Evolving challenges in ocean data management

Exploring recommendations for NOAA and its partners

WORKING DRAFT, May 1, 2020: NOT FOR FURTHER CIRCULATION

Executive Summary

The mission of the National Oceanic and Atmospheric Administration (NOAA) includes a focus on conserving and managing coastal and marine ecosystems and resources while understanding our constantly changing climate, weather, oceans, and coast. This mission requires that the agency collect and analyze data from the atmosphere to the lithosphere.

To achieve this mission, NOAA along with other federal agencies and regional partners collects, manages, analyzes, and disseminates vast amounts of data. These data holdings will only grow larger and become more complex as new technologies make it faster, cheaper, and easier than ever to collect, use, and learn from data. Meanwhile, established industries, emerging startups, academics, and non-governmental organizations (NGOs) are using those same advances to grow their own data holdings and inform their activities.

Ocean data covers a range of interconnected categories. Oceanographic data about the physical, biological, and chemical domains can be combined with socio-economic data for scientific research, planning, natural resource management, and other uses. NOAA facilitates the collection, management, and dissemination of ocean data using a range of programs, products, and partners including federal-regional partnerships such as the Integrated Ocean Observing System (IOOS) and Regional Ocean Partnerships (ROP) that aim to make ocean information available in a manner that responds to varying regional needs. Ocean data is also made available to private and public sector partners through the Marine Cadastre, the National Data Buoy Center, the National Center for Environmental Information, and other sources. Increasingly, NOAA is also exploring new technologies - including cloud computing solutions like the Big Data Project (BDP) - to make its ocean data accessible in new ways to existing and new audiences.

Despite this robust ocean data ecosystem, only 5% of the ocean has been fully explored, leaving limited data to answer scientific questions and help guide management decisions.[1] Those data that are collected often lack interoperability across categories and are not necessarily shared between government and private stakeholders. There are clear opportunities to collect new ocean data while making better use of that data we already have at our disposal.

However, challenges stand in the way:

- Data collection, management, sharing, and use are expensive, and governments and other stakeholders are often resource constrained.

[1] Oceans and Coasts, National Atmospheric and Oceanic Administration, https://www.noaa.gov/oceans-coasts, accessed 30 January 2020.

- Interoperability between different scientific domains and collaboration between stakeholders is not as robust as it could be. Work needs to be done to agree on common standards, vocabularies, and metadata approaches.
- Not all ocean data stakeholders have the same level of technical capacity, making it difficult to achieve broad-based standards adoption or embrace new technologies.
- Different domains and regions have different and sometimes unique data needs.
- There is a lack of incentives and policy frameworks for private industry, research scientists, and other vital stakeholders to share their data.
- Privacy, equity, and confidentiality concerns may limit data availability.
- New technologies are not necessarily integrated into existing data collection and processing workflows, limiting their utility for broader stakeholder groups.

In this paper, potential and actionable solutions to these problems are identified from an ecosystem perspective. Stakeholders can come together in a variety of fora to agree on common standards, budgets can be increased and new approaches can be found that may lower costs, data needs can be prioritized by policymakers, incentives can be identified and implemented to encourage data sharing, and more. The paper reflects a wide ranging literature review, discussions at a February 2020 Roundtable on ocean data hosted by the Center for Open Data Enterprise in partnership with Ocean Conservancy, NOAA, Amazon Web Services, and Microsoft, and interviews with relevant stakeholders.

Purpose and Need Statement

Data on our oceans are vital to ongoing efforts to protect the environment, track and manage changes to the marine ecosystem, ensure continued access to ocean and coastal resources, and grow the Blue Economy in sustainable ways. This paper will explore the current landscape of ocean data - primarily by looking at U.S. sources of ocean data - and identify ways to improve the collection, archiving, dissemination, and application of those data.

The goal of this paper is to help improve our understanding of the oceans to support sustainable use and adaptive management. It explores ways to achieve several objectives to address this goal, including:

- Creating a virtuous cycle of demand and supply for ocean data from public, private, academic, and other sources resulting in better knowledge of the ocean
- Better integrating public and private sources of ocean data
- Improving the technological, policy, and cultural infrastructure for ocean data sharing
- Making data available in a manner that facilitates effective and efficient ocean related decision-making.

The National Oceanic and Atmospheric Administration (NOAA) is responsible for one of the largest data inventories of any federal agency, collecting, managing, and publishing data “from the bottom of the ocean to the surface of the sun." In recent years, NOAA has developed collaborative opportunities with the private sector that promote and enable the public and commercial use of its data, specifically through its Big Data Project (BDP). The BDP uses the Cloud - a system for networked access to shared computing resources[2] - to make it easier to access and use NOAA's large collection of data.[3] So far, users of the BDP have primarily requested satellite and earth observation data, but NOAA sees an opportunity to add more ocean data as well. Additionally, an ongoing dialogue between federal and regional ocean data partners has identified a number of opportunities for improved collection, management, and use of ocean data. NOAA has a strong interest in publishing more ocean data in accessible ways and encouraging its use to address a range of scientific questions and management challenges facing our oceans and coasts.

Simultaneously, established industries, emerging start-ups, academic and research institutions, and other non-federal entities are collecting more data than ever about our oceans. Those data can be used in tandem with NOAA's ocean data to help answer scientific questions and address management challenges. However, there are technical, policy, and cultural hurdles to overcome to ensure integration and use of these data.

There is a clear opportunity to improve NOAA's data collection and dissemination practices while better integrating non-federal ocean data with NOAA ocean data. To do so we need a better understanding of NOAA's current ocean data practices, how they fit in with other sources of ocean data, and how those practices could better meet the needs of the diverse set of ocean data stakeholders. In turn, a better understanding of challenges faced by ocean data stakeholders - including potential barriers to accessing, sharing, or further utilizing those data - will help guide the collective advancement of ocean data and our understanding of the ocean.

Acknowledgements

This paper was written by Matt Rumsey and Nidhisha Philip from the Center for Open Data Enterprise and [Ocean Conservancy Authors]. Editorial direction was provided by Joel Gurin and Paul Kuhne of CODE and [Ocean Conservancy Representatives].

The Center for Open Data Enterprise hosted a roundtable in partnership with Ocean Conservancy, NOAA, Amazon Web Services, and Microsoft in February 2020 to discuss the use of ocean data and development of data driven strategies to improve ocean health and promote

---

[2] The most common definition of 'cloud computing' is the NIST definition: Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. The essential characteristics of such a model are on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. *See, Mell, P., and Grance, T. (2011). The NIST Definition of Cloud Computing. Gaithersburg, MD: National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-145*

[3] Big Data Project, https://www.noaa.gov/big-data-project, accessed 27 January 2020.

the blue economy. While the Roundtable had a special focus on leveraging the Big Data Project, participants also discussed challenges faced by federal and regional partners in collecting and disseminating data. The authors would like to thank the hosts and participants in that Roundtable as well as individuals who participated in informational interviews before and after the event as well as those that read and provided feedback on various drafts of this paper.

Introduction

Data about America's oceans come from a wide range of sources, cross a diverse array of domains, and have nearly endless applications. They are being used to manage ocean ecosystems, enable the sustainable growth of the Blue Economy, protect endangered marine species, help the global community react to and prepare for climate change, and much more. For example, ocean observations and models help us protect endangered species by understanding food sources and migration patterns, identify appropriate sites for offshore wind energy production facilities, develop better hurricane prediction models, and plan for the impacts of climate change on coastal communities.

Data sources and technology:

NOAA is the federal agency with primary responsibility for ocean data. The agency applies those data towards its mission of conserving and managing coastal and marine ecosystems and resources.

Along with other federal agencies, regional organizations, academic partners, and others, NOAA manages a robust system of data on America's oceans that has evolved significantly since its inception. Additionally, private companies, citizen scientists, non-governmental organizations (NGOs) and more are leveraging new, inexpensive technologies to add to the ever-expanding pool of ocean data.

Rapid changes in technology and a growing desire to use data of all types present a host of opportunities and challenges for public actors working to create, manage, and share ocean data. Over the coming years the federal ocean data ecosystem will have to continue its evolution by embracing new technologies, ingesting data from new sources, sharing data with new and growing audiences, and leveraging data to fuel increasingly sophisticated analytical tools to solve new ocean management and conservation problems.

Ocean data, like the ecosystems they represent, are complex. This complexity presents significant baseline challenges for data stewards - individuals who manage data at various points throughout its lifecycle - trying to manage and share ocean data in ways that ensure they are discoverable and useable.[4] The landscape of public ocean data in the U.S. - which features

[4] See https://www.ncddc.noaa.gov/activities/science-technology/data-management/ for more on NOAA's National Centers for Environmental Information's approach to data lifecycle management.

national, regional, and project-based data platforms, overarching data systems and individual datasets, real-time data and archival data assets, and more - reflects this complexity.

This landscape is set to become more complex in the near future as new, inexpensive, and autonomous ocean observation technologies become more widespread. Already, autonomous platforms that can take accurate measurements over the course of years-long deployments "are transmitting as much data in one year as has been acquired in the past century."[5] The increase in real-time data will require a transformation of network architecture and data management capabilities.

As ocean data stewards work to keep pace with this explosion of information, they will face a number of familiar challenges. These include funding and cost concerns, varying levels of technical capacity among stakeholders, data silos, regional data needs, privacy and confidentiality concerns, regional data gaps, a lack of incentives for data sharing, challenges implementing data standards and data integration, and more.

Policy Landscape:

Ocean data has emerged as a federal priority in recent years. In 2018, the Trump Administration updated the federal ocean policy with a specific focus on making federal, unclassified data available to states and regions in a timely manner.[6] As part of this ocean policy frame, the White House Office of Science and Technology Policy (OSTP) and the Council on Environmental Quality (CEQ) hosted the White House Summit on Partnerships in Ocean Science and Technology in late 2019. The goal of the Summit was to engage a cross-section of the U.S. ocean community to discuss how to elevate, empower, and transform how we work together to build and sustain partnerships, and to lay the foundation for a broadly defined but common direction to advance marine science, promote new technologies, and explore the unknown ocean.[7]

These recent policy frames make it clear that building a robust ocean data ecosystem is just one part of a larger effort to fill gaps in our understanding of the ocean.[8] Meanwhile, regional

---

[5] Toste Tanhua, Sylvie Pouliquen, Jessica Hausman, Kevin O'Brien, Pip Bricher, Taco de Bruin, Justin Buck, Eugene Burger, Thierry Carval, Kenneth Casey, Steve Diggs, Alessandra Giorgetti, Helen Glaves, Valerie Harscoat, Danie Kinkade, Jose Muelbert, Antonio Novellino, Benjamin Pfeil, Peter Pulsifer, Anton Van de Putte, Erin Robinson, Dick Schaap, Alexander Smirnov, Neville Smith, Derrick Snowden, Tobias Spears, Shelley Stall, Marten Tacoma, Peter Thijsse, Stein Tronstad, Thomas Vandenberghe, Micah Wengren, Lesley Wyborn and Zhiming Zhao (2019) Ocean FAIR Data Services. *Front. Mar. Sci.* 6:440, 3.

[6] President Donald Trump, Executive Order Regarding the Ocean Policy to Advance the Economic, Security, and Environmental Interests of the United States, the White House, 19 June 2018.

[7] Summary of the 2019 White House Summit on Partnerships in Ocean Science and Technology, Ocean Policy Committee, November 2019, https://www.whitehouse.gov/wp-content/uploads/2019/12/Ocean-ST-Summit-Readout-Final.pdf

[8] Amy Trice, How to Build Partnerships in Ocean Science and Technology, Ocean Conservancy, 12 November 2019,

ocean data stewards have come together in recent months to discuss ways to boost data coverage and improve data sharing across the U.S.[9] Efforts to identify data gaps and needs at the regional level are working to inform actions within federal policy and vice versa. Beyond the ocean policy momentum surrounding federal data, the Foundations for Evidence-Based Policymaking Act of 2018 (FEBPA) imposes a legislative mandate on federal agencies to promote open data policies and inter-agency sharing of data. The cross-agency priority goal, Leveraging Data as a Strategic Asset, and the Federal Data Strategy further reinforce this mandate.[10]

The BDP, a NOAA initiative launched in 2015 to improve the "discoverability, accessibility, and usability" of NOAA's data resources, represents one approach to improve ocean data sharing. The BDP was initially framed as a research project to investigate if the inherent value of NOAA's data could underwrite the costs of commercial cloud storage and as an attempt to drive innovation and new business opportunities for U.S. industry.[11] Under the original set of agreements, NOAA and the five cloud service providers collaborated to identify and publish select datasets of high value, ultimately publishing around 75 NOAA datasets to the cloud.

The initial research project demonstrated promising results. In an early analysis of the effect of hosting one major data dataset - Next Generation Weather Radar (NEXRAD), considered to be one of the most important observation systems - on one of the providers, it was shown that user access through the cloud service provider increased substantially in the months following the transfer. Before its migration to the cloud the NEXRAD dataset had been extremely difficult to share due to size and limitations on bandwidth.[12] In addition to increased data access and higher levels of service to users, the reduction of load on NOAA's systems is an added benefit of migrating data management systems onto the cloud.[13] It is thought that similar benefits can

---

https://oceanconservancy.org/blog/2019/11/12/build-partnerships-ocean-science-technology/, accessed 16 January 2020.

[9] See: Regional Data Platform Scoping Study: Federal Data Task Report, Dewberry Engineers, NOAA OCM, BOEM, 19 October 2018 and Regional Data Sharing Network Meeting Report (draft), NOAA Office for Coastal Management, 7 January 2020.

[10] See The **Foundations for Evidence-Based Policymaking Act** (**Evidence Act**; Pub. L. 115–435), https://www.performance.gov/CAP/leveragingdata/, and https://strategy.data.gov/ for more on government-wide open data efforts.

[11] Steve Ansari, Stephen Del Greco, Edward Kearns, Otis Brown, Scott Wilkins, Mohan Ramamurthy, Jeff Weber, Ryan May, Jed Sundwall, Jeff Layton, Ariel Gold, Adam Pasch, and Valliappa Lakshmanan, Unlocking the Potential of Nexrad Data through NOAA's Big Data Partnership (2018), *Bull. Amer. Meteor*. Soc.,99, 189-190.

[12] While the NEXRAD data was publicly available, the NCEI had to place limitations on time series or large spatial download of the data. Order sizes were limited to 250GB to accommodate limited bandwidth and web server saturation. It was possible to order the data offline at the option of 0.5TB per day and $753 per TB. Prior to BDP, NOAA estimates that to download the NEXRAD Level II archive containing 270TB of data, a single user would have had to pay $203,310 over 540 days. *See* Unlocking the Potential of Nexrad Data through NOAA's Big Data Partnership at pg. 194.

[13] Tiffany C. Vance, Micah Wengren, Eugene Burger, Debra Hernandez, Timothy Kearns, Encarni Medina-Lopez, Nazila Merati, Kevin O'Brien, Jon O'Neil, James T. Potemra, Richard P. Signell, Kyle Wilcox, From the Oceans to the Cloud: Opportunities and Challenges for Data, Models, Computation and

be realized for ocean data producers, managers, and users by incorporating more ocean data into the BDP.

The research phase of the project began in April 2015 under Cooperative Research and Development Agreements. In December of 2019, NOAA moved beyond the research phase and operationalized the BDP through agreements with Microsoft, Amazon Web Services, and Google Cloud.[14]

This paper outlines the current landscape of ocean data, with a particular focus on broad categories of data and how the U.S. federal government, focusing on NOAA, collects, manages, and distributes these data. It then explores the opportunities and challenges associated with efforts to better liberate and leverage ocean data for management decisions, highlighting potential solutions to some of those challenges and recommended paths forward.

Ocean Data Categories

The universe of ocean data is broad, crossing numerous geopolitical boundaries and scientific disciplines. While there is no definitive system for categorizing ocean data, most major types of ocean data can be addressed through four categories. Biological, physical, and chemical and biogeochemical data are all part of a broader set of oceanographic data. These data cover everything from the various organisms that live in the ocean, to the physical properties and processes of the ocean, to the chemical makeup of ocean waters. The fourth category, socio-economic data, relates to Blue Economy uses of ocean data and integrates the other data categories often spatially into decision making processes around ocean planning and human use activity.

Collection and use of all data types happens across a number of U.S. agencies, often driven by statutory mandates that may specify the content, format, and sharing requirements. Federal agencies that collect and use ocean data include NOAA, Bureau of Ocean Energy Management, U.S. Coast Guard, U.S. Navy, Environmental Protection Agency, National Science Foundation, U.S. Army Corps of Engineers, Bureau of Transportation Statistics, and National Aeronautics and Space Administration. The discussion below focuses on government data programs, but many industries collect ocean data for internal purposes that may be shared, in part, with the government through the regulatory process and could be valuable to the broader community. For example, the oil and gas industry has a long history of investing substantial resources into data collection throughout the life cycle of an oil field, from exploration to decommissioning. The renewable energy industry - specifically offshore wind farms - is also emerging as a potential source of oceanographic and geological data that can serve public

---

Workflows (2019), Frontiers in Marine Science,*Front. Mar. Sci.* 6:211,12 https://www.frontiersin.org/articles/10.3389/fmars.2019.00211/full at pg. 3.

[14] NOAA Media Release (December 19, 2019), Cloud platforms unleash full potential of NOAA's environmental data, accessed 26 January 2020.

purposes.[15] We discuss data sharing partnerships - and opportunities to enhance them - later in this paper.

Biological

Biological ocean data applies specifically to marine organisms and how they interact with the ocean environment. These data can be used to track and protect endangered species; achieve, maintain and expand sustainable fisheries; boost ecosystem health; and more.[16]

Biological data have traditionally been collected and managed separately from other types of oceanographic data. This has led to some data interoperability challenges - including incompatible data standards and different data formats - that will be discussed later in this paper.[17] In general, there is a substantial delay in biological data publication (often up to 5 years) due to the processing difficulties associated with identification of samples and consultation of experts. Except for aquatic telemetry, systems enabling automatic identification and sampling are at the beginning stages of development. However, standardization in marine biological data through the Darwin Core standard has enabled structured information on sampling protocols and events making it possible for users to model population monitoring, simultaneous counting, and capture-recapture schemes.[18]

While most observing infrastructure programs include physical sensors, the use of biological sensors are more rare. There is concerted international efforts directed at making progress in biological observations.[19] Biological data already appear in many data sources managed by the U.S. federal government - including the Integrated Ocean Observation System (IOOS) - as well as regional and international platforms like the Ocean Biogeographic Information System (OBIS). These data, on their own or in combination with data from other disciplines, can be used

[15] U.S. Department of Energy and U.S,. Department of the Interior, *National Offshore Wind Strategy (2016)* https://www.boem.gov/sites/default/files/renewable-energy-program/National-Offshore-Wind-Strategy-report-09082016.pdf

[16] Science and Data, NOAA Fisheries, https://www.fisheries.noaa.gov/science-and-data, accessed 10 January 2020.

[17] Ocean FAIR Data Services, 5.

[18] Ocean FAIR Data Services,8-9.

[19] Frank E. Muller-Karger, Patricia Miloslavich, Nicholas J. Bax, Samantha Simmons, Mark J. Costello, Isabel Sousa Pinto, Gabrielle Canonico, Woody Turner, Michael Gill, Enrique Montes, Benjamin D. Best, Jay Pearlman, Patrick Halpin, Daniel Dunn, Abigail Benson, Corinne S. Martin, Lauren V. Weatherdon, Ward Appeltans, Pieter Provoost, Eduardo Klein, Christopher R. Kelble, Robert J. Miller, Francisco P. Chavez, Katrin Iken, Sanae Chiba, David Obura, Laetitia M. Navarro, Henrique M. Pereira, Valerie Allain, Sonia Batten, Lisandro Benedetti-Checchi, J. Emmett Duffy, Raphael M. Kudela, Lisa-Maria Rebelo, Yunne Shin and Gary Geller, Advancing Marine Biological Observations and Data Requirements of the Complementary Essential Ocean Variables (EOVs) and Essential Biodiversity Variables (EBVs) Frameworks Front. Mar. Sci. 5:211. doi: 10.3389/fmars.2018.00211

to assess marine animals' habitat use, changes in migratory patterns due to deoxygenation, warming ocean temperatures, energy industry activity, and much more.[20]

Use Case: When the New England Fishery Management Council (NEFMC) was considering new coral management areas in an effort to protect deep-sea corals that live in the waters off New England while balancing the needs of the fishing industry, they used data to inform fishermen, other stakeholders, and Council members in an effort to provide detailed maps and obtain useful feedback on their proposals. The corals provide a habitat for numerous fish and invertebrates. Protecting this habitat can have positive effects on the ecosystem and provide benefits to the commercial fishing industry. The NEFMC presented proposed management areas, overlayed with data on fishing vessel activity, on the Northeast Ocean Data Portal which also allowed the NEFMC to incorporate more public feedback and create a more useful final product.[21]

Physical

Physical ocean data represents the "physical properties and dynamic processes of the oceans," including how the ocean interacts with the atmosphere, ocean temperature, currents, coastal dynamics, and more.[22] These data are captured through variables including sea surface temperature, subsurface temperature, surface currents, sea surface salinity, subsurface salinity, ocean surface heat flux, sea state, ocean surface stress, and sea ice.[23]

For example, sea surface temperature (SST) is used extensively in weather prediction models such as forecasting the *El Niño-Southern Oscillation* cycle and its associated effects on weather patterns, ocean conditions, and marine fisheries. SST measurements are collected through different types of sensors as well as through a sustained operational stream of satellite imagery data.[24] SST is a key indicator in understanding marine ecosystem fluctuations as the growth and

[20] Animal Telemetry Network, Integrated Ocean Observing System, https://ioos.noaa.gov/project/atn/, accessed 10 January 2020.
[21] Case Study: Balancing Deep-Sea Coral Protection and Commercial Fisheries, Northeast Ocean Data Portal, accessed 14 January 2020, https://www.northeastoceandata.org/case-studies/balancing-deep-sea-coral-protection-and-commercial-fisheries/
[22] Physical Ocean, NASA Science, Research and Analysis Program, https://science.nasa.gov/earth-science/oceanography/physical-ocean, accessed 10 January 2020.
[23] Global Ocean Observing System, Essential Observing Variables. Available at https://www.goosocean.org/index.php?option=com_content&view=article&id=170&Itemid=114
[24] Anne O'Carroll, Edward Armstrong, Helen Beggs, Marouan Bouali, Kenneth Casey,Gary Corlett, Prasanjit Dash, Craig Donlon, Chelle Gentemann, Jacob Høyer, Alexander Ignatov, Kamila Kabobah, Misako Kachi, Yukio Kurihara, Ionna Karagali, Eileen Maturi, Christopher Merchant, Salvatore Marullo, Peter Minnett, Matthew Pennybacker, Balaji Ramakrishnan, RAAJ Ramsankaran, Rosalia Santoleri,

reproduction of many species depend on their thermal tolerance. A predicted increase in SST over the next century may result in poleward migration of fish species and have a profound effect on marine ecosystems. [25] Scientists also predict that, thanks to new forecasting techniques made possible by machine learning, it will soon be possible to combine SST with data like sea surface salinity for improved sub-seasonal and seasonal forecasts.[26] The ability to conduct this sort of extended time-scale forecasting will lead to better drought planning as well as reduced weather-related human and economic losses.[27]

Use Case: NOAA Fisheries uses SST measurements in a product called TurtleWatch that provides up-to-date information to prevent bycatch of loggerhead sea turtles, an endangered species. Fishermen on longline fishing vessels pursuing swordfish in the Pacific Ocean north of the Hawaiian islands often are able to use this information to avoid catching turtles by mistake. In this way, SST measurements are deployed in dynamic fisheries management.[28]

Use Case: The U.S. Coast Guard (USCG) uses physical ocean data on the speed and direction of ocean surface currents collected using IOOS' high frequency (HF) radar network to predict where oil or other material may flow during a disaster situation, track water quality at local beaches, and more. In 2009, the USCG began using high frequency radar data from IOOS to help it carry out search and rescue operations. USGS initially made these valuable data available in the mid-Atlantic region, before eventually providing them to all of USCG's search and rescue teams through its Environmental Data Server (EDS).[29]

Swathy Sunder, Stephane Saux Picart, Jorge Vázquez-Cuervo and Werenfrid Wimmer (2019) Observational Needs of Sea Surface Temperature. *Front. Mar. Sci.* 6:420. doi: 10.3389/fmars.2019.00420

[25] Michael Alexander, James D. Scott, Kevin D. Friedland, Katherine Mills, Janet Nye, Andrew Pershing, Andrew Thomas, Projected sea surface temperatures over the 21st century: Changes in the mean, variability and extremes for large marine ecosystem regions of Northern Oceans (2018). Elem Sci Anth, 6: 9. DOI: https://doi.org/10.1525/elementa.191

[26] Robert Weller, James Baker, Mary Glackin, Susan Roberts, Raymond W. Schmitt, Emily Twigg, and Daniel Vimont (2019) The Challenge of Sustaining Ocean Observations. *Front. Mar. Sci.* 6:105. doi: 10.3389/fmars.2019.00105

[27] National Academies of Sciences, Engineering and Medicine (NASEM) (2016). *Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts*. Washington, DC: The National Academic Press.

[28] Turtle Watch, NOAA Fisheries, https://www.fisheries.noaa.gov/resource/map/turtlewatch, accessed April 6, 2020

[29] HF Radar, Integrated Ocean Observing System, https://ioos.noaa.gov/project/hf-radar/, accessed 10 January 2020. See also, High-Frequency Radar: Supporting Critical Coastal Operations with Real-time Surface Current Data, The COMET Program/MetEd, https://www.youtube.com/watch?v=Ii83ob2cwhE&feature=youtu.be, 19 January 2017

Chemical and Biogeochemical

Chemical ocean data relates to the chemical makeup, processes, and cycles of ocean waters as well as how seawater interacts with the atmosphere and the seafloor.[30] Biogeochemical data relates to the cycling of nutrients from the biotic environment, or biosphere (i.e., living organisms) to the abiotic environment, which includes the atmosphere, lithosphere, and hydrosphere, and vice-versa.

The ocean's role as a major carbon sink is a particularly relevant lens through which to view ocean data on chemical and biogeochemical variables. The deep ocean in particular potentially plays a critical role in the global carbon cycle by acting as a long-term reservoir of most of the earth's carbon and inorganic nutrients. The deep ocean acts as a biological pump to move energy from sunlight into the ocean by converting it into carbon dioxide, although this pump may be at risk due to broader changes in ocean conditions now occurring.

In order to assess the biological pump and its ability to contain carbon dioxide from industrial and other emissions, it is important to measure how carbon is transferred from living organisms and sequestered in deep ocean waters and sediments. Some of the chemical data needed to understand this process include variables like oxygen, nutrients, inorganic carbon, transient tracers, particulate matter, nitrous oxide, stable carbon isotopes, dissolved organic carbon, and ocean color.[31] These observations should be collected and analyzed over time with sufficient frequency to capture patterns on subseasonal, seasonal, and even longer timescales.[32]

Geological observations overlap with physical and chemical data, but are specifically related to the ocean seafloor. By sampling these data, scientists are able to glean insights on seafloor spreading, plate tectonics, volcanic processes, magma genesis, and other phenomena.

Use Case:
Ocean acidification occurs as a result of increasing absorption of atmospheric carbon dioxide. It is measured by a decrease in pH levels of seawater and can adversely affect coral reefs, marine plankton, and survival of larval marine species. Ocean acidification has a real impact on the marine shellfish industry because of the reduced growth of certain species or slowing down of calcification of shellfish. The University of Washington and the Pacific Coast Shellfish Growers Association have collaborated to disseminate ocean acidification data through the Northwest Association of Networked Ocean Observing System (NANOOS) web portal.

---

[30] What does an oceanographer do? NOAA National Ocean Service, https://oceanservice.noaa.gov/facts/oceanographer.html, accessed 16 January 2020
[31] Global Ocean Observing System, GOOS-EOV Specification Sheets, http://goosocean.org/index.php?option=com_oe&task=viewDoclistRecord&doclistID=168
[32] The Challenge of Sustaining Ocean Observations, 6.

Shellfish hatcheries in Washington State and elsewhere are able to use this real-time data to improve oyster production in the context of changing water conditions.[33]

Socio-Economic

The overall Blue Economy could be worth $3 trillion and employ 40 million people around the world by 2030.[34] Socio-economic indicators for ocean based industries include turnover, employment, exports, number of enterprises, density, poverty, and unemployment rates. This category includes data from and about ocean-based industries like shipping, fishing, and offshore renewable energy production as well as the ocean's own natural resources and ecosystem service benefits which include fish, carbon dioxide sequestration, and more. It is not limited to socio-economic indicators, but can also include data from the previously described oceanographic categories which can be displayed spatially and used for planning and in other ocean management decisions.

At the federal level, NOAA and the Bureau of Ocean Energy Management (BOEM) developed the Marine Cadastre to share data to "meet the needs of the offshore energy and marine planning communities."[35] The Cadastre has developed over time to include data that can be used for planning as well as a wide range of other areas and has produced a number of clear use cases. Data hosted on the Marine Cadastre have been used for projects ranging from an effort to understand how vessel noise impacts marine mammals to a state level offshore wind energy development project.[36] Additionally, NOAA's ENOW Explorer provides access to employment data and other economic information for American counties that border the ocean and great lakes.[37]

Use Case: Data on ocean infrastructure and economic activity are used to plan new development and reduce potential conflicts among multiple ocean uses. For example the American Waterway Operators (AWO), a trade association representing the tug and barge industry, is responding to increasing traffic and development by using data from the

---

[33] Monitoring and Adaptation to Ocean Acidification in the Shellfish Industry, Washington Ocean Acidification Center, https://environment.uw.edu/wp-content/uploads/2015/02/Pages-from-2015_0129_WOAC_one-pagers_MonitorinaAdaptation_FINAL.pdf, accessed 30 January 2020.

[34] Ralph Rayner, Claire Jolly and Carl Gouldman, Ocean Observing and the Blue Economy, *Frontiers in Marine Science*, (2019) *Front. Mar. Sci.* 6:330. doi: 10.3389/fmars.2019.00330 , 1-3.

[35] Bureau of Ocean Energy Management and National Oceanic and Atmospheric Administration (BOEM/NOAA). (2020). Marine Cadastre. NOAA Office for Coastal Management. https://marinecadastre.gov/. Accessed 14 January, 2020.

[36] Uses, Marine Cadastre, https://marinecadastre.gov/uses/, accessed 14 January 2020.

[37] ENOW Explorer, NOAA Office for Coastal Management, https://coast.noaa.gov/enowexplorer/#/, accessed April 4, 2020

Mid-Atlantic Ocean Data Portal to plan routes that avoid new offshore wind energy sites and other potential obstacles.[38]

From an ocean data management perspective, these four categories present a number of interesting opportunities and challenges that will be explored later in the paper, specifically as they apply to inconsistent data standards and formats across categories, data silos, and data sharing. Overcoming these challenges will make it possible to derive significant value from combining these data sources to find cross-category insights.

Mapping the Ocean Data Landscape

Technological improvements in the ocean observing infrastructure over the last few decades have increased the scale of ocean observations. The collection of data on our coasts, fisheries, and deep seas has evolved to include a wide variety of technologically advanced ocean observation systems. More recent observation systems include underwater cables with fixed-point ocean observation infrastructure, manned submersibles, autonomous underwater vehicles, and more.

NOAA maintains key pieces of the ocean observation infrastructure. The IOOS - a 'national regional' partnership led by NOAA - is a critical component.[39]

Other key pieces of federal infrastructure include:

- NOAA's research fleet, including sixteen large oceanographic vessels and more than 400 small boats.[40]
- The National Data Buoy Center (NDBC), which distributes on-site and remote observational data from NOAA's network of buoys, piers, bottom-mounted sensors, and volunteer observational ships.
- The National Centers for Environmental Information (NCEI), which host archival data. The NOAA OneStop project was developed to search NCEI data.[41]

The table below presents a sample of federal and regional ocean data products and elements of key NOAA and regional organizational infrastructure along with information about who manages

[38] Portal use example: Charting a course for the tug and barge industry, Mid-Atlantic Ocean Data Portal, accessed 14 January 2020, http://portal.midatlanticocean.org/documents/6/tugs_and_barges.pdf
[39] Governance Milestones and Lessons From Two Decades of Growth, 9.
[40] Office of Marine and Aviation Operations, https://www.omao.noaa.gov/learn/marine-operations/ships, accessed 30 January 2020.
[41] Peng, G., Milan, A., Ritchey, N.A., Partee II, R.P., Zinn, S., McQuinn, E., Casey, K.S., Lemieux III, P., Ionin, R., Jones, P., Jakositz, A. and Collins, D., 2019. Practical Application of a Data Stewardship Maturity Matrix for the NOAA *OneStop* Project. *Data Science Journal*, 18(1), p.41. DOI: http://doi.org/10.5334/dsj-2019-041

them, what categories of data they include, and some of their key data sources. It is not a comprehensive presentation, but serves to highlight the complexity of the ocean data ecosystem. This complexity causes a number of challenges. For example, multiple agencies within NOAA currently produce similar data which makes it difficult for end users to collect and collate disparate datasets or evaluate their quality.[42] These challenges - which will be explored further later on in this paper - make it clear that more coordination and collaboration is needed across the ocean data ecosystem.

[42] Key Takeaways, 6.

Name	Controlling Organization(s)	Data Categories/ Products	Key Data Sources
Marine Cadastre	NOAA and BOEM	Biological and physical oceanographic variables, jurisdiction and boundaries, ocean uses and planning areas, physical and oceanographic.	Federal agencies like NOAA, BOEM, Bureau of Indian Affairs, U.S. Geological Survey, EPA, Department of Energy, academic and research institutions like Duke University, University of New Hampshire, regional portals like the Northeast Ocean Data Portal.
National Data Buoy Center	NOAA	Mostly physical oceanographic data like ocean currents, salinity, sea level pressure, water temperature etc.	Buoys maintained by NOAA and partners such as the IOOS regional associations, oil and gas companies, academic institutions and others.
Comprehensive Large Array-Data Stewardship System	NOAA	Environmental data for land, ocean, and atmospheric applications	NOAA and Department of Defense (DoD) satellites.
National Center for Environmental Information	NOAA	Range of products across ocean data categories with archival or near real time data. Includes buoy data, satellite data products, international projects like the global Argo network, World Ocean Database.	Federal agencies including unclassified data from the DoD, state and local governments, regional portals, private sector. Foreign data through direct bilateral exchanges with other countries and organizations, and through the facilities of the World Data System for Oceanography.
Earth Observing System Data and Information System (EOSDIS)	NASA	Physical and chemical oceanographic variables.	Satellites, aircraft, field measurements, and various other programs from various EOSDIS data centers such as

			Alaska Satellite Facility, Global Hydrology Resource Center, National Snow and Ice Data Center among others.
Integrated Ocean Observing System Regional Associations	NOAA's Integrated Ocean Observing System. Members of regional associations include academic and research institutes, non-governmental organizations, government agencies, and industry.	Real-time observations, models and forecasts, and archival data (to a lesser extent) across different oceanographic categories. Custom made data products reflect the priorities of each regional association.	Platforms and stations maintained by the regional association, NOAA and other federal agencies, academic and research institutions, local and state government etc.
Regional ocean data portals or platforms	Regional Ocean Partnerships defined by region and coordinated when one or more coastal states come together to address ocean management challenges. Eg: Mid-Atlantic Regional Council on the Ocean, West Coast Ocean Partnership, Gulf of Mexico Alliance, and Northeast Regional Ocean Council.	Data made available through the different portals or platform vary depending on the management needs of the Regional Ocean Partnerships. For example, the Northeast Ocean Data Portal contains over 4,500 data layers on marine life, ecosystem function, and human activity for ocean resource management.	Organizations, federal and state agencies, industry, non-profit, IOOS regional associations and other data sources unique to regional needs. Data may be from federal data sources but tailored to scale for regional management needs.

Integrated Ocean Observing System and Regional Ocean Partnerships

The IOOS is a national-regional partnership that provides near-real time as well as archival ocean information. The origin of the IOOS is tied in with the institutionalization of the Global Ocean Observing System (GOOS) in the late 1990s. It was envisaged as a distributed system of regional observation and data management subsystems.[43] The IOOS features 11 regional associations that provide data portals and products that reflect unique regional needs while providing higher resolution observations to complement the federal system.[44]

---

[43] Governance Milestones and Lessons From Two Decades of Growth, 2.
[44] Governance Milestones and Lessons From Two Decades of Growth, 9.

Each regional association has its own governance structure that may include governmental agencies, research institutions, industry, and non-governmental organizations. The associations develop regional networks of ocean observing and data management infrastructure. They often present information through a data explorer which tracks real-time observations from sensor networks and associated archival datasets. These data come from a variety of sources, including federal and state governments. For example, the Alaska Ocean Observation System (AOOS) sources data from NOAA, USGS, the Department of Agriculture, and state government sources such as the Alaska Department of Fish and Game and the Alaska Department of Natural Resources.[45] In addition, the AOOS relies on model data developed at research centers at universities such as University of Alaska Fairbanks and the Alaska Pacific University.[46]

Academic and research institutions play a significant role in the regional associations. For example, the University of Maine maintains the observing systems for the Northeastern Regional Association of Coastal Ocean Observing Systems (NERACOOS). The Gulf of Maine Research Institute has developed the data portal and products for NERACOOS. Similarly, the University of Washington, Oregon State University, and the Oregon Health and Science University have developed the data explorer for the Pacific Northwest Association of Networked Ocean Observing Systems.

Regional Ocean Partnerships (ROPs) are regional organizations convened by governors in collaboration with federal and tribal governments and stakeholders to address ocean and coastal management issues unique to each region. ROPs and regional ocean data portals were recognized in the federal ocean policy.[47] ROPs include the Northeast Regional Ocean Council (NROC), the Mid-Atlantic Regional Council on the Ocean (MARCO), the Gulf of Mexico Alliance (GOMA) and the West Coast Ocean Alliance (WCOA).

Each of these partnerships has identified different strategic issue areas and management challenges based on regional needs and economic interests. NROC, for example, has prioritized ocean and coastal ecosystem health, coastal hazards resilience, and ocean data and planning. MARCO has prioritized climate change adaptation, marine habitats, renewable energy, water quality, and ocean data and planning. GOMA focuses on coastal resilience, data and monitoring, habitat resources, wildlife and fisheries, and ecosystem services. WCOA provides a forum for dialogue on common ocean management priorities including compatible ocean uses, ocean and coastal data, transparent decision-making, and tribal rights, knowledge, and resources.

---

[45] AOOS Federal source Data Inventory (June 30, 2016), http://aoos.org/wp-content/uploads/2016/08/Appendix-B-AOOS-Federal-Source-Data-Inventory.pdf, accessed 16 January 2020.
[46] AOOS Regional Data Stream Inventory (June 30, 2016), http://aoos.org/wp-content/uploads/2016/08/APPENDIX-E-AOOS-regional-streams.pdf, accessed 16 January 2020.
[47] Regional Ocean Partnerships, Bureau of Ocean Energy Management, Department of the Interior, https://www.boem.gov/environment/regional-ocean-partnerships, accessed 17 January 2020.

Each ROP prioritizes ocean and coastal data, however, access, approach, and volume of data vary. For example, the Northeast Regional Ocean Data Portal includes thematically organized data on marine life and habitat, commercial fishing, aquaculture, energy and infrastructure, and more. Data contained on the portal reflect NROC's strategic regional priorities for more than a decade. Many sources of data form part of the workflows for such portals (*see above*), including IOOS regional associations and government agencies. Since underlying workflows for these data portals involve multiple stages of rigorous subject matter review and quality control, there are opportunities for collaboration between federal agencies, the IOOS Regional Associations, and ROPs to collectively advance all interests.

Other sources:

Research centers and academic Institutions play a critical role in the ocean data ecosystem. In addition to maintaining ocean sensors and platforms that contribute to federal data platforms, they also play an important role in the IOOS regional associations and ROPs as described above. Additionally, advocacy groups and other non-governmental organizations play a role in spreading the use of ocean data through platforms such as Resource Watch, Fisheries Solutions Center, and Global Fishing Watch. Citizen science platforms like iNaturalist are crowdsourcing observations of marine life that could be integrated into larger assessments. Finally, private industry collects data for business related purposes and as part of specific projects. These data are often unavailable or siloed, but could be very useful if more widely shared.

Use Case: Global Fishing Watch. Global Fishing Watch is a unique collaboration between corporate and nonprofit partners to collect and present data assets about global fishing from multiple sources. They aggregate vessel tracking data from Automatic Identification System (AIS), Vessel Monitoring Systems (VMS), and other sources to track roughly 65,000 vessels with a 72 hour time delay.[48] Global Fishing Watch's platform has supported Argentina's effort to establish its first Marine Protected Areas (MPA's), helped Indonesia seize a notorious illegal fishing vessel, been used to flag potential illegal fishing activities in numerous jurisdictions, and provided data for dozens of published research articles.[49]

[48] Global Fishing Watch, https://globalfishingwatch.org/, accessed 29 January 2020.
[49] Michelle Winowatan, Andrew Young, and Stefaan G. Verhulst, A Data Collaborative Case Study: Global Fishing Watch, Pooling Data and Expertise to Combat Illegal Fishing (Gov Lab, January 2020) http://thegovlab.org/wordpress/wp-content/uploads/2020/01/Global-Fishing-Watch-Data-Collab-Case-Study-FINAL-3.pdf, 7-8

Use Case: iNaturalist. iNaturalist is an app and community that helps citizen scientists identify the plants and animals they see in the wild and share their knowledge and observations with a large network of scientists and naturalists.[50] iNaturalist data has been used for a variety of scientific purposes including to identify new ocean species. For example in 2019, as part of an annual "BioBlitz" event, graduate students from Northeastern University spent one day collecting and identifying new species in Friday Harbor, Washington. They deposited their samples to Northeastern's Ocean Genome Legacy (OGL) collection and uploaded them to iNaturalist for crowdsourced review and confirmation. In 2019, the project uploaded 60 samples, 25 of which represent new species in OGL's collection.[51]

Use Case: Gulf of Mexico Coastal Ocean Observing System and Corporate Data Providers. Integrating private sector data into open ocean data ecosystems is an ongoing issue. The Gulf of Mexico Coastal Ocean Observing System has had success ingesting data from energy companies including Shell, BP, Chevron, and others.[52] The companies collect data that, when combined with public data, can lead to improved models of hurricane intensity and other issues that are of great interest to regional stakeholders.[53]

Challenges

As ocean data stewards work to keep pace with this explosion of information and unleash the full potential of ocean data, they will face a range of familiar challenges that include:

- Funding and cost concerns,
- Varying levels of technical capacity among stakeholders,
- Integrating new data sources,
- Data processing,
- Challenges with respect to data interoperability,
- Lack of incentives for data sharing,
- Privacy and equity concerns,
- Confidentiality concerns,
- Domain and region specific data needs,

---

[50] About, iNaturalist, https://www.inaturalist.org/pages/about, accessed 29 January 2020.
[51] Hannah Appiah-Madson, Friday Harbor Bioblitz 2019: Northeastern students sample the marine biodiversity of the Pacific Northwest, Northeastern University Ocean Genome Legacy Center, 9 June 2019, https://www.northeastern.edu/ogl/friday-harbor-bioblitz-2019-northeastern-students-sample-the-marine-biodiversity-of-the-pacific-northwest/ accessed 29 January 2020
[52] Data Partners, Gulf of Mexico Coastal Ocean Observing System, https://gcoos.org/get-engaged/data-partners/, accessed 29 January 2020
[53] Interview with Josie Quintrell, Derrick Snowden, Kyle Wilcox, Tim Kearns, et. al. 23 January 2020

These challenges will be explored in further detail in this section. Potential solutions to these challenges along with specific recommendations to help solve them will be presented in the following section.

Funding and Cost Concerns:

Funding is a consistent challenge when it comes to data development and management over time and across complex domains, and the oceans data space is no different. Overall, federal funding for ocean observation initiatives has been below the levels recommended in an independent analysis conducted by NASA's Jet Propulsion Laboratory.[54]

Data management - which is often funded at much lower levels than data acquisition - is particularly impacted by the consistent underfunding of ocean observation initiatives. Data may be collected, but not processed or formatted for interoperability.[55] Additionally, as the volume of ocean data explodes, storage costs will become a more significant concern. Some earth observing organizations are already predicting that data will become "prohibitively expensive and complex to host within their own data centers." [56]

It may be easier to find funding for new initiatives or technologies, but ocean data stewards like the IOOS' have to strike a balance between investment in maintenance and storage and emerging technologies and opportunities.[57] Further, the IOOS regional associations have to balance investment in infrastructure for new data collection with spending on data management processes such as quality assessment and quality control. The IOOS and broader ocean data community should make it a priority to Identify new and innovative ways to partner across sectors to address these data management challenges. With good coordination, a single data collection activity can often address multiple issues.

Emerging cloud technologies provide new opportunities for long-term efficiency but may pose high short-term costs. Data stewards may ultimately achieve significant benefits by moving to the cloud through lower costs associated with capital improvements, storage, and ongoing maintenance as well as potentially increased security, reliability, and computing power. However, in the short term, cloud migration for ocean data can be seen as a cost-prohibitive option without a clear Return On Investment (ROI) or instructive use cases.[58] Even though cloud computing is expected to generate long-term cost savings, transitioning to the cloud might make more economic sense for smaller organizations than larger ones with highly efficient data

---

[54] The U.S. Integrated Ocean Observing System, 8.
[55] Ocean FAIR Data Services, 9.
[56] Tiffany C. Vance, Micah Wengren, Eugene Burger, Debra Hernandez, Timothy Kearns, Encarni Medina-Lopez, Nazila Merati, Kevin O'Brien, Jon O'Neil, James T. Potemra, Richard P. Signell, Kyle Wilcox, From the Oceans to the Cloud: Opportunities and Challenges for Data, Models, Computation and Workflows (2019), Frontiers in Marine Science,*Front. Mar. Sci.* 6:211,12 doi: 10.3389/fmars.2019.00211.
[57] U.S. IOOS Enterprise Strategic Plan 2018-2022, Integrated Ocean Observing System, https://cdn.ioos.noaa.gov/media/2018/02/US-IOOS-Enterprise-Strategic-Plan_v101_secure.pdf, 9.
[58] Opportunities and Challenges for Data, Models, Computation and Workflows, 4-8

centers.[59] Roundtable participants recognized the potential for smaller ocean modelling projects to realize cost savings by hosting data on the cloud instead of relying on supercomputers for all predictive forecasting.[60]

Additionally, simply accessing and using certain sources of ocean data can be cost-prohibitive for some stakeholders because of interoperability issues. For example, some Vessel Monitoring System (VMS) data is provided in formats that require further processing and significant investment by resource constrained end users to fully leverage.[61] VMS is atool to enforce fishery closures or other spatial management regulations by transmitting a vessel's location at mandated time intervals. In addition to its use for enforcement by NOAA, NMFS, US Coast Guard and fisheries councils,[62] the data from such systems can be used for scientific research and management. However, due to technical and other challenges, VMS data from U.S fishing vessels has only been used sparsely in research.[63]

The lack of adequate funding for federal ocean observation initiatives has a ripple effect that ultimately limits the impact of ocean data for researchers who collect the data, managers who process them, and stakeholders who use them.

Varying Levels of Technical Capacity Among Stakeholders:
Some data users cannot access or leverage data as easily as others. Scientists collecting data may lack the necessary technical skills to convert them into interoperable formats while potential end users may be intimidated when faced with raw data downloads. Accounting for these varying levels of technical capacity among stakeholders is a significant challenge for ocean data stewards working to make data more available to interested stakeholders.

These challenges can include a lack of technical capacity for data cleaning, software development, or standards adoption. For example, despite the advantages of the NetCDF format, many science groups lack the capacity to use it consistently.[64] This challenge is even more difficult for users who are not academics or researchers. More broadly, it is often difficult to encourage stakeholders to adopt common standards, due to a lack of understanding of their importance or lack of ability to adopt and use them. For example, researchers in the field have to run custom software routines - often multiple times - to convert and validate data sets.

---

[59] Patricia Moloney Figliola & Eric A. Fischer, Overview and Issues for Implementation of the Federal Cloud Computing Initiative: Implications for Federal Information Technology Reform Management, *Congressional Research Service* (January 2014).
[60] Ocean Data Roundtable, February 10, 20202
[61] Regional Data Platform Scoping Study: Federal Data Task Report, Dewberry Engineers, NOAA OCM, BOEM, 19 October 2018, C-10, C-22.
[62] NOAA, National Vessel Monitoring System Privacy Impact Assessment Statement (2012).
[63] Watson JT, Haynie AC. Using Vessel Monitoring System Data to Identify and Characterize Trips Made by Fishing Vessels in the United States North Pacific. *PLoS One*. 2016;11(10):e0165173. Published 2016 Oct 27. doi:10.1371/journal.pone.0165173

[64] Ocean FAIR Data Services, 8.

Researchers often lack the necessary technical skills and project data management plans rarely account for this complexity.[65]

Additionally, ocean data stewards at the regional level have described challenges associated with stakeholder engagement and the need to build products with and for user groups with varying levels of technical capacity or interest. For example, data stewards in the Pacific Northwest serve a variety of end users ranging from shellfish growers to surfers. These user groups have unique data needs.[66] The Northwest Association of Networked Ocean Observing Systems (NANOOS) has worked with their various stakeholders to develop a range of tailored apps, serving specific data sets to the users who need them most.[67] Similarly, the Carribean Coastal Ocean Observing System (CARICOOS) is focused on getting their data out to non-scientists who may face challenges finding the data or intimidated when faced with raw data.[68]

Other capacity challenges include spotty Internet access or lack of specialized software. For example, ocean data stakeholders in Alaska cited a wide variance in Internet bandwidth in different areas of the state as a major challenge for accessing large quantities of data. Meanwhile, data users in the Pacific Islands cited a lack of access to proprietary GIS mapping software and a need for open source data formats as a key challenge.[69]

Capacity issues can be addressed in a number of ways. Cloud technologies can limit the impact of spotty internet connections or low download speeds while more targeted funding for training, data management, and user engagement can help data stewards ensure interoperable data and tailor products directly to end user needs.

Integrating New Data Sources:

As ocean data stakeholders struggle to engage data users and overcome technical challenges, the volume and diversity of ocean data are skyrocketing thanks to the increasing availability and deployment of new ocean observing technologies.[70] These new technologies include autonomous sensors, Unmanned Aerial Systems (UAS), high-frequency radar, gliders, imaging bots, and eDNA species identification technologies. The technology is evolving on multiple fronts ranging from miniaturization of sensors and the evolution of power harvesting systems for the platforms to improved data transmission systems through better acoustic modems and

---

[65] Derrick Snowden, Vardis Tsontos, Nils Olav Handegard, Marcos Zarate, Kevin O' Brien, Kenneth Casey, Neville Smith, Helge Sagen, Kathleen Bailey, Mirtha Lewis, Sean Arms, Data Interoperability Between Elements of the Global Ocean Observing System, *Front. Mar. Sci.* 6:442, 11-12 doi: 10.3389/fmars.2019.00442.
[66] Interview with Jan Newton, Executive Director of the Pacific Northwest Association of Networked Ocean Observing Systems, March 18, 2020.
[67] Products, Pacific Northwest Association of Networked Ocean Observing Systems, http://www.nanoos.org/products/products.php, accessed March 19, 2020
[68] Interview with Julio Morell, March 12, 2020
[69] Regional Data Platform Scoping Study, C-12, C-33.
[70] Ocean Fair Data Services, 2.

fiber-optic cables.[71] At the same time, new technical approaches to data analysis are emerging, including cloud computing, big data analytics, machine learning, artificial intelligence, and more.[72]

<u>Use Case:</u> Scientists are increasingly using autonomous technologies like profiling floats, gliders, and Saildrones to complement data collected by ship-based researchers. These technologies are unlikely to replace ship-based research, but they can add additional context while saving money and time. These technologies are helping significantly expand the amount of data available related to carbon dioxide and other chemicals in the ocean. For example, NOAA's Pacific Marine Environmental Laboratory (PMEL) uses a Saildrone to take readings in the Arctic Ocean, providing a much broader picture of carbon dioxide levels there than would otherwise be available. Saildrones, which are essentially autonomous sailboats, can cover significantly more ground than larger manned ships and other autonomous vehicles - a significant advantage in the Arctic where unpredictable weather provides smaller time windows for scientists to conduct research than in other ocean areas.[73]

These new technologies present significant opportunities including the potential to improve existing data collections and add new ones in areas as diverse as water chemistry and human use.[74] However, they also come with challenges as ocean data stewards and users work to manage data flowing from these new technologies and ensure that they are interoperable with existing ocean data systems and standards.[75] Integrating old data with new, ongoing data flows could provide substantial benefits by expanding the scope and timescale of data for analysis, but doing so has proven difficult.

Additionally, data collected by certain industries could be better integrated into the larger ocean data ecosystem. Prime examples include oil and gas, offshore wind, and fisheries. There is strong user demand for data from these industries, but opening and integrating them into the larger data ecosystem will present challenges.[76]

For example, while there have been successful short-term collaborations between the oil and gas industry and academia to investigate important environmental questions, there are fewer examples of long-term partnerships. Overall, industry can justify sharing data if they have

---

[71] Technology, Data and New Models for Sustainably Managing Ocean Resources, pages 3-5
[72] Regional Data Sharing Network Meeting Report (draft), NOAA Office for Coastal Management, 7 January 2020, 12, 13.
[73] Sam Lemonick, Podcast: How Robots are Revolutionizing Chemical Oceanography, Chemical & Engineering News, 22 May 2019, https://cen.acs.org/environment/water/Podcast-robots-revolutionizing-chemical-oceanography/97/i21, accessed 17 January 2020.
[74] The U.S. Integrated Ocean Observing System, 14.
[75] Ocean Fair Data Services, 3.
[76] Roundtable Key Takeaways Document, 7

already been submitted as part of regulatory requirements or if companies are able to realize a tangible benefit like improved nautical or helicopter charts.[77] For example, energy companies operating in the Gulf of Mexico share oceanographic data with the regional IOOS association which help create better hurricane prediction models.[78] However, individual companies are hesitant to lose their competitive advantage or risk incurring reputational damage by sharing non-regulatory data. Another factor inhibiting data sharing include organizational cultures that are steeped in interia and that do not commit resources towards data sharing[79]

In comparison with other ocean data systems, fisheries systems remain relatively compartmentalized; fishermen often lack ready access to real-time oceanographic, market, and other data while at sea and scientists and managers tend to have limited access to timely information about what is caught and discarded.[80] Fisheries managers at the federal, state, and tribal levels will benefit from increased access to such data. Increased access to data will help fishermen comply with regulations, expand precision fishing, reduce business costs, and increase market power.[81]

Electronic monitoring (EM) and electronic reporting (ER) provide a concrete opportunity to improve the sustainability of fishery management by improving data reliability, accuracy, and sharing. In many EM programs, there is a significant time lag between when data are collected, often on hard drives on boats, and when they are reviewed by managers. Improvements can be achieved through updated technologies, such as improved sensor and camera capabilities that can capture clearer images of what is caught and discarded at sea as part of EM systems.[82]

As electronic monitoring and reporting programs develop, there is opportunity to use the data beyond enforcement and provide direct benefit to fishermen in the seafood marketplace or allow scientists to use the information for fishing and management options.[83] An example of how modern EM hardware and software can improve data collection is the collaboration between the Alaska Fisheries Science Center and NOAA's Fisheries Information System Program to automate video analysis of Pacific halibut discards from longline vessels and improve overall catch accounting .[84] Beyond EM and ER, increasing NOAA's ability to process stock assessment data and integrate findings to fisheries management contexts would significantly improve fisheries management.

---

[77] Data challenges and opportunities for environmental management of North Sea, page 135-137.
[78] Interview with Derrick Snowden, Kyle Wilcox, Tim Kearns, Tom Shyka, and Josie Quintrell, January 23, 2020
[79] Data challenges and opportunities for environmental management of North Sea, page 135-137.
[80] Environmental Defense Fund, *Smart boats and networked fishers: New paths to sustainable fishing in the digital age*, 9 https://www.edf.org/sites/default/files/documents/SmartBoatVision.March2019.web_.pdf, accessed 16 January 2020.
[81] Smart boats and networked fishers, 20.
[82] Smart boats and networked fishers, 13.
[83] Smart boats and networked fishers, 13.
[84] NOAA Fisheries, 9 April 2018, https://www.fisheries.noaa.gov/feature-story/expanding-electronic-monitoring-technologies-north-pacific-fisheries, accessed 31 January 2020.

Overall, integrating new data sources is possible, but will not happen consistently without efforts around data processing and data interoperability.

Data processing

Data processing is a key link in the ocean data management chain. Ocean data can be used to help make real-time or near real-time management decisions as well as to understand long-term trends and answer major scientific questions. Unfortunately, challenges with data processing can limit the utility of data that could otherwise be used for real-time and long-term decision making. Data that are not properly processed throughout their lifecycle may lack quality, live in non-interoperable formats, or have other faults that limit their utility to end users.

As discussed earlier in this section, research and other data gathering projects often underfund data management and processing and create data that does not align with widely used formats or standards. Additionally, policies and procedures around data collection and processing can be designed in ways that make real-time sharing impossible. EM systems, described above, provide an example of a policy designed for longer-term regulatory and oversight purposes that, with some processing tweaks, could also produce useful data for real-time decision making.

Finally, ocean data are often large and resource intensive to process. With larger volumes of data being collected every day institutions may face increasing difficulty with data processing and access.[85] For example, data generated from AIS systems could be leveraged for real-time decision making, but only with concerted inter-agency effort given its size and processing requirements.[86] Recognition of this need is in line with recent recommendations of the U.S. Committee on the Marine Transportation System, a federal interagency policy coordination committee on marine transportation systems. Cloud technology, including projects like the BDP, could better facilitate such efforts since AIS data typically needs high computing power and suffers from quality control challenges given the number of agencies involved. NOAA currently obtains AIS data from the U.S. Coast Guard and makes them available through the Marine Cadastre with significant time delays. Currently, data are available between the years 2009 and 2017 on the Marine Cadastre. Current processing time for any given year of AIS data is around 6 months in the following calendar year. The Task Force recommended that the processing periods be reduced to shorter timeframes i.e., 6 months, 1 month, with the ultimate goal of near real time availability. Reduced processing periods and enhanced accessibility would allow for safer, more efficient, and more environmentally responsible uses of U.S. waterways.[87]

---

[85] Oceans Roundtable Key Findings, Center for Open Data Enterprise, 6,8 (current draft, update to track with final draft)
[86] Ocean Data Roundtable, February 10, 2020
[87] U.S. Committee on the Marine Transportation System, *Enhancing Accessibility and Usability of Automatic Identification System Data Across the Federal Government and for the Benefit of Public Stakeholders,* (March 2019), https://www.cmts.gov/downloads/Accessibility_and_Usability_of_AIS_Data.pdf, 33, 5.

Adequate data processing will help alleviate challenges associated with interoperability that are particularly relevant to ocean data given the range of disciplines, collection technologies, and data types that are involved in the space..

Challenges with respect to data interoperability:

Data interoperability can be defined as the "*degree to which two or more systems, products or components can exchange information and use the information that is exchanged*".[88] Achieving interoperability in the context of ocean observations and data presents a unique challenge with multiple scientific disciplines and different types of sensors or platforms. Some of the necessary elements to achieve interoperability for ocean data include file standards, common data and metadata models, controlled vocabularies, and ontologies that define the terms and relationships.[89]

In the ocean data context, it is important that scientific data formats are 'self-describing' or formatted to include metadata that describes the data as well as the file structure. The adoption of self-describing file standards and common data models is not simple and requires significant resources. In the last two decades, the netCDF data model and Climate and Forecast (CF) conventions have emerged as the most widely used self-describing file formats. However, non-experts and under-resourced scientists often struggle to apply these formats to their own data. Additionally, losing metadata while transferring historical data to new formats and systems has been flagged as an ongoing challenge.

The need for controlled vocabulary for metadata arises from the fact that different datasets may use different terms to describe a variable (eg., salinity can be described as psal, salinity, Salinity, sal, etc.).[90] Controlled vocabularies are vital to ensure valid interpretation of values by human users and to enable correct compilation of datasets.[91] Ocean observations span different disciplines including physics, chemistry, biology, and geology, making it difficult to adopt metadata standards and vocabularies. Further, related science domain metadata frameworks use different metadata models to represent the exact same types of data and are not necessarily interoperable (for example, the biodiversity domain uses the Darwin Core standard).[92]

---

[88] ISO/IEC/IEEE. (2017). *24765:2017 Systems and Software Engineering – Vocabulary.* Geneva: ISO.
[89] Data Interoperability Between Elements of the Global Ocean Observing System, 5.
[90] Justin JH Buck, Scott Bainbridge, Eugene Burger, Alexandra Kraberg, Matthew Casari, Kenneth Casey, Louise Darroch, Joaquin Rio, Katja Metfies, Eric Delory, Philipp Fischer, Thomas Gardner, Ryan Heffernan, Simon Jirka, Alexandra Kokkinaki, Martina Loebl, Pier Buttigieg, Jay Pearlman and Ingo Schewe (2019) Ocean Data Product Integration Through Innovation-The Next Level of Data Interoperability. *Front. Mar. Sci.* 6:32. doi: 10.3389/fmars.2019.00032
[91] Data Interoperability Between Elements of the Global Ocean Observing System, 5.
[92] Data Interoperability Between Elements of the Global Ocean Observing System,13.

The implementation of data standards is relevant to both the platforms collecting observations as well as the portals which distribute data. Platform manufacturers do not necessarily build their products with data standards in mind so the underlying software does not necessarily produce data in commonly accepted standards.[93] On the distribution end, data management workflows for data portals need to support evolving data standards.

For example, ocean data are often presented and used either in real-time or across longer time series. Combining these data would have great value, but has proven difficult. For example ROPs tend to aggregate time series data while the IOOS regional associations focus more on bringing in standardized ocean observations in real-time. These aggregated time series datasets may have different metadata catalogs or lack interoperability with real-time data.[94] Another challenge that data portals managed by ROPs face are the technical challenges associated with integrating biological or ecological data standards into their existing metadata catalogs.[95]

Interoperability issues represent a technical hurdle to increased ocean data sharing, but cultural hurdles stand in the way as well. Specifically, there are currently limited incentives for researchers, private industry, and even individuals to share ocean data.

Lack of Incentives for Data Sharing:
A large amount of ocean data are collected by scientists in “research mode” working on specific projects with short-term funding. Additional data is collected by private industry, but withheld due to concerns over privacy confidentiality, and competitive advantage.

Scientists are highly incentivized to publish papers, but not necessarily to publish the corresponding data.[96] As a result of this lack of incentives, only a small portion of potentially available ocean data is actually used, with a much larger amount still trapped in notebooks and laptops.[97]

In many grant and funding programs, resources are provided to collect data but not to support the platforms that house data and make them more publically accessible. As a result, some data sharing platforms are under-resourced or outdated. Sometimes data does not even make it onto public access platforms. Efforts are already being made to address this problem, but could go further. For example, NOAA’s Public Access to Research Results (PARR) plan lays out a path to comply with requirements of a White House OSTP memo targeted at increasing access to the

---

[93] Data Interoperability Between Elements of the Global Ocean Observing System,12.
[94] Key Takeaways, 1o
[95] Interview with Andy Lanier.
[96] Ocean FAIR Data Services, 13.
[97] Linwood H Pendleton, Hawthorne Beyer, Estradivari, Susan O Grose, Ove Hoegh-Guldberg,Denis B Karcher, Emma Kennedy, Lyndon Llewellyn, Cecile Nys, Aurélie Shapiro, Rahul Jain, Katarzyna Kuc, Terry Leatherland, Kira O’Hainnin, Guillermo Olmedo, Lynette Seow, Mick Tarsel (2019) Disrupting Data Sharing for a Healthier Ocean, *ICES Journal of Marine Science* 76:6, https://doi.org/10.1093/icesjms/fsz068, 1416.

results of federally funded scientific research.[98] The OSTP is currently exploring updates to its original memo that could prove useful to future efforts in this area.[99]

Additionally, private industry collects ocean data but, with the exception of some data sharing arrangements, these data are not widely available.[100] For offshore oil and gas industries, there are identifiable opportunities to increase data sharing. In the Gulf of Mexico, private industry shares certain data to help improve hurricane prediction models.[101] This data sharing could be expanded through language in lease agreements requiring companies to share data throughout the time that they operate in an area. There have already been successful efforts to increase data sharing while reducing costs related to site decommissioning. Sharing of industry data in the United Kingdom on the shallow southern North Sea has helped companies develop better post-decommissioning monitoring programs. In the context of decommissioning, it is possible to build incentives to share data into regulation through relaxation of infrastructure removal requirements.

Further, while federal agencies and data platforms are collaborating more effectively around ocean resource management, work is still needed related to proactive data sharing. For example, the U.S. Army Corps of Engineers (USACE) has the capacity to provide advanced Automatic Identification System (AIS) analytics, which would allow for enhanced real-time navigation.[102] While these data are provided on request to other federal agencies, they are not made publicly available.[103]

Privacy and Equity Concerns:

Privacy and equity should be considered as part of any data sharing program.

On the privacy front, while government or academic oceanographic research data rarely captures personally sensitive information, issues can arise with smartphone applications and photos that include revealing information in metadata. Displaying the exact location of rare or

[98] Public Access to Research Results (PARR), NOAA, National Centers for Environmental Information, https://www.ngdc.noaa.gov/parr.html, accessed 4 February 2020

[99] Request for Information: Public Access to Peer-Reviewed Scholarly Publications, Data, and Code Resulting from Federally Funded Research, Federal Register Notice, White House Office of Science and Technology Policy, March 5, 2020, Accessed March 19, 2020, https://www.federalregister.gov/documents/2020/03/05/2020-04538/request-for-information-public-access-to-peer-reviewed-scholarly-publications-data-and-code

[100] Regional Data Platform Scoping Study, C-33.

[101] Interview with Derrick Snoweden, et. al., January 23, 2020

[102] AIS is a communication protocol that is intended as a situational awareness tool and a means to exchange navigation information in near real-time. In addition to an integrated system of AIS data maintained by the U.S. Coast Guard, federal agencies like the USACE use AIS data for navigation planning studies and enhanced reporting. Analytical tools developed by the USACE enable users to visualize vessel tracks, generate summary statistics of vessel activity, etc. See, Enhancing Accessibility and Usability of AIS Data, 19-20.

[103] Regional Data Platform Scoping Study: Federal Data Task Report, Dewberry Engineers, NOAA OCM, BOEM, 19 October 2018, C-22.

endangered species in real time could attract attention and put them at risk. Vessel Trip Reporting (VTR) data and Vessel Monitoring System (VMS) data pose privacy concerns as well. VMS are satellite based programs installed on vessels in a fishing fleet that track vessel movement in real time. Such data are collected and monitored by the Office of Law Enforcement at the National Marine Fisheries Service. VTRs help document catch information from licensed fishermen, but VTR data, released in aggregate to protect the identity of individual fishermen, are less useful to decision-makers at such a coarse resolution.

Learning from other sectors: The Center for Open Data Enterprise has explored strategies for protecting individual privacy while maximizing data sharing both broadly - across federal agencies - and in specific industries - healthcare. A number of best practices from other areas should be considered in the ocean data space, including:

- Balancing Tests - strategies to balance the risks of releasing data against the potential for public good. This is the approach taken by the Consumer Financial Protection Bureau in carrying out its mandate to release data under the Home Mortgage Disclosure Act, which can be used to show whether mortgage lenders are discriminating in their loans.
- Differential Access - gradations of openness under different circumstances. For example, some kinds of data could be made "open" only for sharing between federal agencies under certain conditions, or sharing only with qualified and vetted researchers.[104]
- Data Minimization - data collection and the amount of data used for any particular project is only what is necessary to accomplish the needed tasks.
- De-identification and anonymization - seek to remove sensitive personally identifiable information (PII) from individual-level and population-level data or otherwise make it difficult to identify the source of the data.[105]

Equity concerns come into play in two ways. First, open data programs are not necessarily designed to ensure that all stakeholders are able to benefit equally. There is always a risk that communities with more resources, better connections, or a longer history may gain advantage and gain the most from open data. More specifically, there are a number of underrepresented groups with a stake in the ocean data ecosystem, particularly indigenous communities with both traditional and ongoing interest in fisheries and coastal resources. It is important to design open data and data sharing programs that account for these potential inequities.

Confidentiality Concerns:

While there are some examples of private industry sharing ocean data with government, it is not a common practice as confidentiality concerns limit businesses' willingness to freely share the

[104] Open Data Roundtable on Privacy: Key Takeaways, Center for Open Data Enterprise, 2016, http://reports.opendataenterprise.org/KeyTakeawaysonOpenDataandPrivacy.pdf, 2

[105] Balancing Privacy with Health Data Access Roundtable Report, Center for Open Data Enterprise, September 2019, http://reports.opendataenterprise.org/RT2-Privacy-Report-Final.pdf, 14

information they collect. For example, locations and activities of individual fishing vessels may be protected as confidential business information and only released in aggregate, in anonymized formats, or with significant time delays. At the same time, there is growing willingness by fishermen to share data with respect to the physical variables like SST. Industry data submitted through the regulatory process, such as oil and gas lease bidding, may be treated as confidential information even if the lease is not secured. Unless there is an immediate economic gain or a regulatory requirement, companies find it difficult to justify sharing data with specific business value or that were expensive to acquire.[106]

However, both public and private stakeholders are deriving benefit from sharing some data. Private data can be combined with public data to improve models and derive insights, adding value for both stakeholders. For example, oil and gas companies drilling in the Gulf of Mexico share certain oceanographic data as part of their lease agreements.[107] This model has not yet been replicated in the offshore wind industry, which is less mature with most sites still in the research and development phase, but may be useful as that industry matures.[108] Additionally, existing data sharing agreements are not necessarily comprehensive. Even in the gulf, data on oceanographic variables like water temperature are available, however seismic data which are considered commercially sensitive are not. Successful models which have been implemented globally (for example, the Marine Data Exchange and the UKBenthos database) point to the possibility of implementing data sharing mechanisms with offshore energy companies.[109]

All of the issues highlighted above ensure that, while the universe of ocean data is large and growing, many valuable data aren't available for some or all stakeholders that may be interested in them. The next section will briefly lay out some of those data gaps and needs.

Domain and Region Specific Data Gaps and Needs:

Due to the nature of ocean data collection and management a variety of data gaps exist across regions in the U.S and even across states within those regions. Just a few examples include a variance in the amount of data on recreational and commercial fishing between the Northeast, Mid-Atlantic, and South-Atlantic regions, a lack of bathymetry data for the Gulf Coast states, and inconsistent and non-standardized data on species across states.[110]

---

[106] Fiona Murray, Katherine Needham, Kate Gormley, Sally Rouse, Joop Coolen, David Billett, Jennifer Dannheim, Silvana Birchenough, Kieran Hyder, Richard Heard, Joseph Ferris, Jan Holstein, Lea-Anne Henry, Oonagh McMeel, Jan-Bart Calewaert, J.Murray Roberts (2018). Data challenges and opportunities for environmental management of North Sea oil and gas decommissioning in an era of blue growth. Marine Policy, 97, 130, 133 https://doi.org/10.1016/j.marpol.2018.05.021.
[107] Data Portal, Gulf of Mexico Coastal Ocean Observing System, http://data.gcoos.org/fullView.php; National Data Buoy Center, https://www.ndbc.noaa.gov/
[108] Interview with Eoin Howlett, Executive Director, RPS North America, March 11, 20202
[109] ABPmer, (2015). A Review of Access to Industry Environmental Data.
A report produced by ABP Marine Environmental Research Ltd for Productive Seas Evidence Group, November 2015
[110] Regional Data Platform Scoping Study, 27, C-6, C-26.

Further, there are specific data needs and gaps across the different ocean data categories. For example, ecosystem and habitat data have been identified as lacking adequate standardization, preventing comparisons and analyses of the data across regions and federal agencies (although the Coastal and Marine Ecological Classification Standards may address these concerns). Meanwhile, while bathymetry data is generally available - with some gaps - it is located in various places across multiple agencies.[111]

Additionally, there are a number of known sources of data that are not currently available due to national security issues or concerns over competitive advantage. For example, the Department of Defense (DoD) collects data on bottom habitat, bathymetry, acoustic imagery, and more that they do not currently share.[112] Ocean data stakeholders also have a need for data that has been collected by private companies for commercial purposes but is not being shared more widely.[113]

Solutions/Paths Forward

While the challenges outlined above are numerous, many can be met through better collaboration, targeted funding, and the application of emerging technologies. The following section will lay a path towards overcoming the challenges while making specific recommendations for policymakers and data stewards.

New Data Sharing Agreements and Incentives:

While examples of data sharing between private industry and the public are relatively rare, there are examples globally that may prove instructive. In the international context, examples of scientific collaborations include the Deep-ocean Environmental Long-term Observatory System and the Lofoten-Vesterålen observatory which provide publicly accessible data.[114] The limited data sharing by the oil and gas industry described earlier in this paper may provide the most ready-made platform for future improvements in this area.

---

[111] Regional Data Sharing Network Meeting Report (draft), NOAA Office for Coastal Management, 7 January 2020, 8, 21.

[112] Regional Data Sharing Network Meeting Report (draft), 21.

[113] See: Regional Data Platform Scoping Study, C-33 and Regional Data Sharing Network Meeting Report (draft), 23.

[114] Lisa A. Levin, Brian J. Bett, Andrew R. Gates, Patrick Heimbach, Bruce M. Howe, Felix Janssen, Andrea McCurdy, Henry A. Ruhl, Paul Snelgrove, Karen I. Stocks, David Bailey, Simone Baumann-Pickering, Chris Beaverson, Mark C. Benfield, David J. Booth, Marina Carreiro-Silva, Ana Colaço, Marie C. Eblé, Ashley M. Fowler, Kristina M. Gjerde, Daniel O. B. Jones, K. Katsumata, Deborah Kelley, Nadine Le Bris, Alan P. Leonardi, Franck Lejzerowicz, Peter I. Macreadie, Dianne McLean, Fred Meitz, Telmo Morato, Amanda Netburn, Jan Pawlowski, Craig R. Smith, Song Sun, Hiroshi Uchida, Michael F. Vardaro, R. Venkatesan and Robert A. Weller (2019), Global Observing Needs in the Deep Ocean, *Front. Mar. Sci.* 6:241

Recommendation: BOEM and NOAA should collaborate on strengthening data sharing requirements in lease agreements with energy companies (including oil, gas, and wind) that want to operate offshore facilities in American waters. Companies should agree to share all non-proprietary oceanographic data from sea floor to sea surface from the start of their projects in exchange for a lease. This data should be made publicly available.

As new technology companies and NGOs enter the community, they may be able to facilitate multi-party licensing and access agreements across data creators and users in government, academia, and industry. Global Fishing Watch offers an example of a data sharing partnership between Google, Skytruth (an NGO), and country-level governing bodies.[115]

Collaborations between various regional entities may also represent useful models for sharing moving forward. While ecosystems vary by region, the challenges data stewards in those regions often share similarities, leading to effective collaborations.[116] For example, an ocean acidification data portal that was originally built for the West Coast is being expanded to cover the entire U.S.[117] Less formal partnerships and collaborations exist across regions as well. For example, CARICOOS has worked with ocean data stewards in Maine and New Jersey to access ocean observation technology and data. Formal collaboration should be expanded where possible and informal partnerships should be encouraged, expanded, formalized, and funded.

Recommendation: NOAA should collect existing examples of collaboration across regional entities, identify best practices, and use them as basis for more formal data sharing agreements and partnership structures. Existing data sharing agreements and MOU's should be updated with an eye towards replicability and/or scale to easily take advantage of emerging partnership or data sharing opportunities

Several data sharing agreements manage ongoing collaborations at the federal level. These include a Memorandum of Understanding (MOU) between NOAA and the National Science Foundation (NSF) as well as a long-term agreement between NOAA and the Department of the Interior's Bureau of Ocean Energy Management (BOEM) which helps power the Marine Cadastre.There are a number of opportunities for additional collaboration and extension of these types of agreements. Additionally, the Big Data Project could serve as a vehicle to ease data sharing across regions and agencies. For example, participants at a recent ocean data Roundtable identified the BDP's potential to help local and regional stakeholders access and

[115] Global Fishing Watch, Pooling Data and Expertise to Combat Illegal Fishing.
[116] Interview with Julio Morell, Executive Director, CARICOOS, March 12, 2020
[117] IOOS Partners Across Coasts Ocean Acidification, http://www.ipacoa.org/, accessed March 19, 2020

use ocean data. They also explored the potential for the BDP to serve as a catalyst to improve collaboration across federal agencies.[118]

Recommendation: NOAA should explore ways to coordinate with an existing partnership between BOEM and the U.S. Coast Guard to promote communications and harmonize standards. NOAA's Office of Coastal Management and the Coast Guard should collaborate to share and make better use of operational oceanography data.

Recommendation: The BDP should work with regional and local stakeholders to pilot a "digital sandbox" which would help enable collaboration across data silos, improve overall data interoperability between high value datasets (e.g. geographic and socioeconomic data), and enable faster and more accurate modeling. This sandbox could be funded through the Ocean Technology Transfer program.

Leveraging BDP to address existing Data Gaps and priorities:

The universe of ocean data available in the U.S. is robust, but there are a number of specific data gaps at the regional and national level that can and should be filled. Specifically, stakeholders from the various regional ocean platforms and portals recently (as part of a scoping study led by NOAA and BOEM) identified a number of high priority data sets in need of improvement.[119] Further, participants at the February 2020 Roundtable identified opportunities to use cloud computing to more efficiently manage and disseminate several key data sets.[120] Significant priority data categories that emerged across these conversations include:

- Abundance and distribution of marine species
- Commercial fishing effort - Vessel Monitoring System (VMS), Commercial fishing effort - Vessel Trip Report (VTR); and Vessel traffic - Automatic Identification System (AIS)[121] and
- Bathymetry

[118] Key Findings Report, Center for Open Data Enterprise, 8-9
[119] Regional Data Platform Scoping Study: Federal Data Task Report, 2-4. See Appendix IV for a more detailed description.
[120] Other data categories recognized at the Roundtable for potential hosting on the Big Data Project include acoustic doppler current profile, synthetic aperture radar derived high resolution data, imagery response data, glider data, real-time tide gauge, migratory fish and abundance of species, and drinking water quality.
[121] These categories were identified by the regions in response to the Administration's commitment to direct resources in FY 2020 towards four data themes: vessel traffic, marine species, fishing, and offshore infrastructure. Also see, Ocean Policy Committee, *Ocean Resource Management Subcommittee Implementation Plan to Increase Public Access to Marine Data and Information* (September 2019) https://www.whitehouse.gov/wp-content/uploads/2017/11/20191009-FINAL-ORM-Marine-Data-IP-Sep2019.pdf

NOAA's Big Data Project has been identified as a potential tool to address these data gaps and priorities. The BDP is a cloud-based public data dissemination service provided by three Cloud Service Providers (CSPs): Amazon Web Services (AWS), Google, and Microsoft. The BDP was designed to ease access to the terabytes of data produced by NOAA satellites, radar, ships, and weather models every single day.[122] The Big Data Project has already improved service for users of atmospheric data and satellite data,[123] but oceanic data are currently underrepresented on the platform. For example, the AWS open data registry currently has just two ocean-relevant data sets available.[124] The future viability of the Big Data Project was recently confirmed with NOAA awarding multiyear contracts to AWS, Google, and Microsoft.[125] Adding more ocean data to the Big Data Project could prove valuable as it moves into this new phase.

Recommendation: NOAA's Big Data Project team and the BDP Cloud Service Providers should work with data managers at the federal and regional levels to incorporate marine species, VMS, VTR, AIS, and bathymetry data into the BDP.

There are a number of ways in which the BDP could play a bigger role in dissemination of data. The BDP could help potential ocean data users who lack internet bandwidth to download huge datasets or server space to store them. By giving users access to the data and enabling them to analyze them in the cloud, these services make big data both more accessible and more computable. With a recent renewed focus on ocean exploration and mapping,[126] data collected through these efforts could prove valuable to add to the Big Data Project. The cloud can also host 'labelled' training data to improve predictive ability, a recognized priority in NOAA's recently released Artificial Intelligence Strategy. The strategy also recognizes the fact that standards in training data will inform technical guidelines that underscore the reliability and credibility of NOAA's AI products.

---

[122] Big Data Project, National Atmospheric and Oceanic Administration, https://www.noaa.gov/big-data-project, accessed January 16 2020.
[123] For example, following the hosting of the NEXRAD dataset on the BDP, users were able to access more than double the monthly maximum of data that NCEI was able to facilitate. *See*, Unlocking the Potential of Nexrad Data through NOAA's Big Data Partnership at pg. 198..
[124] Registry of Open Data on AWS, Amazon Web Services, https://registry.opendata.aws/, accessed 16 January 2020.
[125] Dave Nyczepir, NOAA partners with 3 big cloud providers to disperse environmental data, FedScoop, 23 December 2019, https://www.fedscoop.com/noaa-cloud-providers-environmental-data/
[126] President Donald J. Trump, Memorandum on Ocean Mapping of the United States Exclusive Economic Zone and the Shoreline and Nearshore of Alaska, the White House, 19 November 2019, https://www.whitehouse.gov/presidential-actions/memorandum-ocean-mapping-united-states-exclusive-economic-zone-shoreline-nearshore-alaska/

Recommendation: The BDP should serve as a central location for navigation of marine species observation data. Centralizing this data will enable the integration of state and federal data and provide a convenient location for training data that can help improve AI/ML tools deployed in unmanned systems for data collection.

Recommendation: The BDP should serve as a central repository of bathymetric data currently collected by various agencies in various file formats. Hosting on a single platform can facilitate standardization. This data should be integrated with data collected by the EEZ mapping project which will be discussed further in a later recommendation.

Additionally, it has been suggested that the BDP can be a platform to improve search capabilities by helping classify data on a temporal basis or the BDP could enable creation of standards conventions that are interoperable with commonly accepted CF conventions but are more end user-friendly.[127]

Recommendation: The BDP team and CSP's should work with the IOOS regional associations and ROP's to identify and engage data end users

Recommendation: The BDP team should work with NOAA and non-NOAA agencies like BOEM, NASA, and the U.S. Coast Guard to explore the possibility of integrating non-NOAA ocean data into the BDP.

Funding:

More funding would go a long way towards an improved ocean data ecosystem that can be leveraged for economic gain, environmental conservation, and more. This could take the form of overall funding boosts for NOAA programs that produce, use, or manage ocean data, and funding for specific projects that attempt to tackle some of the challenges identified in this paper.

Recommendation: In response to the Presidential Memorandum directing the Ocean Policy Committee and NOAA to develop a strategy to ensure complete mapping of the EEZ,[128] NOAA has sought dedicated funding of $8.51 million in its budget estimates for 2021.[129]

[127] Ocean Data Roundtable, February 10, 2020
[128] President Donald J. Trump, Memorandum on Ocean Mapping of the United States Exclusive Economic Zone and the Shoreline and Nearshore of Alaska, the White House, 19 November 2019, https://www.whitehouse.gov/presidential-actions/memorandum-ocean-mapping-united-states-exclusive-economic-zone-shoreline-nearshore-alaska/
[129] NOAA Budget Estimates Fiscal Year 2021.

Congress should fully fund this request as well as all future appropriations related to EEZ mapping. In line with best practices, at least 10% of this amount should be dedicated towards data management. Congress should also mandate that data from this effort be made publicly available.

Recommendation: Congress should appropriate funding to allow the BDP to serve as a host for the EEZ mapping project as well as other efforts to facilitate regional data sharing.

More broadly, funding for NOAA ocean observation and data stewardship activities should be increased, with specific attention paid to funding for data collection, metadata management, data storage, and the adoption and use of cloud technologies. Further, regional ocean associations recognize that there is an opportunity to integrate duplicative data collection processes. However, the associations lack the necessary funding required to achieve this goal.

Recommendation: Congress and NOAA should Increase funding for regional ocean data portals. This should include dedicated funding for projects related to data coordination between ROPs and IOOS associations. NOAA should pursue a pilot project to explore the feasibility of hosting real-time and time series data on IOOS and ROP websites.

Recommendation: Congress and NOAA should increase funding for grants to IOOS regional associations while tying awards to data management and integration programs.

Financial incentives may be appealing to the research community, but non-financial incentives could prove to be even more powerful. For example, scientists currently are not strongly incentivized or required by funders to follow data management best practices or share their data openly. They are heavily incentivized to publish and have their research cited by colleagues, but are not rewarded for having their data cited or reused in the same way. The Digital Object Identifier system, which provides infrastructure to register and use persistent interoperably identifiers on digital networks, may represent one path forward in this area.[130] Other federal agencies are also grappling with this issue. For example, the National Institutes of Health recently proposed a draft plan that would require all NIH grantees to share their scientific data and - as of April 2020 - the OSTP is reviewing their existing policy on increasing access to the results of federally funded scientific research.[131]

---

130 Digital Object Identifier System, the International DOI Foundation, https://www.doi.org/index.html, accessed 4 February 2020

131 Draft NIH Policy for Data Management and Sharing, National Institutes of Health Office of Science Policy, https://osp.od.nih.gov/wp-content/uploads/Draft_NIH_Policy_Data_Management_and_Sharing.pdf, November 2019, accessed 4 February 2020 and Request for Information: Public Access to

Recommendation: NOAA should update its PARR plan to align with updates to the OSTP policy on public access to peer-reviewed scholarly publications, data, and code resulting from federally funded research. Accountability measures in the PARR plan should be strengthened and noncompliant researchers should be barred or limited from receiving NOAA funds in the future. Data released through the PARR plan should be made publicly available - where there are no privacy or security restrictions on release - through an existing NOAA data portal.

Data Management:

Data management through dedicated staffing and project-based investment as well as large-scale data management efforts are specific areas that could use more funding. Dedicated staffing and budget would ensure that data management practices are not ignored at the point of collection while broader data management efforts will boost overall data quality and usability.

As a best practice, 5 to 10% of a scientific research project's budget should be dedicated to data management.[132] This should serve as minimum investment and apply across all projects funded by the federal government, not just grants for scientific research. Additionally, research projects as well as programs tasked with producing, managing, distributing, or using ocean data should be staffed with a full time data manager. That staffer should be involved throughout the project to ensure that data is being collected and managed with an eye towards reuse, interoperability, and long-term preservation.[133]

Recommendation: NOAA should review all of its data programs to ensure that at least 10% of project budgets are apportioned towards data management. Data programs should include budget for at least one full time data manager. Where necessary, NOAA and Congress should make policy changes to ensure that this goal is met.

More broadly, ocean data stakeholders should look for opportunities to participate in cross-enterprise data management efforts. Current efforts at adopting this approach are reflected in NASA's Common Metadata Repository (CMR) and NOAA's OneStop system, which are data management systems that integrate standards-compliant metadata across distributed data sources, enabling unified search and access to a wide range of scientific data.[134] The ability

Peer-Reviewed Scholarly publications, data, and code resulting from federally funded research, White House Office of Science and Technology Policy, February, 2020, accessed April 8, 2020, https://www.federalregister.gov/documents/2020/02/19/2020-03189/request-for-information-public-access-to-peer-reviewed-scholarly-publications-data-and-code

[132] Ocean FAIR Data Services, 9.

[133] Data Interoperability Between Elements of the Global Ocean Observing System, 10.

[134] Data Interoperability Between Elements of the Global Ocean Observing System, 7.

to improve search mechanisms for end users is expected to evolve with greater interoperability and emerging technology such as artificial intelligence. Despite multiple data portals by federal, regional, academic, and other organizations, end users often point to the lack of a single resource that can help them find data across these sources.

Leverage Existing Collaborative Structures and Explore New Opportunities

The ocean data ecosystem has robust and mature structures in place to handle regional to federal data sharing and collaboration. For example, the IOOS structure facilitates data flows between federal, regional and local stakeholders, provides opportunity for collaboration, and much more. Additionally, there are structures in place to help coordinate across federal agencies including the Interagency Ocean Observation Committee (IOOC) which includes 19 bureaus and 12 federal agencies and has co-chairs from NOAA, NASA, the National Science Foundation (NSF), and the White House Office of Science and Technology Policy (OSTP). The IOOC reports up to the Subcommittee on Ocean Science and Technology (SOST), which is itself a subcommittee of the President's Council on Science and Technology.. However, these structures do not always function as well as they could.[135] There are significant opportunities for improved collaboration among federal agencies, regional associations, vendors, scientists, and data users.[136] Other structures, like the National Oceanographic Partnership Program (NOPP), exist to bring together government and non- governmental stakeholders.[137] Increased collaboration can lead to increased data availability, better data quality, more widely adopted standards, sharing of best practices, and more.

To achieve the goal of increased collaboration existing structures should be reexamined to ensure that they are being properly leveraged and new structures may have to be developed. For example, there is no formal mechanism inside NOAA to coordinate the National Environmental Satellite, Data, and Information Service (NESDIS), the National Ocean Service (NOS), and the Office of Oceanic and Atmospheric Research (OAR), even though they work on similar issues and often coordinate informally.[138]

Recommendation: OSTP should conduct a review and inventory of all existing interagency bodies that serve as fora for agencies with responsibility for creating, collecting, processing, managing, distributing, or using ocean data to come together. Changes should be made, where necessary, to ensure that these fora are fulfilling their function and facilitating collaborative conversation and work.

[135] Table One Notes, Ocean Data Roundtable, February 10, 2020.
[136] Key Findings Report, Center for Open Data Enterprise, 7
[137] Summary of the 2019 White House Summit on Partnerships in Ocean Science and Technology, Ocean Policy Committee, November 2019, https://www.whitehouse.gov/wp-content/uploads/2019/12/Ocean-ST-Summit-Readout-Final.pdf, 5
[138] Table One Notes, Ocean Data Roundtable, February 10, 2020

Recommendation: NOAA should create a formal structure to coordinate the National Environmental Satellite, Data, and Information Service (NESDIS), the National Ocean Service (NOS), the Office of Oceanic and Atmospheric Research (OAR), and other NOAA offices with responsibility for data programs.

Recommendation: OSTP should explore opportunities to expand and strengthen existing partnership mechanisms that already bring together government, industry, academic, and philanthropic organizations like the NOPP.

There are also opportunities to expand data collection and sharing in specific ocean industries. For example, Fishery Management Councils (FMC) can work with the fishing community to expand EM and use geolocation and catch data more efficiently.

Recommendation: FMCs and the fishing community should explore opportunities to expand EM and use geolocation and catch data more efficiently

Additionally, new stakeholders need to be engaged in collaborative processes. For example, manufacturers of ocean observation technologies may only coordinate with their direct customers, leading to issues with data standards alignment and limiting data sharing. To ensure that data from new ocean observation technologies are fully integrated into the larger ocean data ecosystem, manufacturers of those instruments should be engaged with a wider array of stakeholders. These manufacturers can ensure that their instruments produce data natively in formats that are commonly used and metadata that conforms with commonly accepted standards.[139]

Recommendation: OSTP, NOAA, NIST, and relevant non-governmental stakeholders should explore the creation of a voluntary consensus standards or standards coordination process aimed at aligning standards across the ocean data space. This process should engage platform and sensor manufacturers to ensure that automatically generated data is aligned with common standards.

[139] Ocean Fair Data Services, 15-16.

Conclusion

TO BE WRITTEN

Appendix I: List of Acronyms

AIS Automatic Identification System
AOOS Alaska Ocean Observation System
AWS Amazon Web Services
AWO American Waterway Operators
BDP Big Data Project
BOEM Bureau of Ocean Energy Management
CEQ Council on Environmental Quality
CF Climate and Forecast
CMR Common Metadata Repository
CSP Cloud Service Provider
DOD Department of Defense
DOI Department of the Interior
EDS Environmental Data Server
EHR Electronic Health Record
EM Electronic Monitoring
ER Electronic Reporting
FEBPA Foundations for Evidence-Based Policymaking Act
GOMA Gulf of Mexico Alliance
GOOS Global Ocean Observing System
HIT Healthcare Information Technology
IOOS Integrated Ocean Observation System
MARCO Mid-Atlantic Regional Council on the Ocean
NANOOS Pacific Northwest Association of Networked Ocean Observing System
NDBC National Data Buoy Center
NCEI National Centers For Environmental Information
NEFMC New England Fishery Management Council
NERACOOS Northeastern Regional Association of Coastal Ocean Observing Systems
NESDIS National Environmental Satellite, Data, and Information Service
NEXRAD Next Generation Weather Radar
NGO non-governmental organization
NOAA National Oceanic and Atmospheric Administration
NROC Northeast Regional Ocean Council
OBIS Ocean Biogeographic Information System
OCADP Ocean Carbon and Acidification Data Portal
OSTP White House Office of Science and Technology Policy
PARR Public Access to Research Results

PMEL Pacific Marine Environmental Laboratory
ROP Regional Ocean Partnership
SST Sea surface temperature
TEK Traditional Ecological Knowledge
UAS Unmanned Aerial Systems
USACE U.S. Army Corps of Engineers
USCG The United States Coast Guard
VMS Vessel Monitoring System
VTR Vessel Trip Reporting
WCOA West Coast Ocean Alliance

Appendix II: Key Technical Issues and Concepts

Ocean data are collected, maintained, and shared using a range of technologies, formats, and standards depending on the scientific domain, type of data collection, and other factors. While the goal of this paper is not to conduct a technical assessment of the ocean data ecosystem, it is important to understand some of its major technical drivers.

Data interoperability can be defined as the "*degree to which two or more systems, products or components can exchange information and use the information that is exchanged*".[140] In the context of ocean observations and data, interoperability presents a unique challenge as the ecosystem involves multiple scientific disciplines and different types of sensors or platforms Specifically, there are a number of file standards, common data and metadata models, controlled vocabularies and ontologies, Web services, and network protocols for data exchange that undergird the ocean data ecosystem. They often lack interoperability.

Plain text and CSV formats are the simplest and have the benefit of being human-readable. However, even though they are widely used in biogeochemical and biological communities, they have very limited technical functionality because they lack proper structure and formatting. ASCII formats are also widely used with popular software like Ocean Data View (ODV).

NetCDF and the Climate and Forecast (CF) convention have emerged as the mostly widely used, "self-describing" file formats. They produce metadata that describes the data itself as well as the file's data structures. Their metadata describes what each variable represents, including physical units and the geospatial location of each value. The associated conventions for Climate and Forecast metadata were designed to promote the processing and sharing of files created with the netCDF API.[141]

Other sets of relevant specifications are the Open Geospatial Consortium's Sensor Web Enablement (SWE) and the OpenGIS Web Map Service Interface (WMS) standards. The standards define conceptual models, web services and XML encoding frameworks.[142]
The SWE standards are geared towards sensor or platform manufacturers and allow developers to make data acquired through these sources discoverable and accessible on the Web. The WMS standards enable an HTTP interface to request geo-registered maps from distributed geospatial databases.[143]

[140] ISO/IEC/IEEE. (2017). *24765:2017 Systems and Software Engineering – Vocabulary.* Geneva: ISO.
[141] NetCDF Climate and Forecast (CF) Metadata Conventions. Accessed 10 January 2020.
[142] Hankin et. al 2010
[143] Standards, Open Geospatial Consortium, https://www.opengeospatial.org/standardas/wms, accessed 10 January 2020.

NOAA developed the ERDDAP software which functions across various custom data servers. It brokers data between data centers that do not have separate dedicated infrastructure.[144] The Thematic Realtime Environmental Distributed Data Services (THREDDS) is another such server developed by Unidata that allows for access to datasets with different data servers or protocols.

[144] Ocean Data Product Integration Through Innovation-The Next Level of Data Interoperability, 11.

Appendix III: Data Portal Links and Information

Marine Cadastre
https://marinecadastre.gov/
The cadastre is an integrated marine information system that provides data, tools, and technical support for ocean planning. Originally designed to support renewable energy efforts on the outer continental shelf, it now supports other ocean management and conservation efforts.

National Data Buoy Center (NDBC)
https://www.ndbc.noaa.gov/
The NDBC is a unit of the National Weather Service's Office of Operational Systems in NOAA. It operates ocean observing networks of data collecting buoys and coastal stations.

Comprehensive Large Array-Data Stewardship System (CLASS)
https://www.bou.class.noaa.gov/saa/products/welcome
CLASS provides a repository of environmental data from a variety of ground-based and remotely-sensed observing systems. It is a multi-site system which ingests data from a number of satellites including the Geostationary Operational Environmental Satellites (GOES) and the Polar-orbiting Operational Environmental Satellites (POES). It also contains data from continuing operating reference stations and derived products.

National Center for Environmental Information (NCEI)
https://www.ncdc.noaa.gov/data-access
NCEI hosts and provides access to archival oceanic, atmospheric, and geophysical data. It maintains a number of datasets and portals including the World Ocean Database and the World Ocean Atlas.

Earth Observing System Data and Information System (EOSDIS)
https://earthdata.nasa.gov/
EOSDIS is an end-to-end data management system for NASA's earth science data from various sources such as satellites, aircraft, field measurements etc. It operates Distributed Active Archives Centers which produce and archive the earth science data products.

Integrated Ocean Observing System Regional Associations
The IOOS consists of 11 regional associations in addition to several federal agencies. Among other objectives, regional associations aim to develop and host data portals that integrate data from multiple sources and build tailored products specific to the unique characteristics of the region. The design of the portals and tools and nature of archival datasets published vary among the different regional associations.

- Alaska Ocean Observing System

https://portal.aoos.org/

- Caribbean Coastal Ocean Observing System

https://www.caricoos.org/

- Central and Northern California Ocean Observing System https://data.cencoos.org/

- Great Lakes Observing System

http://portal.glos.us/

- Gulf of Mexico Coastal Ocean Observing System

http://data.gcoos.org/

- Mid-Atlantic Regional Association Coastal Ocean Observing System

https://oceansmap.maracoos.org/

- Northeastern Regional Association of Coastal Ocean Observing Systems

http://www.neracoos.org/datatools

- Pacific Islands Ocean Observing System

http://www.pacioos.hawaii.edu/

- Pacific Northwest Association of Networked Ocean Observing Systems

http://nvs.nanoos.org/

- Southeast Coastal Ocean Observing Regional Association

https://portal.secoora.org/

- Southern California

http://sccoos.org/observations/

Regional Ocean Partnerships:

Regional Ocean Partnerships (ROPs) are regional organizations voluntarily convened by governors to address ocean and coastal issues of common concern among states and in collaboration with federal agencies, tribes, academic institutions, and ocean stakeholders. While many ROPs have worked to address ocean and coastal management challenges for over a decade, the federal ocean policy, Executive Order 13840 (July 2018), recognized the function of ROPs and their associated regional ocean data portals as providing interagency collaboration on cross-jurisdictional ocean and coastal matters. Several ROPs have developed ocean data portals to provide a common platform where spatial ocean data can be displayed for planning and resource management.

Data found on these platforms come from a variety of sources including federal agency data sources (like the Marine Cadastre), individual agencies, states, industry, IOOS Regional Associations, universities, and non-governmental entities. Platforms and data sets are unique to the needs of the states and region but often similar in the spatial nature of the data displayed. In some regions that do not currently have ROPs, the IOOS observing system is also used to provide information on resource management in addition to the other services the system provides. Those regions include: Alaska, Caribbean, Southeast, Great Lakes, and Pacific Islands.

Northeast Regional Ocean Council (NROC) – Maine, New Hampshire, Massachusetts, Rhode Island, Connecticut, Department of Interior, Environmental Protection Agency, National Oceanic and Atmospheric Administration, Department of Agriculture, Army Corps of Engineers, and Coast Guard
Northeast Ocean Data Portal: https://www.northeastoceandata.org/

Mid-Atlantic Regional Council on the Ocean (MARCO) – New York, New Jersey, Delaware, and Virginia
Mid-Atlantic Ocean Data Portal: https://portal.midatlanticocean.org/

Gulf of Mexico Alliance (GOMA) – Alabama, Florida, Louisiana, Mississippi, and Texas
Gulf of Mexico Alliance manages a number of tools and portals to support management:
https://gulfofmexicoalliance.org/our-priorities/priority-issue-teams/data-and-monitoring-team/

West Coast Alliance (WCOA) – California, Oregon, Washington, and 11 tribal governments
West Coast Ocean Data Portal: https://portal.westcoastoceans.org/

Appendix IV: High Priority Data Categories and Needed Improvements

These categories and improvements were identified in the Regional Data Platform Scoping Study: Federal Data Task Report conducted by Dewberry Engineers for the NOAA Office for Coastal Management.

Data Requirement	Needed Improvements
Jurisdictions and regulated areas	• Boundaries are currently being digitized from descriptions published in Acts, Code of Federal Regulations (CFR), treaties, and permit documents. Authoritative agencies should be publishing geospatial data in addition to the published documents. • Additional details about regulatory restrictions, when or why changed, duration of regulation or agreement, status of permit (e.g. proposed, planned, approved), etc. should be included as attributes in the geospatial data. • Thresholds for project size, update frequency, etc. should be agreed upon with the authoritative agency. • More detailed data are needed for military areas instead of broad areas of restriction (e.g. unexploded ordnance) if such information can be released.
Abundance and distribution of marine species	• Synthesis of observation data is needed from the multiple entities that collect data. • Modeled data, derived products, and documentation of methodology are also needed (e.g. time series, heat or density maps, and trends over time, etc.) • Dependable and continuous updates to models and products are needed.
Synthesized oceanographic parameters	• Synthesis of monitoring and observation data is needed from the multiple regional entities that collect data. • Derived products from the raw data (e.g.) forecasts, change over time, etc.) are needed.

	● In some regions, densification or winterizing of monitoring devices would greatly improve usability of the collected data. ● Standardized, seasonal, annual or decadal products, as applicable, at an ocean-basin scale for temperature, salinity, oxygen, biomass, and productivity are needed.
Commercial fishing effort - Vessel Monitoring System (VMS)	● Processing and publication of data derived from VMS is conducted by regional partners at considerable cost and effort. Annual agency sponsored products are needed, and in more regions than are currently available. ● Consultation with the Fisheries Management Council (FMC) and regional experts by NOAA National Marine Fisheries Service (NMFS) is needed to define appropriate planning products compatible with existing efforts. ● Improvements are needed to the consistency and completeness of declaration, gear type, and other codes. ● Data on recreational fishing, including locations and type of fish caught, are needed. ● Having access to economic ata so the economic importance of fishing areas can be quantified would add considerable value to derived products.
Vessel traffic - Automatic Identification System (AIS)	● Publication of data derived from raw AIS data is currently performed by the Marine Cadastre. Stronger efforts by The U.S. Coast Guard (USCG), Maritime Administration (MARAD) and the U.S. Army Corps of Engineers (USACE) could stabilize, expand and improve this resource for the broader ocean community. ● Improvements are needed to the identity and characteristics of vessels,

	higher frequency access, and ready-to-use products. Better access to satellite AIS data is needed where land-based receiver are not available.
Human and cultural use areas	• Uniform and complete data are not readily available and data gathering is intensive. • Derived products (e.g. summary of use, hot spots, recreation patterns, etc.) are needed. • Data on Tribal Protected Areas need to be updated and made publicly available. • National Historic Preservation Act data need to be updated. • Improved documentation of provenance and procedures is also needed in the metadata.
Commercial fishing effort - Vessel Trip Report (VTR)	• Processing and publication of data derived from VTR is conducted by regional partners at considerable cost and effort. Annual agency sponsored products are needed, and in more regions than are currently available. • Consultation with the FMC and regional experts by NMFS is needed to define appropriate planning products. • Improvements are needed to the consistency and completeness of original codes, documentation, and products interpolated at a spatial resolution to support energy and aquaculture leasing (i.e. ~2.5nm x ~2.5nm or less). • Data on recreational fishing, including location, type of fish caught, and shore-based access location are needed.
Bathymetry	• Bathymetry data are collected and distributed in a patchwork form and are difficult to find and use at scales beyond individual surveys. Additional high resolution/full bottom surveys are

	needed for complete coverage in priority areas of interest, especially near shore. Seamless 'best available' products and more up to date bathymetry data products are needed. ● Seafloor characterization by sediment texture and physiographic zones is also needed.
Sand and borrow sites	● Current information on sand and borrow sites is not complete, is not synthesized, and can be difficult to find. The forthcoming BOEM Marine Minerals Information System will address many issues when published. ● Historic data may not be available in digital format.
Species and habitat locations, including benthic habitat	● Synthesis and normalization of data is needed from the multiple entities that collect data. ● Modeled data, derived products, and documentation of methodology are also needed (e.g. seasonality of occurrence, gear types, high use areas, endangered species, etc.). ● Interpretation of bathymetry into bottom habitat information.