

NSF EarthCube Workshop for Shipboard Ocean Time Series Data



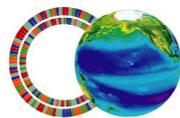
Meeting Report

NSF EarthCube Workshop for Shipboard Ocean Time Series Data

September 13-15, 2019
C-MORE Hale Center, Honolulu, HI



EARTH CUBE



center for microbial oceanography: research and education

c.more *linking genomes to homes*

AUTHORS

Heather Benway

Ocean Carbon and Biogeochemistry (OCB) Project Office, Woods Hole Oceanographic Institution

Justin Buck

British Oceanographic Data Centre, National Oceanography Centre

Lance Fujieki

Department of Oceanography, University of Hawai'i at Manoa

Danie Kinkade

Biological and Chemical Oceanography Data Management Office, Woods Hole Oceanographic Institution

Laura Lorenzoni

Ocean Biology and Biogeochemistry Program, National Aeronautics and Space Administration (NASA)

Mark Schildhauer

National Center for Ecological Analysis and Synthesis (NCEAS), Univ. of California, Santa Barbara

Adam Shepherd

Biological and Chemical Oceanography Data Management Office, Woods Hole Oceanographic Institution

Angelicque White

Department of Oceanography, University of Hawai'i at Manoa

EDITORS

Heather Benway

Woods Hole Oceanographic Institution

Mai Maheigan

Wood Hole Oceanographic Institution

BIBLIOGRAPHIC CITATION:

Benway, H., J. Buck, L. Fujieki, D. Kinkade, L. Lorenzoni, M. Schildhauer, A. Shepherd, A. White. February 2020. NSF EarthCube Workshop for Shipboard Ocean Time Series Data Meeting Report. 59 pp. DOI 10.1575/1912/25480

COVER IMAGE:

EarthCube workshop group at the C-MORE center in Honolulu, HI.

BACK COVER IMAGE:

Image credit: Andrew Collins/Bermuda Institute of Ocean Sciences

Table of Contents

- 1. EXECUTIVE SUMMARY 1**
- 2. INTRODUCTION AND BACKGROUND..... 3**
- 3. PLENARY SESSIONS..... 5**
- 4. BREAKOUT SESSIONS..... 9**
 - 4.1 Breakout 1. Establishing a common data model 9**
 - 4.2 Breakout 2. Improving interoperability 9**
 - 4.3 Breakout 3. Establishing standardized metadata reporting guidelines and required fields to facilitate time series data discovery and re-use .. 10**
 - 4.4 Breakout 4. Streamlining time series data submission..... 11**
 - 4.5 Breakout 5. Data citation and crediting 12**
 - 4.6 Breakout 6. Broadening users of time series data 13**
 - 4.7 Breakout 7. Developing a functional and flexible time series data interface 15**
- 5. ACTIONABLE RECOMMENDATIONS..... 17**
- 6. REFERENCES 19**
- 7. ACKNOWLEDGMENTS 20**
- APPENDIX 1. MEETING PARTICIPANTS 21**
- APPENDIX 2. MEETING AGENDA..... 24**
- APPENDIX 3. COMMUNITY SURVEY RESULTS 25**



Executive Summary

Prior to the OceanObs'19 Meeting, the Ocean Carbon and Biogeochemistry (OCB) Project Office planned and hosted an NSF EarthCube Workshop focused on shipboard ocean time series data (<https://www.us-ocb.org/earthcube-workshop-ocean-time-series-data/>). Data synthesis and modeling efforts across ocean time series represent important and necessary steps forward in broadening our view of a changing ocean, and maximizing the return on our continued investment in these programs. Despite the scientific insights and technology advances of the past couple of decades, significant barriers remain that hinder important synthesis work across time series. This workshop convened 37 participants, including seagoing oceanographers, data managers, and experts in data science and informatics. The goal of the workshop was to identify key ocean time series data challenges related to access and discoverability, metadata reporting, interoperability across databases, and broadening users; and developing recommendations to address those challenges. The workshop adopted the FAIR (Findable, Accessible, Interoperable, Reusable; Wilkinson et al., 2016) Guiding Principles to frame these issues, and included presentations on existing data models and use of controlled vocabularies, guidelines and frameworks for conducting data synthesis and establishing community best practices, and existing and planned ocean time series data products. The following seven small group discussions provided opportunities for participants to prioritize key issues and brainstorm to identify potential short- and long-term solutions towards:

- Group 1. Establishing a common data model
- Group 2. Improving interoperability
- Group 3. Developing standardized metadata reporting guidelines
- Group 4. Streamlining time series data submission
- Group 5. Data citation and crediting
- Group 6. Broadening users of time series data
- Group 7. Developing a functional and flexible time series data interface

Actionable outcomes emerging from the workshop include developing a pilot data model test case with well-established time series programs and a limited set of variables; pursuing a longer-term focused community activity (e.g., NSF RCN, SCOR or National Center for Ecological Analysis and Synthesis (NCEAS) Working Group, etc.) that builds on the recommendations of workshop participants (e.g., vocabularies, data citation guidelines, metadata reporting, data interface design, etc.) to increase data discoverability and interoperability and enable synthesis and broader applications – the implementation of these FAIR data management practices for shipboard ocean time series will require dedicated resources at the individual grant level; forming a centralized international coordination body for shipboard ocean time series to facilitate development of the data

and metadata standards and best practices needed to be recognized as a formal component of the Global Ocean Observing System (GOOS); and writing a series of high-profile briefs and visualizations that highlight the role of sustained ocean time series in detecting and understanding important climate (e.g., expansion of “the Blob”) and marine ecosystem events and changes (HABs, coastal acidification, coral bleaching events, hypoxic events, etc.).

2

Introduction and Background

Decades of research have demonstrated that the ocean varies over a range of time scales, with anthropogenic forcing contributing an added layer of complexity. In a growing effort to distinguish between natural and human-induced variability, sustained ocean time series measurements have taken on renewed importance, representing one of the most valuable tools scientists have to characterize and quantify ocean carbon fluxes and biogeochemical processes and their linkages to changing climate. They provide the long-term, temporally resolved datasets needed to characterize ocean climate, biogeochemistry, and ecosystem change. While individual time series represent tremendously valuable ocean observing assets, there is extraordinary, unexploited strength in numbers when multiple time-series are combined to obtain a broader spatial perspective of the linkages between climate variability and ocean ecosystem changes (e.g., O'Brien et al., 2017).

Since ocean time series data represent an observational cornerstone of research on climate and ocean change, the OCB Project Office has a long history of supporting ocean time series efforts, including facilitating communication and coordination between time series data producers, data users, and funding bodies to ensure continued engagement and scientific participation by the oceanographic community. OCB has organized and supported multiple activities in support of ocean time series science. In 2010, a workshop was held to bring together scientists and PIs involved in three large NSF-supported time series – Bermuda Atlantic Time-series Study (BATS), Hawai'i Ocean Time-series (HOT), and CARbon Retention In A Colored Ocean (CARIACO) to synthesize scientific insights gained over two decades of observations across these three very different marine ecosystems, and to discuss and prioritize future scientific investigations (Church et al., 2013). In 2012, OCB partnered with the Intergovernmental Oceanographic Commission of UNESCO (IOC) and the International Ocean Carbon Coordination Project (IOCCP) to convene an international workshop to discuss the methodologies used in sample collection and analysis across shipboard time series programs, aiming for consensus on a more standardized approach to improve data intercomparability across sites (Benway et al., 2013; Lorenzoni and Benway, 2013). This workshop seeded follow-on efforts such as the International Group for Marine Ecological Time Series (IGMETS), an extended cross-time series synthesis effort (2013-2017) to integrate a suite of *in situ* biogeochemical variables from time series stations, together with satellite-derived information, to look at holistic changes within different ocean regions and explore underlying reasons and potential connections to regional climate indices and anthropogenic change (O'Brien et al., 2017). At the 2016 Ocean Sciences Meeting, OCB led a Town Hall to gather input from the oceanographic community on formulating a longer-term vision for ocean time series, including improved data access and intercomparability, a broader community of data users, new cross-time series synthesis products, development and dissemination of education and outreach tools, and better communication and collaboration across time series (Neuer et al., 2017; Benway et al., 2019).

Data synthesis and modeling efforts across ocean time series represent important and necessary steps forward in broadening our view of a changing ocean and improving the return on our

sustained investment in ocean time series observations. In order for shipboard biogeochemical time series sites to become an operational component of GOOS, well-defined protocols for uniform metadata standards, data processing, and flow need to be developed. Despite the advances achieved over the past decade, significant barriers remain that hinder work across time series, including issues related to data access, discoverability, and metadata reporting. Furthermore, incorporation of ocean time series data into ocean and earth system models is currently limited due to the lack of a standardized data formats, semantics, and user interfaces. To begin addressing this problem, OCB convened a small workshop (Appendices 1 and 2) with funding from NSF EarthCube in September 2019 (in conjunction with OceanObs19). The objective of the workshop was to conduct a gap analysis to identify missing data infrastructure that would increase time series data availability, comparability, and use. This workshop aimed to provide a much-needed forum for discussion of key issues and barriers surrounding data discovery, access, and interoperability.

Prior to the workshop, the organizing committee developed and launched a survey to query the users and generators of shipboard ocean time series data on key challenges of accessing and working with these data sets. The organizing committee used the survey data (Appendix 3) to help develop the content and structure of the meeting so as to foster discussions about how to most effectively address these challenges.

During the workshop, participants strategized a path forward on the development of a common framework for shipboard ocean time series data and metadata reporting, and data management resources. The overarching objective of this workshop was to move the shipboard ocean time series community toward a FAIR model (Wilkinson et al., 2016). Topics of discussion included (but were not limited to) the following ocean time series cyber-infrastructure challenges, which were ultimately condensed into breakout groups:

- Establishing a common data model for reporting core time series variables (e.g., definitions, vocabulary, units, precision, associated errors, etc.)
- Establishing standardized metadata reporting guidelines and required fields to facilitate data discovery and re-use
- Improving interoperability among different databases/portals
- Mechanisms to streamline and simplify data submission to oceanographic data management entities (e.g., automation of the time series data acquisition/upload process)
- Vision and framework for a dedicated time series data interface that also includes user-friendly visualization and computation options
- Meeting the needs of a broader range of users for data synthesis and information products emerging from ocean time series
- Application of unique identifiers (such as DOIs) to data sets to enable citation and crediting of data providers

3

Plenary Sessions

The workshop opened with a series of plenary sessions to set the stage for more focused small-group discussions later on. Presenters first provided 1) background information about the community-driven EarthCube initiative, which seeks to address the data cyberinfrastructure needs of the geoscience community, and 2) an overview of the FAIR data model, its founding principles (Fig. 1), and ongoing international FAIR data initiatives. The application of FAIR principles offers an important opportunity to facilitate integration and synthesis of oceanographic data in impactful ways, such that the return on our investment in the collection of these integrated multi-disciplinary data sets is fully realized.

FAIR principles

(<https://www.force11.org/group/fairgroup/fairprinciples>)

<p>Findable:</p> <p>F1. (meta)data are assigned a <u>globally unique and eternally persistent identifier</u>.</p> <p>F2. data are described with <u>rich metadata</u>.</p> <p>F3. (meta)data are <u>registered or indexed in a searchable resource</u>.</p> <p>F4. metadata <u>specify</u> the data identifier.</p>	<p>Interoperable:</p> <p>I1. (meta)data use a <u>formal, accessible, shared, and broadly applicable language</u> for knowledge representation.</p> <p>I2. (meta)data use <u>vocabularies that follow FAIR principles</u>.</p> <p>I3. (meta)data include <u>qualified references</u> to other (meta)data.</p>
<p>Accessible:</p> <p>A1 (meta)data are <u>retrievable by their identifier</u> using a <u>standardized communications protocol</u>.</p> <p>A1.1 the <u>protocol</u> is open, free, and universally implementable.</p> <p>A1.2 the <u>protocol</u> allows for an authentication and authorization procedure, where necessary.</p> <p>A2 <u>metadata are accessible</u>, even when the data are no longer available.</p>	<p>Re-usable:</p> <p>R1. meta(data) have a <u>plurality of accurate and relevant attributes</u>.</p> <p>R1.1. (meta)data are released with a <u>clear and accessible data usage license</u>.</p> <p>R1.2. (meta)data are associated with their <u>provenance</u>.</p> <p>R1.3. (meta)data <u>meet domain-relevant community standards</u>.</p>

<https://www.nature.com/articles/sdata201618>





Figure 1. Overview of the FAIR data principles (J. Buck, BODC/NOC, from Wilkinson et al. 2016, Box 2).

Shipboard ocean time series pose several unique operational challenges, including aging infrastructure (research vessels), methodological heterogeneity that hinders data comparability, incorporation of evolving technology into multi-year core data sets, flat or declining funding, effective technology and knowledge transfer in long-term data collection efforts, and lack of data and metadata reporting standards in the face of numerous disconnected data management efforts, making it hard to track ocean time series data users and applications. Currently, there is limited potential for users outside the immediate ocean time series community (e.g., modelers, managers, educators, and other stakeholders) to easily find and gain access to ocean time series data. International coordination efforts such as the [G7 Future of the Seas and Ocean Working Group](#)

(ongoing, action area 3 objective), the [GOOS Observations Coordination Group \(OCG\)](#) (ongoing), and the [Ocean Data Interoperability Platform \(ODIP\)](#) (2012-2018) have sought to address the issue of disconnected data management efforts and facilitate the development of a common global framework for ocean data management. However, a more deliberate investment of resources in leveraging existing frameworks (e.g., vocabularies, infrastructures, etc.) to connect ongoing data management efforts rather than continuing to invest in new ones to is a critical step toward a more FAIR model that makes oceanographic data available to a much broader community of users. To streamline access to scientific data and address interoperability issues, [ERDDAP](#) is becoming an increasingly prominent data provision tool that can translate data from many different types of file formats, databases, and services and provide access via a single standardized interface (Fig. 2). However, adoption of ERDDAP servers does not address the need to harmonize, disambiguate, and enrich the vocabularies (terms) used to describe data and metadata. Such efforts are still critically needed, as well as support for globally unique and persistent identifiers for datasets, that are dereferenceable through the Web (i.e. through HTTP IRIs). These additional frameworks and services are necessary to making data FAIR.

ERDDAP services and FAIR principles

- F**
 - ERDDAP can easily augment metadata with missing elements for compliance
 - ERDDAP can generate ISO metadata documents on demand (harvestable)
 - ERDDAP supports schema.org metadata
- A**
 - ERDDAP provides access to data in multiple formats
 - ERDDAP supports RESTful services for machine to machine access (harvestable)
 - ERDDAP provides a platform upon which data-specific applications can be built
- I**
 - All of these capabilities lead to improved interoperability
- R**
 - ERDDAP supports creation of [Baalt](#) packages for easy archival (DOI)
 - ERDDAP supports Climate and Forecast metadata, [json-ld](#), [schema.org](#)

NSF [EarthCube](#) Workshop for Ocean Time Series Data

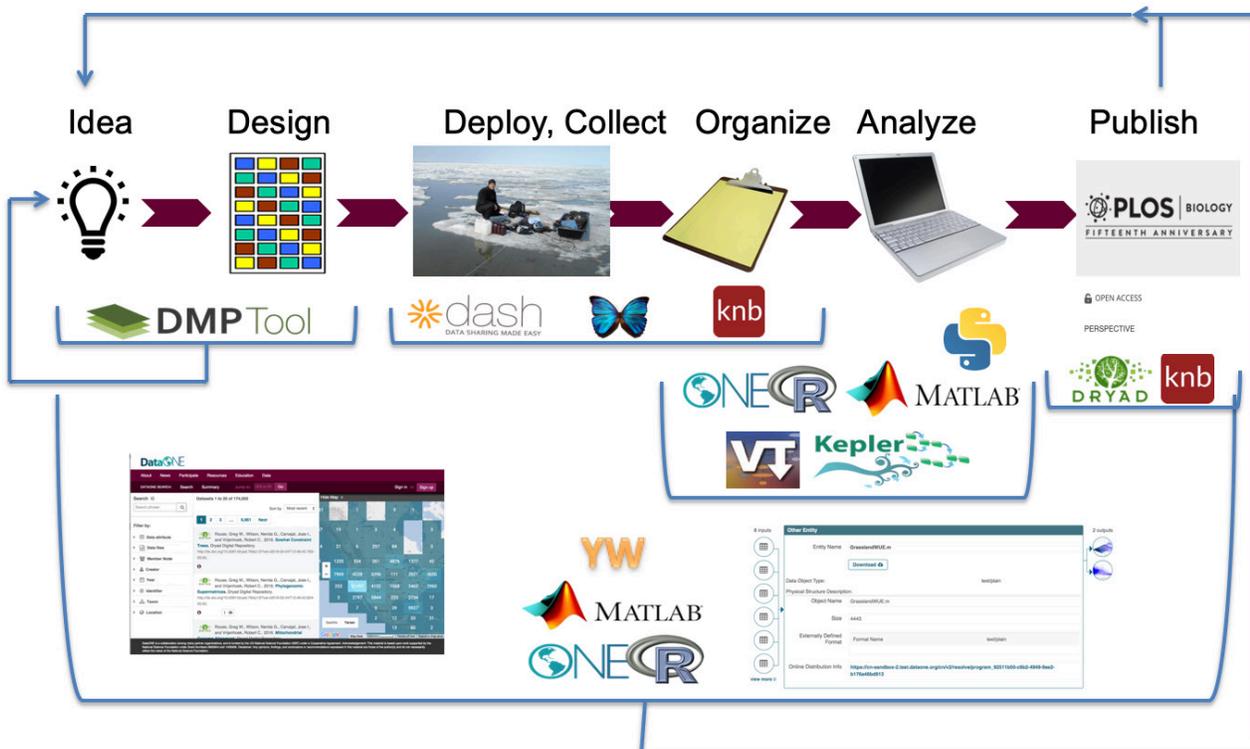
Figure 2. How ERDDAP supports FAIR data principles (K. O’Brien, NOAA)

Presentations on existing multi-platform ocean networks such as the [Ocean Observatories Initiative \(OOI\)](#), [OceanSITES](#), and the [Integrated Carbon Observing System \(ICOS\) Ocean Thematic Centre \(OTC\)](#) provided information on existing oceanographic data frameworks, how they implement FAIR principles, and how they might be adapted for shipboard ocean time series data. The adoption of standardized vocabularies such as schema.org represents a further promising step forward in enabling consistent description of structured data, thus improving the efficacy of searching for

desired information over the Web. [A presentation on controlled vocabularies and the application of Uniform Resource Identifiers \(URIs\)](#) to data and metadata demonstrated the utility of this approach and associated best practices for facilitating computer-based data searches. For many of the parameters that are unique to shipboard ocean time series, there is an urgent need to develop consensus on community-adopted data and metadata reporting guidelines. The recently established [Ocean Best Practices System](#) provides a centralized and trusted mechanism to support the collaborative development, sharing, and adoption of best practices (methodological, analytical, data, etc.) across the oceanographic community.

Shipboard ocean time series data sets have primarily been used to support the goals of individual stations and ancillary projects. An important opportunity and challenge for this community is an increased investment of time and resources in data synthesis efforts that span multiple time series sites and generate products that serve the needs of a broader community of users. Many [data frameworks and tools](#) now exist to facilitate data synthesis, starting from idea inception, experimental design, and data collection all the way through to publication and data sharing (Fig. 3).

Research: Open-science, FAIR-enabled



Open, Transparent, Reproducible, Reusable, Dynamic

Figure 3. A more dynamic, FAIR approach to sharing data and scientific workflows, enabled by accessible data frameworks, interoperable tools, and standardized naming conventions (M. Schildhauer, NCEAS/UCSB, from DataONE training materials).

The final workshop presentations described recent and emerging examples of shipboard time series synthesis products. The [International Group for Marine Ecological Time Series \(IGMETS\)](#) (Fig. 4) has synthesized data and analyzed trends in *in situ* biogeochemical variables from >300 globally distributed time series stations (O'Brien et al., 2017) against a backdrop of satellite-derived ocean temperature and chlorophyll data, revealing holistic changes in marine ecosystems within different ocean regions over the past one to three decades. Another synthesis effort is currently underway to develop a [carbon relevant shipboard time series pilot data product](#) that will leverage the data framework and analysis tools of current carbon synthesis products (Global Ocean Data Analysis Project, GLODAP and the Surface Ocean CO₂ Atlas, SOCAT) using a well-established subset of time series stations and focusing on a core set of carbon relevant biogeochemical variables.

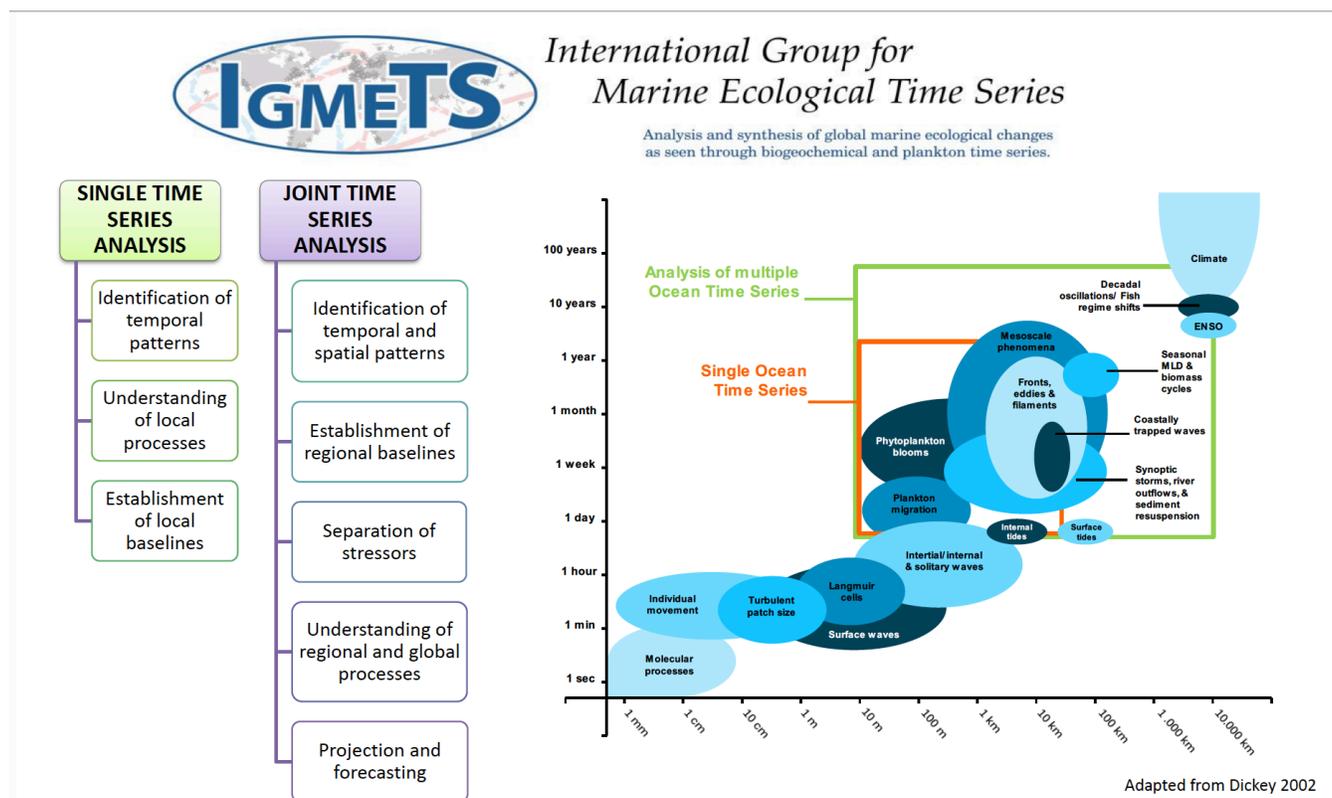


Figure 4. Rationale behind multi-site ocean time series studies such as IGMETS, demonstrating the enhanced spatiotemporal perspective and associated scientific applications that are possible (L. Lorenzoni, NASA/USF).

4

Breakout Sessions

4.1 Breakout 1. Establishing a common data model

Participants: Buck (chair), Whiteaker, Cael, Gum, Izett, Caldwell, Bourbonnais, Currie, Abella-Gutiérrez, Clemente, Ruef, Fiedler, Pfeil, Hunt, Goericke, Walsh, Benway, Nicholson

This group discussed the most desirable features of a shipboard ocean time series data model, including definitions, vocabulary, units, precision, associated errors, etc., with the ultimate goals being **ease of use** (data submission and access) and **flexible formats** to meet the needs of a broad range of users. While we cannot dictate what data formats are served and used by different communities, smart design of data systems can ensure that different formats are interoperable by using tools such as ERDDAP. Group members participated in a sticky note prioritization activity to identify immediate items, follow-on actions, and potential champions to help move the community forward. The development of a minimum metadata model emerged as the most urgent near-term priority. For the pilot data model, participants recommended starting with a few well-established sites (e.g., HOT, BATS, etc.) and focusing on a core subset of key variables (e.g., carbon, oxygen, rates) that would be of interest to a wide range of stakeholders. Group members suggested that an international coordination committee, perhaps under the auspices of the Observations Coordination Group, could facilitate development of community-adopted best practices and increased recognition as a key component of GOOS.

4.2 Breakout 2. Improving interoperability

Participants: Schildhauer (chair), O'Brien, Pacheco, Potemra, Fujieki, Rintala, Shepherd, Lorenzoni, Pfeil

In a brief (~45 minutes) but wide-ranging discussion, consensus was reached that several key principles, mechanisms, and approaches will be critical for improving interoperability among existing, distinct shipboard ocean time series database portals, and assuring interoperability with future efforts:

1. The FAIR model provides a good set of principles for guiding development of interoperability of oceanographic time series data
2. **Explicit use cases and requirements for interoperability must be developed**, since these concretize the specific capabilities expected from "interoperability." Participants recognized that interpretation of the term "interoperability" can vary widely among different researchers due to their varying needs and expectations. For example, one participant said *finding* data is not a problem for them, but *integrating* them is; while other participants doing synthesis studies across database resources typically confronted challenges in *finding*, *interoperating*, and *reusing* data.
3. Based upon the Use Cases and Requirements, a "**minimal metadata**" specification for

shipboard ocean time series data should be developed. If based on a common syntax and semantics, such a metadata specification can serve to “interoperate” databases that have distinct underlying data models.

4. Any prospective Metadata standard, minimal or otherwise, should embrace the **use of controlled vocabularies** for both its fields (e.g., “Person”) and their potential fillers (e.g., “some ORCID ID”) as much as possible. Community agreement on what constitute best-of-class vocabularies for this effort is recommended so that applications are built around these shared semantics. Existing candidate vocabulary(ies) could include e.g., NERC, EnvO, or CF. In addition, any recommended or “endorsed” vocabularies should be accessible through expression in the schema.org vocabulary or extensions thereof into community-vetted RDF/OWL ontologies, and compliant with current W3C Semantic Web recommendations (serialized into JSON-LD, TTL, RDF/XML, etc. for discovery, integration, and interpretation).
5. **Standard templates for collecting data** should be developed that simplify the process of creating consistent and “endorsed” metadata. The current practice is problematic for interoperability—e.g., when individual researchers can devise their own names for column headers and variables, and might forget to document units used in measurements, etc.. This situation can be remedied either by researchers acknowledging that they are using specific standards (such as defined in endorsed, controlled vocabularies, and referenced by name, i.e. a URI); or by affording researchers data entry templates embedded in their primary tools (e.g., MATLAB scripts, Excel, etc.).
6. **Standardize QC flags' definitions and naming conventions** wherever possible, and codify and make these terms web-accessible in some Semantic vocabulary. There are “standard” qualifier flags contained in the NERC vocabulary that could be used as a starting point.
7. **Standardize digital representation of date/time** (e.g., W3C Time, and/or ISO 8601), and provide cross-walking codes to conform variants. This could be effected through templates or convergence in community practice supported by relevant professional societies or publications; or by *requiring* standardized time formats in order to be acceptable for contributing to repositories, or as addenda in publications.
8. **Use ERDDAP or other existing frameworks to support easy file format conversions** (e.g. Excel “xls” to CSV, NetCDF to (CSV + Metadata), etc.)
9. Encourage **Web database portals** to “expose” their holdings for search or harvest, using schema.org terms serialized as JSON-LD (e.g. in the <HEAD>)

4.3 3 Breakout 3. Establishing standardized metadata reporting guidelines and required fields to facilitate time series data discovery and re-use

Participants: White (chair), Cael, Izett, Schildhauer, Bourbonnais, Pacheco, Currie, Abella-Gutiérrez, Clemente, Ruef, Fiedler, Hunt, Plueddemann, Fujieki, Goericke, Walsh, Lorenzoni, Shepherd

Metadata are critical to data discoverability and interoperability and yet, this aspect of data reporting is often inconsistent and at times missing, particularly across shipboard ocean time series. In this breakout session, participants discussed current metadata standards and vocabularies for ocean time series. Using the rate of ¹⁴C-production measured by HOT and BATS as an example, the group worked

through the differences in vocabulary, metadata, and other reporting details. This example was starkly illustrative of the varied ways in which time series report core data, specifically biological data. It is understood that there are currently no common vocabularies nor minimum metadata reporting requirements, hence the multitude of approaches towards metadata in the field.

As an exercise, the participants worked through what a minimum set of metadata might be for ¹⁴C-production. Some but not all of these minimum requirements were included in the respective example datasets. The working group agreed that in order for real progress to be made, time series PIs need to convene to agree upon common vocabularies (borrowing from nearest neighbors such as NERC) and minimum metadata standards. This need is more imperative for biological variables than physical variables. NSF program managers would need to weigh in on which, if any, funding mechanisms would be appropriate to facilitate this necessary task of agreeing upon common metadata standards. Proposed routes include NSF RAPID grants and Research Coordination Networks. Importantly, it was also acknowledged that time series have limited time, personnel, and resources to implement broad-scale changes in data reporting, and that the data science field is somewhat in flux, so if there are changes in standards, time series may not be agile in compliance. For this reason, the path forward would need to be strongly linked to BCO-DMO (for the USA) and include international data managers to help advise the time series organizers. In summary, there was strong consensus that additional effort and resources must be invested to build consensus on metadata standards for time series so that FAIR data principles can be reasonably attainable.

4.4 Breakout 4. Streamlining time series data submission

Participants: Kinkade (chair), Whiteaker, O'Brien, Caldwell, Buck, Lange, Pfeil, Potemra, Gum

Breakout group 4 focused on topics related to streamlining shipboard ocean time series data submission to appropriate repositories. Discussion began by highlighting the current state of time series data sharing and any existing challenges or barriers, followed by brainstorming on ideal data submission workflows and recommendations.

Data sharing among US time series projects occurs through highly variable workflows with known bottlenecks and redundancy in reporting (e.g., BATS and HOT data are served by two to three repositories, in addition to being accessible from the projects' websites and potentially being submitted to NOAA's National Center for Environmental Information (NCEI) multiple times). Currently, there is no centralized oversight body that governs and coordinates data submissions. Submission processes are heavily dependent upon the individual data models of each time series project. Reported variables differ in their descriptions and levels of granularity, and are often dependent upon the PI/project focus. Data collectors are not always motivated to share data, and unfunded or esoteric variables complicate the reporting process (i.e. there is less incentive for PIs to invest time in preparing and submitting unfunded data). Effective data preparation and submission require time and dedicated resources, both of which are often inadequate. Subsequently, data are made available on variable timescales ranging from months to multiple years.

The group then discussed what an ideal data submission workflow might look like, what community

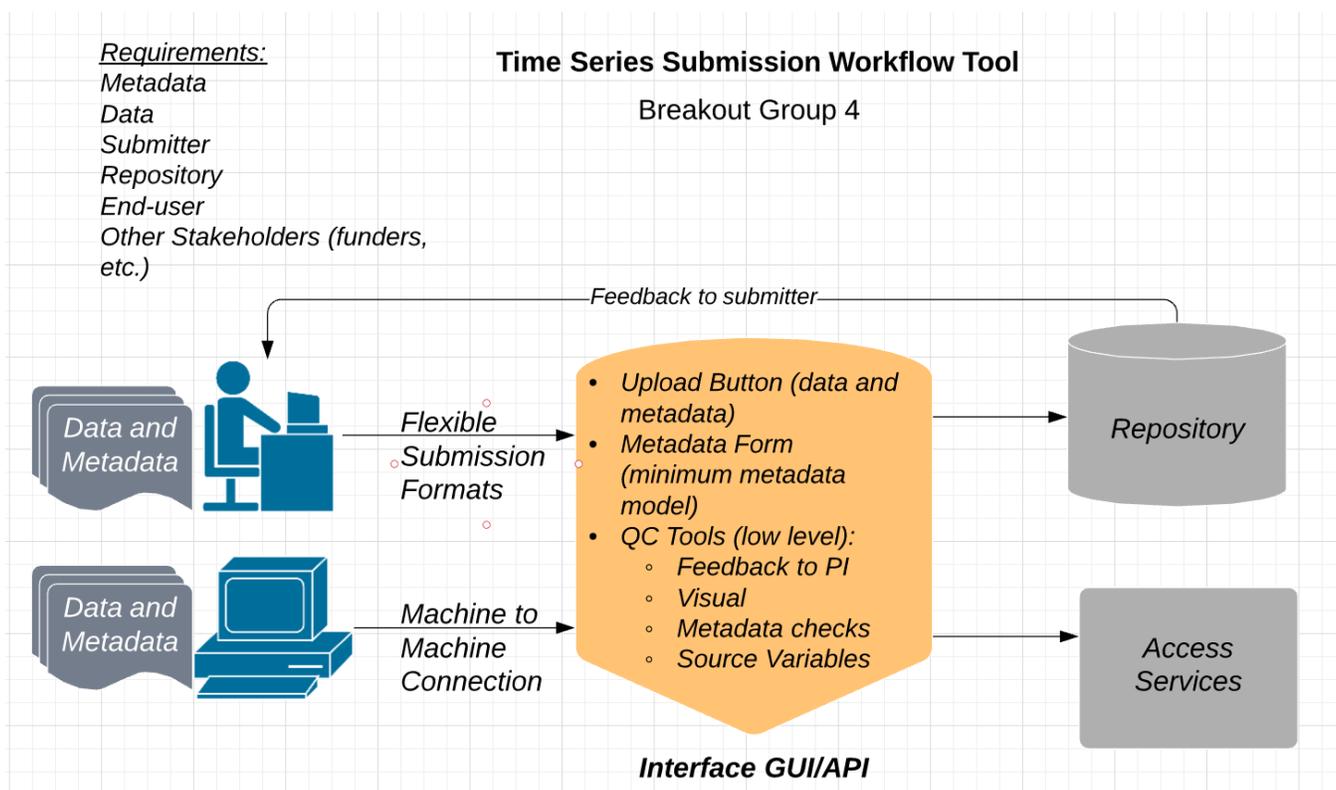


Figure 5. A potential data submission workflow for shipboard ocean time series.

requirements would need to be addressed, and if examples of workflow tools or systems currently exist. At the top of the community requirements list was development of a common time series data model. Additional requirements included the ability to assign DOIs to data submissions, with consideration of granularity and versioning strategies (i.e. whether to assign DOIs to a collection/project, cruise, or station; temporally assign DOIs to yearly granules or collection; how to handle streaming data, etc.). The need for both human and automated machine upload of datasets and metadata was highly desired, as were quality control capabilities and user feedback at the point of submission. The [SOCAT](#) tool was described as one existing model that addresses some of the discussed requirements and may be adapted by this community. The group brainstormed a “SOCAT-like” submission workflow tool (Fig. 5) that addressed community requirements in addition to submitter feedback, deposition of data into appropriate repositories, and data access services.

4.5 Breakout 5. Data citation and crediting

Participants: Shepherd (chair), Caldwell, Bourbonnais

Shipboard ocean time series observations pose unique challenges for proper data citation and attribution of credit. The nature of time series data suggests newer versions of the data with an evolving set of authors. Over its duration, the time series may also undergo changes in methodologies and variables measured. Breakout 5 sought to understand how persistent identifiers

(PIDs) such as DOIs for datasets and ORCID identifiers for authors could rectify some of these challenges.

The group started by defining what successful citation and attribution would resemble and concluded on the following truths: 1) Upon discovering a time series dataset within a journal article, the version used by the journal's authors will be clear to the reader; 2) data versioning [with DOIs] will be easy and hassle-free for data producers; 3) if DOIs for data versions exist, metrics across all versions can be aggregated for holistic assessment of impact; 4) methodologies are also citable, such that they are linked to the data DOI; and 5) authors/contributors are easy to identify.

The group discussed the drivers behind these five scenarios to arrive at the following takeaways: 1) The challenges presented by time series data are not unique to time series data; 2) expertise should be sought from the PID and data management communities to provide guidance for data publishers, producers, and data re-users; 3) journals are an important partner in the solution and should be engaged for arriving at a holistic strategy; and 4) the shipboard ocean time series community needs a champion to engage these communities for information and formulating best practices in citation.

Finally, the group discussed the following action items in relation to these takeaways: 1) the PIDapalooza conference and the Research Data Alliance are two communities that unite the identified stakeholders and should be engaged for aligning time series data needs to best practices in data citation; and 2) the group identified Reyna Jenkins of Ocean Networks Canada, an ocean data specialist, and Martin Fenner, technical lead of Datacite, as people to engage on the topic.

4.6 Breakout 6. Broadening users of time series data

Participants: Lorenzoni (chair), Whiteaker, Cael, Izett, O'Brien, Buck, Abella-Gutiérrez, Ruef, Lange, Fiedler, Pfeil, Hunt, Plueddemann, Potemra

Ocean time series data currently serve a wide range of scientists studying ocean ecology, biogeochemical cycles, as well as physical and atmospheric dynamics. To bolster continued investment in these programs, we must broaden the community of time series data end-users. These sustained observations are highly valuable assets that enable important and often transformative discoveries and directly support applied research, advisory services (e.g., ecosystem-based management and policy), education, and technology development. Facilitating an ongoing dialog with a broader end-user community and documenting outcomes of these exchanges is necessary to optimize the regional and global knowledge gained from sustained time series measurements. To help refine and inform time series sampling efforts and identify opportunities for leveraging, add-ons, and product development, we need more effective mechanisms of collaborating and communicating with different time series data end-users. Ultimately, a robust network of sustained time series observations can be an effective avenue to facilitate collaborations between research and management communities.

In this breakout discussion, participants agreed that *finding* and *accessing* ocean time series data currently represent the biggest hurdles for enabling broader applications of the data, suggesting

the need for a new approach to serving the data. To avoid misuse of data, the datasets and accompanying metadata must be well documented enough for users who are not trained to work with these data. Group members discussed the “fit-for-purpose” data paradigm, emphasizing that every user’s needs will be different, so output formats must be flexible enough to support a broad spectrum of computational, statistical, and visualization needs (see Section 4.7).

USERS	MECHANISM/ACTIVITY/PRODUCT
Modelers and statisticians/data analysts	Community activities focused on facilitating an ongoing dialog among ocean time series generators, data scientists, statisticians, and modelers (e.g., CarboOcean) to foster the development of products like GLODAP and SOCAT; development of community-driven, open source code for extracting, quality-controlling, and gridding time series data
Educators and students	Curricula and modules to get classrooms working with time series data streams; teacher-at-sea programs; virtual classrooms, blogs, and social media for cruises; mentoring programs for early career researchers
Policy makers	Incorporate time series data into continually updated global synthesis products with high visibility (e.g., Global Carbon Project with published annual budgets, IPCC reports, etc.)
Resource managers	Initiate and sustain a dialog among time series scientists, stakeholders, and decision makers to co-generate knowledge and develop targeted products, activities, and visualization tools to inform decision-making using regional coastal time series data sets to monitor threats to coastal ecosystems such as HABs, coastal acidification, eutrophication, hypoxia, warming, invasive species, and biodiversity changes
General public	Content development for informal learning centers, radio interviews, newspapers, and other popular media; citizen scientists at sea or visiting the vessel
Private industry	Utilize time series as testing platforms for new sensors and equipment
Media	Publish in local/regional/national news outlets on time series documentation of ocean changes (e.g., the “Blob,” a Keeling curve for the ocean, etc.)

Table 1. A summary of potential users of ocean time series data and suggested mechanisms or activities to support these users.

The discussion then evolved toward identifying a broader community of users of ocean time series data and potential mechanisms and leveraging opportunities to support broader applications of these datasets (Table 1). While the primary mandate of ocean time series is to support scientific research, there are many untapped partnerships and opportunities to broaden impact and increase visibility. Several time series representatives shared ideas about how they access different user communities. Querying these different users about their needs and providing opportunities for a two-way dialog to exchange ideas and articulate needs will be essential to this process. This could be especially effective at the local/regional level (e.g., the [annual State of the Pacific Ocean Meeting](#)). There are also untapped opportunities to leverage complementary observational resources and networks. For example, shipboard ocean time series serve as reference points and critical sources of calibration/validation samples for remote (satellite) and autonomous (e.g., Biogeochemical-Argo floats) measurements, and conversely, these remote assets expand the spatial footprint of ocean time series sites. These multi-platform interactions provide opportunities for collaboration and co-development of data products to serve a broad range of users.

In summary, shipboard ocean time series datasets are incredibly valuable scientific assets, and there are several engagement, community building, and outreach strategies and products that could broaden the user-base for these data. Moving forward, however, generators of time series data must weigh the relative costs and benefits of these efforts in the face of practical challenges of funding, time, and personnel. Data Repositories, through their curators and information architects, will play a critical role in broadening users of shipboard ocean time series. This will, however, require their gaining sufficient support and ongoing community input to build, sustain, and enhance the data preservation and access frameworks that are needed by the community.

4.7 Breakout 7. Developing a functional and flexible time series data interface

Participants: Fujieki (chair), Schildhauer, Pacheco, Currie, Clemente, Goericke, Walsh, Lange

Breakout Group 7 focused on designing the ultimate time series data interface that also includes user-friendly visualization and computation options. Discussion started with an overview of who the target audience should be and how the *Ultimate Interface* could be structured. Group members then developed a wish list of the various features it could have.

The *Ultimate Interface* should serve a broad audience, including researchers, decision-makers, policy makers, educators, students, and the general public. It should be Web-based, work with any modern Browser, and should not require any additional add-ons such as Python. Finally, it should be layered from simple to complex, depending on the needs and applications of the target audience. For instance, policy and decision-makers may only be interested in generating plots to address specific questions (e.g., change in ocean $p\text{CO}_2$ or Sea Surface Temperature over time). Educators may wish to illustrate the relationships between temperature, nutrients, and chlorophyll. Researchers may wish to access a full suite of statistical features.

At the very minimum, the *Ultimate Interface* should be able to export data in some standard format

(e.g., CSV). It should be able to generate property-property plots, trend plots, ratios, and contour plots. It should be able to provide simple statistics such as averages, medians, standard deviations, and standard error. Additional statistical tools would include coefficients of variation (CVs), trend-lines, splines, and lags. More advanced features could include data interpolation to span temporal gaps, "sliding" contour plots, a geospatial-temporal map interface for the selection of data, and the capacity to query the dataset (e.g., "before sunset"). Finally, the *Ultimate Interface* should support data download via code (API call), and direct dataset import and visualization into desktop-based or other analytical frameworks.

5

Actionable Recommendations

On the final day of the workshop, participants divided into two groups, one with a focus on *data and metadata needs* and one with a focus on *broadening data users*. The purpose of these group discussions was to generate 2-3 actionable recommendations that could be implemented over the next couple of years to address the challenges and priorities laid out in the breakout sessions. The following recommendations emerged from these group discussions.

There is an urgent need for a **prolonged effort and investment in a sustained interaction between time series data generators, data users, and data scientists** to address issues of data discoverability (F), flexible formats and interfaces (A), common vocabularies (F, I), data citation guidelines (F, R), and data and metadata reporting (I, R), all with an eye toward making shipboard ocean time series data FAIR (Tanhua et al., 2019). In the near-term, participants determined that any dataset should have a globally unique and persistent identifier (DOI, URL) such that a data user can always obtain metadata (such as the protocol used to analyze the samples) if they so choose. For the longer term, workshop participants identified several potential formats for such an activity, including NSF RCNs, SCOR Working Groups, NCEAS Working Groups, and Research Data Alliance. One or more of these opportunities will be explored in the coming year. Following on the establishment of community-vetted best practices for shipboard ocean time series data, the **implementation of FAIR data management practices must be supported by dedicated resources at the individual grant level**.

A **pilot data model and product that includes a limited set of time series sites and core variables** may be helpful in informing the best paths forward for effective time series data management frameworks. Workshop participants Lange and Fiedler have a small amount of EuroSea funding to develop a [pilot product](#) over the next 1-2 years. This product will be focused on carbon-relevant parameters and will leverage existing frameworks of other global carbon products (SOCAT, GLODAPv2).

Developing a common set of data standards for shipboard ocean time series will require a **high-level international leadership group to provide scientific oversight and coordination**. The group's activities would be guided by a common set of objectives and a unifying framework for data collection, analysis, and reporting. This leadership body and the network that it supports would facilitate the development of community-vetted data standards and best practices and elevate recognition of shipboard ocean time series as a key component of GOOS. Integration with existing efforts such as the GOOS Observations Coordination Group (OCG) and the Ocean Best Practices System (OBPS) (Pearlman et al., 2019) is the most sensible path forward. There is potentially much to be gained by collaborations between shipboard and autonomous (OceanSITES) observing networks and having those networks integrated under the OCG. Furthermore, the Ocean Best Practices repository represents an opportunity to codify and distribute shipboard data standards and best practices as broadly as possible. This leadership body would have rotating membership, including

time series scientists and data generators, data scientists, information managers, technology experts, modelers, statisticians, educators, and communication experts. We would envision close coordination and collaboration with OceanSITES.

One of the ideas that emerged from the *broadening end users* discussion was to develop **high-level communication briefs and visualizations targeting media and policy makers** to illustrate how shipboard ocean time series are capturing important climate and marine ecosystem events and changes. One example cited at the workshop is the recent reappearance of the marine heat wave coined as “the Blob” now spanning from Ocean Station Papa in the northeast Pacific down the coast of California and also detected in the North Pacific subtropical gyre at the HOT site, with associated marine ecological impacts such as coral bleaching, HABs, and fisheries disruptions. We hope to use these types of communication products to boost the visibility of these ocean observing assets and initiate a dialog with an extended community of users.

6

References

- Benway, H. M., L. Lorenzoni, A. E. White, B. Fiedler, N. M. Levine, D. P. Nicholson, M. D. DeGrandpre, H. M. Sosik, M. J. Church, T. D. O'Brien, M. Leinen, R. A. Weller, D. M. Karl, S. A. Henson, R. M. Letelier (2019). [Ocean time series observations of changing marine ecosystems: An era of integration, synthesis, and societal applications](https://doi.org/10.3389/fmars.2019.00393). *Frontiers in Marine Science*, <https://doi.org/10.3389/fmars.2019.00393>.
- Benway, H. M., M. Telszewski, L. Lorenzoni (2013). [Improving intercomparability of marine biogeochemical time series](https://doi.org/10.1029/2013AGU17000), *Eos Trans. AGU* 94(17), 160.
- Church, M. J., M. W. Lomas, F. Muller-Karger (2013). [Sea change: Charting the course for biogeochemical ocean time-series research in a new millennium](https://doi.org/10.1029/2013AGU17000). *Deep-Sea Research Part II: Topical Studies in Oceanography* 93, 2-15 (Report from [OCB Scoping Workshop on biogeochemical ocean time-series](https://doi.org/10.1029/2013AGU17000)).
- Lorenzoni, L., H. M. Benway (Eds.) (2013). [Global intercomparability in a changing ocean: An international time-series methods workshop](https://doi.org/10.25607/OBP-12), November 28-30, 2012, Ocean Carbon and Biogeochemistry (OCB) Program and International Ocean Carbon Coordination Project (IOCCP), 61 pp. DOI: <http://dx.doi.org/10.25607/OBP-12>.
- Neuer, S., H. Benway, N. Bates, C. Carlson, M. Church, M. DeGrandpre, J. Dunne, R. Letelier, M. Lomas, L. Lorenzoni, F. Muller-Karger, M. J. Perry, P. Quay (2017). [Monitoring ocean change in the 21st Century](https://doi.org/10.1029/2017EO080045). *Eos* 98, doi:10.1029/2017EO080045 (published on 08 September 2017).
- O'Brien, T. D., L. Lorenzoni, K. Isensee, L. Valdés (Eds) (2017). [What are marine ecological time series telling us about the ocean? A status report](https://doi.org/10.1029/2017EO080045). IOC-UNESCO, IOC Technical Series, No. 129: 297 pp.
- Pearlman J, et al (2019). [Evolving and Sustaining Ocean Best Practices and Standards for the Next Decade](https://doi.org/10.3389/fmars.2019.00277). *Frontiers in Marine Science* 6:277. doi: 10.3389/fmars.2019.00277
- Tanhua T, et al (2019). [Ocean FAIR Data Services](https://doi.org/10.3389/fmars.2019.00440). *Frontiers in Marine Science* 6:440. doi: 10.3389/fmars.2019.00440.
- Wilkinson, M. D. et al. (2016). [The FAIR Guiding Principles for scientific data management and stewardship](https://doi.org/10.1038/sdata.2016.18). *Scientific Data* 3:160018 doi: 10.1038/sdata.2016.18.

7 Acknowledgments

The authors wish to acknowledge NSF EarthCube for the workshop funding and Michael Sieracki (NSF Biological Oceanography) and Hedy Edmonds (NSF Chemical Oceanography) for connecting us with this opportunity. We thank David Karl, Angelicque White, Jennifer Kondo, and all of our wonderful hosts at the C-MORE Hale Center at the University of Hawaii. EarthCube Science Committee representatives Emma Aronson and D. Sarah Stamps provided valuable input throughout the planning of the workshop. Although unable to attend, organizing committee member Rod Johnson also provided assistance with the content and organization of the workshop. We thank the workshop participants for stimulating discussions.

Appendix I. Meeting Participants

Jose Abella-Gutiérrez
jabella@cigom.org

Andrew Barna
abarna@ucsd.edu

Heather Benway
hbenway@whoi.edu

Annie Bourbonnais
aboutbonnais@seoe.sc.edu

Justin Buck
juck@bodc.ac.uk

B. B. Cael
bbcael@hawaii.edu

Pat Caldwell
patrick.caldwell@noaa.gov

Fernando Carvalho Pacheco
fcarvalho.pacheco@gmail.com

Tara Clemente
tclement@hawaii.edu

Kim Currie
kim.currie@niwa.co.nz

Björn Fiedler
bfiedler@geomar.de

Lance Fujieki
fujieki@hawaii.edu

Ralf Goericke
rgoericke@ucsd.edu

Adriana Gonzalez Silvera
adriana.gonzalez@uabc.edu.mx

Joseph Gum
jgum@ucsd.edu

Dana Hunt
dana.hunt@duke.edu

Robert Izett
rizett@eoas.ubc.ca

David Karl
dkarl@hawaii.edu

Johannes Karstensen
jkarstensen@geomar.de

Danie Kinkade
dkinkade@whoi.edu

Nico Lange
Nico.Lange@uib.no

Ricardo Letelier
letelier@coas.oregonstate.edu

Laura Lorenzoni
laura.lorenzoni@nasa.gov

Mai Maheigan
mmaheigan@whoi.edu

David Nicholson
dnicholson@whoi.edu

Kevin O'Brien
kevin.m.o'brien@noaa.gov

Benjamin Pfeil
benjamin.pfeil@uib.no

Al Plueddemann
aplueddemann@whoi.edu

James Potemra
jimp@hawaii.edu

Janne-Markus Rintala
janne-markus.rintala@icos-ri.eu

Wendi Ruef
wruef@uw.edu

Mark Schildhauer
schild@nceas.ucsb.edu

Adam Shepherd
ashepherd@whoi.edu

Jim Todd
james.todd@noaa.gov

Ian Walsh
iwalsh@seabird.com

Angelicque White
aewhite@hawaii.edu

Timothy Whiteaker
whiteaker@utexas.edu

Appendix 2. Meeting Agenda

Ocean Time Series Data Workshop

September 13-15, 2019

C-MORE Hale Center, Honolulu, HI

Friday, September 13

- 8:00-9:00 Breakfast
- 9:00-9:30 Welcome and introductions around the room

PLENARY SESSION

Introduction/overview talks *(15 min. talks with 5 mins. for Qs)*

- 9:30 EarthCube overview (Danie Kinkade, BCO-DMO/WHOI)
- 9:50 FAIR data principles and initiatives (ENVRI-FAIR, GO-FAIR, Enabling FAIR data, etc.) (Justin Buck, BODC/NOC and Steve Diggs, SIO)
- 10:10 Time series data challenges that hinder science (Angelique White, UH)
- 10:30-10:45 Break

Insights from current data models *(12 min. talks with 3 mins. for Qs)*

- 10:45 Interconnecting ocean time series efforts and frameworks (Laura Lorenzoni, NASA)
- 11:00 Moving towards FAIR data principles with ERDDAP (Kevin O'Brien, NOAA/PMEL)
- 11:15 The Ocean Observatories Initiative (OOI) approach to data and metadata (Wendi Ruef, Univ. Washington)
- 11:30 The international OceanSITES Eulerian Observing network (Johannes Karstensen, GEOMAR)
- 11:45 Use of controlled vocabularies: Potential applications to time series data (Justin Buck, BODC/NOC)
- 12:00 Integrated Carbon Observation System (ICOS) Ocean Thematic Centre (OTC) (Benjamin Pfeil, Univ. Bergen)
- 12:15 Ocean Best Practices (Johannes Karstensen, GEOMAR)
- 12:30-1:30 Lunch

Time series data integration and products *(15 min. talks with 5 mins. for Qs)*

- 1:30 Tools and approaches to facilitate data synthesis (Mark Schildhauer, NCEAS/UCSB)
- 1:50 International Group for Marine Ecological Time Series (IGMETS) overview and challenges (Laura Lorenzoni, NASA/USF)
- 2:10 Carbon relevant ship-based time series station synthesis – A data product pilot (Nico Lange, Univ. Bergen)

- 2:30-5:00 Small group discussions (participants self-select, coffee will be available if needing a break)
 Group 1. Establishing a common data model for reporting core time series variables (e.g., definitions, vocabulary, units, precision, associated errors, etc.) (Chair: Justin Buck, Auditorium)
 Group 2. Improving interoperability among different databases/portals for time series data (Chair: Mark Schildhauer, Agora Room)
- 5:00 Adjourn for the day, dinner on your own

Saturday, September 14

- 8:00-9:00 Breakfast
- 9:00 Group 1-2 report outs (*15 mins. for each chair*)
- 9:30 Group discussion
- 10:00-12:30 **Small group discussions** (*participants self-select*)
Group 3. Establishing **standardized metadata reporting guidelines** and required fields to facilitate time series data discovery and re-use (Chair: Angel White, Auditorium)
Group 4. Streamlining and simplifying **time series data submission** (Chair: Danie Kinkade, Agora Room)
- 12:30-1:30 Lunch
- 1:30 Group 3-4 report outs (*15 mins. for each chair*)
- 2:00 Group discussion
- 2:30-5:00 **Small group discussions** (*participants self-select*)
Group 5. Application of unique identifiers (such as DOIs) to enable time series **data citation and credit** to be attributed to data providers (Chair: Adam Shepherd, Table Outside)
Group 6. Meeting the needs of a **broader range of users** for data synthesis and information products emerging from ocean time-series (Chair: Laura Lorenzoni, Auditorium)
Group 7. Designing the **ultimate time-series data interface** that also includes user-friendly visualization and computation options (Chair: Lance Fujieki, Agora Room)
- 5:00 Adjourn for the day
- 6:30-8:30 Group dinner (Tiki's Grill and Bar, 2570 Kalakaua Ave)

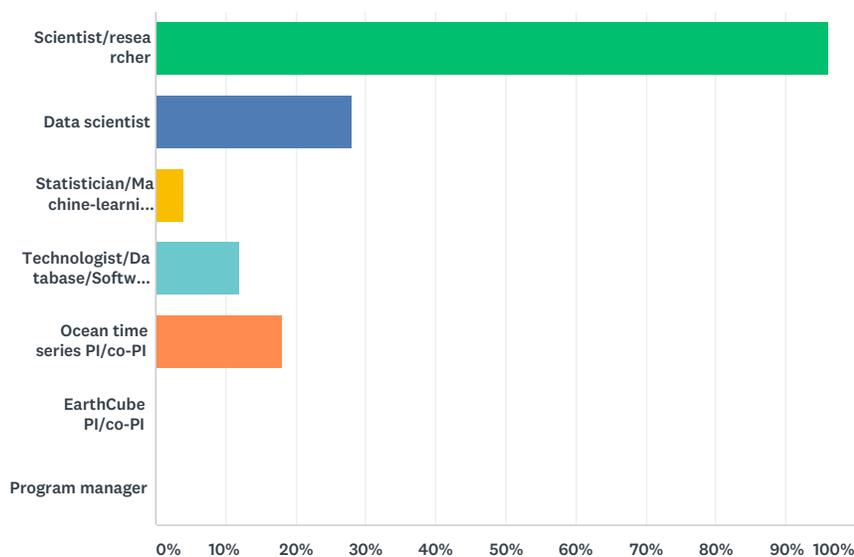
Sunday, September 15

- 8:00-9:00 Breakfast
- 9:00 Group 5-7 report outs (*15 mins. for each chair*)
- 9:45 Group discussion
- 10:15 Break
- 10:30 Participants assist chairs to finish compiling recommendations from small groups and discuss format/outcomes (*participants self-select for one breakout topic to assist with*)
- 11:30 Full group discussion (issues and remaining Qs, next steps, assignments, etc.) – considering the following
- Format of recommendations (best practices) – Ocean Best Practices to codify once finalized
 - EarthCube guidelines/requirements?
 - Dissemination to time series community for feedback
 - Opportunities for peer-reviewed publication(s)
- 12:00 Closing remarks and adjourn
- 12:30 Lunch

Appendix 3. Survey Results

Q1 What are your primary roles/responsibilities (please check all that apply)?

Answered: 50 Skipped: 0

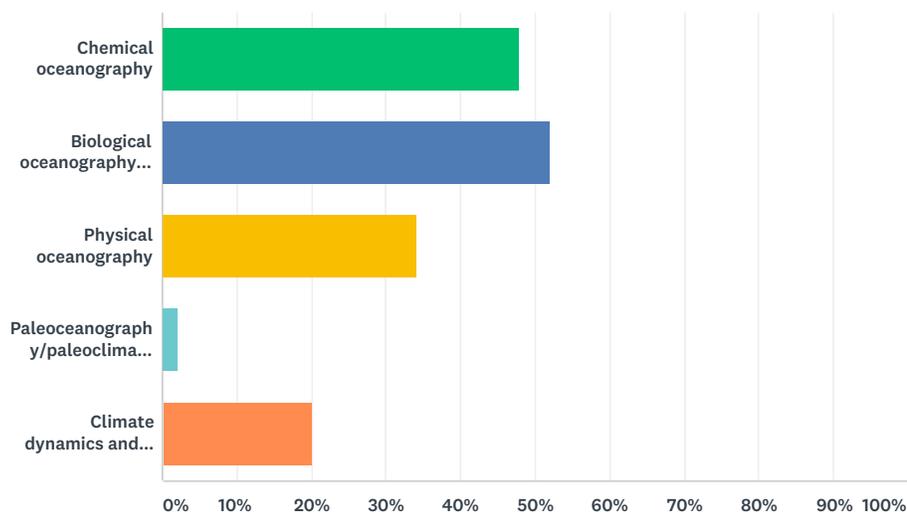


ANSWER CHOICES	RESPONSES
Scientist/researcher	96.00% 48
Data scientist	28.00% 14
Statistician/Machine-learning specialist	4.00% 2
Technologist/Database/Software developer	12.00% 6
Ocean time series PI/co-PI	18.00% 9
EarthCube PI/co-PI	0.00% 0
Program manager	0.00% 0
Total Respondents: 50	

#	OTHER (PLEASE SPECIFY)	DATE
1	Model technician/analyst	8/2/2019 12:43 AM
2	Atmospheric time series ex-PI	6/27/2019 1:17 PM

Q2 What is (are) your primary discipline(s)? (please check all that apply)

Answered: 50 Skipped: 0

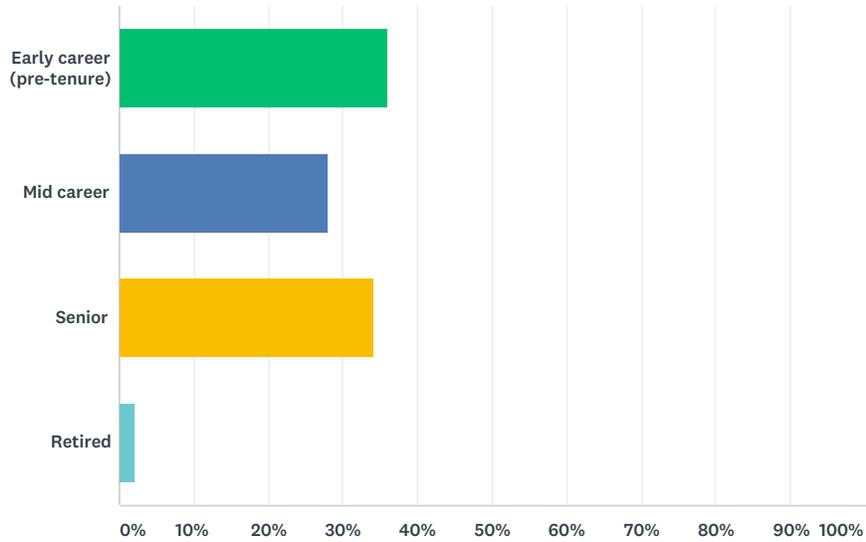


ANSWER CHOICES	RESPONSES
Chemical oceanography	48.00% 24
Biological oceanography/marine biology	52.00% 26
Physical oceanography	34.00% 17
Paleoceanography/paleoclimatology	2.00% 1
Climate dynamics and/or atmospheric science	20.00% 10
Total Respondents: 50	

#	OTHER (PLEASE SPECIFY)	DATE
1	Hydrology, water resources engineering	8/7/2019 11:10 AM
2	Data assimilation	7/25/2019 7:25 AM
3	sensor development	7/9/2019 5:18 AM
4	Biogeochemistry	6/27/2019 7:12 PM
5	Biogeochemistry	6/27/2019 7:34 AM
6	Coupled physical-biogeochemical modeling	6/24/2019 7:39 AM

Q3 What is your career stage?

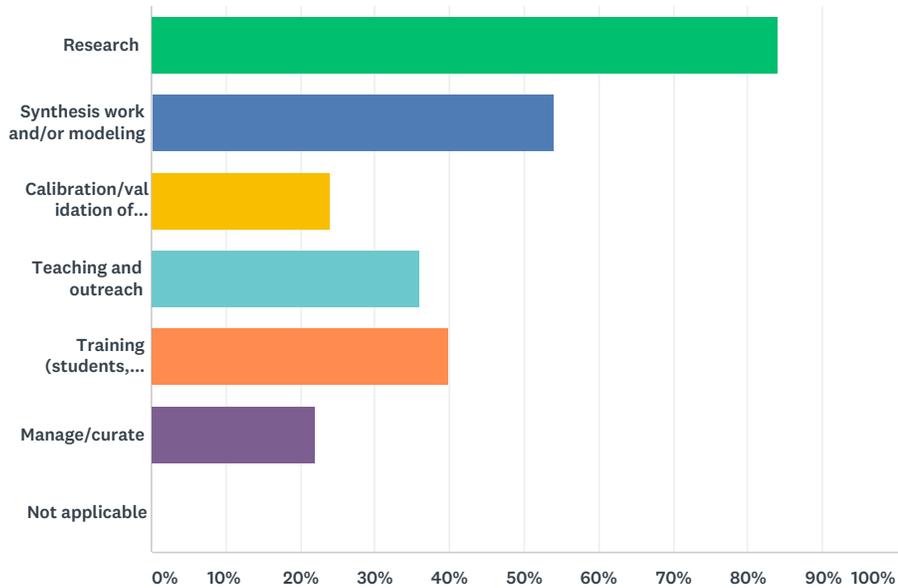
Answered: 50 Skipped: 0



ANSWER CHOICES	RESPONSES	
Early career (pre-tenure)	36.00%	18
Mid career	28.00%	14
Senior	34.00%	17
Retired	2.00%	1
TOTAL		50

Q4 How do you use ocean time series data? (please check all that apply)

Answered: 50 Skipped: 0



ANSWER CHOICES	RESPONSES
Research	84.00% 42
Synthesis work and/or modeling	54.00% 27
Calibration/validation of satellite data and/or autonomous sensors	24.00% 12
Teaching and outreach	36.00% 18
Training (students, postdocs, etc.)	40.00% 20
Manage/curate	22.00% 11
Not applicable	0.00% 0
Total Respondents: 50	

#	OTHER (PLEASE SPECIFY)	DATE
1	scientific assistance to coastal water managers	8/23/2019 5:01 AM
2	assimilation of observations for and verification of operational oceanography products	7/25/2019 1:50 AM
3	Same for atmospheric time series	6/27/2019 1:17 PM
4	Not doing modeling anyway	6/14/2019 6:58 AM

Q5 Please list your primary sources of funding for ocean time series-related work (e.g., NSF, NOAA, NASA, private foundations, etc.)

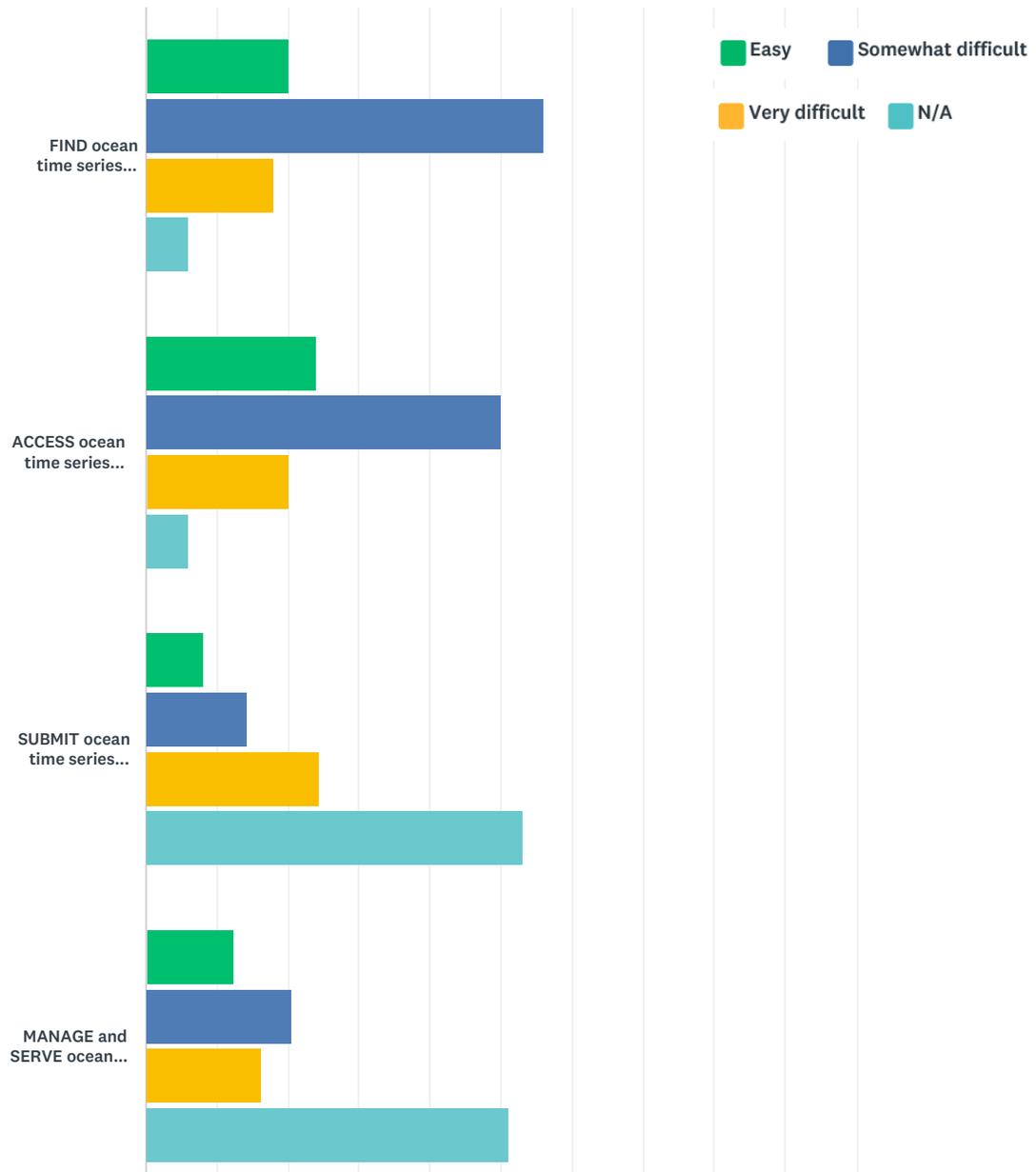
Answered: 48 Skipped: 2

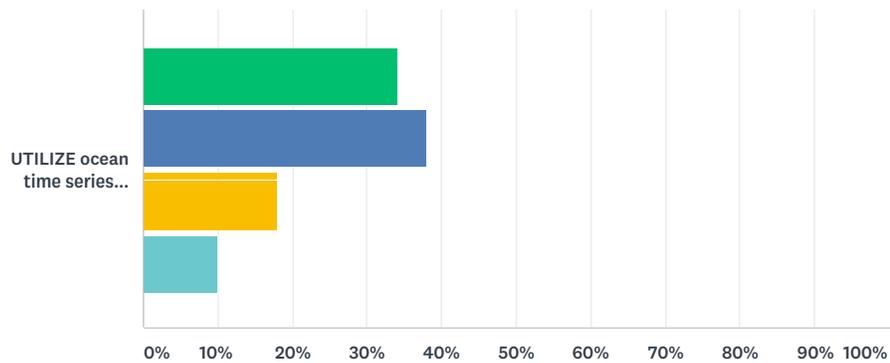
#	RESPONSES	DATE
1	private foundations	9/10/2019 6:29 PM
2	NSF,NASA, NOAA,foundations	9/3/2019 9:33 AM
3	University Grants Commission, India; ICAR, (NICRA) India	9/3/2019 9:10 AM
4	Spanish Ministry for Science and Education	9/2/2019 7:22 AM
5	CNPq (Brazil) CAPES (Brazil)	9/2/2019 6:52 AM
6	ICOS, H2020, national funding	8/27/2019 12:23 AM
7	french water agency	8/23/2019 5:01 AM
8	NSF, NOAA	8/22/2019 10:39 AM
9	NOAA	8/22/2019 8:20 AM
10	UIB/GEOMAR/EuroSea/AtlantOS	8/9/2019 12:54 AM
11	BOEM, NSF	8/7/2019 11:10 AM
12	CMEMS	8/2/2019 12:43 AM
13	-	7/25/2019 7:25 AM
14	National Weather Services, Copernicus Marine	7/25/2019 1:50 AM
15	NSF, NASA, NOAA	7/18/2019 2:28 PM
16	national funding	7/17/2019 11:59 PM
17	NSF	7/16/2019 9:00 AM
18	DOST-funded project	7/14/2019 2:05 AM
19	NIWA SSIF funding (New Zealand Govt funding)	7/10/2019 6:52 PM
20	Currently I am funded by a Simons Fellowship to do work at Station ALOHA, but by the time this workshop happens I will be funded by two grants (one NERC, project: CLASS, the other EU, project: COMFORT) to be doing global climate-relevant biogeochemistry data analysis, using time series & model output.	7/10/2019 6:30 PM
21	Non	7/10/2019 11:28 AM
22	NASA, The Gordon and Betty Moore Foundation	7/10/2019 5:17 AM
23	NSF	7/9/2019 1:15 PM
24	ERDDAP	7/9/2019 12:18 PM
25	NSF	7/9/2019 11:52 AM
26	NSF	7/9/2019 5:18 AM
27	NSF, private foundations	7/5/2019 7:59 AM
28	Instituto Español de Oceanografía (IEO, Spain) Galicia Regional Government (Xunta de Galicia, Spain) European Research Funds (various EU research and innovation programmes and Interreg funds) Spanish Government Research Funds (various research projects and infrastructure funds)	6/28/2019 3:22 AM
29	NSF	6/27/2019 7:12 PM
30	NASA, NSF	6/27/2019 1:17 PM
31	private foundations, NSF	6/27/2019 8:28 AM

32	NOAA, NASA, EPA, NSF	6/27/2019 7:34 AM
33	NSF, NASA	6/27/2019 6:30 AM
34	NSF	6/27/2019 6:27 AM
35	NOAA	6/24/2019 9:47 AM
36	NSF, NOAA	6/24/2019 8:22 AM
37	NASA, NOAA, NSF, ONR (in that order)	6/24/2019 7:39 AM
38	Instituto Nacional de Investigacion y Desarrollo Pesquero (Argentina)	6/21/2019 11:27 AM
39	BOEM, NSF	6/21/2019 8:03 AM
40	Mexican government funding (Conacyt), Nippon Foundation and POGO, IAI	6/19/2019 10:53 AM
41	NOAA	6/19/2019 8:25 AM
42	NSF	6/17/2019 10:13 AM
43	NASA, private foundations	6/14/2019 2:05 PM
44	Currently no sources of funding	6/14/2019 6:58 AM
45	NSF	6/14/2019 4:30 AM
46	EU funded projects, i.e. EU COST ACTION Ocean Governance for Sustainability (OceanGov CA 1527) www.oceangov.eu - Working Group Ocean, Climate Change and Acidification	6/13/2019 11:38 PM
47	USGS, state of California	6/13/2019 5:01 PM
48	IMARPE, NOAA and NASA	6/13/2019 3:21 PM

Q6 We are trying to identify key barriers to working with ocean time series data. How difficult is it for you to:

Answered: 50 Skipped: 0





■ Easy
 ■ Somewhat difficult
 ■ Very difficult
 ■ N/A

	EASY	SOMEWHAT DIFFICULT	VERY DIFFICULT	N/A	TOTAL
FIND ocean time series data you need	20.00% 10	56.00% 28	18.00% 9	6.00% 3	50
ACCESS ocean time series data you need	24.00% 12	50.00% 25	20.00% 10	6.00% 3	50
SUBMIT ocean time series data within 12-24 months of collection to a data management entity	8.16% 4	14.29% 7	24.49% 12	53.06% 26	49
MANAGE and SERVE ocean time series data once submitted	12.24% 6	20.41% 10	16.33% 8	51.02% 25	49
UTILIZE ocean time series data sets once obtained (formatting, data quality, etc.)	34.00% 17	38.00% 19	18.00% 9	10.00% 5	50

Q7 Please comment on specific aspects of your difficulties with ocean time series data

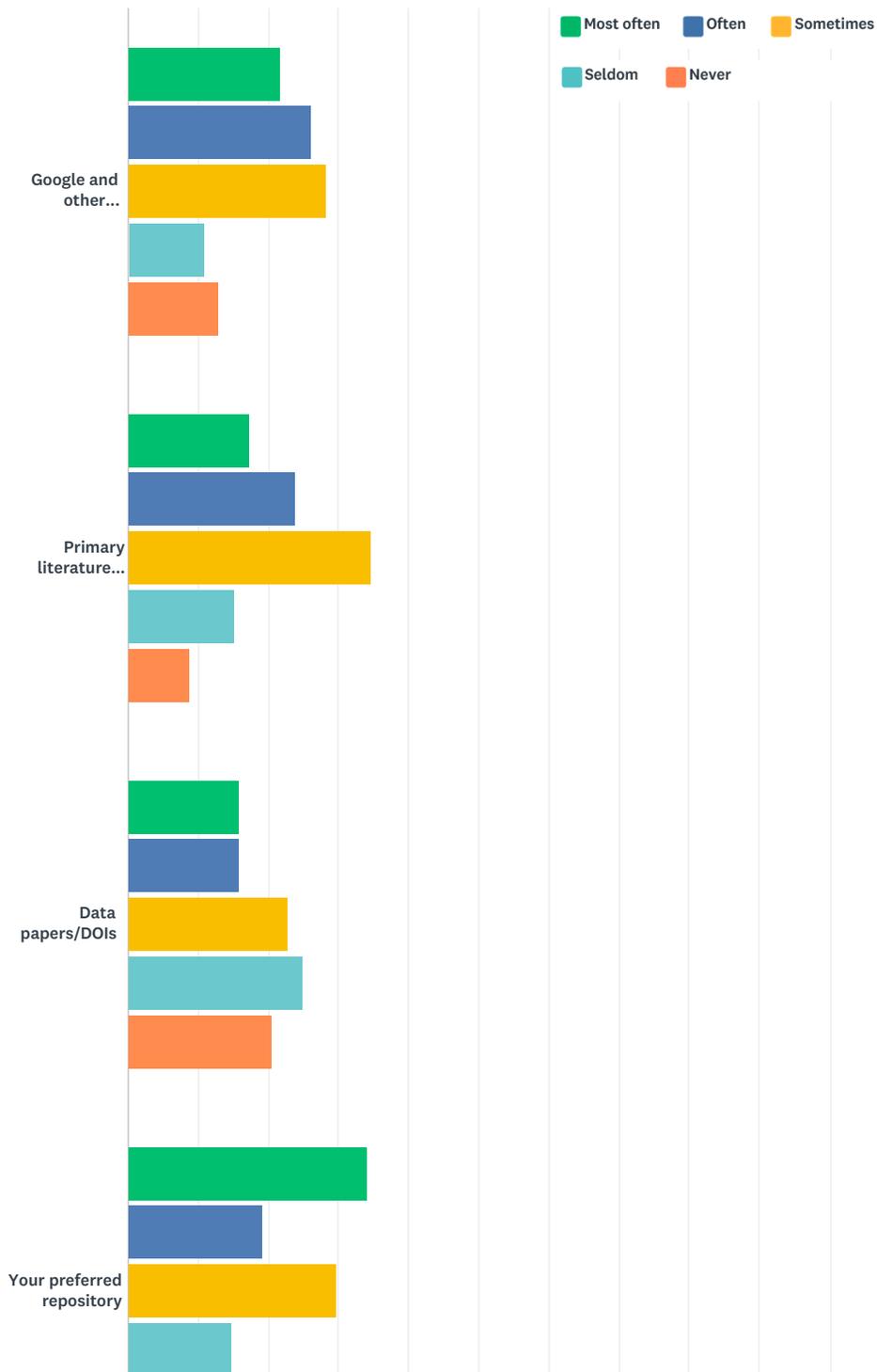
Answered: 41 Skipped: 9

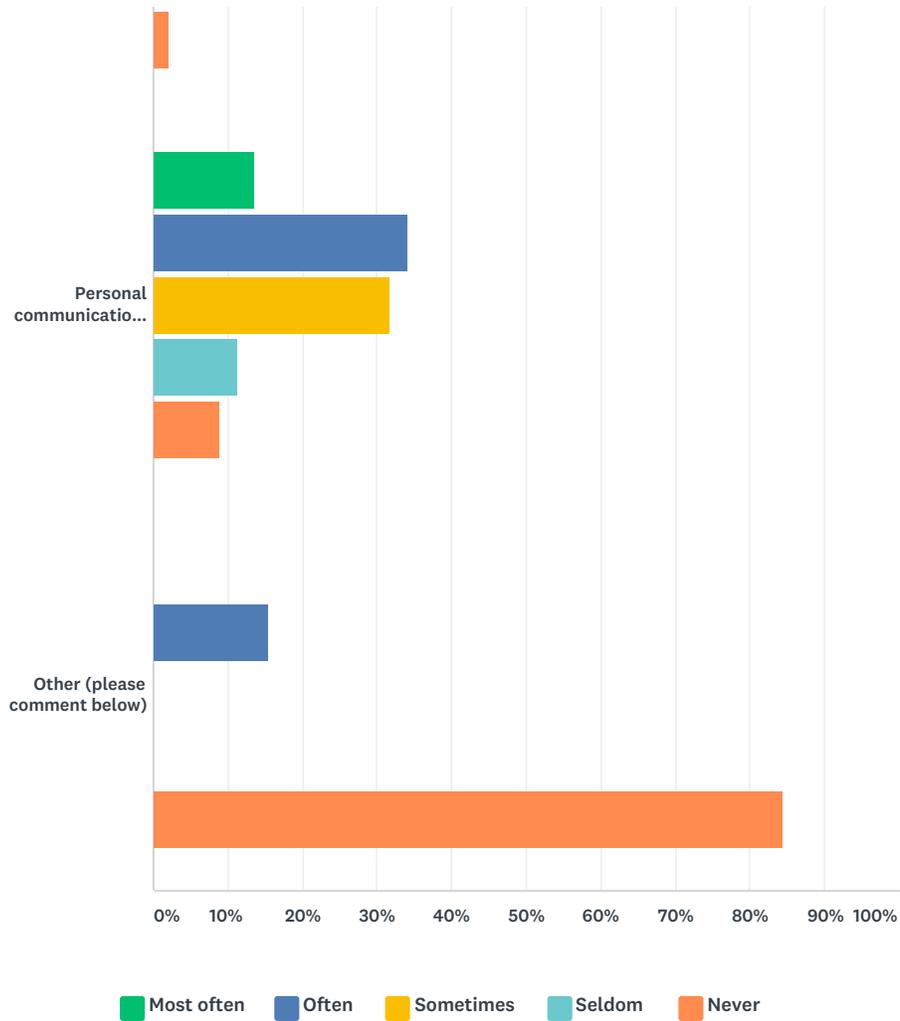
#	RESPONSES	DATE
1	When I first find time series data, sometimes I got some difficulties to process the data because of the limiting tools and methods that I am good at. Then sometimes I do not find the appropriate data to my research regarding spatial-temporal resolution.	9/10/2019 6:29 PM
2	sites change constantly, available data changes, format changes constantly, different web sources for same or different version of the same data. difficulty obtaining a time series, usually it is easy to obtain a short period or a day, but not all the time series	9/3/2019 9:33 AM
3	none	9/2/2019 6:52 AM
4	While getting access to physical oceanographic data it is rather difficult to get access to bgc data	8/27/2019 12:23 AM
5	to use data treatments more complicate than classical statistic (spectral analysis the follow-up of the use of the data series	8/23/2019 5:01 AM
6	Often, there is insufficient metadata - or non-standard terms or jargon from another discipline	8/22/2019 8:20 AM
7	Data gaps an inconsistent units are my main problems.	8/13/2019 4:57 AM
8	I often have to ask researchers directly for where their data are, and then I have to deal with various proprietary sensor formats.	8/7/2019 11:10 AM
9	Finding, and downloading time-series has been time consuming and sometimes unsuccessful. Data is often in non standardised .txt, .dat, .csv files rather than netcdf.	8/2/2019 12:43 AM
10	Inconsistent access locations, formats, metadata.	7/25/2019 7:25 AM
11	- finding a good inventory of what is available for a given time period is difficult - there are more and more data portal with visual maps to help find data and this is a good direction, but sometimes data that are visualized as available only have meta data - some data set exist on multiple platforms, but do not have a common references that allow easily to see overlap.	7/25/2019 1:50 AM
12	Lack of standardization and lack of technical "best practices" documentation (ie what database should I use, what API is preferred).	7/18/2019 2:28 PM
13	As they are multidisciplinary.. getting first the analysis, then a merged and formatted file is difficult	7/17/2019 11:59 PM
14	BCO-DMO / NSF partitions time-series data into different projects based on funding. Therefore we cannot have a continuous project/dataset, leading to numerous requests for data. And comments about missing data. Maintaining project funding to avoid any data gaps is becoming increasing difficult.	7/16/2019 9:00 AM
15	accessibility and format	7/14/2019 2:05 AM
16	I am primarily working with my own data, from voyage planning, equipment maintenance, sample analysis, data analysis, data management, data submission. The data QA/QC submission process generally is time consuming, particularly dealing with the back data (ie data from some years ago). I don't often use data from other time series.	7/10/2019 6:52 PM
17	Mostly I think the challenge is finding time series data that I'm not explicitly aware of but know are out there, & determining to what extent different datasets are comparable.	7/10/2019 6:30 PM
18	There are so many requirement in some data bases that for a new user it becomr very difficult to cite the area/location of interest	7/10/2019 11:28 AM
19	Data sets all have different formats, very little information on measurement errors, data can be difficult to assess regarding techniques and associated errors and what the measurement actually means. Techniques vary with little metadata. Little repeat measurements for assessing errors. Little to no QA/QC. The GOV data repositories not data bases so they are hard to access/filter.	7/10/2019 5:17 AM

20	implementing QARTOD and other QC/QA checks to meet community standards, building trust in time series we acquire	7/9/2019 1:15 PM
21	Find the description of the data, metadata and units	7/9/2019 12:18 PM
22	Finding up to date figures prepared by others, e.g. BATS CO2 time-series data	7/9/2019 5:18 AM
23	HOT and BATS should manage their own data - they both do a good job of making their data available in a timely fashion. BCO-DMO is not set-up well for serving these types of data, at least not in a user-friendly way like currently available through systems like HOT-DOGS. NSF should provide funding to the time series programs themselves for data management and service.	7/5/2019 7:59 AM
24	Lack of a specific unit to deal with multidisciplinary ocean data (particularly integrating biology - plankton- data with hydrographic and biogeochemistry data)	6/28/2019 3:22 AM
25	standard products and data layout.	6/27/2019 7:12 PM
26	There is a difference between long term "time series" data, such as from BATS, HOTS, etc and shorter time series that only have their metadata submitted to a data base or data portal. It is the latter that are difficult to find or manage at times; sometimes only the integrated data, rather than the original profiles or rates, are submitted. And I understand why NetCDF but I hate NetCDF!	6/27/2019 1:17 PM
27	I don't really have any difficulties. The data are easy to access for a modeler, and usually have all the meta-data I need. Time-series data are generally much easier to work with for a modeler than data from individual process studies.	6/27/2019 7:34 AM
28	Not always clear what considerations need to be taken into account when using the data (methodological concerns, gaps, sensitivity, etc).	6/27/2019 6:30 AM
29	Usually the problems arise when trying to aggregate deployment data and/or comparing two different data sets. These problems all come down to the way time is represented (e.g., matching up "hours since ..." with calendar day)	6/24/2019 8:22 AM
30	[FAIR relevance indicated in ()]: The discovery (F) to access (A) step is problematic, especially for biogeochemical data. Sites like BCO-DMO and MACAN/MARCO offer data searches, but really only deliver metadata about projects and PIs that have data (F, but ~A). They use tedious interactive map interfaces (~I). We have existing (for many years) tools like ERDDAP where geospatial/temporal extent searches are trivial (F), and the download formats are rich and flexible (A,R). If more data sets were made available via ERDDAP, even in a distributed sense, the Advanced Search capability and ability to have virtual catalogs would address all the FAIR issues with these.	6/24/2019 7:39 AM
31	We use data from our own time series, or from other time series obtained directly from the PIs	6/21/2019 11:27 AM
32	Researchers at UAF have worked in isolation for years, and have a collection of MATLAB scripts that generate data in formats we are used to working with. But we don't comply with standards etc so sharing our data is easy for us (hard for those who work with it). Also, our institution doesn't have the resources to support data management.	6/21/2019 8:03 AM
33	Finding the resources to maintain ocean time series data collection]	6/19/2019 10:53 AM
34	Getting the data from the original data collectors. Even if they give it, they do not want it shared. So "managing time series data" is not a challenge because there is not much to "manage" (at least in terms of making a public database of it).	6/19/2019 8:25 AM
35	N/A	6/17/2019 10:13 AM
36	As someone who mostly submits data, I struggle with versioning. I work in part with DNA sequencing, and I've found that the database managers who want to upload species ID and counts as for organismal data don't understand why, once my data are "analyzed," they aren't static and may change with, e.g., newer reference sequences. I have not found a good solution here.	6/14/2019 2:05 PM
37	Easiness in accessing is a challenge and sometimes complications of the language used in the website (not use a friend)	6/14/2019 6:58 AM
38	Every time series dataset is accessed slightly differently and it can be difficult for a user that doesn't know the system/site to easily access these data. It would great if we could be more consistent in the platforms used to access time series data so that it could be easier for users to find what they need.	6/14/2019 4:30 AM
39	Gaining timely (within 12-24 months of collection) access to the most updated data sets; data formatting; unit of measurement	6/13/2019 11:38 PM
40	unwillingness to share data; incomplete metadata; uncertainties about data quality; formatting that gives the impression data 'owners' are not highly motivated to share	6/13/2019 5:01 PM
41	It is difficult in wind data	6/13/2019 3:21 PM

Q8 How do you most commonly discover your data?

Answered: 49 Skipped: 1





	MOST OFTEN	OFTEN	SOMETIMES	SELDOM	NEVER	TOTAL
Google and other commercial search engines	21.74% 10	26.09% 12	28.26% 13	10.87% 5	13.04% 6	46
Primary literature (science publications)	17.39% 8	23.91% 11	34.78% 16	15.22% 7	8.70% 4	46
Data papers/DOIs	15.91% 7	15.91% 7	22.73% 10	25.00% 11	20.45% 9	44
Your preferred repository	34.04% 16	19.15% 9	29.79% 14	14.89% 7	2.13% 1	47
Personal communication with PIs	13.64% 6	34.09% 15	31.82% 14	11.36% 5	9.09% 4	44
Other (please comment below)	0.00% 0	15.38% 2	0.00% 0	0.00% 0	84.62% 11	13

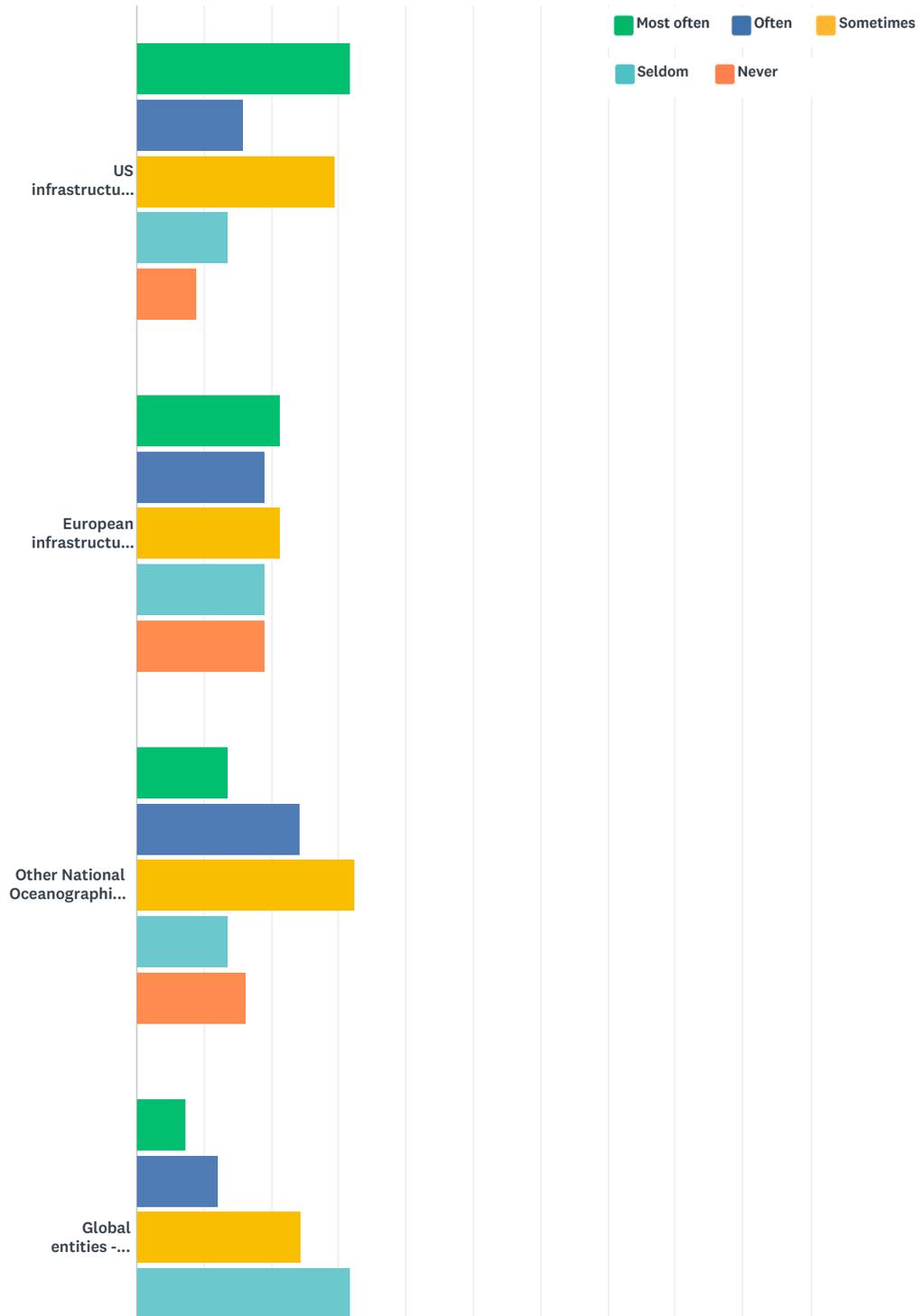
Q9 If you chose "Other" in the previous question, please comment further here on how you most commonly discover ocean time series data.

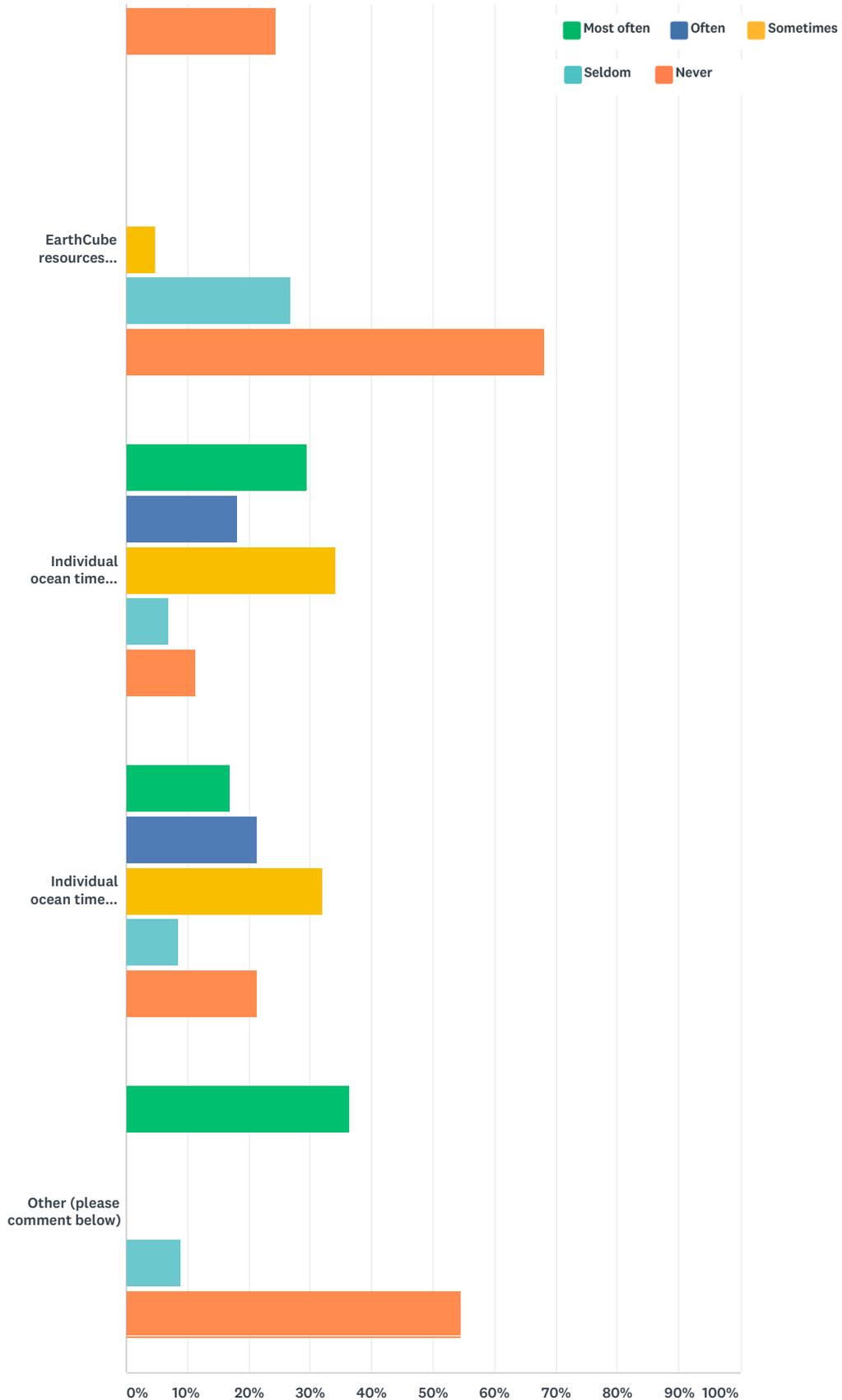
Answered: 9 Skipped: 41

#	RESPONSES	DATE
1	Word of mouth from other researchers	8/13/2019 4:57 AM
2	NA	7/18/2019 2:28 PM
3	N/A	7/10/2019 6:52 PM
4	Through the ocean time series websites, which do a good job of making the data accessible and "viewable".	7/5/2019 7:59 AM
5	I access ocean related data using ERDDAP servers that solve most of these issues	6/24/2019 9:47 AM
6	N/A	6/24/2019 8:22 AM
7	I discover existence through papers/metadata searches, but then try various comprehensive catalogs like ERDDAP at Coastwatch or PacIOOS, or some other OOSs, hoping to find the data mirrored there.	6/24/2019 7:39 AM
8	N/A	6/17/2019 10:13 AM
9	NA	6/14/2019 4:30 AM

Q10 Where do you most commonly access ocean time series data?

Answered: 49 Skipped: 1





■ Most often
 ■ Often
 ■ Sometimes
 ■ Seldom
 ■ Never

	MOST OFTEN	OFTEN	SOMETIMES	SELDOM	NEVER	TOTAL
US infrastructures - e.g., Biological and Chemical Oceanography Data Management Office (BCO-DMO), National Centers for Environmental Information (NCEI), DataONE, etc.	31.82% 14	15.91% 7	29.55% 13	13.64% 6	9.09% 4	44
European infrastructures and data centres - e.g., SEADATANET, PANGAEA, British Oceanographic Data Centre (BODC), Integrated Carbon Observation System (ICOS), European Multidisciplinary Seafloor and water-column Observatory (EMSO), etc.	21.43% 9	19.05% 8	21.43% 9	19.05% 8	19.05% 8	42
Other National Oceanographic Data Centres	13.51% 5	24.32% 9	32.43% 12	13.51% 5	16.22% 6	37
Global entities - e.g., GEO/GEOSS, IODE Ocean Data Portal, OceanSITES	7.32% 3	12.20% 5	24.39% 10	31.71% 13	24.39% 10	41
EarthCube resources and/or data infrastructure	0.00% 0	0.00% 0	4.88% 2	26.83% 11	68.29% 28	41
Individual ocean time series websites/data interfaces - e.g., HOT-DOGS	29.55% 13	18.18% 8	34.09% 15	6.82% 3	11.36% 5	44
Individual ocean time series PIs/staff	17.02% 8	21.28% 10	31.91% 15	8.51% 4	21.28% 10	47
Other (please comment below)	36.36% 4	0.00% 0	0.00% 0	9.09% 1	54.55% 6	11

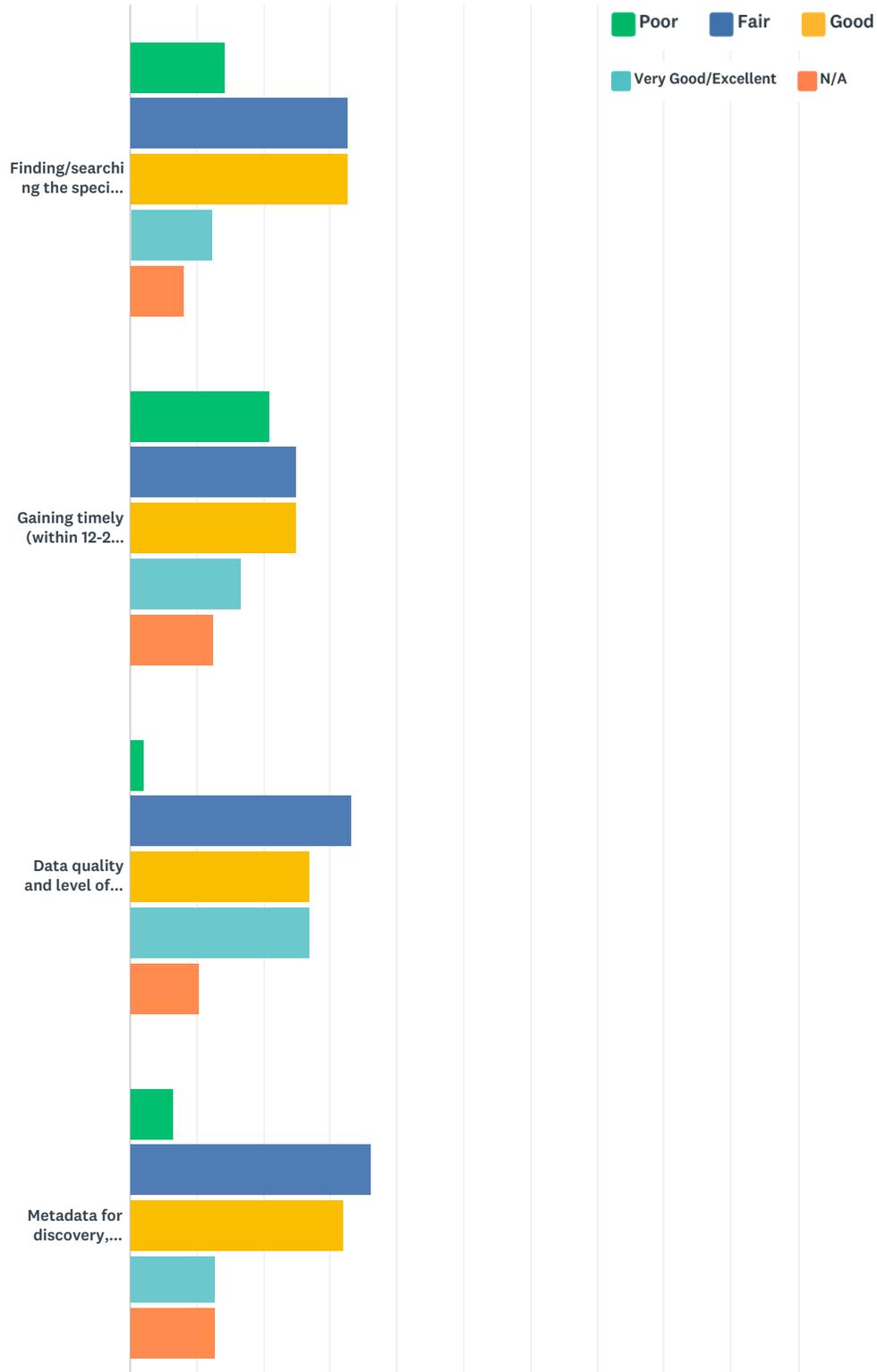
Q11 If you chose "Other" in the previous question, please comment here on where you access your data.

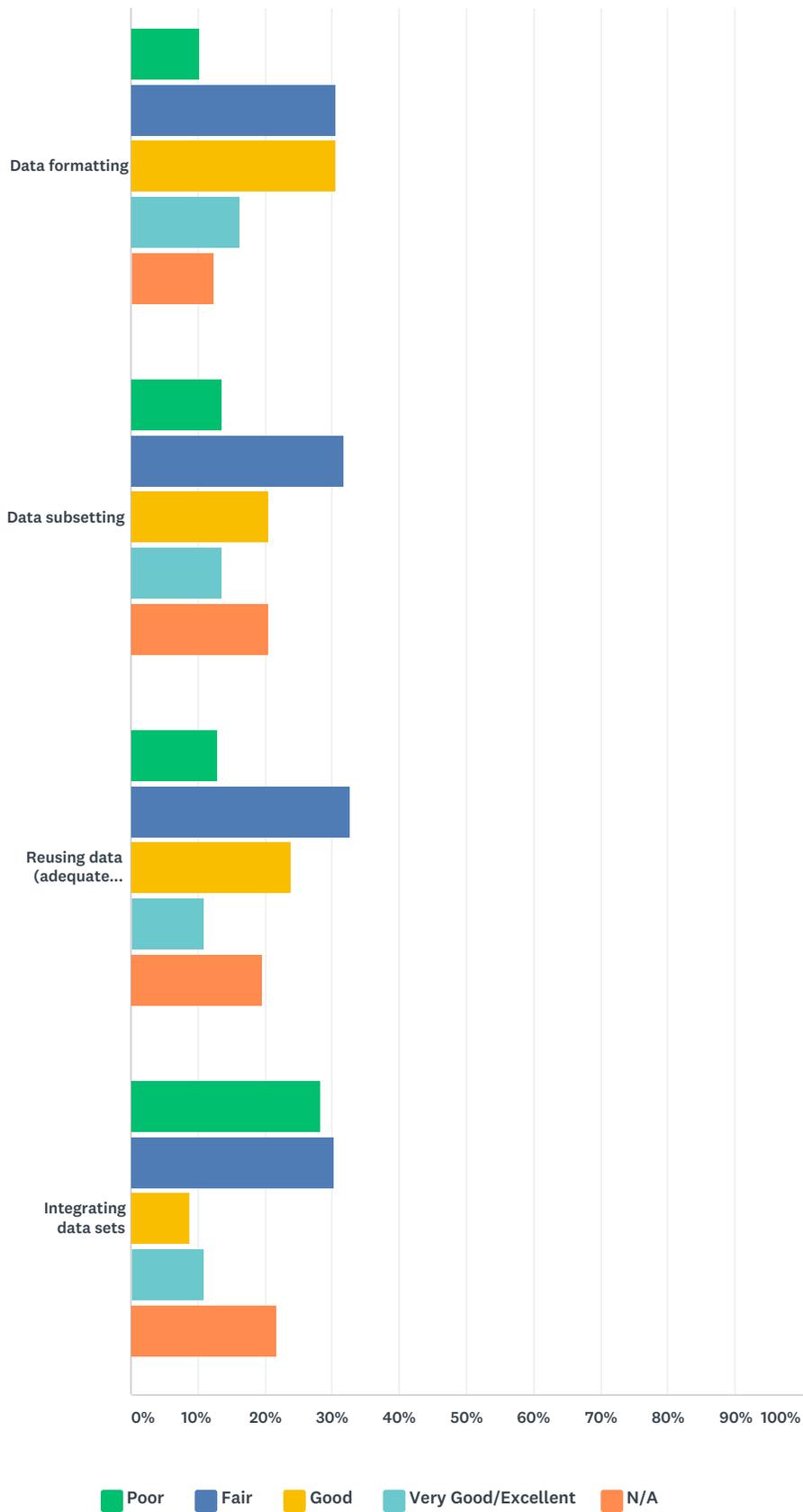
Answered: 9 Skipped: 41

#	RESPONSES	DATE
1	from the WMO GTS	7/25/2019 1:50 AM
2	NA	7/18/2019 2:28 PM
3	N/A	7/10/2019 6:52 PM
4	For coastal time series, EPA and NOAA sites.	6/27/2019 7:34 AM
5	Various ERDDAP servers	6/24/2019 9:47 AM
6	N/A	6/24/2019 8:22 AM
7	https://coastwatch.pfeg.noaa.gov/erddap/info/index.html?page=1&itemsPerPage=1000 http://erddap.sensors.ioos.us/erddap/info/index.html?page=1&itemsPerPage=1000	6/24/2019 7:39 AM
8	N/A	6/17/2019 10:13 AM
9	NA	6/14/2019 4:30 AM

Q12 Please indicate the quality of the following aspects of your data access experience with these ocean data portals:

Answered: 50 Skipped: 0





	POOR	FAIR	GOOD	VERY GOOD/EXCELLENT	N/A	TOTAL
Finding/searching the specific data sets you want (precision/completeness of search process)	14.29% 7	32.65% 16	32.65% 16	12.24% 6	8.16% 4	49
Gaining timely (within 12-24 months of collection) access to the most updated data sets	20.83% 10	25.00% 12	25.00% 12	16.67% 8	12.50% 6	48
Data quality and level of quality control	2.08% 1	33.33% 16	27.08% 13	27.08% 13	10.42% 5	48
Metadata for discovery, provenance	6.38% 3	36.17% 17	31.91% 15	12.77% 6	12.77% 6	47
Data formatting	10.20% 5	30.61% 15	30.61% 15	16.33% 8	12.24% 6	49
Data subsetting	13.64% 6	31.82% 14	20.45% 9	13.64% 6	20.45% 9	44
Reusing data (adequate documentation)	13.04% 6	32.61% 15	23.91% 11	10.87% 5	19.57% 9	46
Integrating data sets	28.26% 13	30.43% 14	8.70% 4	10.87% 5	21.74% 10	46

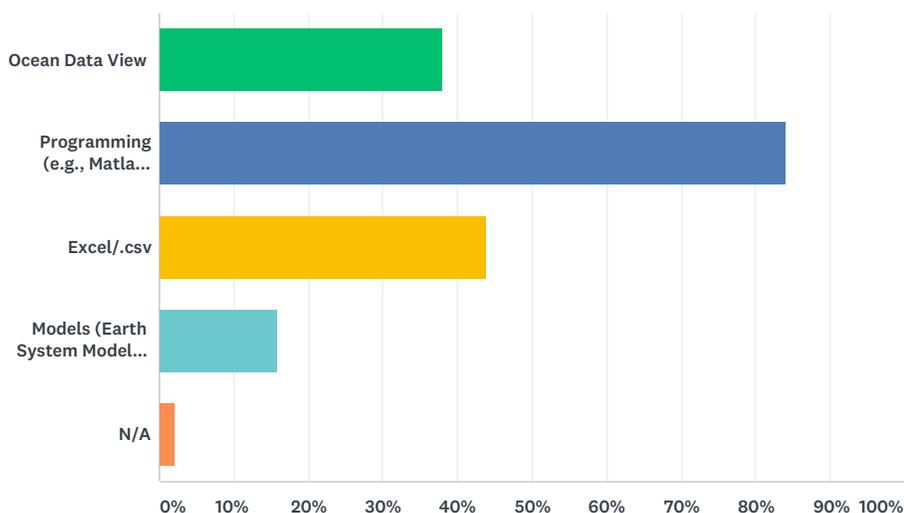
Q13 Please comment on specific issues you've had with obtaining data.

Answered: 23 Skipped: 27

#	RESPONSES	DATE
1	Not sure if I'm commenting on servers I use or ones I don't use because of difficulties!	8/22/2019 8:20 AM
2	Searching for what I need, particularly finding biogeochemical data when the majority of observational data (time series and other) is physical.	8/13/2019 4:57 AM
3	No consistent data product containing all time series data at the same QC level.	8/9/2019 12:54 AM
4	accessibility or the lack thereof	7/14/2019 2:05 AM
5	Usually a formatting issue,	7/10/2019 6:52 PM
6	Problem in accessing data bases	7/10/2019 11:28 AM
7	Need to develop data-set specific QA/QC and difficult time in trying to filter out repeat observations in different archives as there are no unique data identifiers.	7/10/2019 5:17 AM
8	Information quality is not present.	7/9/2019 12:18 PM
9	Synthesizing paleoceanographic timeseries from NOAA Paleo or PANGEA is difficult.	7/9/2019 11:52 AM
10	In the past it has been hard to get the most updated BATS time-series data, e.g. within 1 year of collection.	7/9/2019 5:18 AM
11	The biggest issues are those associated with finding and using time series data from secondary sites (like BCO-DMO or Ocean SITES).	7/5/2019 7:59 AM
12	Difficulties when integrating datasets with limited information on specific quality controls applied to each dataset	6/28/2019 3:22 AM
13	None.	6/27/2019 7:34 AM
14	N/A	6/24/2019 8:22 AM
15	Of them all, the NSF Ocean Observatories Initiative (OOI) access to Pioneer Coastal Array data is the worst. It is barely functional so we have written our own circumvention to the online portal and scape all the data into a private ERDDAP at http://www.myroms.org:8080/erddap/info/index.html?page=1&itemsPerPage=1000 .	6/24/2019 7:39 AM
16	Not applicable.	6/21/2019 11:27 AM
17	Metadata from individual researchers is non-existent.	6/21/2019 8:03 AM
18	It depends on the type of data but many times the problems are formatting.	6/19/2019 10:53 AM
19	Most data (and/or the most recent or complete version of the data) is only in the hands of the original collectors. Data still in the hands of individual (the collectors) is in a plethora of different formats. (Each person has a different format.)	6/19/2019 8:25 AM
20	N/A	6/17/2019 10:13 AM
21	Metadata is key - for example, to not inappropriately compare measurements made with different techniques.	6/14/2019 2:05 PM
22	The language used in the data source is not user-friendly. Eg if you need data for a specific location, how to set or locate properly the area in the data source/base is a challenge to new users	6/14/2019 6:58 AM
23	It can be difficult to find the datasets I'm interested in, again, because they come from different sites that have different ways to access their data so it makes it difficult to integrate the datasets in the end.	6/14/2019 4:30 AM

Q14 What analysis and visualization tools and/or platforms do you most frequently use to analyze and/or visualize ocean time series data? (select all that apply)

Answered: 50 Skipped: 0

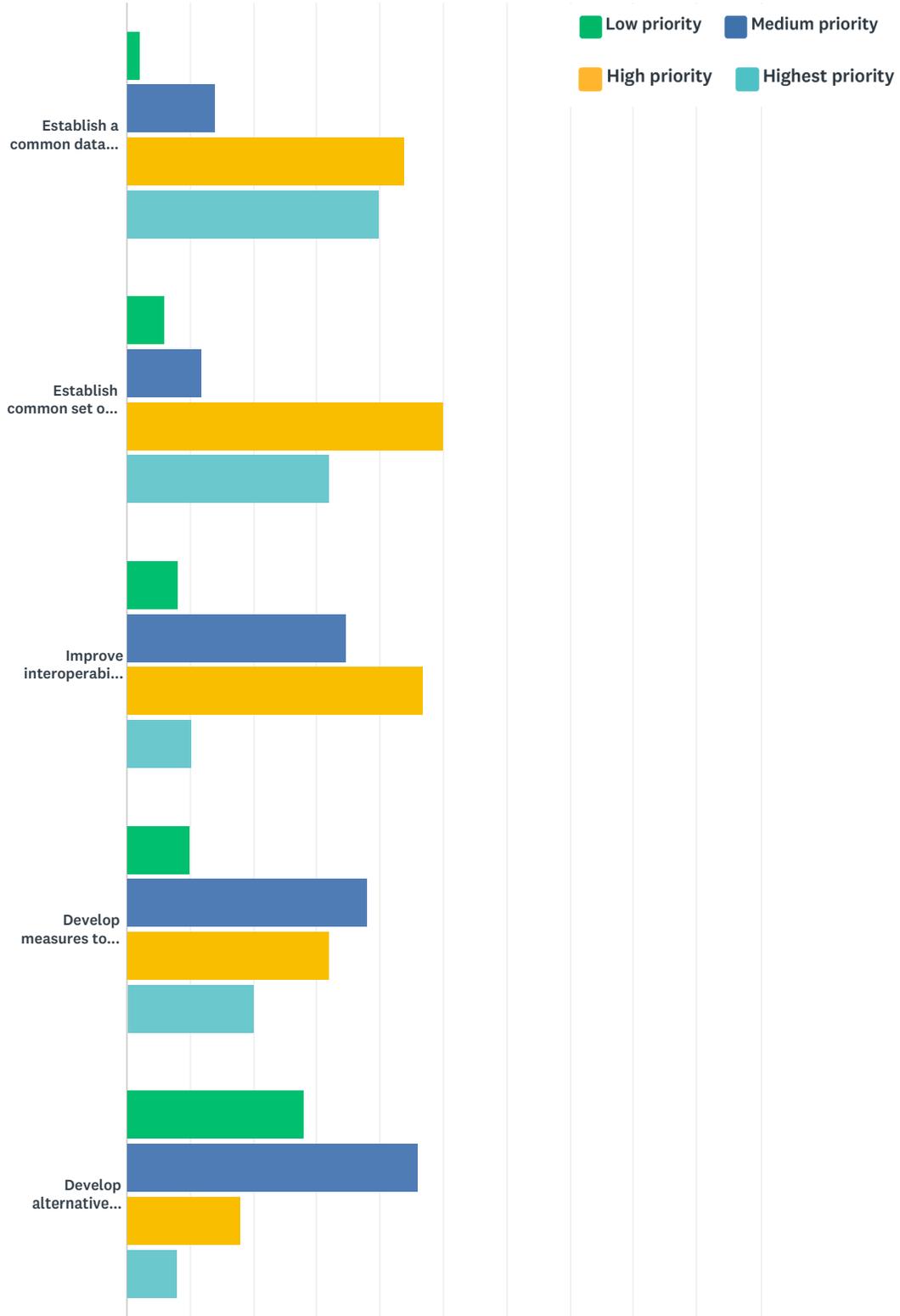


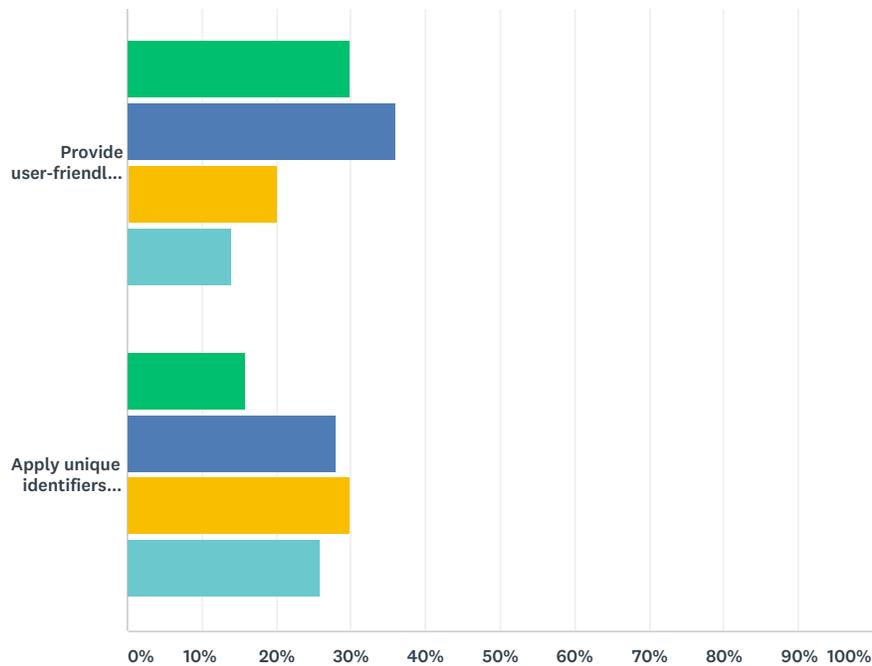
ANSWER CHOICES	RESPONSES
Ocean Data View	38.00% 19
Programming (e.g., Matlab, Python, R, etc.)	84.00% 42
Excel/.csv	44.00% 22
Models (Earth System Models, biogeochemical models, ecological models, etc.)	16.00% 8
N/A	2.00% 1
Total Respondents: 50	

#	OTHER (PLEASE SPECIFY)	DATE
1	I'm a very low level user, so use simple products as my programming skills are low.	7/10/2019 6:52 PM
2	HOT-DOGS is a fantastic resource - it allows you to visualize the data before downloading and using - this saves time and makes for an excellent teaching resource.	7/5/2019 7:59 AM
3	JMP	6/27/2019 1:17 PM
4	Matlab	6/27/2019 7:34 AM
5	R	6/27/2019 6:30 AM
6	GrADS	6/24/2019 8:22 AM
7	COPEPODITE	6/19/2019 8:25 AM

Q15 How important do you perceive the following for improving ocean time series data systems?

Answered: 50 Skipped: 0





■ Low priority
 ■ Medium priority
 ■ High priority
 ■ Highest priority

	LOW PRIORITY	MEDIUM PRIORITY	HIGH PRIORITY	HIGHEST PRIORITY	TOTAL
Establish a common data model for reporting core ocean time series variables (e.g., vocabulary, units, error reporting, etc.)	2.00% 1	14.00% 7	44.00% 22	40.00% 20	50
Establish common set of metadata reporting guidelines	6.00% 3	12.00% 6	50.00% 25	32.00% 16	50
Improve interoperability across different data portals	8.16% 4	34.69% 17	46.94% 23	10.20% 5	49
Develop measures to streamline and simplify data submission to ocean data portals	10.00% 5	38.00% 19	32.00% 16	20.00% 10	50
Develop alternative output formats and products to support a broader range of applications (e.g., synthesis, decision support, modeling, etc.)	28.00% 14	46.00% 23	18.00% 9	8.00% 4	50
Provide user-friendly visualization and computational tools as part of data interface	30.00% 15	36.00% 18	20.00% 10	14.00% 7	50
Apply unique identifiers (such as DOIs) to data sets to enable citation and better credit data providers	16.00% 8	28.00% 14	30.00% 15	26.00% 13	50

Q16 Please comment further on these and other priorities for improving the experience of ocean time series data users.

Answered: 20 Skipped: 30

#	RESPONSES	DATE
1	Some existing standards are great, please do not reinvent the wheel. See OceanSITES data format, Climate and Forecast convention, Attribute Convention for Data Discovery (at ESIP).	8/22/2019 8:20 AM
2	See point 13	8/9/2019 12:54 AM
3	The problem of "getting people to submit data" is best solved by adding value in the form of context-empowered visualization and analysis enabled by similar data.	7/18/2019 2:28 PM
4	As a data provider rather a data user, I identified the data submission process as the highest priority, then I think an on-line metadata editor. I like the SOCAT portal	7/10/2019 6:52 PM
5	None	7/10/2019 11:28 AM
6	I think this process would be best served if a specific data product was chosen as a test data set to develop the methods and then once the standard was created and tested it would be used on others. I also think the effort should triage the different data types so that the effort can focus on the higher needs first.	7/10/2019 5:17 AM
7	The data should be incorporated a data quality flag	7/9/2019 12:18 PM
8	no comment	7/9/2019 5:18 AM
9	It is imperative that data managers were close to data originators. Ideally, one data expert will be required to ensure data flow (including metadata and formatting) between ca. 10-20 data generators and the data base or repository.	6/28/2019 3:22 AM
10	Highest priority is getting the actual data into the data portal; if only metadata, ensuring the metadata are current and accurate. Nad getting all the data in in a timely manner; if it is after 2 yrs of collection, then have all the parameters, not only some. Then one has to contact the PI who most often will gladly send the data files personally. As a data generator, ensuring that the users recognize the data originators. This step is essential with the funders, even long after the project data collection phase is over.	6/27/2019 1:17 PM
11	It's most important to allow the users to download the raw data (hopefully as netCDF files) and let us all use our favorite software to plot and analyze the data. We don't need software and visualization online of the data for scientists (perhaps for the public though.)	6/27/2019 7:34 AM
12	1) I believe we already have a common data model (netCDF) and convention (CF) 2) Metadata is a long, tedious process that is typically ignored by users (especially those accessing data via service). I find the typically workflow is download data, look for familiar things like "time", "temperature", etc., then email p.o.c. (PI) for questions rather than reading metadata (data equivalent of appliance manual) 3) the hang up with submitting data is it is always changing (delayed-mode QC, refinement, etc.), so it rarely is "final"	6/24/2019 8:22 AM
13	There are already a number of communities who have grappled with these issues and developed excellent solutions, e.g. Climate-Forecast Conventions for metadata (including a vigorous active online community adding to and improving these standards), UNIDATA THREDDS/OPeNDAP services for providing easy access and subsetting, ERDDAP with Advanced Search capabilities to scan multiple data sets to find occurrence in requested interval/spaces (excellent for data discovery) and an active international user community. The panels of GOOS, GCOS and JCOMM regularly debate these issues and are knowledgeable about solution. If NSF ignores this community of practise and goes it alone on Earthcube, repeating the failures and waste of the OOI CyberInfrastructure effort, that would be unfortunate to say the least.	6/24/2019 7:39 AM
14	Not applicable.	6/21/2019 11:27 AM

15	We do not need more databases. We need more data ... both in terms of access to existing data and in more sampling programs. Change "publish or perish" to "share [data] or get shunned". If a person's productivity is ONLY weighed on their publications, sharing data (and spending tons of time documenting and reformatting it for submission) takes time away from more publishing. There is no incentive to share data in this model.	6/19/2019 8:25 AM
16	N/A	6/17/2019 10:13 AM
17	If there is an effort to generate online visualization tools, there also have to be provisions to check them occasionally and make sure they're working as intended.	6/14/2019 2:05 PM
18	The use of the map to locate the area may be of good help to a new user or simple software where you may fill your GPS locations etc in a simple language	6/14/2019 6:58 AM
19	Again, establishing a common data model is key for accessibility and will aid in the other priorities as well.	6/14/2019 4:30 AM
20	easy access and search tools	6/13/2019 11:38 PM

Q17 What are your pressing needs for short term data management of ocean time series data in the near term?

Answered: 28 Skipped: 22

#	RESPONSES	DATE
1	data quality control for data, as there are many users that do not have the time or knowledge to deal with it	9/3/2019 9:33 AM
2	OPEN PORTAL FOR UPDATE OF OCEAN TIME SERIES DATA AND EASY RETRIVAL FOR THE RESEARCH GROUP THOSE WHO ARE ALL WORKING THE CORE DATA OF OCEAN TIME SERIES.	9/3/2019 9:10 AM
3	Applying standardised QC and assembling (hopefully) a TS data product	8/27/2019 12:23 AM
4	establishingg Template for data submission	8/23/2019 5:01 AM
5	DOIs for real time data, where NetCDF is updated daily (or more frequently). Identify a 'primary source' for a data set, so if it proliferates on line it can be kept in sync	8/22/2019 8:20 AM
6	Access to consistent, long-term, biogeochemical time series, with metadata.	8/13/2019 4:57 AM
7	An easy usage of multiple time series station data through consistent QC protocols/handling structures and data formats.	8/9/2019 12:54 AM
8	Guidelines for metadata when submitting data to an archive, to better enable discovery and reuse	8/7/2019 11:10 AM
9	I need a standard time-series API I can connect into grafana. I am currently scraping `csv`s into a graphite database, but haven't seen graphite anywhere else in the community.	7/18/2019 2:28 PM
10	"Small" projects rarely get funding for a data management person, yet submission of data is becoming increasing complex.	7/16/2019 9:00 AM
11	An meta data editor integrated to data submission process. Standard processing and QA methods	7/10/2019 6:52 PM
12	Establishing emperical OA data like pH and alkalinity also SST as part of climate change and compare with other data sources for improving reliability	7/10/2019 11:28 AM
13	would love a roll-up capability to something like GitHub that would allow specific data sets to always be updated and refreshed from the various data hubs.	7/10/2019 5:17 AM
14	validation of data, improving data access	7/9/2019 1:15 PM
15	faster availability (right now I see the BATS CO2 time-series figures on-line are updated to 2015, this is too slow).	7/9/2019 5:18 AM
16	The US NSF funded time series programs (HOT and BATS), which are not part of a larger network (like LTER) do a great job with data management - fund them to continue this effort.	7/5/2019 7:59 AM
17	Lack of experts in data management (preparing metadata, formatting, uniformizing quality-flags,...)	6/28/2019 3:22 AM
18	dataset interconnectivity	6/27/2019 7:12 PM
19	None.	6/27/2019 7:34 AM
20	Requisite funding and recognition of the time needed to manage the data correctly.	6/24/2019 8:22 AM
21	Access to a comprehensive historical data set of ocean acidification observations for U.S. coastal waters.	6/24/2019 7:39 AM
22	To finalize our institutional database.	6/21/2019 11:27 AM
23	help standardizing our formats to our data can be shared	6/21/2019 8:03 AM
24	LOL! I need more public data to actually "manage" (e.g., more people willing to share their data).	6/19/2019 8:25 AM
25	N/A	6/17/2019 10:13 AM
26	Aragonite saturation and alkalinity data in the western Indian Ocean.	6/14/2019 6:58 AM
27	More intuitive data management infrastructure, particularly for students that may be utilizing datasets for class projects.	6/14/2019 4:30 AM
28	apply unique doi, user-friendly visualization and computational tools;	6/13/2019 11:38 PM

Q18 What are your expectations for long-term data management and dissemination of ocean time series data?

Answered: 22 Skipped: 28

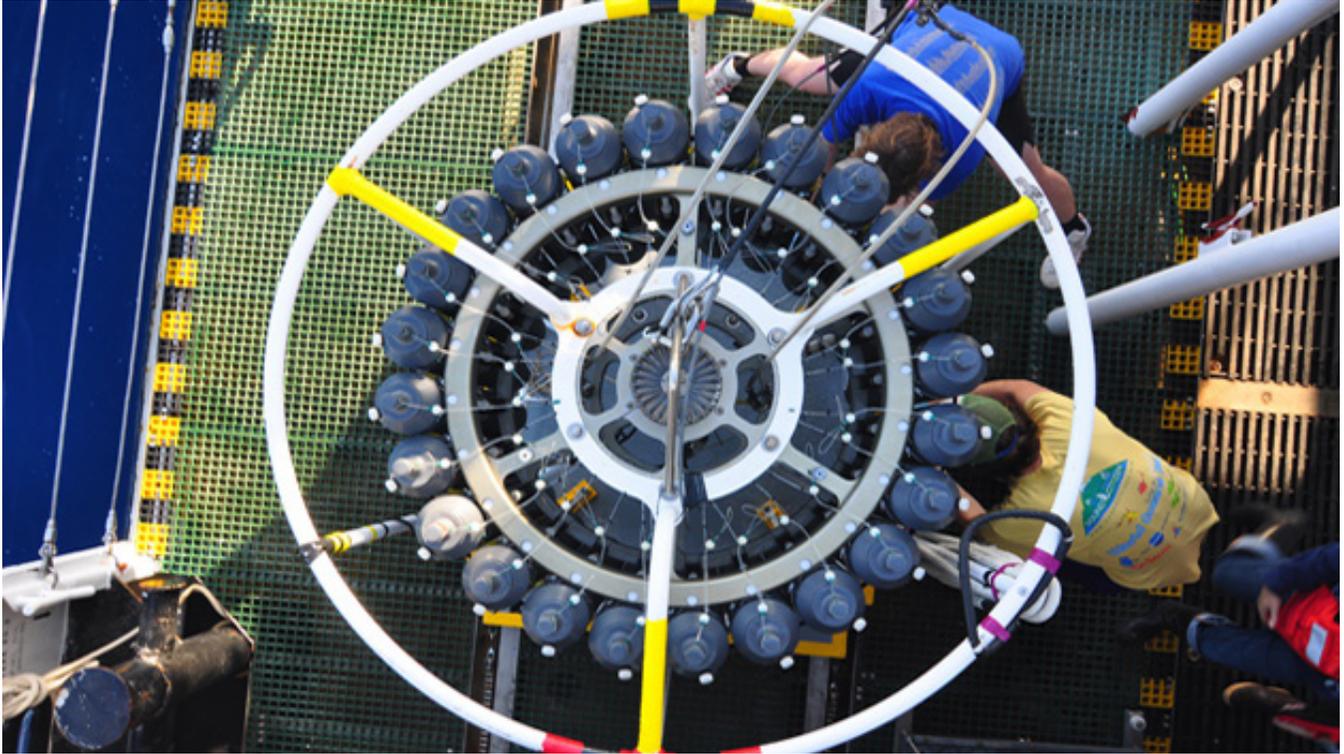
#	RESPONSES	DATE
1	a more standardize format and enough meta-data	9/3/2019 9:33 AM
2	Allowing re-usability and serving the needs of the scientific community. Integration of data from various sources. Currently we have too much redundancy.	8/27/2019 12:23 AM
3	Need funding to keep the conversation going between disciplines, so standards are shared and don't become barriers.	8/22/2019 8:20 AM
4	Easy submission and upload (near real-time) of new data and metadata within a consistent framework. Further, to enhance related automatization processes (e.g. QC procedures).	8/9/2019 12:54 AM
5	Tools to enable archiving and discovery that don't require the average user to know how to program	8/7/2019 11:10 AM
6	A solution must include federated/distributed hosting with replication, versioning, "permanent" URL & API access, and RESTful subsetting.	7/18/2019 2:28 PM
7	A one-stop shop! For submitting QC'ed data and metadata, to a common format, that is then archived, managed (doi), and accessible by users. The ability to include additional parameters. And also , somehow compatible with other OTS platforms (floats, repeat glider sections etc),	7/10/2019 6:52 PM
8	To have local emperical data that may be modified then uploaded to any data bases	7/10/2019 11:28 AM
9	A closure on the metadata requirements and the QA/QC procedures. For instance, I would argue that the data providers not truncate the observations. In practice this is arbitrary and leads to significant uniform errors in the data sets across a wide range of observations.	7/10/2019 5:17 AM
10	would like to see things be accommodated as much as possible in common repositories such as NCEI. Would like to provide a way for users to see validation data	7/9/2019 1:15 PM
11	Operation of a network of specific repositories for ocean (multidisciplinary) data	6/28/2019 3:22 AM
12	All data will be available in netCDF.	6/27/2019 7:34 AM
13	software solutions will outpace any efforts to standardize data	6/24/2019 8:22 AM
14	I hope that data owners see the value of contributing to unified data portals, rather than holding data in various specialists formats on their own data portals.	6/24/2019 7:39 AM
15	To link our institutional database to other relevant international portals.	6/21/2019 11:27 AM
16	We used to provide data management at the institutional level (websites, ftp, internal databases), but are looking more and more to national repositories.	6/21/2019 8:03 AM
17	We need clear logging/tracking of the original-source identifications and usage. How do "Joe" or "Suzy" report to their bosses/funders who is using their data if the large database entities don't track usage *and* automatically report it back to "Joe" and "Suzy".	6/19/2019 8:25 AM
18	N/A	6/17/2019 10:13 AM
19	As molecular techniques (eDNA, etc.) increase in popularity, the field needs to figure out a better way to integrate both analyzed results and info about raw sequence accessibility (NCBI SRA accession numbers, for example) into databases. We need to preserve both the results used in published literature and the possibility of re-analysis and re-interpretation within the context of the associated physicochemical time-series data.	6/14/2019 2:05 PM
20	Looking for funding to install loggers or buoys with carbonate, alkalinity and CO2 sensors in the western Indian Ocean to generate data	6/14/2019 6:58 AM

21	A single portal for all time series datasets that is formatted the same and easy to select what datasets and variables you want. I think BCO-DMO is doing a good job with this and it would be useful to follow their model. Also, if the different funding agencies could collaborate to be willing to have their data on the same platform that would be helpful.	6/14/2019 4:30 AM
22	improve interoperability across different data portals	6/13/2019 11:38 PM

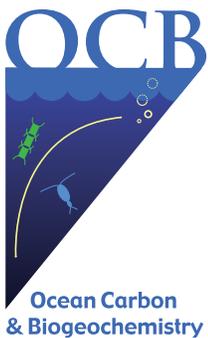
Q19 Is there anything else you wish to add/comment on related to ocean time series data?

Answered: 12 Skipped: 38

#	RESPONSES	DATE
1	Although I will not be able to attend, please add my email to any relevant listservs or other asynchronous discussions: xxxxxxx	7/18/2019 2:28 PM
2	This is a great initiative - thank you.	7/10/2019 6:52 PM
3	Need scientists to work together and funding too	7/10/2019 11:28 AM
4	Establishment of simple focussed effort. Remember that the more general the solution the more complex. One solution will not support all needs.	7/10/2019 5:17 AM
5	Please help make sure the time-series data continue! :)	6/27/2019 7:34 AM
6	As I have said, many solved by existing ERDDAP servers. It would be a shame to reinvent the wheel.	6/24/2019 9:47 AM
7	The challenge to maintain a time series in a developing country is far higher than in well developed centers. Hence, an organized database is desirable, but it takes longer to achieve.	6/21/2019 11:27 AM
8	Again, if everything is based only on the number-of-publications produced, where is the incentive for spending time on data management/documenting/sharing.	6/19/2019 8:25 AM
9	N/A	6/17/2019 10:13 AM
10	Welcome scientists to work together in the western Indian ocean and Tanzania in particular so as to generate more data (xxxxxxx)	6/14/2019 6:58 AM
11	I'm glad you're having this workshop, it's needed!	6/14/2019 4:30 AM
12	NA	6/13/2019 11:38 PM



For more information visit
[https://www.us-ocb.org/
earthcube-workshop-ocean-time-series-data/](https://www.us-ocb.org/earthcube-workshop-ocean-time-series-data/)



Ocean Carbon & Biogeochemistry Program
Woods Hole Oceanographic Institution
266 Woods Hole Road MS#25
Woods Hole, MA 02543
508.289.2838 | hbenway@whoi.edu
www.us-ocb.org | @us_ocb

OCB acknowledges support from these US agencies:



This report was developed with federal support of NSF (OCE-1558412) and NASA (NNX17AB17G). Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsoring agencies.