of a submission process which I believe is the closest thing they provide to a defined Submission Information Package. This walkthrough outlines everything needed to complete the PANGAEA submission form: basic information like title, authors, keywords, abstract, and licenses; references as full citations; applicable projects and grants; data file upload and file description with advice on different data sizes; and final submission. (Schumacher[a] 2024)

Along with the submission walkthrough, the wiki guide also provides submitters with instructions and the specific requirements for metadata about the data and preparing data in the tabular format preferred by PANGAEA. The Data Policy page states that PANGAEA follows "metadata fields according to ISO19115, GCMD-DIF and DublinCore standards" and they also state in several places, including the repository's Mission and Scope page, that they operate in accordance with the FAIR principles for scientific data management and stewardship. (Schumacher[c] 2024; Felden, et. al. 2023) There are no 'live' human assistants or submission consultations available, but again, PANGAEA provides a substantial amount of supplemental material including metadata templates, YouTube workshops and tutorials, best practices manuals, and community workshops in the winter and spring. In addition to all of the above mentioned tools for submitters, it is also important to mention that PANGAEA is very clear about the

post-submission editorial process as well as how long the turnaround is for final publication and availability in the repository. (Schumacher[a] 2024)

While creating an account and logging in is required for submission of data, PANGAEA does not require the same for downloading and viewing data. However, creating an account does not require any verification more complicated that what can usually be expected from registering any online account: you create a username, provide your email, your full name, and a few other optional details after which an email verification is needed to finalize your account. General users can search the repository, refine their search, view single entries and data review statuses, see a portion of the metadata, and download data all without having an account. Some of the data and parameters are viewable in-page in tabular format and/or HTML text, but downloadable data - sometimes in compressed ZIP form - is available is a few different formats: UTF-8 Unicode (PANGAEA's preferred format); ISO 8859-1; windows-1252; x-MacRoman; IBM437; IBM850. (Felden, et. al. 2023) There are also dozens of other characters encoding formats available for download but PANGAEA does not recommend using them because they may change some characters. The Dissemination Information Package does not differ too significantly from the information required for the SIP, but visual formatting of each data entry does occur during the editorial process after submission and, as can be expected, the metadata that is displayed differs from entry to entry depending on the details of the data.

Additional Information

Recommended Data Citation

Barlow, D., M.E. Jorge, H. Klinck, L.G. Torres. 2022. Shaken, not stirred: blue whales show no acoustic response to earthquake events. *Figshare*. https://doi.org/10.6084/m9.figshare.20001752. Accessed 2026-2-25.

Long-Term Preservation Considerations

Considerations for long-term preservation of these data should largely involve things like fixity and technology checks. The technology required to analyze, view, and work with these data is free, and not particularly difficult to acquire, but it should be kept "up to date." The file format of the tabular files (.csv) is one of the most commonly used spreadsheet file formats but preservationists should still perform scheduled checks to make sure that the files are still able to be opened, or if the file format needs to be updated. In the same vein, the R programming language and R software used to write and run the data analysis code is a free software that is available from numerous institutions, but should also still be checked at regular intervals to see if the data analysis code written by the authors needs to be updated or not.

Copyright License

The authors of these data have licensed it under the CC (Creative Commons) BY 4.0 license, which is a license that allows other users and/or researchers to use these data in other places and even change or add to it, as long as proper credit in the correct format is given.

Human Subject Considerations

       The research these data come from involves monitoring blue whale vocalizations and seismic activity, and does not involve any human subjects or sensitive personal information.

Reference List

Barlow, D. R., et. al. (2022). *Shaken, not stirred: blue whales show no acoustic response to earthquake events*. Royal Society Open Science, Volume 9, Issue 7. https://doi.org/10.1098/rsos.220242

Barlow, D.R., et. al. (2022). *Shaken not stirred: blue whales show no acoustic response to earthquake events.* Figshare. https://doi.org/10.6084/m9.figshare.20001752

Felden, Janine; Möller, Lars; Schindler, Uwe; Huber, Robert; Schumacher, Stefanie; Koppe, Roland; Diepenbroek, Michael; Glöckner, Frank Oliver. (2023). *PANGAEA – Data Publisher for Earth & Environmental Science.* Scientific Data, 10(1), 347, https://doi.org/10.1038/s41597-023-02269-x

GNS Science. (2019). GeoNet Aotearoa New Zealand Stations Metadata Repository [Data set]. GNS Science, GeoNet. https://doi.org/10.21420/0VY2-C144

Research Data Alliance. (2025) The RDA Metadata Standards Catalog: Marine biology; Oceanography. https://rdamsc.bath.ac.uk/subject/Geology

Schumacher, S.[a] (2024). *PANGAEA Wiki: Authors Guides*. MediaWiki. https://wiki.pangaea.de/w/handler?title=Main_Page&oldid=16361

Schumacher, S.[b] (2024). *PANGAEA Wiki: Format Guide*. MediaWiki. https://wiki.pangaea.de/wiki/Format

Schumacher, S.[c] (2024). *PANGAEA Wiki: Data policy*. MediaWiki. https://wiki.pangaea.de/wiki/Data_policy