Taylor & Francis
Taylor & Francis Group

Big Earth Data

ISSN: 2096-4471 (Print) 2574-5417 (Online) Journal homepage: www.tandfonline.com/journals/tbed20

Big data, sound science, lasting impact: A framework for passive acoustic monitoring

Carrie C. Wall, Megan F. McKenna, Leila T. Hatch, Sofie M. Van Parijs, Rob Bochenek, Peter Dugan, Clea Parcerisas, John Ryan, Charles D. Anderson, Kyle Becker, Catherine Berchok, Mathew Biddle, Olaf Boebel, Adrienne Canino, Gabrielle Canonico, Danelle Cline, Genevieve E. Davis, Kaitlin Frasier, Jason Gedamke, Samara M. Haver, Karina Khazmutdinova, Niels Kinneging, Anurag Kumar, Alyssa Marian, Jennifer L. Miksis-Olds, Eric W. Montie, Dimitri Ponirakis, Aaron N. Rice, Timothy J. Rowell, Carlos Rueda, Emily Shumchenia, Thomas Shyka, Erica Staaterman & Karolin Thomisch

To cite this article: Carrie C. Wall, Megan F. McKenna, Leila T. Hatch, Sofie M. Van Parijs, Rob Bochenek, Peter Dugan, Clea Parcerisas, John Ryan, Charles D. Anderson, Kyle Becker, Catherine Berchok, Mathew Biddle, Olaf Boebel, Adrienne Canino, Gabrielle Canonico, Danelle Cline, Genevieve E. Davis, Kaitlin Frasier, Jason Gedamke, Samara M. Haver, Karina Khazmutdinova, Niels Kinneging, Anurag Kumar, Alyssa Marian, Jennifer L. Miksis-Olds, Eric W. Montie, Dimitri Ponirakis, Aaron N. Rice, Timothy J. Rowell, Carlos Rueda, Emily Shumchenia, Thomas Shyka, Erica Staaterman & Karolin Thomisch (04 Dec 2025): Big data, sound science, lasting impact: A framework for passive acoustic monitoring, Big Earth Data, DOI: 10.1080/20964471.2025.2583505

To link to this article: https://doi.org/10.1080/20964471.2025.2583505

Published online: 04 Dec 2025.

BIG EARTH DATA
https://doi.org/10.1080/20964471.2025.2583505

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Big data, sound science, lasting impact: A framework for passive acoustic monitoring

Carrie C. Wall [a,b], Megan F. McKenna [a,b], Leila T. Hatch [c], Sofie M. Van Parijs [d], Rob Bochenek [e], Peter Dugan [f,g], Clea Parcerisas [h,i], John Ryan [j], Charles D. Anderson [a,b], Kyle Becker [k], Catherine Berchok [l], Mathew Biddle [m], Olaf Boebel [n], Adrienne Canino [e], Gabrielle Canonico [m], Danelle Cline [j], Genevieve E. Davis [d], Kaitlin Frasier [o], Jason Gedamke [p], Samara M. Haver [q,r], Karina Khazmutdinova [e], Niels Kinneging [s], Anurag Kumar [t], Alyssa Marian [u], Jennifer L. Miksis-Olds [v], Eric W. Montie [u], Dimitri Ponirakis [g], Aaron N. Rice [g], Timothy J. Rowell [d,w], Carlos Rueda [j], Emily Shumchenia [x], Thomas Shyka [y], Erica Staaterman [z] and Karolin Thomisch [n]

[a]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA; [b]NOAA National Centers for Environmental Information, Boulder, CO, USA; [c]NOAA Office of National Marine Sanctuaries, Silver Spring, MD, USA; [d]NOAA Northeast Fisheries Science Center, Woods Hole, MA, USA; [e]Axiom Data Science, a Tetra Tech company, Anchorage, AK, USA; [f]US Naval Undersea Warfare Center, Middletown, RI, USA; [g]K. Lisa Yang Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University, Ithaca, NY, USA; [h]Flanders Marine Institute, Ostend, Belgium; [i]Department of Information Technology, Ghent University, Ghent, Belgium; [j]Monterey Bay Aquarium Research Institute, Moss Landing, CA, USA; [k]Office of Naval Research, Arlington, VA, USA; [l]NOAA Alaska Fisheries Science Center, Seattle, WA, USA; [m]NOAA Integrated Ocean Observing System, Silver Spring, MD, USA; [n]Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Ocean Acoustics Group, Bremerhaven, Germany; [o]Scripps Institution of Oceanography, La Jolla, CA, USA; [p]NOAA National Marine Fisheries Service, Office of Science and Technology, Silver Spring, MD, USA; [q]Cooperative Institute for Marine Ecosystem and Resources Studies, NOAA Pacific Marine Environmental Laboratory and Oregon State University, Newport, OR, USA; [r]Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Corvallis, OR, USA; [s]Rijkswaterstaat, Ministry of Infrastructure and Water Management, Netherlands; [t]US Navy Living Marine Resources, Port Hueneme, CA, USA; [u]Department of Natural Sciences, University of South Carolina Beaufort, Bluffton, SC, USA; [v]Center for Acoustics Research and Education, University of New Hampshire, Durham, NH, USA; [w]NOAA Southeast Fisheries Science Center, Beaufort, NC, USA; [x]Regional Wildlife Science Collaborative for Offshore Wind, USA; [y]Northeastern Regional Association of Coastal Ocean Observing Systems, Portsmouth, NH, USA; [z]Bureau of Ocean Energy Management Center for Marine Acoustics, Sterling, VA, USA

ABSTRACT

Marine passive acoustic monitoring (PAM) has produced petabytes of data that are used by researchers, resource managers, industry, and regulators to understand how marine animals use sound and the impacts of anthropogenic noise on species and ecosystems throughout the global ocean. These big data provide unprecedented opportunities to study underwater soundscapes but also enormous challenges to efficiently extract information. To address these challenges, the Sound Cooperative (*SoundCoop*) project built community-

ARTICLE HISTORY
Received 20 May 2025
Accepted 16 October 2025

KEYWORDS
Passive acoustic monitoring; big data analytics; ocean soundscape; cyberinfrastructure; FAIR principles; open science;

CONTACT Carrie C. Wall carrie.bell@colorado.edu Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, 216 UCB, Boulder, CO 80309, USA

Supplemental data for this article can be accessed online at https://doi.org/10.1080/20964471.2025.2583505

focused cyberinfrastructure to promote improved, scalable and sustainable processing and access of marine PAM data for management, science, industry and military applications. Driven by cross-institutional participation representing a diversity of data collection methods and conditions, *SoundCoop* established guidance for standardized processing of ocean sound level metrics using freeware software toolkits and developed core tools and processes that support open science. Four comparative analyses that connect PAM monitoring efforts and integrate non-acoustic data illustrate how interoperable sound level metrics support a more coherent and synoptic perspective on ocean soundscapes using methods that current and future PAM projects can leverage. Such a framework around PAM big data offers the opportunity to revolutionize large-scale marine ecology and oceanography in similar ways to other transformative approaches for understanding environmental patterns and processes at global scales.

analysis-readydatasets; AI/ ML-ready datasets

1. Introduction

Big data, characterized by its high volume, velocity, veracity, variety, and value, have revolutionized how we collect, analyze, and interpret information across diverse fields (Li et al., 2023; Vance et al., 2024). Passive acoustic monitoring (PAM), an example of big data, provides unprecedented opportunities to study soundscapes and their ecological and anthropogenic drivers across terrestrial, freshwater, and marine environments (Desjonquères et al., 2020; Duarte et al., 2021; Greenhalgh et al., 2020; Rasmussen et al., 2024; Ross et al., 2023). Specifically, marine PAM is employed throughout the global ocean by academic researchers, resource managers, industry, and military to understand how marine animals use sound and the impacts of anthropogenic noise on species and ecosystems, including how ocean soundscapes are affected by change (Affatati et al., 2022; Caiger et al., 2020; Davis et al., 2020; Delarue et al., 2022; Duarte et al., 2021; Hatch et al., 2016; Mariani et al., 2024; McKenna et al., 2024; Miksis-Olds et al., 2013; Pearson et al., 2023; Possenti et al., 2024; Rice et al., 2021; Ryan et al., 2021; Transue et al., 2023; Tribble et al., 2023; Van Parijs et al., 2023; ZoBell et al., 2024) (Table 1). These diverse drivers for data collection coupled with cheaper and thus more accessible instrumentation, longer deployment periods and an increasing variety of deployed sensors have led to an exponential growth of PAM data over the past decade (Gibb et al., 2019; Howe et al., 2019; Kowarski & Moors-Murphy, 2021).

With this large variety of PAM users and uses (Table 1), it is not surprising that tremendous progress has been made by individual projects, spanning regional to national scales and across multiple years, to develop practices for data curation, management, integration, and visualization. While not exhaustive, Box 1 provides examples of the progress made in these areas. As the PAM community moves toward international standardization for soundscape measurements, several best practices documents were published (Consortium for Ocean Leadership, 2018; International Quiet Ocean Experiment, 2019; International Whaling Commission, 2014) and have been implemented among long-term monitoring projects (e.g., Miksis-Olds et al., 2025).

Previous analyses demonstrated the value of exploring data from separate monitoring projects by leveraging their centralized availability and a standardized soundscape measurement (Wall et al., 2021); however, scalable methods for interpretation and comparison that are generalizable for broader meta-analysis or community adoption lag, and thus hindering

Table 1. Diverse applications of PAM, including examples of the user-community and their applications of the data.

User Community	Applications for Passive Acoustic Monitoring Big Data	Example References
Offshore industries	Offshore energy companies: siting, permit applications, monitoring plan designs, protected species compliance Ports and maritime transport organizations: routing, traffic assessments	ZoBell, Vanessa M., et al. "Retrofit-induced changes in the radiated noise and monopole source levels of container ships." *PLOS ONE* 18.3 (2023): e0282677.
Federal action agencies	BOEM: NEPA environmental assessment for offshore development U.S. Navy and U.S. Coast Guard: Testing/training and port access	Van Parijs, Sofie M., et al. "NOAA and BOEM minimum recommendations for use of passive acoustic listening systems in offshore wind energy development monitoring and mitigation programs." *Frontiers in Marine Science* 8 (2021): 760840.
State and federal natural resource agencies	NOAA Fisheries: Integrated Ecosystem Assessment, essential fish habitat/habitat of particular concern-spawning fish monitoring, Endangered Species Act/Marine Mammal Protection Act-cetacean monitoring NOAA ONMS: condition reports, management plans; California Ocean Protection Council, West Coast Regional Alliance New York State Department of Environmental Conservation: New York Bight Whale Monitoring Program	Bolgan, Marta, et al. "Use of passive acoustic monitoring to fill knowledge gaps of fish global conservation status." *Aquatic Conservation: Marine and Freshwater Ecosystems* 33.12 (2023): 1580–1589.
Academic and industry big data users	Machine learning and artificial intelligence techniques applied to oceanographic and ecological studies	Frasier, Kaitlin E. "A machine learning pipeline for classification of cetacean echolocation clicks in large underwater acoustic datasets." *PLoS Computational Biology* 17.12 (2021): e1009613.
Cross-federal to international organizations	Council on Environmental Quality-SOST Interagency Ocean Sound and Marine Life Task Force COVID soundscape study United Nations International Maritime Organization's ship quieting initiatives International Quiet Ocean Experiment Science Plan initiatives Global Ocean Observing System (GOOS) Essential Ocean Variable implementation plan	Miksis-Olds, Jennifer L., et al. "Envisioning a global multi-purpose ocean acoustic network." *Marine Technology Society Journal* 55.3 (2021): 78–79.
Academic acousticians	Project PIs who collect PAM data for research purposes and/or teaching students, or for science communication	Parsons, Miles JG, et al. "A Global Library of Underwater Biological Sounds (GLUBS): An online platform with multiple passive acoustic monitoring applications." The Effects of Noise on Aquatic Life: Principles and Practical Considerations. Cham: Springer International Publishing, 2024. 2149–2173.
Environmental consultants	Companies who provide services that could encompass characterizing industrial noise, biological sounds, and shipping activity	Austin, Melanie E., S. Bruce Martin, and Craig R. McPherson. "Measurements of underwater radiated noise from mobile offshore drilling units." *The Effects of Noise on Aquatic Life: Principles and Practical Considerations*. Cham: Springer International Publishing, 2024. 1683–1696.
Non-profit organizations	Organizations who collect and use passive acoustic data, typically—but not exclusively—for marine mammal observations and/or science popularization	https://www.raincoast.org/noisetracker/ Vergara, Valeria, et al. "Effects of vessel noise on beluga (*Delphinapterus leucas*) call type use: ultrasonic communication as an adaptation to noisy environments?" *Biology Open* 14.3 (2025): bio061783.

Box 1. Passive Acoustic Monitoring Big Data Projects

This is not an exhaustive list and intends to highlight efforts in the two decades that the collaborative framework presented here has been built from. This list only includes projects that have a significant PAM data collection component of the project. Other projects not listed (e.g., International Quite Ocean Experiment (IQOE): https://iqoe.org/) serve as more of a coordination role in PAM Big Data. All websites last accessed on Jan. 15, 2025

NOAA Fisheries Ocean Acoustics Program | 2006–current | https://www.fisheries.noaa.gov/national/science-data/ocean-noise

Regional ocean noise activities and programs are conducted across NOAA Fisheries offices to support a multitude of monitoring objectives to better understand fish and marine mammal populations, and to inform how different anthropogenic noise sources might impact protected species. The result is petabytes of PAM data and a wealth of marine mammal detections collected over decades and throughout U.S. waters.

Ocean Noise Reference Station (NRS) | 2014–current | https://www.pmel.noaa.gov/acoustics/noaanps-ocean-noise-reference-station-network

The objective of the project is to establish a network of ocean noise reference stations in U.S. waters to monitor long-term changes and trends in the underwater ambient sound field. The identical autonomous acoustic recording systems were developed and built in-house at Pacific Marine Environmental Laboratory to ensure proper calibration and consistency of the collected data sets. This unique network of hydrophones is a collaborative effort between OAR's Pacific Marine Environmental Laboratory, all NMFS Science Centers, the NOS National Marine Sanctuary System, and the National Park.

Sanctuary Soundscape Monitoring Project (SanctSound) | 2018–2021 | https://sanctsound.ioos.us/

This project was led by NOAA's Office of National Marine Sanctuaries and the U.S. Navy to better understand underwater sound within the National Marine Sanctuary System. The agencies worked with numerous scientific partners to study sound within seven national marine sanctuaries and one marine national monument, which included waters off Hawai`i and the east and west coasts. The project generated large volumes of raw audio, standardized sound level metrics, and single sound source detections for all sites.

Office of National Marine Sanctuaries Sound Monitoring Program (ONMS Sound) | 2022–current | https://sanctuaries.noaa.gov/science/monitoring/sound/

This program coordinates monitoring across the National Marine Sanctuary System to study sound within ten existing and two proposed national marine sanctuaries off the US East Coast, in the Gulf of Mexico, off the West Coast and in the Pacific Islands region. ONMS Sound generates audio recordings and standardized sound measurements

Scripps Institution of Oceanography (Scripps) Whale Acoustic Lab (SWAL) & Marine Bioacoustic Research Collaborative (MBARC) | 2004–current | https://www.cetus.ucsd.edu/index.html **&** https://acoustics.ucsd.edu/

This long-term monitoring program aims to advance the study of marine animal population abundance, seasonality and behavior using acoustic techniques. In addition to generating petabytes of data, the team develops new acoustic instrumentation and analytical approaches to improve understanding of the ecology of marine animals and the impact of anthropogenic sounds on them.

Cornell University K. Lisa Yang Center for Conservation Bioacoustics | 2004–current | https://www.birds.cornell.edu/ccb/

This program conducts a broad range of terrestrial, aquatic, and marine bioacoustic research, often at large geographic scales. Many projects are focus on the conservation of endangered species, soundscape ecology, and the development of acoustic metrics to assess biodiversity and ecosystem health, generating multiple petabytes of PAM data.

Estuarine Soundscape Observatory Network in the Southeast (ESONS) | 2013–current | https://secoora.org/observations-systems/marine-life/

This network provides, long-term monitoring of soundscapes in estuaries of South Carolina, USA. These recordings provide information on the behavior of snapping shrimp, spawning patterns of fish, foraging patterns and communication of bottlenose dolphins, and noise levels associated with human activity. The long-term goal is to 'eavesdrop' on key behaviors of marine animals that can change rapidly or gradually in response to environmental changes and human impacts, thus providing a measure of resilience or shifting baselines in a globally changing environment.

Monterey Accelerated Research System (MARS) | 2015–current | https://www.mbari.org/technology/monterey-accelerated-research-system-mars/

This cabled observatory hosted at the Monterey Bay Aquarium Research Institute facilitates long-term and real-time analyses from data collected in Monterey Bay. MARS' omni-directional hydrophone records continuously within and outside of the human hearing range. MBARI staff developed methods to facilitate data access and use: raw data are processed to daily files of decimated data at 2 kHz and 16 kHz sample rates, high-resolution spectrograms for the full bandwidth of the recordings are accessible through a visual browsing interface, and all data are available freely through the Amazon Web Services Open Data Program, together with tutorials for research applications using open-source software.

Atlantic Deepwater Ecosystem Observatory Network (ADEON) | 2017–2021 | https://adeon.unh.edu/

This observatory network generated long-term measurements of both the natural and human factors active in this region, thus informing the ecology and soundscape of the U.S. Mid- and South Atlantic Outer Continental Shelf. The integrated data, including PAM, provided a mechanistic understanding of the cumulative impacts these factors have on marine resources and provided insight for ecosystem-based management efforts. Long-term observations of living marine resources and marine sound is intended to assist Federal agencies, including BOEM, ONR, and NOAA, in complying with mandates in the Endangered Species Act (ESA), Marine Mammal Protection Act (MMPA), and Sustainable Fisheries Act (SFA).

Acoustic and Environmental Observation Network (AEON) | 2022–current | https://eos.unh.edu/aeon
This network provides simultaneous, long-term monitoring of soundscapes and multiple acoustically relevant parameters such as marine mammal behavior and prey concentration at key locations where the effects of changes in the Labrador and Gulf Stream currents are projected to impact the Gulf of Maine.

Alfred Wegener Institute (AWI) Passive Acoustic Monitoring | 2006–current |
The Alfred Wegener Institute (AWI) in Germany conducts long-term PAM projects in the Arctic and the Antarctic including Frontiers in Arctic Marine Science (FRAM). FRAM focused on recording sound in the northern Greenland Sea to monitor for whales, seals, fish and invertebrates between Spitsbergen and Greenland.

Joint Monitoring Programme for Ambient Noise North Sea (JOMOPANS) | 2018–2021 | https://northsearegion.eu/jomopans/
This project operationalized a coordinated monitoring of ambient noise in the North Sea across multiple European countries and agencies, which resulted in new documentation and tools for stakeholders to incorporate ambient noise into assessments as well as public access to the decidecade sound level data on the International Council for Exploration of the Seas (ICES) database.

Open Portal for Underwater Soundscapes (OPUS) | 2020–current | https://opus.aq/
OPUS was developed by the International Quiet Ocean Experiment's (IQOE) Data Office at AWI. This interactive visualization portal enables exploration and listening to large volumes of passive acoustic data. OPUS relies on metadata and assurance of quality control throughout the data processing and visualization steps for submitted datasets.

World Oceans Passive Acoustic Monitoring Day | 2023–current | https://www.wo-pam.com/
This global initiative coordinates the volunteer collection and submission of underwater sounds recorded on June 8 of each year to create a snapshot of the cacophony of noises generated by wildlife and humans. Data from over 400 sites all over the world were contributed to the 2023 collection. With these recordings, the WOPAM community will collaboratively explore what ocean sound means for animal behavior, biodiversity and ecosystem integrity.

delivery of actionable metrics to meet diverse applications, exposing a need to strategically expand collaborative efforts.

The Sound Cooperative (*SoundCoop*) project was created to address challenges in PAM curation, processing, and sharing by developing innovative, community-oriented tools and processes that make it easier and more efficient to extract information from large datasets and compare across monitoring programs. Funded by three U.S. federal government agencies (National Oceanic and Atmospheric Administration [NOAA], Bureau for Ocean Energy Management [BOEM] and U.S. Navy), *SoundCoop* leveraged existing national and international datasets and management efforts, identified remaining gaps, and built capacity towards community-prioritized functionality. The project's framework focused on cyberinfrastructure that 1) calculated a comparable, standardized sound level metric using freeware software for a variety of recording equipment; 2) developed a standardized file format for an interoperable metric output; 3) established workflows to access data from separate cloud repositories; 4) co-visualized sound levels and integrated them with environmental data in a performant, interactive public portal; and 5) published open-source tools for the community to replicate the project's processing and visualization methods. The success of such applications relies on the ability to develop standard and broadly accessible data management mechanisms, and to continue to ensure that these mechanisms can scale effectively with increasing global participation. Here, we showcase the progress achieved through *SoundCoop* in addressing big data challenges for PAM, presenting the project as a model framework for tackling similar challenges across scientific fields, and distributing the best practices and approaches to a broader earth science community. We also present four analytical comparisons that apply those best practices to demonstrate their utility and versatility in exploring ocean soundscapes.

2. Materials and methods

2.1. Broad community input

Given the international scope, multiple U.S. federal agency applications, and diverse applications (Table 1), the *SoundCoop* effort was guided by an advisory committee and several stakeholder engagement events. The advisory committee represented experts across federal agencies, academic institutions, and industry, namely the U.S. Navy's Living Marine Resources Program and Office of Naval Research, BOEM, NOAA National Marine Fisheries Service (NMFS), NOAA Office of National Marine Sanctuaries, NOAA Integrated Ocean Observing System (IOOS), IOOS' Northeastern Regional Association of Coastal Ocean Observing System, University of Colorado Boulder/NOAA National Centers for Environmental Information (NCEI), Axiom Data Science, Monterey Bay Aquarium Research Institute, the Regional Wildlife Science Collaborative for Offshore Wind, Alfred Wegener Institute (AWI) in Germany, International Quiet Ocean Experiment, and Rijkswaterstaat Ministry of Infrastructure and Water Management in Netherlands.

To avoid duplication and divergent products, we sought to promote the use of centralized assets (i.e., data and tools) where appropriate, leverage what has already been built for a larger group of stakeholders, and coordinate further development opportunities. Therefore, during the first phase of community input, the project's steering committee collated and reviewed information on the status of existing national and international PAM data management tools in support of *SoundCoop*'s focus on enhancing longer-term and large-scale curation, access and interoperability. Previously collected PAM datasets were selected by the team that represent 12 disparate monitoring efforts with varying collection purposes and applications, 10 recording instruments deployed across 7 geographic regions spanning 17 years and varying sampling regimens (Figure 1).

Datasets were grouped into four categories:

(1) Long-time series recorded in the Arctic. Data were contributed from the NOAA NMFS Alaska Fisheries Science Center Arctic Long-Term Integrated Mooring Array (AFSC-ALTIMA), NOAA-National Park Service Ocean Noise Reference Station Network (NRS), and U.S. Navy-funded Scripps Whale Acoustic Laboratory (SWAL) projects (NOAA Alaska Fisheries Science Center Marine Mammal Laboratory [NOAA], 2023; NOAA OAR Pacific Marine Environmental Laboratory et al., 2023; Scripps Institution of Oceanography, 2023).
(2) Variation in recordings throughout 2021 across three geographic regions: Gulf of Maine, Southeast U.S., and Central California Coast. Gulf of Maine datasets were contributed from the Acoustic and Environmental Observation Network (AEON) program, NOAA-Navy Sanctuary Soundscape Monitoring (SanctSound) project, and NOAA NMFS Northeast Fisheries Science Center (NEFSC) monitoring project (NOAA Northeast Fisheries Science Center [NOAA], 2023; NOAA Office of National Marine Sanctuaries & U.S Navy, 2023; University of New Hampshire & Office of Naval Research, 2023). The Southeast U.S. dataset was recorded in the May River from

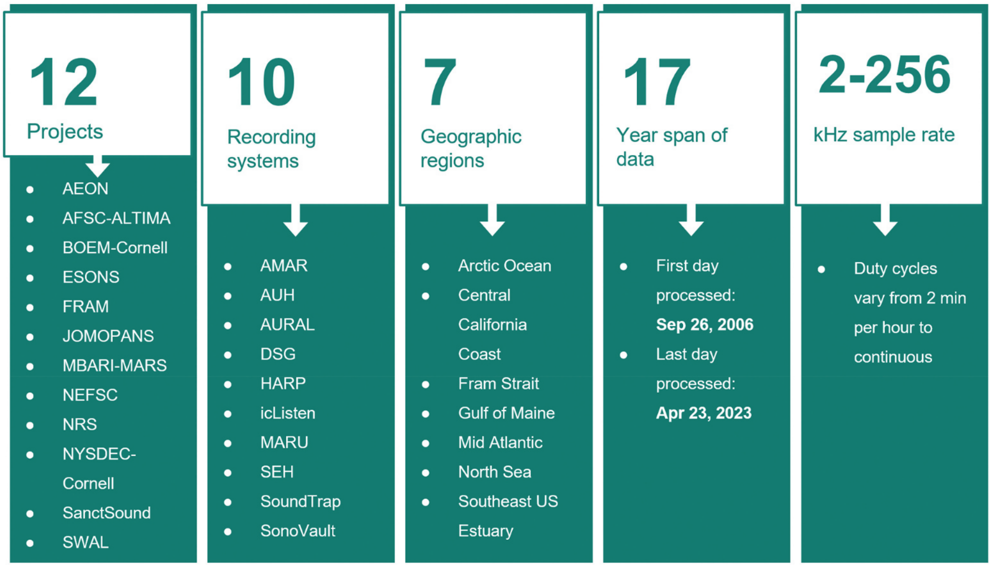

Figure 1. Scope of this *SoundCoop* project to build the collaborative framework for PAM big data. Breakdown of the datasets incorporated into this effort representing a variety of projects, recording systems, geographic regions, non-continuous time span of recording, and sampling regimens.

the Estuary Soundscape Observatory Network in the Southeast (ESONS) project (Marian et al., 2021; Monczak et al., 2022; University of South Carolina Beaufort & IOOS SECOORA, 2023). The Central California Coast datasets were from MBARI Monterey Accelerated Research System (MARS) and the NRS project (Monterey Bay Aquarium Research Institute, 2023; NOAA OAR Pacific Marine Environmental Laboratory et al., 2023).

(3) Offshore wind monitoring. Sound data were contributed from a site off Virginia (Mid Atlantic) that was part of a BOEM-funded, Cornell University-led project as well as North Atlantic right whale (*Eubalaena glacialis*) upcalls detected off the New York Bight from a New York State Department of Environmental Conservation-funded, Cornell University-led (NYSDEC-Cornell) project (Cornell University & Bureau for Ocean Energy Management, 2023; Dugan et al., 2013; Estabrook et al., 2021).

(4) International comparisons. Two international (non-U.S.) datasets were contributed from AWI's Frontiers in Arctic Marine Monitoring (FRAM) and the multiple European Union agency-led Joint Monitoring of Ambient Noise in the North Sea (JOMOPANS) projects (ICES, 2024; Knust, 2017; Soltwedel et al., 2013; Thomisch et al., 2023a).

The location of all *SoundCoop* sites is illustrated in Figure 2.

An in-person workshop held a year into the project served as a key opportunity to check in with *SoundCoop* participants. Since many who shared data were not directly involved in the mechanics of the project, it was important to be transparent about progress towards the project goals while also allowing participants to access the tools being developed and see data product visualizations. Through guided discussion and hands-on activities, participants were able to share feedback to the steering committee.

Figure 2. Location of all *SoundCoop* recording sites. Sites that are labeled are included in the Case study analyses presented in this paper. The colored dots for the sites in the Gulf of Maine inset correspond to the colored lines in the power spectrum plots in Figures 3 and 4. https://www.ncei.noaa.gov/products/passivepacker.

During the third phase of community input, a hybrid-format workshop was held to allow broader participation especially from those working more broadly on big data challenges (e.g., *NCEI World Ocean Database, NOAA Center for Artificial Intelligence, IOOS-funded Reaching for the Cloud project*). The first session summarized the successes of this large undertaking, highlighting the enormous community efforts as well as key advances in metadata and data quality, and robust and repeatable methods. Sharing with a broader big data community raised awareness of complementary efforts and helped ensure longevity and scalability of the progress made in *SoundCoop*. The final session focused on future pathways and included use and application demonstrations from different user communities (Table 1). These groups are currently building upon what was established in *SoundCoop* to operationalize the creation of standardized sound level metrics in an interoperable self-describing format for their programs. Lastly, we see this collaborative publication as the final phase to broad community input by creating a collective vision with immediately available tools and guidance we can all build from.

2.2. Standard outputs from open-source tools

Standard outputs from open-source tools benefit big data methodologies by ensuring consistency, accessibility, and efficiency in data analysis, and promote collaboration and sharing (Vance et al., 2024). Leveraging existing documentation and associated software for a broader community of PAM practitioners, all *SoundCoop* data (except for the NYSDEC-Cornell and JOMOPANS datasets) were processed into daily files of one-minute resolution hybrid millidecade (HMD) spectra (Martin et al., 2021a, 2021b). This method balances reduced data volume with suitable resolution of acoustic spectra for many soundscape analyses, focusing on finer resolution at lower frequencies with 1 Hz bins from 0 to 435 Hz and millidecade bins above 435 Hz (Martin et al., 2021a, 2021b). This soundscape metric was used here as a way to provide a high level, comparison-ready overview of data that are too large to explore directly. Further, open-source and freeware processing tools were chosen to calculate the HMD spectra to ensure the broader community can replicate the work without the barrier of paid licensing. The *Making Ambient Noise Trends Accessible* (MANTA) Matlab tool and the *Python passive acoustic analysis tool for Passive Acoustic Monitoring* (PyPAM) with PyPAM-Based Processing (PBP) wrapper were both used for processing (Miksis-Olds et al., 2021; Parcerisas, 2023; Rueda et al., n.d.). Collectively, over 14 million minutes of recordings were processed into HMD spectra, equating to nearly 30 years of sound levels. Participants preferred multiple software options to provide flexibility for the community with diverse processing capabilities and technical expertise. As a result, careful evaluation throughout multiple phases of the processing, from calibration to content of the output were required. Summaries of the processing methods and decisions are found in Appendix I (see 1. Software and 2. Evaluation).

To ensure interoperability in the software output and facilitate data sharing, a standardized netCDF file with complete, self-describing metadata was developed. Content on the netCDF structure is found in Appendix I (see 3. Standard Outputs) and metadata details are described in Appendix II. This new standardized output for PAM sound level metrics provides a common, standards-driven format that simplifies integration across diverse datasets, reducing the time and effort required to process and analyze data from multiple sources. Included within is an effort variable that documents the number of seconds of acoustic recording that went into each HMD minute and a data quality matrix that provides a machine-readable way to mask frequency and/or time bins when HMD results are present but not calibrated or otherwise of degraded quality (see 4. Data Quality in Appendix I). The netCDF also includes a digital object identifier (DOI) for the respective project provided by the NOAA NCEI Passive Acoustic Data Archive, where the datasets are stewarded. DOIs are persistent identifiers that facilitate transparency, long-term stewardship, and citability of the datasets. Workflows built using open-source tools, which are generally widely accessible, that input these standard, machine-readable netCDFs further lower barriers for different user groups (de Visser et al., 2023; Tjernström et al., 2024; Wilkinson et al., 2024). Moreover, standardized outputs facilitate transparency and reproducibility, critical for fostering collaboration in big data projects (Kanterakis et al., 2018; Stoudt et al., 2024).

2.3. Public repositories and data portals

Public repositories and data portals provide critical infrastructure for maximizing the utility of big data by ensuring accessibility and facilitating open science (Lnenicka & Nikiforova, 2021; Ramachandran et al., 2021). By hosting globally accessible data in standardized formats with complete, standards-driven metadata, data repositories enhance data discoverability and interoperability. This enables integration with other publicly accessible datasets and open-source tools, and ultimately meets FAIR (Findable-Accessible-Interoperable-Reproducible) principles (Jacobsen et al., 2020; Wilkinson et al., 2016). The comparable HMD spectra produced through *SoundCoop* were distributed across three freely accessible platforms: NCEI Passive Acoustic Data Archive Google Cloud Platform (GCP) bucket (https://console.cloud.google.com/storage/browser/noaa-passive-bioacoustic), MBARI Amazon Web Services (AWS) S3 bucket (https://registry.opendata.aws/pacific-sound/), and Axiom Data Science Research Workspace (https://researchworkspace.com/intro/). Automated workflows leveraging Python libraries were developed to programmatically access files from the GCP, AWS, and the Axiom's Workspace, and published as a community resource (https://github.com/ioos/soundcoop). Ultimately, these repositories democratize data access, empowering a broader range of stakeholders —including policymakers, educators, and community organizations—to leverage big data for informed decision-making and innovative solutions.

To promote discoverability and demonstrate the value of standardized products for PAM housed in public repositories, an interactive data portal was developed that visualizes the *SoundCoop* datasets and integrates non-acoustic data from marine environmental sensors (https://soundcoop.portal.axds.co/). Specifically, wind speed and wave height were accessed from the IOOS environmental sensor map (https://sensors.ioos.us/) and integrated into the *SoundCoop* data portal. The one-minute HMD spectra were divided based on wind speed and wave height bins that align with sea state based on the Beaufort scale: wind speed bins were <10 kts (Beaufort scale 0–3), 10–20 kts (Beaufort scale 4–5), and greater than 20 kts (Beaufort scale 6+) and significant wave height bins were <1 m (Beaufort scale 0–3), 1–3 m (Beaufort scale 4–5), and greater than 3 m (Beaufort scale 6+). One-degree resolution sea surface temperature values for a given site and month were extracted from the NOAA World Ocean Atlas 2023, WOA23 (Reagan & NOAA National Centers for Environmental Information, 2023; Reagan et al., 2024) to compute sea surface temperature climatologies for each site. Sea surface temperature anomaly was then calculated by subtracting the WOA 2023 derived temperature climatology from each measured sea water temperature value at that site for the recorded month.

A schematic of the processing workflow developed by *SoundCoop*, showing the steps from raw audio data to comparable and standardized sound level metrics that are cloud-hosted and accessible through the interactive *SoundCoop* portal, is provided in Appendix III. This appendix also includes links to specific code bases, access points and web pages with guiding principles. Appendix IV presents a schematic of SoundCoop's workflow for accessing HMD spectra from different cloud and open-access repositories, leveraging the portal's exploratory capabilities, and facilitating interpretation. It also references the project's Github site, which hosts Python-based Jupyter Notebooks with tutorials and README files that support the workflow in Appendix III and enable additional processing of the calculated HMD spectra.

Table 2. Datasets used to demonstrate value from comparative analyses. Details include project name, region where data were recorded, recording site name, latitude (Lat), and longitude (lon), depth of the instrument (Instr.) and seafloor, start and end of the analysis period, effective frequency bandwidth of the calibrated audio data, how long sound was recorded (duty cycle), type of recording instrument, software used to process the audio data into HMD spectra, and the repository (repo) from where the HMD spectra were accessed. The asterisks by instrument and bottom depth indicate values may vary slightly due to multiple deployments represented within the analysis period. When sites are not recording continuously (Cont.), the recording duration is listed first, followed by the interval between the start of two consecutive recordings. For example, the DSG recorder used in the ESONS project recorded for 2 min every hour. MANTA refers to the making ambient noise trends accessible freeware processing software; PyPAM refers to the Python passive acoustic analysis tool for passive acoustic monitoring open source processing software; NMS is National Marine Sanctuary; GCP refers to the NCEI passive acoustic data archive GCP bucket; AWS refers to the MBARI MARS AWS bucket; and Axiom is Axiom data Science's Research Workspace.

Case Study	Project Name	Region	Site Name	Lat	Lon	Instr. Depth (m)*	Bottom Depth (m)*	Analysis Start	Analysis End	Effective Bandwidth (Hz)	Duty Cycle	Instr. Type	Software	Repo
1	AFSC-ALTIMA	Arctic Ocean Chukchi Plateau	AU_CH01	75.10	−168.00	118.6	166	2008-10-13	2009-10-05	20–4,096/ 20–8,192	8 min/ 20 min	AURAL	MANTA	GCP
2 & 3	AEON	Gulf of Maine Wilkinson Basin	AEON5	42.87	−70.02	133.5	135	2021-02-14	2021-12-31	5–7,072	45 min/ 1 hr	AMAR	MANTA	GCP
2	ESONS	Southeast U.S. May River	37M	32.20	−80.79	5.1	5.3	2021-01-01	2021-12-31	50– 40,000	2 min/ 1 hr	DSG	MANTA	Axiom
2 & 3	MBARI-MARS	Central California Coast Monterey Bay	MARS	36.71	−122.19	890	891	2021-01-01	2021-12-31	10–100,000	Cont.	icListen	PyPAM	AWS
2	NEFSC	Gulf of Maine Coastal	Monhegan Island	43.72	−69.30	58	61	2021-02-05	2021-12-18	20– 24,000	Cont.	Sound Trap	MANTA	GCP
			Petit Manan	44.32	−67.86	58	61	2021-01-01	2021-12-10		Cont.	Sound Trap		
2 & 3	NRS	Central California Coast Cordell Bank NMS	NRS11	37.88	−123.44	500	550	2021-01-01	2021-10-04	10–2,000	Cont.	AUH	MANTA	GCP
2	SanctSound	Gulf of Maine Stellwagen Bank NMS	SB01	42.44	−70.55	47	50	2021-01-26	2021-12-31	20– 24,000	Cont.	Sound Trap	MANTA	GCP
			SB03	42.26	−70.18	42	45	2021-01-26	2021-12-31		Cont.	Sound Trap		GCP
4	BOEM-Cornell	Coastal Mid Atlantic U.S.	BOEM-VA	36.90	−75.69	20	22	2016-08-28	2016-09-10	10–4,000	Cont.	MARU	MANTA	GCP

2.4. Comparative analyses

A subset of the overall *SoundCoop* datasets are used here to provide concise examples of the versatility of HMD spectra and the benefit of comparable sound level metrics (Table 2).

Case Study 1 examines changing sound levels relative to a reference period. When a PAM program has been maintained over a long period of time, the resulting time series can be used to show how sound levels may be changing in a specific location on seasonal, annual, and interannual scales (Jalkanen et al., 2022; McDonald et al., 2006; McKenna et al., 2012; Merchant et al., 2016; Stafford, 2021). The AFSC-ALTIMA project has recorded sound in the Arctic since 2007 to monitor marine mammal distribution and migration timing (Danielson et al., 2022). To control for the influence of seasonal variation, including the influence of sea ice coverage on the soundscape, only sound levels from September months, a constant ice-free period at this location, were included. The project's AU_CH01 site is used here to compare how sound levels between 100 Hz and 1,000 Hz from HMD spectra vary over time. The reference year was defined as the earliest year in which there were over 80% "Good" quality minutes available, specifically 2009. The one-minute HMD spectra were median aggregated to one-day resolution and processed into a median power spectral density (PSD) for each available year's September. A 5-kernel window size, one dimensional median filter was applied to the median PSD to remove impulsive artifacts. Years with "Good" quality minutes less than 80% of the expected total were omitted from the analysis. This threshold was applied to 2008, 2010–2015, and 2021. The September 2009 median PSD was subtracted from subsequent year's September median PSD and plotted to illustrate relative change across years.

Case Study 2 compares sound levels across monitoring projects. A major benefit of producing a robust workflow to create standardized sound level metrics in an interoperable format is that datasets from different monitoring programs across federated repositories are now easily comparable. To illustrate this concept, we co-visualized HMD spectra from eight sites (AEON5, 37 M, MARS, Petit Manan, Monhegan Island, NRS11, SB01 and SB03) recorded across three geographic regions (Gulf of Maine, Southeast U.S., and Central California Coast) and six monitoring projects (AEON, ESONS, MARS, NEFSC, NRS, and SanctSound). The HMD netCDF files were pulled from three data repositories (GCP, AWS, and Axiom Workspace), processed into weekly aggregated median PSD that represent sound levels recorded throughout 2021, and visualized.

Case Study 3 discovers dominant sound sources through indices. Producing standardized data products provides opportunities to target additional analyses that leverage the original quality-controlled product. The result is further insights that can be compared across space and time, although considering context for the analysis becomes even more relevant as further indices are derived (McKenna et al., 2021). To highlight what sources could be contributing to the soundscape and how their presence varies throughout the year at different locations, we leveraged established methods to identify blue whale (*Balaenoptera musculus*) B calls and fin whale (*B. physalus*) 20 Hz calls for deepwater recordings (Haver et al., 2020; Mellinger et al., 2009; Oestreich et al., 2020; Pearson et al., 2023; Širović et al., 2009, 2015). These methods use a call index (CI) to quantify the intensity of the calls relative to background noise applied here using the minute-resolution HMD netCDFs from the *SoundCoop*'s deepwater sites (>100 m): AEON5, MARS, and NRS11. For blue whales, the CI ratio was calculated as [42, 43 Hz]/[37, 50 Hz] where the third and strongest harmonic of B calls occurs between 42 and 44 Hz, and 37 and 50 Hz represent frequency bands that would not include energy from a B call. The results are then binned to one-day resolution using the median. CI was calculated for fin whale calls using [20, 21 Hz]/[12, 34 Hz]. The CI results were then aggregated further to one-week resolution and plotted over time.

Finally, **Case Study 4** advances interpretation of the sound levels through integration of environmental data. While PAM is still emerging in the big data landscape, monitoring of the physical environment at global scales is well established (e.g., wind speed, wave height) (Johnson et al., 2022; Jung & Schindler, 2019; Timmermans et al., 2020; Young et al., 2011; Zheng et al., 2016). These physical conditions provide insight on drivers of changes in soundscape. For example, surface agitation from wind and waves add acoustic energy to the ocean with known effects on the sound levels (Wenz, 1962). To demonstrate the value of data integration, HMD data from BOEM-VA, a PAM site in the North Atlantic near the mouth of Chesapeake Bay, was combined with wind speed, wave height, and sea surface temperature pulled from an open-source data repository ("gov-ndbc-44014" ERDDAP dataset). This buoy is located 82.9 km from BOEM-VA. The HMD spectra were median aggregated to one-hour resolution to match the hourly resolution of the wind speed, wave height and water temperature data for data recorded from August 28 to September 10, 2016. Hourly median PSD was calculated for time periods associated with Beaufort scale 0–3, Beaufort scale 4–5, and Beaufort scale 6+, and temperature anomaly bins of <2 °C; -2 °C < 0° C; 0 °C < 2 °C; >2 °C.

Collectively, these comparisons elucidate important components of our ocean soundscapes, how they vary by region and over time, and the utility of the HMD data product for insights from PAM data. All analyses were performed in Python using Jupyter Notebooks run on a local JupyterLab. The Jupyter Notebooks can be found in the *SoundCoop* project's Github repository (https://github.com/ioos/soundcoop) and visualizations of the datasets can be found in *SoundCoop* data portal.

3. Results

3.1. Case study 1: Change relative to reference period

The comparison presented here for sound levels recorded in AU_CH01 during September for 2009, 2016–2020 and 2022 provides initial insight on the dominant patterns of change in ocean sound (Figure 3). Generally, sound levels recorded above 300 Hz were higher in amplitude for years following 2009 with 2016 and 2018 reaching over 4 dB higher. Lower frequency (100–300 Hz) sound levels were up to 2 dB lower in 2019 and 2022 compared to 2009 and up to 7 dB lower than 2018. Sounds from biological (e.g., migrating whales), physical (e.g., higher sea states due to storms), and anthropogenic (e.g., noise from vessels) sources all contribute to the overall soundscape, therefore, these broad patterns in sound levels over 7 years inform additional analysis, including what frequency bands to focus on.

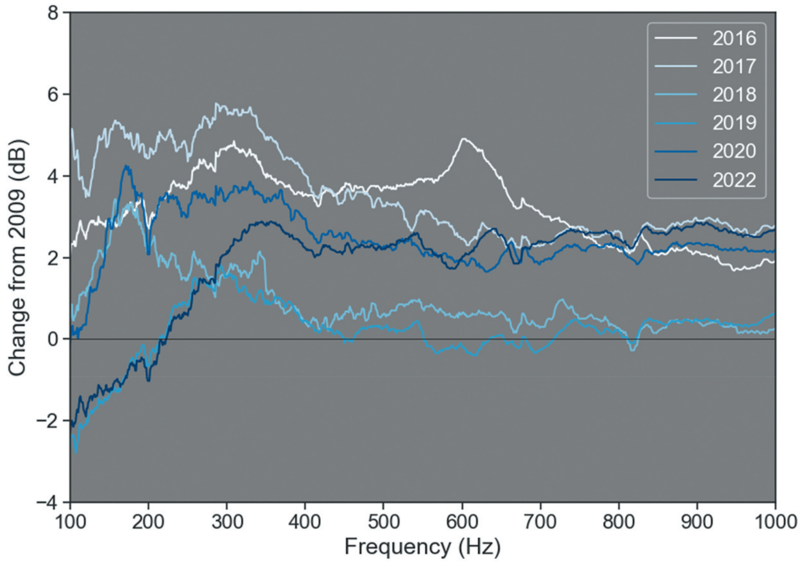

Figure 3. Change relative to reference period at a long-term PAM monitoring location in Arctic Ocean at site AU_CH01. Variation PSD sound levels recorded in September, an ice-free period, of available years from the reference year, 2009, is shown.

3.2. *Case study 2: Comparisons across monitoring projects*

While the value demonstrated here mainly focuses on the interoperability of standard data products, broad differences emerge from this comparison, relating to the oceanographic settings (Figure 4). In the Gulf of Maine, sound levels at the deepwater AEON5 site were consistently higher than the shallower SB01 and SB03 sites, likely due to the more

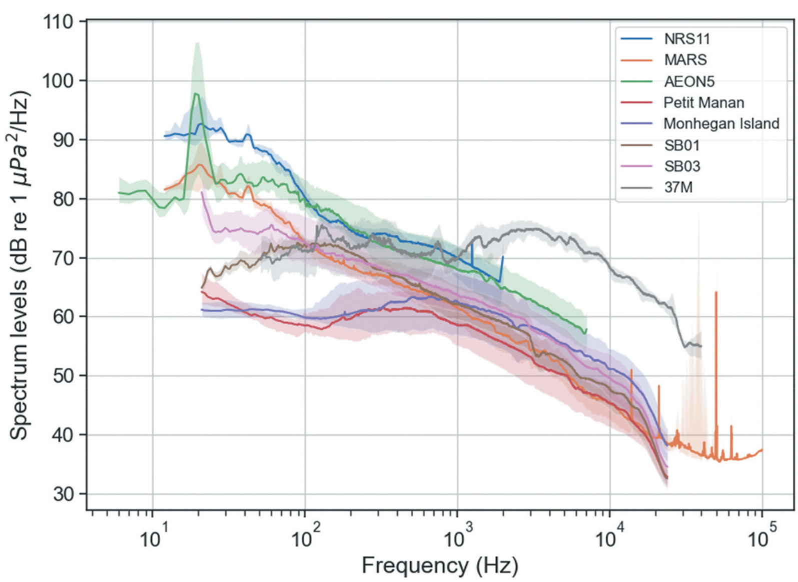

Figure 4. Large-scale comparisons of weekly-aggregated median power spectral density for eight sites recording throughout 2021. Sound levels are shown as the median (solid line) with 25th and 75th percentiles (shaded area).

efficient sound propagation in deeper water and the sites' close proximity to a heavily used shipping channel with traffic from both large and small vessels (Hatch et al., 2008; McKenna et al., 2024). Petit Manan and Monhegan Island sites had the lowest overall sound levels likely attributed to their coastal location away from the shipping channel and large vessel traffic. Located at the mouth of May River where there is heavy small vessel traffic and significant fish sound production (Ji et al., 2024; Marian et al., 2021; Monczak et al., 2017, 2020, 2022; Monczak, Ji, et al., 2019; Monczak, Mueller, et al., 2019; Smott et al., 2018), ESONS' 37 M sound levels were high between 100 and1000 Hz mostly due to snapping shrimp snaps, fish sound production, and vessel noise, and highest across all sites above 1000 Hz due to snapping shrimp snaps as well as vessel noise and silver perch chorusing (Monczak et al., 2020, 2022; Smott et al., 2018). The two Central California Coast sites, NRS11 in the Cordell Bank National Marine Sanctuary and MARS in the Monterey Bay National Marine Sanctuary, show elevated sound levels at lower frequencies, which are mainly attributed to the presence of larger vessels and seasonal migration of whales (Haver et al., 2020; Ryan et al., 2016).

3.3. Case study 3: Dominant features revealed

The blue and fin whale call indexes show seasonal patterns across the sites (Figure 5). For example, blue whale B calls, as determined from the call index, are present in the NRS11 and MARS recordings in winter (January to March) and late summer to winter (August through December). MARS has two peaks later in the year in August and again in November while NRS11 only has a single peak around October. Fin whale calls as identified in the call index were detected across all three sites with the strongest peaks occurring in the fall. These patterns align with previous findings for blue and fin whale sound production (Davis et al., 2020; Haver et al., 2020; Oestreich et al., 2020; Pearson et al., 2023) and highlight the versatility of the HMD spectra to be adopted for existing algorithms due to its high temporal resolution and resolution at low frequencies.

3.4. Case study 4: Integration with environmental data sources

During the two-week period (August 28–September 20, 2016), the integration of environmental data revealed the influence of Hurricane Hermine (NWS, Bragg, 2017) on the soundscape (Figure 6). Specifically, sound levels at this site were highest when Beaufort Sea State was 6+, which is associated with wind speeds greater than 20 knots and wave height greater than 3 m, that occurred when the storm was present in the area from September 3–4, 2016, based on the NOAA National Hurricane Center hurricane/tropical storm track. Sound levels were most strongly influenced by increased wind speed and wave height at lower frequencies (less than 50 Hz) and frequencies above approximately 800 Hz. A decrease in noise from vessels occurs prior to the storm as a result of Tropical Storm Warnings (NWS, 2016), and changes in biological sounds also emerge. During the storm, diurnal fish chorusing events between 100 and 1,000 Hz dramatically increase in amplitude and an additional chorusing event (likely a different species) between 1,500 and 2,700 Hz begins. Beyond this specific event, continuous data integration of the physical

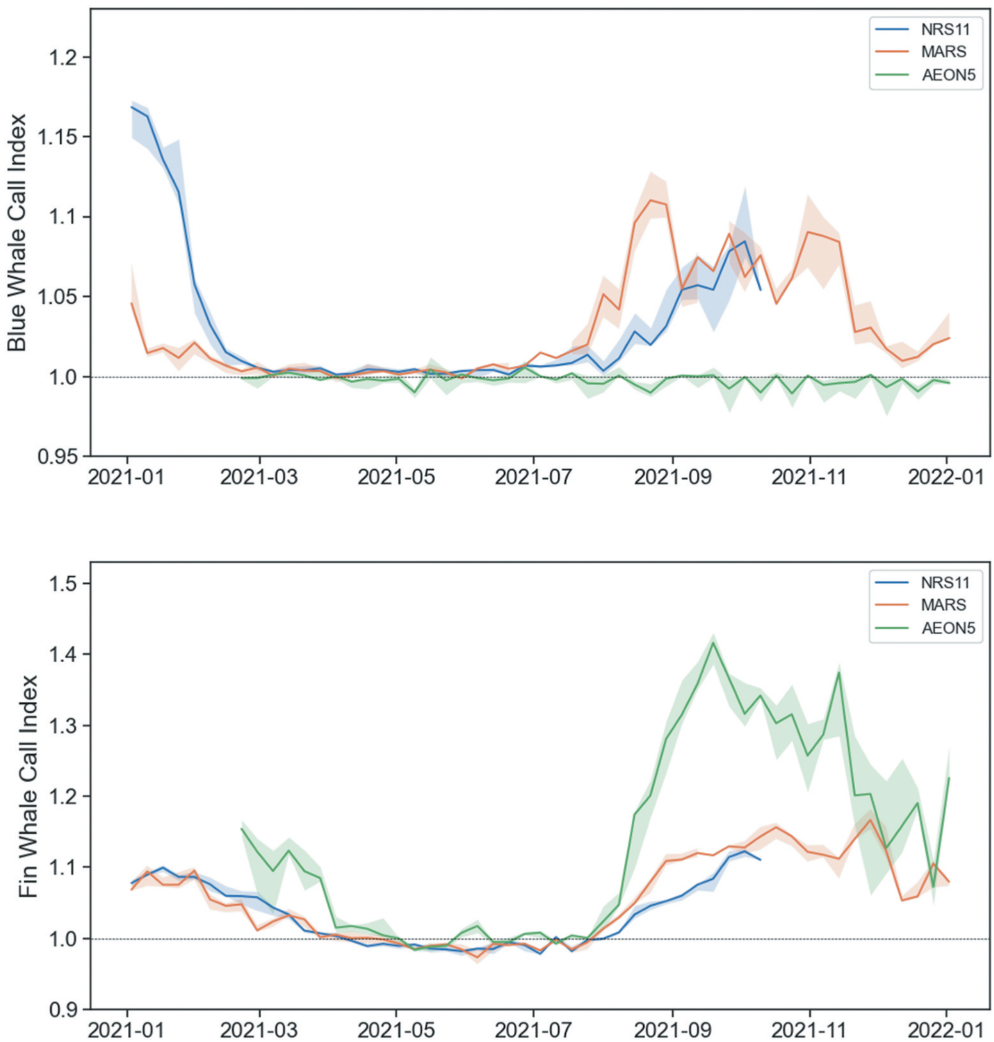

Figure 5. Comparison of blue and fin whale call indices derived from standard products. Weekly whale call index for (top) whale B calls and (bottom) fin whale 20 Hz calls for three deeper water sites. The solid line shows the median (50th percentile) CI value for that week with the shaded envelope delineating the 25th and 75th percentiles. The dotted line at 1.0 denotes where the signal would be above background noise (above the line) and where the signal would be less than the background noise (below the line).

environment with sound levels provides key insights on the seasonality of soundscapes dominated by physical conditions (McKenna et al., 2021) and could help monitor the interplay of ocean sounds and wind-wave dynamics in a changing climate (Casas-Prat et al., 2024; Cheng et al., 2022). As other sources of big data become available, scalable integration with PAM data becomes possible furthering the ability to understand drivers of change in ocean soundscapes.

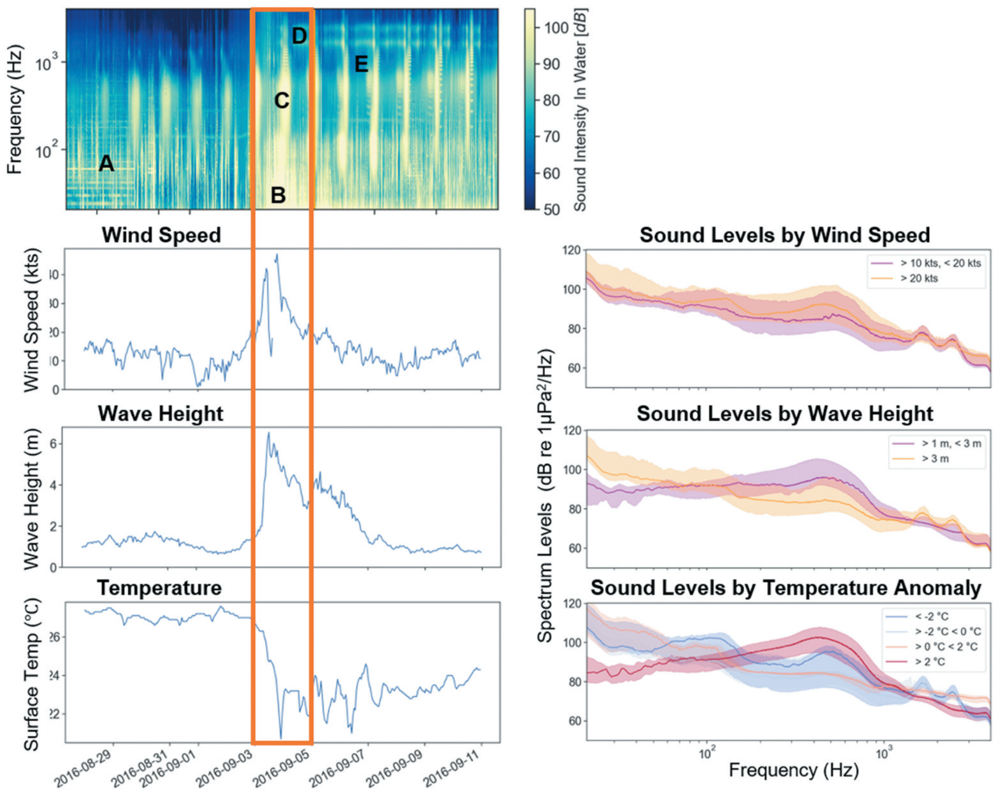

Figure 6. Integrated data sources reveal weather related changes in underwater soundscapes. (left panel) the hourly resolution HMD spectra, wind speed, wave height, and water temperature data from August 28 to September 10, 2016. The orange box indicates the time period when Hermine was present near the recording instrument, September 3–4, 2016. (right panel) hourly median PSD by Beaufort scale 0–3, Beaufort scale 4–5, and Beaufort scale 6+, and temperature anomaly bins for sound levels recorded on September 3–4, 2016. Annotations in the spectrogram denote A) noise from vessel activity; B) low frequency noise from wind associated with the storm; C) nocturnal fish chorus of increasing intensity as storm nears and passes; D) new fish chorusing that begins after the storm passes, and E) third fish chorusing event that arose after the storm.

4. Discussion

The *SoundCoop* project's collaborative approach provided a unique opportunity to address challenges associated with big data in marine PAM by focusing on existing community standards, knowledge, and datasets and thus established a community-driven process. By engaging a broad network of stakeholders, *SoundCoop* fostered collective input to identify key obstacles, and develop practical and scalable solutions. The project emphasized the creation of standard outputs from freeware tools (Miksis-Olds et al., 2021; Parcerisas, 2023), ensuring consistency and accessibility for numerous PAM user groups. Additionally, it championed the use of public repositories and data portals to promote transparency, data sharing, and reproducibility. This framework serves as an example for harnessing collective expertise to identify a meaningful data product to address the heterogeneity in raw data and enable comparison across monitoring projects;

efficiently share and connect data to address the challenges of dissemination associated with large data volumes; integrate data quality information to address noise accumulation; and streamline access to federated datasets, especially cloud-based to enable further processing capabilities in scalable, high compute environments. The case studies presented here demonstrate the versatility of the HMD spectra in exploring scientific questions relevant for many of the PAM user communities outlined in Table 1. There are numerous other applications to which this information-rich and storage-efficient sound level metric could be applied to accelerate learning and drive more complex analyses.

4.1. Challenges overcome

The sheer size of data produced by modern monitoring systems, like PAM, is vast, especially when monitoring over extended periods and across multiple locations (Van Parijs et al., 2009; Wall et al., 2021). *SoundCoop*'s new infrastructure enables transformative, translational research by improving the accessibility of data, information, and methods for standardized metrics and formatting. Free software toolkits facilitated the standardized processing of sound levels. However, extensive testing and revisions were needed to produce the necessary output of predictable and therefore comparable 60-s time bins, ensure accurate second counts within each bin, and handle the variety of recorder types associated with the *SoundCoop*. This effort resulted in a new suite of datasets to test against including different instruments, calibration considerations, "wiggle" in the internal timekeeping, and unexpected recording gaps. Tracking each second of every recording is crucial to produce accurate and reproducible one-minute resolution HMD files. This provides fine-scale accuracy while correct calibration ensures statistical importance to the analytical results. Comparing multiple variables, namely HMD spectra, time, and effort, from the software outputs to a "ground truth" dataset—as was done in this project—is necessary to examine the accuracy of each time bin (see 4. Evaluation in Appendix I).

Over 14,000 netCDF files were produced from the HMD-processed datasets in this project. Managing these many files from different sources in a CSV format would have been brittle and required additional spreadsheets or databases containing each dataset's metadata. The self-describing netCDF embeds all file, deployment-, and project-level metadata within each file, facilitating efficient management, sharing, and interpretation (see 3. Standard Outputs in Appendix I and Appendix II). Aligning the outputs of both software to the same netCDF structure ensured interoperability, comparability of results, and seamless integration into higher-level processing workflows. The integration of publicly available sensor data provided a first-tier interpretation of the sound level data, leading to some initial insights into patterns within the acoustic data. These data portals provide tangible examples of the benefit of standardized data products for big data projects, like PAM, by enabling novice users to explore large volumes of data with ease, fostering curiosity, and supporting learning while also empowering experts to identify areas for deeper analysis.

Tutorials empower practitioners to leverage PAM data more effectively, while visualization tools and standardized workflows simplify large-scale data analysis, reducing the need for individuals to "reinvent the wheel." To ensure transparency and reproducibility of the methods and results, and in turn provide tools that could be easily adapted to existing and future workflows, Python-based Jupyter Notebooks were developed (see Appendices III and IV). These notebooks document how to access, read in, aggregate, and visualize the project's thousands of files using open-source processing in reproducible environments. The notebooks are shared publicly on GitHub along with documentation such as READMES and guidance for different user groups (from novice to expert, science-driven to management-driven). Environment setup files that capture system and library requirements were produced so that the notebooks could be run seamlessly by any user. Depending on the volume of data being processed and the processing needs, Google Colab, JupyterLab, Github Code Space, and Binder are potential options to run the notebooks for the community to explore—though there are assuredly many other environments that are best suited for individual needs including higher performing, cloud-based, and fee-based platforms.

4.2. Where *SoundCoop* leaves off

While significant progress was made and several barriers were overcome through *SoundCoop*, some areas of additional development remain. The *SoundCoop* resulted in community production of a standardized underwater sound level metric using open source/freeware processing software supporting regional, national and international comparisons among PAM recordings. However, throughout the project, the calculation of the HMD spectra for individual datasets was completed in serial on local machines limiting processing power to the capacity of each computer. Processing speed was further constrained when data had to be accessed from USB-connected external drives or through VPN-linked servers. A cloud-based workflow, where processing can be scaled on demand and computation is brought to cloud-hosted data, would substantially improve efficiency (Jannapureddy et al., 2019). Therefore, the next steps for development should focus on developing cloud-based processing pipelines to go from the raw audio data (L0) to HMD spectra netCDFs (L1) and further into cloud-ready advanced products (L2). This includes the development of tools that ingest the L1 as a base variable or further interpret L1 time series with environmental/non-acoustic information and feed into clustering or other machine learning models. Workflows to run PyPAM in a cloud environment and apply clustering to HMD spectra to explore soundscape composition across ecosystems are under active development and show great potential, but the completion of those efforts fell outside of the scope and timeline of the *SoundCoop* project.

The project provided community documentation of standards-driven metadata-rich file format for sharing and interoperability. The *SoundCoop* netCDF was built using Python code run by NCEI and externally provided metadata on data quality to inform applications. Next steps for development should include implementing a data quality matrix in the PyPAM/PBP workflow. Further, the framework developed—including the netCDF structure—was designed exclusively for stationary acoustic recorders. To extend its applicability, additional considerations will be required to support moving instrumentation, such as towed arrays, autonomous surface/underwater vehicles, and drifting systems. This includes ensuring the netCDF structure properly accounts for variations in the hydrophone's spatial location and depth, which has downstream implications for accurate interpretation of the HMD spectra. Feedback and adoption of the file format guidance by international partners would accelerate broader implementation in ongoing monitoring efforts.

The *SoundCoop* developed a centralized visualization platform to compare standardized PAM products. This interactive portal provides internationally distributed, high-volume, and comparable sound level attributes, showcasing data-rich HMD spectra and demonstrating the value of co-plotting with buoy-based wind and wave observations as well as oceanographic model outputs. Further enhancements could include regional portals, tools tailored to offshore wind energy visualization, and NOAA-specific applications at regional and national scales.

In parallel, the *SoundCoop* curated a suite of community tools, including publicly available Python-based resources on GitHub that allow users to generate *SoundCoop* L1 products, and visualize data in ways consistent with the portal (see Appendices III and IV). Ensuring the long-term sustainability of these community notebooks—including where and how future users will run them—will be essential to maintain usability. Expanding the library of notebooks to address broader use cases, expanding test datasets that represent additional scenarios, and developing alternative entry points for non-Python users would further extend their value to a wider community base.

This project was framed around the curation of two software toolkits for community use, chosen for their freeware and open-source status. Sustaining such software remains a persistent challenge that affects both developers and users (Bonaccorsi & Rossi, 2003). Routine maintenance, bug fixes, security updates, and community-requested feature enhancements are expected, yet questions of implementation—who will make changes, on what timeline, and under what governance—often create obstacles (Geiger et al., 2021; Sethanandha et al., 2010; Sun, 2024). The long-term sustainability and advancement of open-source tools such as PyPAM, and more broadly the entire suite of software supporting passive acoustic monitoring, will ultimately depend on the community itself and is best approached at an international scale. Because the passive acoustic community is inherently global, a broad contributor base increases the likelihood of long-term viability. Establishing an international working group dedicated to sustaining open-source software for passive acoustic data analysis could consolidate ongoing efforts, strengthen existing tools, reduce the learning curve for new practitioners by providing even more shared code resources, and increase the overall resilience of community codebases.

4.3. Best practices for data quality

Big data often includes unstructured, noisy, or incomplete data, making analysis complex (Fadlallah et al., 2023). To realize the value of PAM data, we highlight considerations when building big data frameworks to ensure data quality and interoperability between datasets. HMD offers a robust yet versatile foundational sound level metric product and these key requirements are necessary before applying this approach to your dataset:

4.3.1. Calibration

Proper characterization of the calibration is critical for accurate, absolute sound level measurements. Users must understand the calibration process and how the HMD software applies it. Clear documentation, intuitive workflows, and graphical outputs from the software are essential to vet this information (see 1. Software in Appendix I). If calibration is unknown, not possible, or the hydrophone sensitivity degrades over the duration of the recording, processing the data into sound level metrics like HMD for quantitative analyses will not produce accurate results.

4.3.2. Timekeeping

Clock drift occurs in PAM recording units and can accumulate over long deployments depending on internal clock quality and in response to temperature changes (Marchetto et al., 2012). If drift is high, it can noticeably impact the accuracy of one-minute resolution HMD products if not accounted for. Autonomous recording systems are built to be fault-tolerant but errors can occur, leading to timekeeping problems such as lost or extra samples, or variability in true sampling rates. Each scenario needs to be properly handled by the software calculating the HMD, and a thorough examination of where each second is being accounted for in the calculation will ensure the 1-minute time bins are accurate (see 2. Evaluation in Appendix I). In severe cases where timekeeping is highly variable, consider aggregating to a coarser time resolution (e.g., hour) or another metric that is robust to some temporal inaccuracy.

4.3.3. Data integrity

Data quality must be thoroughly documented for proper use or reuse of sound level metrics. Within the frequency dimension, documentation should note where calibration is not available or where recording instrument accuracy limits the quantitative analysis of specific frequencies. This is particularly necessary for data processed with MANTA, which outputs 0 Hz to Nyquist regardless of the calibration range or presence of electrical noise. Within the time dimension, document periods of known issues and the timing of deployment and recovery. As demonstrated in *SoundCoop*, it is recommended to implement a machine-readable method for masking unverified, compromised, or poor-quality times and frequencies. This can be achieved by creating a numeric matrix of data quality tags that matches the dimensions of the power spectral density matrix. The data quality tags could derive from a standardized JSON metadata file, such as the one outputted by the NCEI PassivePacker PAM data documentation tool (https://www.ncei.noaa.gov/products/passivepacker), or built with automated scripts following a quantitative QA/QC review of the processed data (see 3. Standard Outputs and 4. Data Quality in Appendix I). The netCDF file with the HMD spectra results and data quality mask can then be used to programmatically ensure only "good" time periods and frequency bands are used in quantitative analyses. Further, to inform the data quality assessment, the daily graphic of the long-term spectral average based on the HMD spectra outputted by MANTA and PyPAM provides an efficient evaluation method (albeit largely qualitative) for large volumes of data.

These collective advancements foster collaboration and momentum within the PAM community, building partnerships and aligning efforts. By coordinating and standardizing data, the community can better address critical questions about the impacts of industrialization on marine habitats and life, and enable scientists to explore climate effects on marine ecosystems in the most sustainable ways possible. This distributed model, where standardized tools are widely adopted, allows more researchers to contribute to collective solutions and deliver valuable insights to science and management applications.

4.4. Emerging big data considerations

The demand for real-time data analysis is increasing, especially in scenarios requiring immediate decision-making (Hossain et al., 2023). Real-time processing of acoustic data is particularly important in urgent scenarios such as detecting the changing presence of endangered species or responding to illegal activities like poaching or illegal fishing (Baumgartner et al., 2019; Drakopulos et al., 2023; Spaulding et al., 2009; Welch et al., 2024). However, real-time analysis is computationally intensive and requires sophisticated algorithms and hardware to handle the continuous flow of data without delays. Notably, advances discussed here can directly support conservation or resource management decisions that are neither immediate nor decadal but rather responsive to increasingly dynamic inter-annual changes in animal and human use patterns in ocean and coastal waters.

Managing access, privacy, and security concerns is critical when handling sensitive or proprietary big data (O'Brien et al., 2024). Some PAM data may be sensitive, particularly in the context of national security, monitoring protected areas or species, and data sovereignty considerations arise with diverse data collection. Ensuring secure data transmission, storage, and sharing across platforms is important to protect both species and project confidentiality. Screening and filtering of some data may be necessary prior to public dissemination, but such treatments can be surgical and allow the majority of the collected data to inform a broad suite of users.

Providing public access to data, particularly data funded by government agencies, is acknowledged to increase transparency and citizen engagement, and to support economic growth through innovation, evidence-based policy making, and the ability for researchers and businesses to leverage data for new insights and solutions. However, providing public access to data will continue to raise ethical questions about data ownership, usage, and the potential consequences of insights derived from analysis (Carroll et al., 2021). In the context of PAM, although there are many conservation goals that can be advanced significantly through more efficient and interpretable data practices, there is also the potential for data misuse in ways that could harm protected species (e.g., revealing the location of fragile managed fish or mammal concentration). Considering such risks while empowering new applications is crucial for advancing PAM, ensuring it continues to grow as an effective tool for conservation, biodiversity research, understanding ocean systems, and advancing the blue economy (Spalding, 2016). The global accessibility of standards-driven, metadata-rich and analysis-ready data provides a scalable foundation from which AI/ML models can be trained. It further serves as a model for tackling big data challenges for the broader earth-system science community where AI-driven research is of increasing prevalence (Haupt et al., 2022; Verhulst et al., 2025).

5. Conclusion

The value of big data lies in its potential to transform raw values into actionable insights, identify emergent properties that can drive decision-making and innovation (Alsunaidi et al., 2021; Li et al., 2021; Ramachandran et al., 2021). In today's data-driven world, extracting value from big data enables organizations to uncover patterns, predict trends, and advance management objectives across diverse applications (Vance et al., 2024). In the context of PAM, this includes ensuring species protection (Gibb et al., 2019), reducing noise in protected places (Chou et al., 2021), and understanding the blue economy activity in coastal and offshore ecosystems (Spalding et al., 2021). However, realizing this value depends on effective data management, advanced analytical tools, and a clear understanding of how insights align with organizational goals or societal needs. By focusing on value, big data becomes more than just vast quantities of information—it becomes a powerful catalyst for informed decisions and progress.

Acknowledgements

This research was supported by NOAA IOOS, Bureau of Ocean Energy Management, U.S. Navy Living Marine Resources, and Office of Naval Research (award #s: NA21NOS0120100 and N00014-22-1-2601). Brian Stone and Trevor Golden played key roles in developing workflows to access sound levels from three different repositories, integrate acoustic and non-acoustic variables, and the entire *SoundCoop* portal. CP and PyPAM were supported by the Research Foundation Flanders (FWO) as part of the Belgian contribution to the LifeWatch ESFRI, Grant number I002021N. KT and OB were supported by the Helmholtz strategic infrastructure FRAM under grant number AWI-PS100_0. ESONS was supported as a result of work sponsored by the Southeast Coastal Ocean Observing Regional Association (SECOORA) with NOAA financial assistance award number NA21NOS0120097. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of SECOORA or NOAA. JM-O and the AEON project were funded by the Office of Naval Research under award number N00014-20-1-2312. Publication of this article was funded by the University of Colorado Boulder Libraries Open Access Fund.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by NOAA IOOS, Bureau of Ocean Energy Management, U.S. Navy Living Marine Resources, and Office of Naval Research, the Research Foundation Flanders (FWO) as part of the Belgian contribution to the LifeWatch ESFRI, the Helmholtz strategic infrastructure, the Southeast Coastal Ocean Observing Regional Association. Publication of this article was funded by the University of Colorado Boulder Libraries Open Access Fund.

Notes on contributors

Carrie C. Wall is a Senior Research Scientist at the University of Colorado Cooperative Institute for Research in Environmental Sciences where she leads the NOAA National Centers for Environmental Information ocean acoustic archives, encompassing water column sonar and passive acoustic data. Carrie and her data management and software development teams have established standards-based best practices to describe, submit, ingest, query, and access large volumes of acoustic data enabling the archives to grow rapidly to over a petabyte.

Carrie received an M.S. and Ph.D. in Marine Science from the University of South Florida where her dissertation research focused on passive acoustic monitoring of fish on the West Florida Shelf. Following graduate school, she conducted research on deep-sea fish sound production at the University of Victoria and taught undergraduate students at Vancouver Island University in British Columbia, Canada.

Megan F McKenna is an ocean scientist and educator with 15+ years of experience bridging science, policy, and education. She was a passive acoustic data scientist at the University of Colorado Boulder Cooperative Institute for Research in Environmental Sciences and NOAA National Centers for Environmental Information and is faculty and Director of Study Away Program at the Middlebury Institute of International Studies in Monterey.

Leila Hatch is a marine ecologist who worked for NOAA's Office of National Marine Sanctuaries where she coordinated acoustic research across the national marine sanctuary system and worked with her co-leads from NOAA Fisheries to lead the agency's Ocean Noise Strategy.

Sofie Van Parijs leads the Passive Acoustic Branch at the NOAA Northeast Fisheries Science Center. She focuses on management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales.

Rob Bochenek is the founder and president of Axiom Data Science with over 20 years of experience building information architecture and cyberinfrastructure for clients in the earth science and environmental domains. Axiom Data Science was acquired by Tetra Tech in 2022.

Peter Dugan has over 20 years of high-level engineering experience in commercial and defense industries focused on applied analytical software development systems science. His experience encompasses large complex computing systems, and the delivery of technologies to diverse clientele, including the U.S. Navy, Army, and Air Force.

Clea Parcerisas is a postdoctoral fellow at VLIZ Flanders Marine Institute. Her PhD research focused on mapping shallow water acoustic marine ecology using a combination of modelling and measurements.

John Ryan is a biological oceanographer studying the ecology of plankton, fish, and mammals—primarily through integration of remote and in situ sensing technologies.

Charles Anderson is a senior ocean acoustic data manager at the University of Colorado Boulder Cooperative Institute for Research in Environmental Sciences and NOAA National Centers for Environmental Information where he supports the NCEI Ocean Acoustic Archives.

Kyle Becker was a Program Officer for the Office of Naval Research where he developed and managed basic and applied science programs with a global reach. His projects aimed to understand how mammals, including bats, dolphins, and humans, perform scene analysis using active acoustic sensing - or echolocation.

Catherine Berchok has been leading the Passive Acoustics Group at Alaska Fisheries Science Center's Marine Mammal Lab since 2007. After receiving her B.S. in Biology from the University of Pittsburgh at Johnstown, she worked for many years as a researcher at the Mingan Island Cetacean Study in Quebec, Canada, before returning to school to earn her Ph.D. in Acoustics from Penn State University.

Mathew Biddle is the NOAA IOOS Data Management and Cyberinfrastructure (DMAC) community manager where he serves as lead for the marine life DMAC, serves as the IOOS representative to the NOAA NOS Data Governance Board, Chairs the ERRDAP Strategic Insight Group, manages the DMAC Community of Practice, and promotes and communicates IOOS data products and services through management of the IOOS CodeLab.

Olaf Boebel leads the Ocean Acoustics Group at the Alfred Wegener Institute. His research revolves around the development and deployment of novel observation tools, like under ice Argo floats, Pressure Inverted Echosounders for oceanographic studies, and infrared imaging for automatic marine mammal detection and passive acoustic monitoring for marine mammal research and mitigation. He supervises the maintenance of the Weddell Sea's HAFOS (Hybrid Antarctic Float Observing System) long-term marine observatory including the PALAOA listening station and acted as chief scientists for several RV Polarstern expeditions.

Adrienne Canino has over 5 years of experience as a data librarian at Tetra Tech where she oversees the data publication process in collaboration with large research projects, guided by CARE, FAIR, and TRUST principles. She designs and implements data governance strategies for humans and machines, to produce the goal data products.

Gabrielle Canonico is part of the U.S. Integrated Ocean Observing System (IOOS) Program and manager of the US Marine Biodiversity Observation Network (MBON). Her work focuses on integration of marine life observations, data and products into IOOS and the Global Ocean Observing System for use by resource managers, decision-makers and the public, and advancing efforts to ensure sustained monitoring of marine life and biodiversity.

Danelle Cline is a senior software engineer in the Information Engineering Group at the Monterey Bay Aquarium Research Institute. Her work focuses on video, image, and acoustic data analysis with a career in instrumentation development and data science. She has worked on various projects, ranging from DNA sampling instruments and camera systems to developing machine-learning algorithms and workflows for real-time and cloud-based designs. She received a M.S. in Computer Engineering from University of California Santa Cruz.

Genevieve Davis is a research acoustician at the NOAA Northeast Fisheries Science Center. She started working with the Passive Acoustics Branch as a NOAA Hollings scholar in 2010, working on North Atlantic right whale calling behavior across seasons. She then returned to work with the Passive Acoustics Branch as a research biologist focusing on baleen whale acoustics and distribution. She now leads the cetacean passive acoustic projects.

Kaitlin Frasier is a research scientist and program leader at Scripps Institution of Oceanography with 15+ years of experience applying machine learning, signal processing, and large-scale sensor data analysis to complex environmental and acoustic challenges. She received a M.S. and Ph.D. in Acoustics from UCSD Scripps Institution of Oceanography.

Jason Gedamke manages the Ocean Acoustics Program within NOAA Fisheries Office of Science and Technology. His work focuses on underwater sound, its importance for marine life, and the potential impacts of human generated noise on marine life and soundscapes.

Samara Haver is a passive acoustic scientist with the Oregon State University Cooperative Institute for Marine Resources Studies whose research focuses on marine soundscapes and ocean noise within U.S. waters. She works closely with NOAA NMFS and NOS and the National Park Service on collaborative projects to record underwater sound in diverse ecosystems.

Karina Khazmutdinova is a Portfolio manager at Tetra Tech. She previously served as a project manager at Axiom Data Science, a Tetra Tech company, where she led data analysis and data management projects to enhance environmental data reuse and impact. Karina received a PhD in Geophysical Fluid Dynamics from Florida State University.

Niels Kinneging is a senior consultant at Rijkswaterstaat currently involved in underwater noise. He served as project manager of Jomopans, a project aimed at developing a monitoring framework for ambient noise in the North Sea together with specialists from 11 institutes of countries around the North Sea.

Anurag Kumar has served as the Navy Living Marine Resources Program Director since 2014. He has applied his background in bioacoustics, impacts assessments, marine species density estimation, and marine mammal biology with the U.S. Navy since August 2001.He has an MS in Marine Science from Moss Landing Marine Laboratories and B.Sc. in Ecology from California State University Fresno.

Alyssa Marian is a laboratory manager at the University of South Carolina-Beaufort with over 8 years of experience working in a bioacoustic lab deploying PAM arrays as well as performing marine mammal photo-identification surveys. Her work includes manually analyzing acoustic data for multiple species of fish, bottlenose dolphins, manatees, boats, and other underwater sound. She is also experienced at testing and scheduling recorders prior to deployments, operating small vessels around marine mammals, photographing marine mammals for identification, organizing and maintaining the photo-ID catalog, and statistical analysis using R.

Jennifer Miksis-Olds is the Director of the Center for Acoustics Research and Education at the University of New Hampshire. She received an A.B. in biology from Harvard University, an M.S. in biology from the University of Massachusetts Dartmouth, and a Ph.D. from the University of Rhode Island where she studied the connections between sound and manatees.

Eric Montie is a Full Professor of Biology at the University of South Carolina Beaufort, where he directs the Marine Sensory and Neurobiology Lab. He leads two long-term programs: the Estuarine Soundscape Observatory Network in the Southeast (ESONS), which uses passive acoustic monitoring to track marine life and noise in South Carolina estuaries, and the Lowcountry Dolphin Conservation Program, which studies the population dynamics and ecology of bottlenose dolphins in the Port Royal Sound area.

Dimitri Ponirakis is a senior noise analyst and applications programmer for the K. Lisa Yang Center for Conservation Bioacoustics where he develops tools and methods for noise analysis. Software development includes development of an Acoustic Ecology Toolbox to model the loss of communication space and masking of animal vocalizations from anthropogenic and abiotic sound sources. He is responsible for carrying out noise analysis on numerous projects from liquid natural gas terminal construction and operational activities off the East Coast, to seismic prospecting in the Arctic, to looking at the impact of cruise vessels in Glacier Bay National Park in the Pacific Northwest.

Aaron Rice is a research scientist who studies how animals produce and perceive sound, and how we can use these animal sounds to examine changes in populations across time and space. He is also interested in the influence of human activities on animal populations and their habitats. His research focuses on fishes and whales, but also examines similar questions in frogs and birds.

Timothy Rowell is the branch chief of the Advanced Technologies Branch of the Fisheries Assessment, Technology, and Engineering Support Division at the NOAA Southeast Fisheries Science Center. His work is generally focused on advanced survey technologies, passive acoustics, active acoustics, telemetry, optics, fish spawning aggregations, soundscapes, and bioacoustics to better understand and conserve marine resources. His background includes studying fish sound production during reproductive periods and its utility for monitoring populations, spawning behaviors, and essential fish habitats through comparisons with other data sources.

Carlos Rueda is a senior software engineer at the Monterey Bay Aquarium Research Institute, where he applies his expertise in scientific data management, visualization, data integration and interoperability, programming languages, and machine learning. He received a Ph.D. degree in Computer Science from University of California, Davis, in 2007.

Emily Shumchenia is the Director of the Regional Wildlife Science Collaborative. As a marine scientist with fifteen years of experience practicing science and science communication, her work focuses on fostering and maintaining collaborative relationships with leading marine scientists and key ocean stakeholders in New England and beyond, with the goal of integrating the best available science in coastal and ocean management and decision-making. She received a BS in Marine Biology and Biological Oceanography, and a PhD in Oceanography from University of Rhode Island.

Thomas Shyka is the Director of Technology at NERACOOS. Prior to this role, he served as NERACOOS Outreach and Communications Specialist and then Product and Engagement Manager. He earned a BA in Biology and Environmental Science from Colby College and an MS in Marine Estuarine and Environmental Science from the University of Maryland.

Erica Staaterman is the Director for the Bureau of Ocean Energy Management's Center for Marine Acoustics serving as the liaison between the scientific community, lawmakers, and industry representatives, helping to understand the issues and implement policy related to the effects of anthropogenic noise on marine life. She also leads BOEM's efforts in Passive Acoustic Monitoring. She

earned a PhD at University of Miami focused on acoustic communication and navigation in fish larvae.

Karolin Thomisch is a postdoctoral scientist at the Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research (AWI). Her research addresses how the distribution of marine mammals relates to abiotic and biotic environmental factors to explore the ecological role of marine mammals in the Arctic Ocean. As data curator for the AWI's passive acoustic monitoring data, her expertise further includes scientific data management, processing and curation.

ORCID

Carrie C. Wall http://orcid.org/0000-0002-5315-8528

Data availability statement

The data products used in this project are available at the NOAA National Centers for Environmental Information (NCEI) Passive Acoustic Data Archive, https://www.ncei.noaa.gov/products/passive-acoustic-data. All archived SoundCoop datasets on the NCEI data portal can be viewed directly here: https://www.ncei.noaa.gov/maps/passive-acoustic-data/?projectName = SoundCoop. Each project that contributed to the SoundCoop and who's data are archived at NCEI received a citation and DOI:

- Cornell University and Bureau for Ocean Energy Management. 2023. Hybrid Millidecade Spectra at 1 Minute Resolution Recorded by Cornell University for BOEM (BOEM-Cornell). NOAA National Centers for Environmental Information. https://doi.org/10.25921/wzyx-g235
- Monterey Bay Aquarium Research Institute. 2023. Hybrid Millidecade Spectra at 1 Minute Resolution Recorded at the Monterey Accelerated Research System (MARS) Cabled Observatory. NOAA National Centers for Environmental Information. https://doi.org/10.25921/6gvv-dk87
- NOAA Northeast Fisheries Science Center. 2023. Hybrid Millidecade Spectra at 1 Minute Resolution Recorded by the Northeast Fisheries Science Center (NEFSC). NOAA National Centers for Environmental Information. https://doi.org/10.25921/s87x-gb35
- NOAA OAR Pacific Marine Environmental Laboratory and Cooperative Institute for Marine Ecosystem and Resources Studies, National Marine Fisheries Service, NOS Office of National Marine Sanctuaries, and DOI NPS Natural Resource Stewardship and Science Directorate. 2023. Hybrid Millidecade Spectra at 1 Minute Resolution Recorded at the Ocean Noise Reference Station Network (NRS). NOAA National Centers for Environmental Information. https://doi.org/10.25921/tr4q-xa94
- NOAA Office of National Marine Sanctuaries and U.S Navy. 2023. Hybrid Millidecade Spectra at 1 Minute Resolution Recorded for the Sanctuary Soundscape Monitoring Project (SanctSound). NOAA National Centers for Environmental Information. https://doi.org/10.25921/ymf8-5k59
- Scripps Institution of Oceanography. 2023. Hybrid Millidecade Spectra at 1 Minute Resolution Recorded by the Scripps Whale Acoustics Laboratory (SWAL). NOAA National Centers for Environmental Information. https://doi.org/10.25921/ht84-bg61
- Thomisch, Karolin; Spiesecke, Stefanie; Boebel, Olaf. 2023b. Hybrid Millidecade Spectra at 1 Minute Resolution Recorded at the Frontiers in Arctic Marine Monitoring (FRAM) mooring F5-17 of the Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research. NOAA National Centers for Environmental Information. https://doi.org/10.25921/jh84-k867
- University of New Hampshire and Office of Naval Research. 2023. Hybrid Millidecade Spectra at 1 Minute Resolution Recorded at the Acoustic and Environmental Observation Network (AEON) for the Office of Naval Research. NOAA National Centers for Environmental Information. https://doi.org/10.25921/d21t-be25

- University of South Carolina Beaufort and IOOS SECOORA. 2023. Hybrid Millidecade Spectra at 1 Minute Resolution Recorded at the Estuary Soundscape Observatory Network in the Southeast (ESONS) for IOOS SECOORA. NOAA National Centers for Environmental Information. https://doi.org/10.25921/0xbb-xx76

References

Affatati, A., Scaini, C., & Salon, S. (2022). Ocean sound propagation in a changing climate: Global sound speed changes and identification of acoustic hotspots. *Earth's Future*, *10*(3), e2021EF002099. https://doi.org/10.1029/2021EF002099

Alsunaidi, S. J., Almuhaideb, A. M., Ibrahim, N. M., Shaikh, F. S., Alqudaihi, K. S., Alhaidari, F. A., Khan, I. U., Aslam, N., & Alshahrani, M. S. (2021). Applications of big data analytics to control COVID-19 pandemic. *Sensors*, *21*(7), 2282.

Baumgartner, M. F., Bonnell, J., Van Parijs, S. M., Corkeron, P. J., Hotchkin, C., Ball, K., Pelletier, L.-P., Partan, J., Peters, D., Kemp, J., Pietro, J., Newhall, K., Stokes, A., Cole, T. V. N., Quintana, E., & Kraus, S. D. (2019). Persistent near real-time passive acoustic monitoring for baleen whales from a moored buoy: System description and evaluation. *Methods in Ecology and Evolution*, *10*(9), 1476–1489. https://doi.org/10.1111/2041-210X.13244

Bonaccorsi, A., & Rossi, C. (2003). Why open source software can succeed. *Research Policy*, *32*(7), 1243–1258.

NWS, & Bragg, R. (2017). *National Hurricane Center tropical cyclone report: Hurricane Hermine*. National Weather Service. https://www.nhc.noaa.gov/data/tcr/AL092016_Hermine.pdf

Caiger, P. E., Dean, M. J., DeAngelis, A. I., Hatch, L. T., Rice, A. N., Stanley, J. A., Tholke, C., Zemeckis, D. R., & Van Parijs, S. M. (2020). A decade of monitoring Atlantic cod Gadus morhua spawning aggregations in Massachusetts Bay using passive acoustics. *Marine Ecology Progress Series*, *635*, 89–103. https://doi.org/10.3354/meps13219

Carroll, S. R., Herczog, E., Hudson, M., Russell, K., & Stall, S. (2021). Operationalizing the CARE and FAIR principles for Indigenous data futures. *Scientific Data*, *8*(1), 108.

Casas-Prat, M., Hemer, M. A., Dodet, G., Morim, J., Wang, X. L., Mori, N., Young, I., Erikson, L., Kamranzad, B., Kumar, P., Menéndez, M., & Feng, Y. (2024). Wind-wave climate changes and their impacts. *Nature Reviews Earth and Environment*, *5*(1), 23–42. https://doi.org/10.1038/s43017-023-00502-0

Cheng, L., von Schuckmann, K., Abraham, J. P., Trenberth, K. E., Mann, M. E., Zanna, L., England, M. H., Zika, J. D., Fasullo, J. T., Yu, Y., Pan, Y., Zhu, J., Newsom, E. R., Bronselaer, B., & Lin, X. (2022). Past and future ocean warming. *Nature Reviews Earth and Environment*, *3*(11), 776–794. https://doi.org/10.1038/s43017-022-00345-1

Chou, E., Southall, B. L., Robards, M., & Rosenbaum, H. C. (2021). International policy, recommendations, actions and mitigation efforts of anthropogenic underwater noise. *Ocean & Coastal Management*, *202*, 105427. https://doi.org/10.1016/j.ocecoaman.2020.105427

Consortium for Ocean Leadership. (2018). *Report of the workshop on recommendations related to passive ocean acoustic data standards*.

Cornell University & Bureau for Ocean Energy Management. (2023). *Hybrid millidecade spectra at 1 minute resolution recorded by Cornell University for BOEM (BOEM-Cornell)*. NOAA National Centers for Environmental Information. https://doi.org/10.25921/wzyx-g235

Danielson, S. L., Grebmeier, J. M., Iken, K., Berchok, C., Britt, L., Dunton, K. H., Eisner, L., Farley, E. V., Fujiwara, A., & Hauser, D. D. (2022). Monitoring Alaskan Arctic shelf ecosystems through collaborative observation networks. *Oceanography*, *35*(3/4), 198–209.

Davis, G. E., Baumgartner, M. F., Corkeron, P. J., Bell, J., Berchok, C., Bonnell, J. M., Bort Thornton, J., Brault, S., Buchanan, G. A., Cholewiak, D. M., Clark, C. W., Delarue, J., Hatch, L. T., Klinck, H., Kraus, S. D., Martin, B., Mellinger, D. K., Moors-Murphy, H.Nieukirk, S., . . . Todd, S. (2020). Exploring movement patterns and changing distributions of baleen whales in the western North Atlantic using a decade of passive acoustic data. *Global Change Biology*, *26*(9), 4812–4840. https://doi.org/10.1111/gcb.15191

Delarue, J. J.-Y., Moors-Murphy, H., Kowarski, K. A., Davis, G. E., Urazghildiiev, I. R., & Martin, S. B. (2022). Acoustic occurrence of baleen whales, particularly blue, fin, and humpback whales, off eastern Canada, 2015-2017. *Endangered Species Research*, *47*, 265–289. https://doi.org/10.3354/esr01176

Desjonquères, C., Gifford, T., & Linke, S. (2020). Passive acoustic monitoring as a potential tool to survey animal and ecosystem processes in freshwater environments. *Freshwater Biology*, *65*(1), 7–19.

de Visser, C., Johansson, L. F., Kulkarni, P., Mei, H., Neerincx, P., Joeri van der Velde, K., Horvatovich, P., van Gool, A. J., Swertz, M. A., Hoen, P. A., & Niehues, A. (2023). Ten quick tips for building fair workflows. *PLOS Computational Biology*, *19*(9), e1011369. https://doi.org/10.1371/journal.pcbi.1011369

Drakopulos, L., Silver, J. J., Nost, E., Gray, N., & Hawkins, R. (2023). Making global oceans governance in/visible with smart Earth: The case of global fishing watch. *Environment and Planning E: Nature and Space*, *6*(2), 1098–1113.

Duarte, C. M., Chapuis, L., Collin, S. P., Costa, D. P., Devassy, R. P., Eguiluz, V. M., Erbe, C., Gordon, T. A., Halpern, B. S., Harding, H. R., Havlik, M. N., Meekan, M., Merchant, N. D., Miksis-Olds, J. L., Parsons, M., Predragovic, M., Radford, A. N., Radford, C. A.Simpson, S. D., . . . Zhang, X. (2021). The soundscape of the Anthropocene ocean. *Science*, *371*(6529), eaba 4658. https://doi.org/10.1126/science.aba4658

Dugan, P., Pourhomayoun, M., Shiu, Y., Paradis, R. D., Rice, A., & Clark, C. W. (2013). Using high performance computing to explore large complex bioacoustic soundscapes: Case study for right whale acoustics. *Procedia Computer Science*, *20*, 156–162. https://doi.org/10.1016/j.procs.2013.09.254

Estabrook, B., Hodge, K., Salisbury, A., Rahaman, A., Ponirakis, D., Harris, D., Zeh, J., Parks, S., & Rice, A. (2021). *Final report for New York Bight whale monitoring passive acoustic surveys October 2017-October 2020. Contract C009925*. New York State Department of Environmental Conservation East Setauket, NY. https://www.dec.ny.gov/docs/fish_marine_pdf/dmrnybacousticthree.pdf

Fadlallah, H., Kilany, R., Dhayne, H., El Haddad, R., Haque, R., Taher, Y., & Jaber, A. (2023). Context-aware big data quality assessment: A scoping review. *ACM Journal of Data and Information Quality*, *15*(3), 1–33.

Geiger, R. S., Howard, D., & Irani, L. (2021). The labor of maintaining and scaling free and open-source software projects. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–28.

Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, *10*(2), 169–185.

Greenhalgh, J. A., Genner, M. J., Jones, G., & Desjonquères, C. (2020). The role of freshwater bioacoustics in ecological research. *Wiley Interdisciplinary Reviews: Water*, *7*(3), e1416. https://doi.org/10.1002/wat2.1416

Hatch, L. T., Clark, C., Merrick, R., Van Parijs, S., Ponirakis, D., Schwehr, K., Thompson, M., & Wiley, D. (2008). Characterizing the relative contributions of large vessels to total ocean noise fields: A case study using the Gerry E. Studds Stellwagen Bank National Marine Sanctuary. *Environmental Management*, *42*(5), 735–752. https://doi.org/10.1007/s00267-008-9169-4

Hatch, L. T., Wahle, C. M., Gedamke, J., Harrison, J., Laws, B., Moore, S. E., Stadler, J. H., & Van Parijs, S. M. (2016). Can you hear me here? Managing acoustic habitat in US waters. *Endangered Species Research*, *30*, 171–186. https://doi.org/10.3354/esr00722

Haupt, S. E., Gagne, D. J., Hsieh, W. W., Krasnopolsky, V., McGovern, A., Marzban, C., Moninger, W., Lakshmanan, V., Tissot, P., & Williams, J. K. (2022). The history and practice of AI in the environmental sciences. *Bulletin of the American Meteorological Society*, *103*(5), E1351–E1370.

Haver, S. M., Rand, Z., Hatch, L. T., Lipski, D., Dziak, R. P., Gedamke, J., Haxel, J., Heppell, S. A., Jahncke, J., McKenna, M. F., Mellinger, D. K., Oestreich, W. K., Roche, L., Ryan, J., & Van Parijs, S. M. (2020). Seasonal trends and primary contributors to the low-frequency soundscape of the Cordell Bank National Marine Sanctuary. *Journal of the Acoustical Society of America*, *148*(2), 845–858. Article 2. https://doi.org/10.1121/10.0001726

Hossain, M., Tarafder, M., Ahmed, N., Al Noman, A., Sarkar, M., & Hossain, Z. (2023). Integrating AI with edge computing and cloud services for real-time data processing and decision making. *International Journal of Multidisciplinary Sciences and Arts*, *2*(4), 252–261.

Howe, B. M., Miksis-Olds, J., Rehm, E., Sagen, H., Worcester, P. F., & Haralabus, G. (2019). Observing the oceans acoustically. *Frontiers in Marine Science*, *6*, 426. https://doi.org/10.3389/fmars.2019.00426

ICES. (2024). *Continuous noise database* https://underwaternoise.ices.dk/continuous [Dataset].

International Quiet Ocean Experiment. (2019). *Guidelines for observation of ocean sound*. (18). https://scor-int.org/wp-content/uploads/2022/08/IQOE_2019_Standards_Workshop_Report.pdf

International Whaling Commission. (2014). *Draft joint workshop report: Predicting sound fields - global soundscape modelling to inform management of cetaceans and anthropogenic noise*. Paper SC/65b/Rep03. Leiden, Netherlands (Paper SC/65b/Rep03; Issue Paper SC/65b/Rep03).

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C. T., Goble, C., Guizzardi, G., Hansen, K. K., Hasnain, A., Hettne, K., Heringa, J., Hooft, R. W. W., Imming, M.Jeffery, K. G., ... Mons, B. (2020). Fair principles: Interpretations and implementation considerations. *Data Intelligence*, *2*(1–2), 10–29. https://doi.org/10.1162/dint_r_00024

Jalkanen, J.-P., Johansson, L., Andersson, M. H., Majamäki, E., & Sigray, P. (2022). Underwater noise emissions from ships during 2014-2020. *Environmental Pollution*, *311*, 119766. https://doi.org/10.1016/j.envpol.2022.119766

Jannapureddy, R., Vien, Q.-T., Shah, P., & Trestian, R. (2019). An auto-scaling framework for analyzing big data in the cloud environment. *Applied Sciences*, *9*(7), 1417.

Ji, Y., Marian, A. D., & Montie, E. W. (2024). Deep-learning-based detection of recreational vessels in an estuarine soundscape in the May River, South Carolina, USA. *PLOS ONE*, *19*(7), e0302497.

Johnson, G. C., Hosoda, S., Jayne, S. R., Oke, P. R., Riser, S. C., Roemmich, D., Suga, T., Thierry, V., Wijffels, S. E., & Xu, J. (2022). Argo-two decades: Global oceanography, revolutionized. *Annual Review of Marine Science*, *14*(1), 379–403. https://doi.org/10.1146/annurev-marine-022521-102008

Jung, C., & Schindler, D. (2019). Changing wind speed distributions under future global climate. *Energy Conversion and Management*, *198*, 111841. https://doi.org/10.1016/j.enconman.2019.111841

Kanterakis, A., Potamias, G., Swertz, M. A., & Patrinos, G. P. (2018). Creating transparent and reproducible pipelines: Best practices for tools, data, and workflow management systems. In C. Lambert D. Baker, & G. Patrinos (Eds.), *Human genome informatics* (pp. 15–43). Elsevier. https://doi.org/10.1016/B978-0-12-809414-3.00002-4

Knust, R. (2017). Polar research and supply vessel Polarstern operated by the Alfred-Wegener-Institute. *Journal of Large-Scale Research Facilities*, *3*, A119. https://doi.org/10.17815/jlsrf-3-163

Kowarski, K. A., & Moors-Murphy, H. (2021). A review of big data analysis methods for baleen whale passive acoustic monitoring. *Marine Mammal Science*, *37*(2), 652–673.

Li, V. O., Lam, J. C., & Cui, J. (2021). AI for social good: AI and big data approaches for environmental decision-making. *Environmental Science and Policy*, *125*, 241–246. https://doi.org/10.1016/j.envsci.2021.09.001

Li, X., Feng, M., Ran, Y., Su, Y., Liu, F., Huang, C., Shen, H., Xiao, Q., Su, J., Yuan, S., & Guo, H. (2023). Big data in Earth system science and progress towards a digital twin. *Nature Reviews Earth and Environment*, *4*(5), 319–332. https://doi.org/10.1038/s43017-023-00409-w

Lnenicka, M., & Nikiforova, A. (2021). Transparency-by-design: What is the role of open data portals? *Telematics and Informatics*, *61*, 101605. https://doi.org/10.1016/j.tele.2021.101605

Marchetto, P., Strickhart, A., Mack, R., & Cheyne, H. (2012). Temperature compensation of a quartz tuning-fork clock crystal via post-processing. In M. Larsen & M. Bulatowicz (Eds.), *2012 IEEE International Frequency Control Symposium Proceedings* (pp. 1–4). IEEE.

Marian, A. D., Monczak, A., Balmer, B. C., Hart, L. B., Soueidan, J., & Montie, E. W. (2021). Long-term passive acoustics to assess spatial and temporal vocalization patterns of Atlantic common bottlenose dolphins (Tursiops truncatus) in the May River estuary, South Carolina. *Marine Mammal Science*, *37*(3), 1060–1084.

Mariani, A. C. B., Godoy, S. N., & Santos, M. C. O. (2024). The use of passive acoustic monitoring as an auxiliary tool for monitoring marine protected areas. *Ocean and Coastal Research*, *72*, e24081. https://doi.org/10.1590/2675-2824072.24033

Martin, S. B., Gaudet, B. J., Klinck, H., Dugan, P. J., Miksis-Olds, J. L., Mellinger, D. K., Mann, D. A., Boebel, O., Wilson, C. C., Ponirakis, D. W., & Moors-Murphy, H. (2021a). Erratum: Hybrid millidecade spectra: A practical format for exchange of long-term ambient sound data. *JASA Express Letters*, *1*, Article 8. https://doi.org/10.1121/10.0003324

Martin, S. B., Gaudet, B. J., Klinck, H., Dugan, P. J., Miksis-Olds, J. L., Mellinger, D. K., Mann, D. A., Boebel, O., Wilson, C. C., Ponirakis, D. W., & Moors-Murphy, H. (2021b). Hybrid millidecade spectra: A practical format for exchange of long-term ambient sound data. *JASA Express Letters*, *1*, Article 1. https://doi.org/10.1121/10.0003324/

McDonald, M. A., Hildebrand, J. A., & Wiggins, S. M. (2006). Increases in deep ocean ambient noise in the northeast Pacific west of San Nicolas Island, California. *Journal of the Acoustical Society of America*, *120*(2), 711–718.

McKenna, M., Baumann-Pickering, S., Kok, A. C., Oestreich, W. K., Adams, J. D., Barkowski, J., Fristrup, K. M., Goldbogen, J. A., Joseph, J., Kim, E. B., Kügler, A., Lammers, M. O., Margolina, T., Peavey Reeves, L. E., Rowell, T. J., Stanley, J. A., Stimpert, A. K., Zang, E. J.Southall, B. L., . . . Parijs, S. V. (2021). Advancing the interpretation of shallow water marine soundscapes. *Frontiers in Marine Science*, *8*, 719258. https://doi.org/10.3389/fmars.2021.719258

McKenna, M., Katz, S., Wiggins, S., Ross, D., & Hildebrand, J. (2012). A quieting ocean: Unintended consequence of a fluctuating economy. *Journal of the Acoustical Society of America*, *132*(3), EL169–EL175.

McKenna, M., Rowell, T. J., Margolina, T., Baumann-Pickering, S., Solsona-Berga, A., Adams, J. D., Joseph, J., Kim, E. B., Kok, A. C., Kügler, A., Lammers, M. O., Merkens, K., Reeves, L. P., Southall, B. L., Stimpert, A. K., Barkowski, J., Thompson, M. A., Van Parijs, S.Wall, C. C., . . . Zang, E. J. (2024). Understanding vessel noise across a network of marine protected areas. *Environmental Monitoring and Assessment*, *196*(4), Article 4. https://doi.org/10.1007/s10661-024-12497-2

Mellinger, D. K., Küsel, E. T., Thomas, L., & Marques, T. A. (2009). Taming the Jez monster: Estimating fin whale spatial density using acoustic propagation modeling. *Journal of the Acoustical Society of America*, *126*(4_Supplement), 2229–2229.Article 4_Supplement. https://doi.org/10.1121/1.3248985

Merchant, N. D., Brookes, K. L., Faulkner, R. C., Bicknell, A. W., Godley, B. J., & Witt, M. J. (2016). Underwater noise levels in UK waters. *Scientific Reports*, *6*(1), 36942.

Miksis-Olds, J., Ainslie, M., Blair, H. B., Hazen, E. L., Heaney, K. D., Lyons, A. P., Martin, B. S., & Warren, J. D. (2025). Overview of the Atlantic deepwater ecosystem observatory network. *Oceanography*. https://doi.org/10.5670/oceanog.2025.301

Miksis-Olds, J., Dugan, P. J., Martin, S. B., Klinck, H., Mellinger, D. K., Mann, D. A., Ponirakis, D. W., & Boebel, O. (2021). Ocean sound analysis software for making ambient noise trends accessible (MANTA). *Frontiers in Marine Science*, *8*, 703650. https://doi.org/10.3389/fmars.2021.703650

Miksis-Olds, J., Stabeno, P. J., Napp, J. M., Pinchuk, A. I., Nystuen, J. A., Warren, J. D., & Denes, S. L. (2013). Ecosystem response to a temporary sea ice retreat in the Bering Sea: Winter 2009. *Progress in Oceanography*, *111*, 38–51. https://doi.org/10.1016/j.pocean.2012.10.010

Monczak, A., Berry, A., Kehrer, C., & Montie, E. W. (2017). Long-term acoustic monitoring of fish calling provides baseline estimates of reproductive timelines in the May River estuary, southeastern USA. *Marine Ecology Progress Series*, *581*, 1–19. https://doi.org/10.3354/meps12322

Monczak, A., Ji, Y., Soueidan, J., & Montie, E. W. (2019). Automatic detection, classification, and quantification of sciaenid fish calls in an estuarine soundscape in the Southeast United States. *PLOS ONE*, *14*(1), e0209914.

Monczak, A., McKinney, B., Mueller, C., & Montie, E. W. (2020). What's all that racket! Soundscapes, phenology, and biodiversity in estuaries. *PLOS ONE*, *15*(9), e0236874.

Monczak, A., McKinney, B., Souiedan, J., Marian, A. D., Seder, A., May, E., Morgenstern, T., Roumillat, W., & Montie, E. W. (2022). Sciaenid courtship sounds correlate with juvenile appearance and abundance in the May River, South Carolina, USA. *Marine Ecology Progress Series*, *693*, 1–17. https://doi.org/10.3354/meps14109

Monczak, A., Mueller, C., Miller, M. E., Ji, Y., Borgianini, S. A., & Montie, E. W. (2019). Sound patterns of snapping shrimp, fish, and dolphins in an estuarine soundscape of the southeastern USA. *Marine Ecology Progress Series*, *609*, 49–68. https://doi.org/10.3354/meps12813

Monterey Bay Aquarium Research Institute. (2023). *Hybrid millidecade spectra at 1 minute resolution recorded at the Monterey Accelerated Research System (MARS) cabled observatory*. NOAA National Centers for Environmental Information. https://doi.org/10.25921/6gvv-dk87

NOAA Alaska Fisheries Science Center Marine Mammal Laboratory. (2023). *Hybrid millidecade spectra at 1 minute resolution recorded by the Alaska Fisheries Science Center (AFSC)*. NOAA National Centers for Environmental Information. https://doi.org/10.25921/sw1h-g843

NOAA Northeast Fisheries Science Center. (2023). *Hybrid millidecade spectra at 1 minute resolution recorded by the Northeast Fisheries Science Center (NEFSC)*. NOAA National Centers for Environmental Information. https://doi.org/10.25921/s87x-gb35

NOAA OAR Pacific Marine Environmental Laboratory, Cooperative Institute for Marine Ecosystem and Resources Studies, National Marine Fisheries Service, NOS Office of National Marine Sanctuaries, & DOI NPS Natural Resource Stewardship and Science Directorate. (2023). Hybrid millidecade spectra at 1 minute resolution recorded at the ocean noise reference station network. NOAA National Centers for Environmental Information. https://doi.org/10.25921/tr4q-xa94

NOAA Office of National Marine Sanctuaries & U.S Navy. (2023). *Hybrid millidecade spectra at 1 minute resolution recorded for the Sanctuary soundscape monitoring project (SanctSound)* [dataset]. NOAA National Centers for Environmental Information. https://doi.org/10.25921/ymf8-5k59

NWS. (2016, August 31). Tropical Storm Hermine Intermediate Advisory Number 12A. [NOAA National Weather Service]. *National Hurricane Center*. https://www.nhc.noaa.gov/archive/2016/al09/al092016.public_a.012.shtml?

O'Brien, M., Duerr, R., Taitingfong, R., Martinez, A., Vera, L., Jennings, L., Downs, R., Antognoli, E., ten Brink, T., Halmai, N., David-Chavez, D., Carroll, S. R., Hudson, M., & Buttigieg, P. L. (2024). Earth science data repositories: Implementing the CARE principles. *Data Science Journal*, *23*. https://doi.org/10.5334/dsj-2024-037

Oestreich, W. K., Fahlbusch, J. A., Cade, D. E., Calambokidis, J., Margolina, T., Joseph, J., Friedlaender, A. S., McKenna, M. F., Stimpert, A. K., & Southall, B. L. (2020). Animal-borne metrics enable acoustic detection of blue whale migration. *Current Biology*, *30*(23), 4773–4779.

Parcerisas, C. (2023). *Pypam: A package to process long-term underwater acoustics data in chunks*. [Python]. https://doi.org/10.5281/zenodo.7584847

Pearson, E. J., Oestreich, W. K., Ryan, J. P., Haver, S. M., Gedamke, J., Dziak, R. P., & Wall, C. C. (2023). Widespread passive acoustic monitoring reveals spatio-temporal patterns of blue and fin whale song vocalizations in the northeast Pacific Ocean. *Frontiers in Remote Sensing*, *4*, 994518. https://doi.org/10.3389/frsen.2023.994518

Possenti, L., de Nooijer, L., de Jong, C., Lam, F.-P., Beelen, S., Bosschers, J., van Terwisga, T., Stigter, R., & Reichart, G.-J. (2024). The present and future contribution of ships to the underwater soundscape. *Frontiers in Marine Science*, *11*, 1252901. https://doi.org/10.3389/fmars.2024.1252901

Ramachandran, R., Bugbee, K., & Murphy, K. (2021). From open data to open science. *Earth & Space Science*, *8*(5), e2020EA001562.

Rasmussen, J. H., Stowell, D., & Briefer, E. F. (2024). Sound evidence for biodiversity monitoring. *Science*, *385*(6705), 138–140.

Reagan, J. R., & NOAA National Centers for Environmental Information. (2023). *World Ocean Atlas 2023 (NCEI accession 0282728). NOAA National Centers for Environmental Information* [dataset]. https://www.ncei.noaa.gov/archive/accession/0282728

Reagan, J. R., Garcia, H. E., Boyer, T. P., Baranova, O. K., Bouchard, C., Cross, S. L., Dukhovskoy, D., Grodsky, A., Locarnini, R. A., Mishonov, A. V., Paver, C. R., Seidov, D., & Wang, Z. (2024). World Ocean Atlas, 2023: Product documentation. A. Mishonov, technical editor.

Rice, A., Širović, A., Trickey, J. S., Debich, A. J., Gottlieb, R. S., Wiggins, S. M., Hildebrand, J. A., & Baumann-Pickering, S. (2021). Cetacean occurrence in the Gulf of Alaska from long-term passive acoustic monitoring. *Marine Biology*, *168*(5), 72.

Ross, S. R.-J., O'Connell, D. P., Deichmann, J. L., Desjonquères, C., Gasc, A., Phillips, J. N., Sethi, S. S., Wood, C. M., & Burivalova, Z. (2023). Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions. *Functional Ecology*, *37*(4), 959–975. https://doi.org/10.1111/1365-2435.14275

Rueda, C., Cline, D., & Ryan, J. P. (n.d.). *PBP - PyPAM-based processing* [Computer software].

Ryan, J., Cline, D., Dawe, C., McGill, P., Zhang, Y., Joseph, J., Margolina, T., Caillat, M., Fischer, M., & DeVogelaere, A. (2016). New passive acoustic monitoring in Monterey Bay National Marine Sanctuary. In J. Zande & B. Kirkwood (Eds.), *Oceans, 2016 MTS/IEEE Monterey* (pp. 1–8). IEEE. https://doi.org/10.1109/OCEANS.2016.7761363

Ryan, J., Joseph, J. E., Margolina, T., Hatch, L. T., Azzara, A., Reyes, A., Southall, B. L., DeVogelaere, A., Peavey Reeves, L. E., Zhang, Y., Cline, D. E., Jones, B., McGill, P., Baumann-Pickering, S., & Stimpert, A. K. (2021). Reduction of low-frequency vessel noise in Monterey Bay National Marine Sanctuary during the COVID-19 pandemic. *Frontiers in Marine Science*, *8*, 656566. https://doi.org/10.3389/fmars.2021.656566

Scripps Institution of Oceanography. (2023). *Hybrid millidecade spectra at 1 minute resolution recorded by the Scripps Whale Acoustics Laboratory (SWAL). NOAA National Centers for Environmental Information* [dataset]. NOAA National Centers for Environmental Information. https://doi.org/10.25921/ht84-bg61

Sethanandha, B. D., Massey, B., & Jones, W. (2010). Managing open source contributions for software project sustainability. In *Portland International Conference on Management of Engineering and Technology (PICMET)*, Phuket, Thailand (pp. 1–9). IEEE.

Širović, A., Hildebrand, J. A., Wiggins, S. M., & Thiele, D. (2009). Blue and fin whale acoustic presence around Antarctica during 2003 and 2004. *Marine Mammal Science*, *25*(1), Article 1.125–136. https://doi.org/10.1111/j.1748-7692.2008.00239.x

Širović, A., Rice, A., Chou, E., Hildebrand, J. A., Wiggins, S. M., & Roch, M. A. (2015). Seven years of blue and fin whale call abundance in the Southern California Bight. *Endangered Species Research*, *28*(1), 61–76.Article 1. https://doi.org/10.3354/esr00676

Smott, S., Monczak, A., Miller, M. E., & Montie, E. W. (2018). Boat noise in an estuarine soundscape-a potential risk on the acoustic communication and reproduction of soniferous fish in the May River, South Carolina. *Marine Pollution Bulletin*, *133*, 246–260. https://doi.org/10.1016/j.marpolbul.2018.05.016

Soltwedel, T., Mokievsky, V., Rabouille, C., Sauter, E., Volkenandt, M., & Hasemann, C. (2013). Effects of experimentally increased near-bottom flow on meiofauna diversity and community structure in the Arctic Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, *73*, 31–45. https://doi.org/10.1016/j.dsr.2012.11.008

Spalding, M. J. (2016). The new blue economy: The future of sustainability. *Journal of Ocean and Coastal Economics*, *2*(2), 8.

Spalding, M. J., Braestrup, A. E., & Refosco, A. (2021). Greening the blue economy: A transdisciplinary analysis. In R. Lozano, S. Afionis, & C. Desha (Eds.), *Sustainability in the maritime domain: Towards ocean governance and beyond* (pp. 27–59). Springer.

Spaulding, E., Robbins, M., Calupca, T., Clark, C. W., Tremblay, C., Waack, A., Warde, A., Kemp, J., & Newhall, K. (2009). An autonomous, near-real-time buoy system for automatic detection of North Atlantic right whale calls. *Proceedings of Meetings on Acoustics*, *6*(1), 010001.

Stafford, K. (2021). The changing Arctic marine soundscape. *NOAA Technical Report*, *21-14*, 7. https://doi.org/10.25923/jagc-4a84

Stoudt, S., Jernite, Y., Marshall, B., Marwick, B., Sharan, M., Whitaker, K., & Danchev, V. (2024). Ten simple rules for building and maintaining a responsible data science workflow. *PLOS Computational Biology*, *20*(7), e1012232.

Sun, J. (2024). Sustaining scientific open-source software ecosystems: Challenges, practices, and opportunities. In A Paiva & R. Abreu (Eds.), *2024 IEEE/ACM 46th international conference on software engineering, association for computing machinery*, Lisbon, Portugal (pp. 234–236). Association for Computing Machinery. https://doi.org/10.1145/3639478.36398

Thomisch, K., Spiesecke, S., & Boebel, O. (2023a). *Hybrid millidecade spectra at 1 minute resolution recorded at the Frontiers in Arctic Marine Monitoring (FRAM) mooring F5-17 of the Alfred Wegener*

Institute Helmholtz Centre for Polar and Marine Research. NOAA National Centers for Environmental Information. https://doi.org/10.25921/jh84-k867

Thomisch, K., Spiesecke, S., & Boebel, O. (2023b). Passive acoustic monitoring data recorded by 972 recorder SV1088 at mooring F5-17 in Fram Strait in 2016/2017. PANGAEA [Dataset]. 973. https://doi.org/10.1594/PANGAEA.956286

Timmermans, B., Gommenginger, C., Dodet, G., & Bidlot, J.-R. (2020). Global wave height trends and variability from new multimission satellite altimeter products, reanalyses, and wave buoys. *Geophysical Research Letters*, *47*(9), e2019GL086880. https://doi.org/10.1029/2019GL086880

Tjernström, J., Gallagher, M., Holt, J., Svensson, G., Shupe, M. D., Day, J. J., Ferrighi, L., Khalsa, S. J., Hartten, L. M., & O'Connor, E. (2024). Accelerating research through community open source software for a standardized file format to improve process representation in numerical weather prediction models. *EGUsphere [preprint]*, 1–20. https://doi.org/10.5194/egusphere-2024-2088

Transue, L., Monczak, A., Tribble, C., Marian, A., Fair, P., Ballenger, J., Balmer, B., & Montie, E. W. (2023). The biological and anthropogenic soundscape of an urbanized port-the Charleston Harbor estuary, South Carolina, USA. *PLOS ONE*, *18*(4), e0283848. https://doi.org/10.1371/journal.pone.0283848

Tribble, C., Monczak, A., Transue, L., Marian, A., Fair, P., Balmer, B., Ballenger, J., Baker, H., Weinpress-Galipeau, M., Robertson, A., Strand, A., & Montie, E. W. (2023). Enhancing interpretation of cetacean acoustic monitoring: Investigating factors that influence vocalization patterns of Atlantic bottlenose dolphins in an urbanized estuary, Charleston Harbor, South Carolina, USA. *Aquatic Mammals*, *49*(6), 519–549. https://doi.org/10.1578/AM.49.6.2023.519

University of New Hampshire & Office of Naval Research. (2023). *Hybrid millidecade spectra at 1 minute resolution recorded at the acoustic and Environmental Observation Network (AEON) for the Office of Naval Research*. NOAA National Centers for Environmental Information. https://doi.org/10.25921/d21t-be25

University of South Carolina Beaufort & IOOS SECOORA. (2023). *Hybrid millidecade spectra at 1 minute resolution recorded at the estuary soundscape observatory network in the southeast (ESONS) for IOOS SECOORA*. NOAA National Centers for Environmental Information. https://doi.org/10.25921/0xbb-xx76

Vance, T. C., Huang, T., & Butler, K. A. (2024). Big data in Earth science: Emerging practice and promise. *Science*, *383*(6688), eadh9607. https://doi.org/10.1126/science.adh9607

Van Parijs, S., Clark, C. W., Sousa-Lima, R. S., Parks, S. E., Rankin, S., Risch, D., & Van Opzeeland, I. C. (2009). Management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales. *Marine Ecology Progress Series*, *395*, 21–36. https://doi.org/10.3354/meps08123

Van Parijs, S., DeAngelis, A., Aldrich, T., Gordon, R., Holdman, A., McCordic, J., Mouy, X., Rowell, T., Tennant, S., & Westell, A. (2023). Establishing baselines for predicting change in ambient sound metrics, marine mammal, and vessel occurrence within a US offshore wind energy area. *ICES Journal of Marine Science*, *148*(3). https://doi.org/10.1093/icesjms/fsad148

Verhulst, S., Zahuranec, A. J., & Chafetz, H. (2025). Moving toward the FAIR-R principles: Advancing AI-ready data. *Available at SSRN*. https://doi.org/10.2139/ssrn.5164337

Wall, C. C., Haver, S. M., Hatch, L. T., Miksis-Olds, J., Bochenek, R., Dziak, R. P., & Gedamke, J. (2021). The next wave of passive acoustic data management: How centralized access can enhance science. *Frontiers in Marine Science*, *8*, 703682. https://doi.org/10.3389/fmars.2021.703682

Welch, H., Ames, R. T., Kolla, N., Kroodsma, D. A., Marsaglia, L., Russo, T., Watson, J. T., & Hazen, E. L. (2024). Harnessing AI to map global fishing vessel activity. *One Earth*, *7*(10), 1685–1691.

Wenz, G. M. (1962). Acoustic ambient noise in the ocean: Spectra and sources. *Journal of the Acoustical Society of America*, *34*(12), 1936–1956. Article 12. https://doi.org/10.1121/1.1909155

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T.Finkers, R., … Zhao, J. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 1–9. https://doi.org/10.1038/sdata.2016.18

Wilkinson, S. R., Aloqalaa, M., Belhajjame, K., Crusoe, M. R., Kinoshita, B. P., Gadelha, L., Garijo, D., Gustafsson, O. J. R., Juty, N., & Kanwal, S. (2024). Applying the FAIR principles to computational workflows. *ArXiv Preprint ArXiv: 2410.03490*.

Young, I., Zieger, S., & Babanin, A. V. (2011). Global trends in wind speed and wave height. *Science*, *332*(6028), 451–455.

Zheng, C. W., Pan, J., & Li, C. Y. (2016). Global oceanic wind speed trends. *Ocean & Coastal Management*, *129*, 15–24. https://doi.org/10.1016/j.ocecoaman.2016.05.001

ZoBell, V. M., Hildebrand, J. A., & Frasier, K. E. (2024). Comparing pre-industrial and modern ocean noise levels in the Santa Barbara Channel. *Marine Pollution Bulletin*, *202*, 116379. https://doi.org/10.1016/j.marpolbul.2024.116379