EarthCube RCN for Marine Ecological Time Series (METS)

Heather Benway, Angelicque White, Danielle Kinkade

Project Description

INTELLECTUAL MERIT

1. METS RCN Steering Committee: Contributions and Expertise

In addition to representation from the Ocean Carbon and Biogeochemistry (OCB) Project Office and the Biological and Chemical Oceanography Data Management Office (BCO-DMO), the METS RCN steering committee will consist of representatives from well renowned, long-running METS such as Hawai'i Ocean Time-series (HOT), Bermuda Atlantic Time-series Study (BATS), and California Cooperative Oceanic Fisheries Investigations (CalCOFI). Representation from complementary NSF-funded ocean observing efforts such as the Ocean Observatories Initiative (OOI) and marine Long-Term Ecological Research (LTER) sites will provide insights from and build synergies with existing data models. We will include representation from the British Oceanographic Data Centre (BODC), which also manages biogeochemical and ecological datasets emerging from METS, and has developed ocean data vocabularies and tools that will be leveraged in METS cyberinfrastructure design. We are also engaging the expertise of a modeler and a statistician, both of whom have worked extensively with METS data, and a data scientist who specializes in ecological data synthesis. Additional expertise will be sought as necessary. Steering committee affiliations and expertise are as follows:

- **Danie Kinkade** (Woods Hole Oceanographic Institution (WHOI)) – director of BCO-DMO and ocean data expert
- **Adam Shepherd** (WHOI) – technical director of BCO-DMO and data scientist
- **Heather Benway** (WHOI) – chemical oceanographer, lead PI of OCB Project Office
- **Angelicque White** (Univ. Hawai'i) – biological oceanographer and lead PI of HOT
- **Rodney Johnson** (Bermuda Inst. Ocean Sciences) – research oceanographer and co-PI of BATS
- **Ralf Goericke** (Scripps Inst. Oceanography) – research oceanographer and phytoplankton ecologist, co-PI of CalCOFI and California Current Ecosystem LTER
- **Todd O'Brien** (NOAA/NMFS) – oceanographer and developer of statistical and visual methods for analyzing METS data, has led projects such as Coastal and Oceanic Plankton Ecology, Production, and Observation Database (COPEPOD) and International Group for Marine Ecological Time Series (IGMETS)
- **Al Plueddemann** (WHOI) - physical oceanographer and OOI Project Scientist for Pioneer and Global arrays
- **Mark Schildhauer** (UCSB/NCEAS) – data scientist with expertise in data synthesis at National Center for Ecological Analysis and Synthesis (NCEAS) and LTER Network Communications Office
- **Justin Buck** (NOC) – senior marine data manager at BODC
- **Naomi Levine** (USC) – oceanographer with research interests in microbial ecosystem dynamics and biogeochemical cycling and experience using METS data in numerical models; current chair of OCB time series committee

2. Background and Rationale

Marine ecosystems are experiencing unprecedented rates of change associated with rising atmospheric carbon dioxide (CO_2) levels and climate change, including concurrent shifts in temperature, circulation, stratification, nutrient input, oxygen (O_2) content, and ocean acidification. Marine food webs comprise a delicate balance among primary producers, intermediate consumers, and top predators. Since the early to mid-20th century, shipboard marine ecological time series (METS), or sustained repeat measurements of physical, biogeochemical, and biological processes, have documented changes in ocean biogeochemistry (e.g., Bograd et al., 2008; Bates et al., 2014; Tanhua et al., 2015), ecosystem state (e.g., Taylor et al., 2012; Edwards et al., 2013; Barton et al., 2016), organism phenology (Asch, 2015; Hunter-Cevera et al., 2016), and even domoic acid levels in shellfish (McKibben et al., 2017). 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 few long-term, temporally resolved observing assets scientists have to characterize and quantify marine ecosystem response to and feedbacks on a highly dynamic and changing climate system. METS have also served as test beds for the development of new sensors and methodologies. They provide valuable seagoing opportunities and hands-on training for the next generation of ocean scientists, as well as a forum for collaboration among nations to build scientific capacity. In addition to their high temporal resolution, integration of data across multiple time series and with other ocean data assets provides the opportunity to obtain a broader spatial perspective and explore causal links between climate variability and marine ecosystem changes. Data synthesis and modeling efforts across ocean time series represent important and necessary steps forward in broadening our view of a changing ocean and fully realizing the value of these ocean observing assets.

A lack of METS data and metadata reporting standards combined with numerous disconnected data management efforts makes it exceedingly difficult for potential users outside the immediate METS community to find and gain access to these valuable and unique datasets. Despite the wealth of data, scientific insights, and technology advances of the past couple of decades, significant barriers remain that hinder science within and across METS, most prominently issues related to data access, discoverability, reporting, and interoperability. Our investment in these observing assets cannot be fully realized without standardized guidelines and best practices for cyberinfrastructure. Furthermore, in order to be recognized as an operational component of the Global Ocean Observing System (GOOS), well-defined protocols for uniform metadata standards, data processing, and workflows need to be developed.

While observations of physical parameters and associated data processing and quality control procedures are fairly well established, biogeochemical parameters, particularly biological and ecological measurements, are less mature (Lindstrom et al., 2012). Many of the METS were established in the Joint Global Ocean Flux Study (JGOFS) era (1980s) including BATS, European Station for Time-series in the Ocean, Canary Islands (ESTOC), Dynamique des Flux Atmosphériques Méditerranée (DYFAMED), HOT, Kerguelen Point (KERFIX), Kyodo Northwest Pacific Ocean Time-series (KNOT), Ocean Station Papa (OSP) and Southeast Asia Time-series (SEATS). Each of these time-series had their own scientific drivers but all were motivated by a common overarching set of JGOFS objectives and implementation procedures, including standard core measurements and protocols (Karl et al., 2003) that have somewhat

diverged over time. This implementation plan was well before the cyberinfrastructure revolution, and prior to development of community-wide best practices and standards for data and metadata reporting, variable naming conventions, or QA/QC protocols. Using HOT and BATS as an example, variable vocabularies are used for nutrients, specific rate measurements, and particle standing stocks (e.g., ^{14}C primary production rate measurements are variously termed L12 or Lt1-3 and sediment trap carbon export are termed Carbon Flux or M1-3 by the respective programs), which are not likely intuitive to new users of these datasets. Moreover, core parameters are often measured via varying protocols with variable-inherent error and bias, which also may not always be clear to the end user, but are nevertheless critically important for interpretation across time series. These are widespread issues across time series. Simply put, despite the importance of these biogeochemical data to understanding planetary change and ecosystem function, the lack of common vocabularies, minimum metadata reporting requirements, and standardized QA/QC protocols for biological and ecological 'core' parameters hinders interoperability and limits the METS end-user community.

Long-term METS datasets can provide much needed constraints on seasonal to decadal variability in physical, biological, and chemical parameters, and can thus be used to validate model dynamics. However, widespread incorporation of ocean time series data into ocean and earth system models is limited by issues of discoverability and lack of standardized data formats, QA/QC, metadata reporting, and user interfaces. Furthermore, the oceanographic community has seen a proliferation of databases with a broad range of overlapping scientific and programmatic drivers. The lack of connectivity and interoperability among databases translates to a great deal of duplicative effort (e.g., ingestion and formatting of METS data for models) and makes it more challenging for end-users to find comparable datasets of interest from within and across regions. Addressing these issues will require a concerted and collaborative effort among METS data producers, data users, funding bodies, and data scientists to co-develop a more unified and standardized approach to time series data frameworks.

In September 2019, a small NSF EarthCube-sponsored workshop was held to conduct a gap analysis of METS cyberinfrastructure and provide a much-needed forum for discussion of key issues and barriers surrounding data discovery, access, and interoperability for shipboard time series (Benway et al., 2020). The workshop organizers adopted the FAIR (Findable, Accessible, Interoperable, Reusable; Wilkinson et al., 2016) Guiding Principles to frame these issues. This 2.5-day workshop convened a small group (37) of oceanographers, data managers, and experts in data science and informatics to 1) assess the current status of METS cyberinfrastructure challenges; 2) identify existing data resources and frameworks that could be leveraged to address those challenges; and 3) develop short- and long-term actionable recommendations to help the community move toward a FAIR data model for METS. During the workshop, participants identified and discussed the following METS cyberinfrastructure challenges:

- Lack of a common data model for reporting core variables (e.g., definitions, vocabulary, units, precision, associated errors, etc.)
- Lack of standardized metadata reporting guidelines and best practices to facilitate data discovery and re-use
- Lack of interoperability across databases and portals
- Delays in data submission and availability
- Lack of flexible, user-friendly data formats and interfaces that serve a broad range of users
- Lack of data citation standards and best practices

The most urgent recommendation from the workshop was to pursue a larger, longer-term community-driven Research Coordination Network (RCN) to bring together ocean and data science expertise on a larger scale to build community consensus on mechanisms to increase data discoverability (F) and develop flexible formats and interfaces (A), common vocabularies (F, I), data citation guidelines (F, R), and data and metadata reporting standards (I, R) to better support scientific and broader applications of METS data.

3. EarthCube RCN for Marine Ecological Time Series (METS)

3.1 METS RCN Objectives

To build on the momentum of the new collaborations and cross-disciplinary discussions at this 2019 workshop and engage a broader faction of the ocean and data science communities to define cyberinfrastructure needs and broaden data access and applications, we propose to establish a METS RCN with the following objectives:

- Bring together members of the oceanographic, data science, and informatics communities to **build consensus** on key components of a FAIR data model for METS, including common vocabularies, metadata reporting standards, and data citation practices
- **Engage broader METS data users**, including modelers, educators, and decision makers, to facilitate broader applications of METS data and foster collaborations and regional networks to address climate and environmental challenges
- **Identify use cases to develop reference implementation of data workflows** that can be adopted by a range of METS data users
- **Build community capacity** for METS data analysis, statistical methods, and data-model integration

As a direct outgrowth of the September 2019 EarthCube meeting, the proposed RCN will extend these important METS cyberinfrastructure discussions to the broader community of data scientists and managers, METS data producers and users, and representatives of complementary ocean observing programs. RCN activities will also provide the necessary time for extended dialog between ocean and data scientists to identify, co-design, and iterate on FAIR data solutions.

3.1 RCN Coordination

OCB and BCO-DMO will lead the coordination and facilitation of the METS RCN, in collaboration with the RCN Steering Committee. OCB (us-ocb.org) was formed in 2006 and is a network of scientists who work across disciplines to understand the ocean's role in the global carbon cycle and the response of marine ecosystems and biogeochemical cycles to environmental change. OCB brings together scientific disciplines to cultivate new research areas and opportunities in the US and with international partners. OCB community activities have precipitated new collaborations, funded projects, field campaigns, publications, and science plans to inform agency investment. With funding from NSF and NASA, OCB maintains a project office at WHOI. The OCB Project Office plays multiple important support roles for this network, including coordination of workshops and other community activities on emerging research issues; serving as a central communication hub; national and international science planning; development and dissemination of education and outreach materials; and providing training and networking opportunities for early career scientists.

Since METS data represent an observational cornerstone of OCB research, the OCB Project Office has led or collaborated in several US and international METS efforts, including:

- **Scientific subcommittee on ocean time series** (2007-present) - focuses on shipboard ocean time series and their synergies with other ocean observing assets such as autonomous- and satellite-based platforms; works with OCB Project Office to facilitate communication across METS data producers, users, and funding agencies; help plan community workshops and activities; and develop products and publications that increase the visibility of METS
- **Sea Change: Charting the course for ecological and biogeochemical ocean time series research** (2010) – workshop that convened scientists involved in three NSF-supported time series BATS, 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)
- **Global intercomparability in a changing ocean** (2012) - international time series workshop that focused on methodologies used in sample collection and analysis across METS (Benway et al., 2013; Lorenzoni and Benway, 2013)
- **International Group for Marine Ecological Time Series (IGMETS)** (2013-2017, led by IOC) – a cross-time series synthesis effort 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 drivers and potential connections to regional climate indices and anthropogenic changes (O'Brien et al., 2017)
- **Ocean Sciences Town Hall Meeting** (2016) - OCB led a Town Hall at the 2016 Ocean Sciences Meeting to gather input from the oceanographic community on how to improve data access, build a broader community of data users, develop new cross-time series synthesis and education/outreach products, and improve communication and collaboration across time series (Neuer et al., 2016; Neuer et al., 2017)
- **OceanObs19 White paper** (2019) - OCB and its ocean time series committee led the development of a broad decadal vision for METS as a contribution to OceanObs19 (Benway et al., 2019)
- **EarthCube Ocean Time Series Data Meeting** (2019) - OCB led a small EarthCube-sponsored meeting to assess METS data gaps and challenges as described above (Benway et al., 2020)

BCO-DMO (https://www.bco-dmo.org) was created in 2006 to assemble, curate, publicly serve, and archive data and related products resulting from biological, chemical and biogeochemical research conducted in coastal, marine, great lakes, and laboratory environments. The office works closely with NSF-funded and other grant-sponsored oceanographic researchers throughout the full data life cycle, helping them fulfill funder requirements, educating them in basic data skills, and contributing their data to a growing database available for new research. The BCO-DMO database contains >9,000 datasets from highly diverse oceanographic and limnological measurement types, including those from: in situ sampling, moorings, floats and gliders, sediment traps; laboratory and mesocosm experiments; satellite images; derived parameters and model output; and synthesis products from data integration efforts. Since its inception, the office has cultivated a positive and fruitful relationship with its research community, and has worked with over 2,600 data contributors from more than 1,000 projects,

including those of sustained time series efforts such as HOT, BATS, CARIACO, the Ocean Particle Flux (OPF), and the San Pedro Ocean Time-series (SPOT).

BCO-DMO has historically worked with individual time series projects to manage METS data, and curation practices have evolved to incorporate emerging FAIR principles, including: performing gross quality control and reformatting, capture of data processing and provenance into metadata, assembly and curation of standardized robust metadata necessary (for both humans and machines) to discover and reuse the data, and publishing data package versions online with a usage license and a Digital Object Identifier (DOI). A copy of each curated package version is stored locally and deposited at an appropriate national archive (e.g., NOAA's National Centers or Environmental Information), as well as the WHOI and Marine Biological Laboratory (MBL) Institutional Library's Open Access Server, WHOAS) for long-term preservation.

In addition to data curation, BCO-DMO is committed to data management capacity building efforts, improving data literacy and increasing science engagement in data management topics through education, training, and outreach. The office participates in the development and use of open-source, standards-based technologies that enable interoperable data systems of all domains to exchange data and information that will facilitate next-generation research. For example, BCO-DMO is an active member of the Council of Data Facilities, where it is uniquely positioned to communicate and promote domain-agnostic cyber-based solutions to the broader geoscience data and science communities. By collaborating with METS project PIs through FAIR data management practices, BCO-DMO can facilitate the integration of time series data with other diverse datasets in order to enable researchers to achieve a deeper understanding of ocean ecological and biogeochemical systems. With OCB, this office is uniquely poised to facilitate community consensus to advance a unified METS data model.

3.2 RCN Activities

The METS RCN will support coordination efforts that bring together different cross-sections of the METS community (i.e. data producers, users, scientists, and managers) in large- and small-group formats to foster the necessary dialog to develop FAIR data solutions and practices. Specific objectives include: 1) **building consensus on and developing reference implementations of a data model** for adoption by the broader METS community; 2) **identifying the needs of a broader range of METS data end users** and associated data interfaces and tools to meet those needs; 3) **building capacity to ingest, analyze, and integrate METS data** with other disciplinary and cross-disciplinary data to accelerate scientific discovery.

<u>Consensus building to advance METS data cyberinfrastructure</u>

For many of the biogeochemical and biological parameters that are unique to METS, there is an urgent need to develop consensus on community-adopted data and metadata reporting standards that will make these data FAIR. To make progress, the community needs more opportunities to work together with data scientists and database managers to co-develop and adopt a common set of FAIR-enabled best practices for METS cyberinfrastructure.

At the 2019 EarthCube workshop, participants identified existing vocabularies and data frameworks that could be leveraged in this process (Benway et al., 2020). Any recommended or "endorsed" vocabularies should be accessible through expression in schema.org or extensions thereof into community-vetted ontologies, and compliant with current W3C Semantic Web recommendations. Potential existing candidate vocabularies include NERC, EnvO, or CF. The application of Uniform Resource Identifiers (URIs) to data and metadata would enable consistent description of structured data and improve the efficacy of searching for desired information over

the Web. The METS community must agree on a minimum metadata model that defines key identifiers within datasets (parameters, methods, instrumentation, etc.). If based on common syntax and semantics, such a metadata specification can serve to "interoperate" across databases that have distinct underlying data models. Furthermore, smart design of data systems can ensure that different formats are interoperable via tools such as the Environmental Research Division's Data Access Program (ERDDAP), which can translate data from many different types of file formats, databases, and services and provide access via a single standardized interface. The METS community must also build consensus on standardized templates for data submission, QC (quality control) flags, digital representation of date/time, and data citation (e.g., DOI assignment and associated versioning protocols).

The first activity of the METS RCN will be a community workshop focused on consensus building planned by the RCN steering committee that builds on the gaps and resources identified at the September 2019 EarthCube meeting. This workshop will engage members of the METS biogeochemistry, marine ecology, and data science communities, including representatives of well-established US-funded coastal and open-ocean METS that intend to contribute to the development and implementation of a common set of FAIR data solutions. We will also engage participation and input from a couple of long-running non-US METS to ensure connectivity to international METS and ongoing global ocean data practices. The format will consist primarily of small group work that maximizes participant interaction and brainstorming. The workshop will be built around key METS issues that align with challenges identified during the September workshop (see above - e.g., adoption of a common vocabulary, development of a minimum metadata model, data citation guidelines, etc.). All participants will rotate through all of these challenges providing input in mixed groups to initiate the consensus building process. After gathering participant input on each challenge, we will form dedicated **task teams** for each challenge, each with a mix of oceanographic and data science expertise. The task teams will spend the remainder of the workshop identifying the best path forward based on the input gathered (i.e. best practices) and developing prototype solutions to be shared across task teams. Each task team will appoint two leaders, an ocean scientist and a data scientist, to serve on a **METS Data Working Group** that will be sustained throughout the duration of the RCN to continue interfacing with RCN participants and the broader METS community on FAIR data practices for METS.

<u>METS Data Working Group</u>

In the first year of the METS RCN, we will establish a working group of marine scientists, data scientists, and data managers from the **task team** leadership of the community consensus-building workshop. The working group will meet virtually and in person, in conjunction with RCN activities and other national meetings, to identify use cases that will drive development of reference implementations of FAIR METS data framework components.

This effort will leverage BCO-DMO expertise and its ongoing efforts to develop common oceanographic data type models. Here BCO-DMO is employing semantic web technologies to help overcome data integration challenges associated with marine data types. Semantic technologies like ontologies define the schema, axioms and relationships between data. By using ontologies, the working group can openly share the knowledge of how data are related, but also how data could be formatted and shared. A data model for reporting time series parameters, encoded as an ontology, communicates to both humans and machines the expectations of both data submitted as times series, as well as how time series data can be interpreted by users. A data type ontology can express the meaning of these parameters, the expected units for each

parameter, and any other useful knowledge required to properly share time series data across the user and submitter communities.

Many of the targeted NSF-supported METS projects currently work with BCO-DMO to share their data through the BCO-DMO system. However at present, these activities are unique to each project and vary in timing and effort across the METS community. Through this working group, the office will collaborate to identify required and optional time series parameters, along with their associated attributes, to create a community-wide data model for METS data sharing. This work will be incorporated into data model development already underway within BCO-DMO.

Broadening METS data end users

METS data currently serve a wide range of ocean scientists studying marine ecology and biogeochemical cycles. These long-term, consistently sampled ocean observations are valuable multidisciplinary assets that have provided transformative insights into how marine ecosystems function and how they are changing. However, a lack of common data reporting standards not only limits scientific capacity to compare and synthesize data across sites to observe geographic patterns of variability and change, but causes METS data to remain an untapped resource for ocean and earth system modelers, resource managers, educators, students, etc. (Benway et al., 2020). Facilitating an ongoing dialog with broader end-user communities and documenting outcomes of these exchanges is necessary to enhance the regional and global knowledge gained from sustained time series measurements.

As part of the METS RCN, we will develop novel **regional METS user networks** around a small subset of US-funded METS (e.g., HOT, BATS, CalCOFI, etc.) to build and strengthen links to interested scientists (including observationalists and modelers), resource managers (fisheries, water quality, etc.), and educators (informal, K-12, colleges and universities, etc.). We will initiate this work virtually and then as the user networks start to take form, we will convene a workshop of the METS user network participants to share use cases in order to identify data and information needs. Using a *Design Thinking*-inspired approach (Fig. 1), participants will work with data scientists and data managers to identify the distinct information needs of different groups (e.g., modelers, resource managers, educators, etc.), brainstorm on a core set of data formats and data interface capabilities to best meet those needs through use cases from each group.

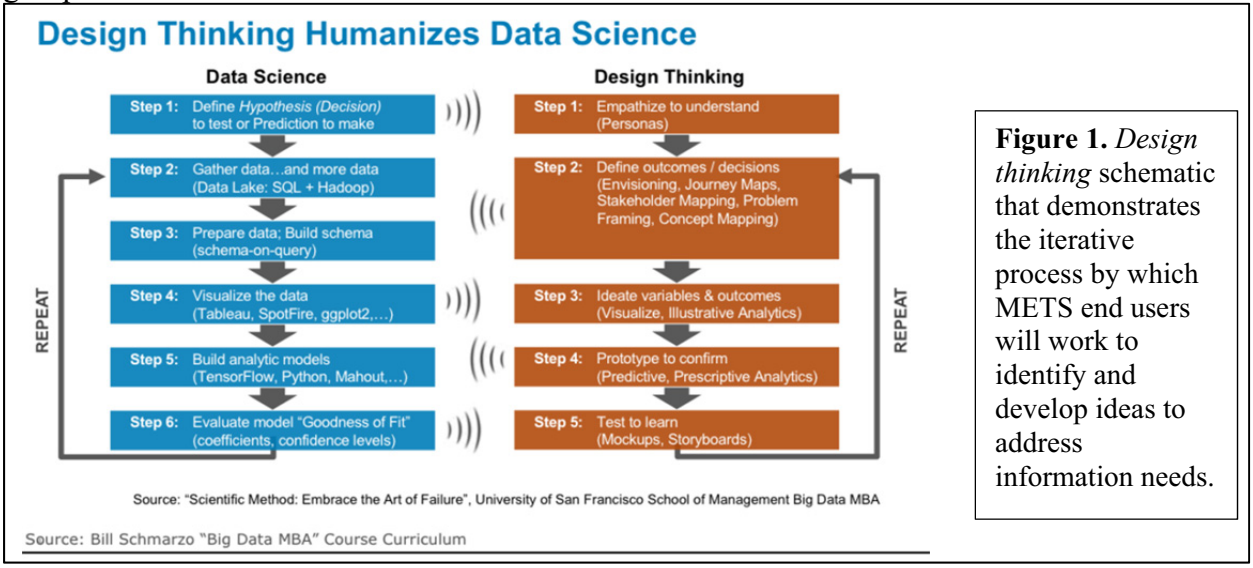

Figure 1. *Design thinking* schematic that demonstrates the iterative process by which METS end users will work to identify and develop ideas to address information needs.

Building data analysis capacity

To date, METS data have primarily been used to address questions and support the scientific goals of individual METS and ancillary projects. An important challenge and opportunity to enhance the scientific impact of METS is to build capacity for data analysis and synthesis across METS and with complementary multi-platform observations to address spatial patterns of biogeochemical variability and change. This RCN's development of a FAIR METS data model framework will be critical in supporting these efforts. To build capacity among emerging leaders in the oceanographic community and foster hands-on instruction, peer learning, and collaboration, the METS RCN will convene a **METS data hackathon** geared toward early career participants (student, postdocs, new faculty) that follows an ***Oceanhackweek*** format, which is described below:

> *"The hackweek model has emerged within the data science community as a powerful tool for fostering exchange of ideas in research and computation by providing training in modern data analysis workflows. In contrast to conventional academic conferences or workshops, hackweeks are intensive and interactive, facilitated by three core components: tutorials on state-of-the-art methodology, peer-learning, and on-site project work in a collaborative environment. This setup is particularly powerful for sciences that require not only domain-specific knowledge, but also effective computational workflows to foster rapid exchange of ideas and make discovery…."*

While a great deal of funding is invested in the collection of METS and other oceanographic data, there is little explicit investment in the follow-on synthesis and data analysis work that is needed to realize the full scientific potential of these data. The oceanography community faces "big data" informatics and cyber-infrastructure challenges, including combining very large oceanographic datasets spanning a wide range of spatiotemporal scales, as well as data derived from multiple platforms and disciplines. The METS RCN will accelerate progress toward *Harnessing the Data Revolution* (NSF Big Idea) by building analysis tools and capacity to integrate diverse datasets. Lead PI Benway has been in contact with Oceanhackweek lead organizers to query alignment of METS RCN capacity building objectives with the Oceanhackweek model and has received a positive response (see letter of collaboration from W.-J. Lee, UW). Potential hackathon modules might include:

- **Statistics and visualization** – Combining data across METS to view temporal trends in a broader spatial context provides insight on marine ecosystem links to regional climate indices, local anthropogenic impacts, etc., which can inform prediction and decision making. For example, T. O'Brien (RCN Steering Committee) has adapted and applied statistical methods to analyze trends in METS variables from globally distributed time series stations against a backdrop of satellite data (O'Brien et al., 2017).
- **METS data for modelers** – It is important for time series data to be assimilated into numerical modeling frameworks that may elucidate cause and effect scenarios not easily perceived through simple statistical analyses. To facilitate increased use of time series data by the modeling community and improve communication and exchanges between observationalists and modelers, the use of shared repositories such as *GitHub* can stimulate the development of community-driven, open source code for extracting, quality-controlling, and gridding time series data. Simply creating shared post-processing scripts that can be tailored for each individual modelers' needs will significantly decrease duplicative efforts, increase access to time series data, and improve validation of numerical models.

- **Data integration** - Monitoring ocean change requires a sustained, globally distributed network of observatories that integrates shipboard, autonomous, and remote sensing platforms, which is generating increasingly complex data streams. Many METS sites have actually incorporated autonomous assets, including gliders, profiling floats, and wirewalkers to name a few, and many ancillary project participants deploy complementary instrumentation on an ad hoc basis. The integration of data across these different platforms is an important area for building computational capacity.

3.3 METS RCN Participants and Diversity

The METS RCN will include participation from a broad cross-section of the inherently interdisciplinary ocean and data sciences communities, as well as new METS data users such as educators, resource managers, and modelers. Given OCB's history of coordinating METS activities, a large network of METS scientists has already been amassed, representing a good starting point for the METS RCN. Activities will largely focus on US-based METS, but will include a small number of non-US collaborators who have been strong contributors to FAIR data discussions (e.g., Sept. 2019 workshop) and a larger METS vision (Benway et al., 2019), which is critical to ensure connectivity to international METS data and science efforts. A preliminary list of METS RCN participants that have either participated in previous OCB-led METS activities or are representatives of US-funded METS that we intend to engage in the RCN activities is included in the supplementary personnel list. METS RCN activities will not be invitation-only; they will be open to the community, so participation from additional METS scientists is anticipated and will be strongly encouraged. To achieve RCN goals, we will also use our networks and RCN Steering Committee members to engage broader end users (TBD) and early career scientists (students, postdocs TBD) as the RCN progresses. As with all of its activities, OCB will promote balanced representation with regard to gender, career stage, ethnicity, etc. in METS RCN activities, with an emphasis on engagement of early career participants in the capacity building activity.

3.4 METS RCN Outcomes and Deliverables

The proposed RCN activities will address several METS grand challenges of the coming decade (Benway et al., 2019), including building consensus on key elements of a FAIR data model, broadening users and applications of METS data, and building METS data synthesis and analysis capacity through immersive, hands-on training and collaboration. All information and products of the METS RCN, including publications, best practices documents, code-sharing repositories, data and networking templates and recommendations, etc. will be vetted as necessary by the community (open review/commenting periods) and openly shared with academic and broader user communities via the large OCB network (web, eNewsletter, social media) and METS-focused community list-serves, as well as US and international OCB partner program networks. Specific anticipated outcomes of the RCN activities are described below.

During the first 18 months of the RCN, FAIR METS data model recommendations generated during the first RCN activity, the **Consensus Building Workshop**, will be vetted with the broader METS community during an open comment period and then developed into a peer-reviewed publication in an open access journal (e.g., *Earth System Science Data*). All METS data standards and best practices will also be compiled in the recently established Ocean Best Practices System, which provides a centralized and trusted mechanism to support the collaborative development, sharing, and adoption of best practices across the global

oceanographic community. BCO-DMO will work with the **METS Data Working Group** to build and iterate on a METS community data model with members of the RCN.

The regional user networks initiated as part of the **Broadening Data Users** RCN activity will broaden METS data applications and serve as a model of regional engagement for other METS. Based on use cases developed at this workshop, METS data technician L. Fujieki (HOT) will develop draft data interface tools to facilitate broader METS data applications (trend analysis, property-property plots, contour plots, data querying capacity, etc.) that can be tested by workshop participants and the broader community before being implemented by other METS user networks. This expanded METS user community and "toolbox" will greatly increase the impact and applications of METS data.

The **METS Data Hackathon** will foster new collaborations and development of new tools for analyzing and visualizing METS data. Participant interactions will facilitate the compilation of open source code to streamline the assimilation of METS data into ocean and earth system models. All new tools and code repositories will be shared broadly via the OCB and METS communities, and submitted to the EarthCube Resource Registry for discovery by the broader EarthCube and geoscience communities.

4. METS RCN Management

In consultation with the full RCN Steering Committee and BCO-DMO, OCB (PI: H. Benway, WHOI) will oversee the management and implementation of the METS RCN, including the planning and coordination of RCN activities. OCB has 14 years of experience bringing interdisciplinary groups together to address important questions, build consensus, and achieve high-impact outcomes, making it an effective incubator for new ideas and activities. The OCB Project Office at WHOI is a coordination and communication hub for a large oceanographic network, with established infrastructure and dedicated staff capacity (scientific leadership, communications, logistics) to support community building activities, as well as an extensive network of METS scientists with whom we have worked for over a decade to address challenges related to data, methods, and science. The OCB Office is equipped to manage RCN participant communications, including regular web/teleconferences of the METS RCN Steering Committee and METS Data Working Group. OCB will establish and maintain a web and social media presence for the RCN and provide logistical support for RCN activities.

The METS RCN will also leverage the infrastructure, expertise, and extensive METS data handling experience of the co-located BCO-DMO office (Co-PI: D. Kinkade, WHOI), a world-renowned data management facility for biogeochemical and biological datasets. BCO-DMO brings a wealth of expertise in data science and technology, as well as experience working with heterogeneous oceanographic data. As the designated repository for multiple NSF-funded METS, BCO-DMO will interface with RCN participants to design and develop a FAIR METS data model.

Co-PI A. White (Univ. Hawai'i) brings a wealth of oceanographic expertise and hands-on experience in the collection and scientific applications of METS data. As the lead PI of one of the longest running and well respected METS in the world, she will provide leadership to METS community discussions of data challenges and solutions, and oversee the development of new data interfaces and products at HOT to serve as a template for other METS projects.

5. METS RCN Evaluation

Information about the establishment and activities of the METS RCN will be broadly disseminated to OCB and international METS networks to ensure connectivity to complementary

efforts. This collective network of >3,000 scientists will be engaged to test and evaluate RCN products. In addition to surveying participants of each RCN activity on the effectiveness and value of the activity, we will query these broader networks annually on the RCN's overall progress, broader engagement, and efficacy of products in meeting the stated goals and objectives of the RCN. This will enable METS RCN leadership to change course as necessary throughout the duration of the RCN to address perceived deficiencies. Anonymous survey results will be shared via the METS RCN website. If possible, METS RCN leadership will engage a representative from another EarthCube RCN in some of its activities and virtual meetings to provide external evaluation from an unbiased party.

6. METS RCN Participation in EarthCube

The METS RCN will bring a novel ocean data type to the EarthCube landscape. RCN goals and objectives align well with EarthCube's long-term vision of a community-driven cyberinfrastructure to facilitate data interoperability that can lead to interdisciplinary research and help educate scientists in the emerging practices of data stewardship and open science.

RCN leadership is keen to interface with other RCNs and EarthCube activities, particularly time series data, to discuss common data challenges and share approaches and insights. METS RCN Co-PI and Steering Committee member D. Kinkade has been active within the EarthCube program, having participated: in EarthCube governance by serving on the Leadership Council, as lead of a Domain End-User Workshop, co-PI or collaborator on several EarthCube-funded projects, and as member of the Technology and Architecture Committee (TAC). Co-PI Kinkade is currently serving as Assembly Chair of the EarthCube Council of Data Facilities (CDF) and has represented METS efforts at previous EarthCube meetings (e.g., Sept. 2019 workshop). Lead PI Benway has been in contact with EarthCube governance liaisons from the Sept. 2019 workshop Emma Aronson (Science Committee Chair) and D. Sarah Stamps (Science Committee Co-Chair) about this RCN proposal, and is aware of ongoing discussions about streamlining EarthCube governance by combining the Liaison and Engagement Teams with the Science Committee into a single governance body, the Science Engagement Team (SET). Regardless of the final decision, co-PIs Benway and/or White are prepared to participate in Earthcube governance as appropriate. The RCN leadership also plans to reach out to the EarthCube Council of Funded Projects to explore points of potential collaboration with other EarthCube projects (e.g., possible synergies with the Cloud-Hosted Real-time Data Services for the Geosciences (CHORDS) project for services applicable to real time METS data streams).

In addition, the RCN will participate in sharing its progress and output through EC Annual Meetings, scheduled webinars, and by registering any subsequent RCN-facilitated infrastructure in the EC Resource Registry (e.g., interactive data processing or analysis notebooks as a result of hackathons) and Tools Inventory as appropriate. By virtue of leveraging BCO-DMO (a CDF member repository), this project will ensure METS data will comply with FAIR principles and that its domain community practices align with the EC position on promoting FAIR within the NSF geosciences. METS data stewarded by BCO-DMO will comply with EC strategies for discovery as developed and promoted by project P418/P419 (i.e. published using schema.org).

BROADER IMPACTS

The METS RCN will bring together a diverse cross-section of the ocean and data sciences for a sustained dialog to address long-standing METS challenges, most prominently the lack of a consistent and FAIR data model. This centralized coordination effort will facilitate development of community-driven METS cyberinfrastructure standards and best practices that improve

interoperability and integration of METS data. The application of semantic technologies will enhance scientific and broader applications of METS data as we enter the United Nations Decade of Ocean Science for Sustainable Development (2021-2030). The RCN will also provide networking and training opportunities for scientists across career stages and disciplines of data and ocean science. Tangible outcomes will include community-vetted best practices publications, shared data analysis and interface tools for scientific and broader applications, and network models for regional stakeholder engagement.

A METS RCN Consensus Building Workshop and METS Data Working Group will foster community discussion and build consensus on METS data model elements that will contribute to the broader geoscience data infrastructure, with reusable technologies that are applicable across domains. The Broadening Data Users RCN activities will create regional networks of METS data users (modelers, educators, decision makers, etc.) focused on querying user-specific information needs and developing a suite of METS data interface tools for testing and broad distribution. Strengthening ties to a broader end user community will form the basis for new collaborations and products to enhance regional knowledge, climate prediction, and decision making, which will greatly increase return on investment in these observing programs. The METS Data Hackathon will promote capacity building in the data and ocean sciences communities, equipping participants with novel statistical, visualization, and data analysis tools to integrate across METS and with other multi-platform ocean datasets and support scientific synthesis, process understanding, and prediction.

RESULTS FROM PRIOR NSF SUPPORT

Heather M. Benway: *Ocean Carbon Biogeochemistry Project Office* (NSF OCE 1850983, $4,410,308, August 2019 to July 2024)

Intellectual merit: OCB is a network of scientists who work across disciplines of oceanography to understand the ocean's role in the global carbon cycle and the response of marine ecosystems and biogeochemical cycles to environmental change. OCB works with scientists and federal agency managers to cultivate new research areas and opportunities in the U.S. and with international partners. OCB is a bottom-up organization that responds to the continually evolving research priorities and needs of its network and engages marine scientists at all career stages. OCB plays multiple important support roles for its network as follows: Coordinates workshops and other community activities on emerging research issues; serves as a communication hub to broadcast scientific news, funding and job opportunities, new OCB-relevant research, etc.; engages with relevant national and international science planning initiatives; develops and disseminates education and outreach activities, tools, and products to share ocean carbon science with broader audiences; facilitates training and networking opportunities and engages early career scientists in OCB and partner program activities. Through its previous and current suite of interdisciplinary activities, OCB has built strong collaborative networks of scientists that are poised to contribute to several of NSF's 10 Big Ideas. To promote cross-fertilization and enhance its community building efforts, OCB will continue to collaborate with partner programs to advance progress on scientific areas of overlapping interest such as coastal carbon cycling; ocean carbon uptake; planning for more integrated chemical-biological observing campaigns; capacity building around new ocean observing technologies and associated training opportunities; and air-sea exchange of greenhouse gases and associated impacts on marine ecosystems and

biogeochemical cycles. OCB will also continue to provide coordination and leadership on community planning initiatives in support of OCB science and relevant observing programs.

Broader Impacts: OCB is a bottom-up organization that serves in coordination, communication, training and outreach roles that ultimately benefit its entire network, the broader oceanographic community, and the public. OCB will continue to engage scientists and broader audiences on issues pertaining to the ocean carbon cycle through early career engagement and travel support; development and distribution of education and outreach products; and high-impact communication platforms, including a new, more navigable and attractive website (us-ocb.org), topical informational websites (e.g., ocean fertilization), a highly subscribed (>1,800 members) eNewsletter, and social media platforms.

Product(s) and Data Archive: All publications from OCB activities have been broadly disseminated via OCB communication channels and are archived on our website. They include the five publications indicated with a § in the References.

Angelicque E. White: *Collaborative Research: Measuring Ocean Productivity from the Diurnal Change in Oxygen and Carbon*. (NSF OCE 1536866, $359,108, Sep 2015 to Aug 2020)

Intellectual merit: The diel changes of dissolved O_2, O_2/Ar, and Particulate Carbon (PC, and optical proxies thereof) were measured during several research cruises (HOT294, KM1712, and KM1713) in the N. Pacific. Key findings included a 3-fold increase in GPP, CR, and export from subtropical to subpolar regions and enigmatically consistent ratios of GPPo/GPPc (=1.4±0.2), CRo/CRc (=1.0±0.2) and GPP/NPP (=1.8±0.2) for both the subtropical and subpolar regions. To our knowledge this is the first robust comparison of in situ, incubation-based, shipboard, and autonomous O_2 and C based diel-measurements of gross and net production.

Broader impacts: With PI White, 2 female postdoctoral scholars (Fernanda Henderikx-Freitas and Macarena Burgos) were mentored and partially supported by this project. As part of our postdoctoral mentoring plan, Henderikx- Freitas has been given a leadership role in dissemination of results and has presented project results at several national meetings. The results of this research were used in an undergraduate honors course (OC407H: Oceans, Coasts and People) led by AW.

Product(s) and Data Archive: Four peer-reviewed publications have been submitted or accepted detailing results from this work: *White et al. (2017), Janssen et al. (2020), Henderikx Freitas et al. (2020a) and Henderikx Freitas et al. (2020b*). One additional manuscript is in preparation detailing observed changes in community structure relative to productivity trends. Research results have been presented at Ocean Sciences in 2016, 2018, and 2020, and at Optics Conference in 2019. All data from our field campaign have been deposited in BCO-DMO (#792396). Publication from this grant include the four indicated with a † in the References.

Danielle Kinkade: NSF OCE- 1924618 - PIs: M. Saito, D. Kinkade, A. Shepherd, P. Wiebe; 9/1/19 to 8/31/24; $ 11,580,954. *BCO-DMO: Accelerating Scientific Discovery through Adaptive Data Management.*
Intellectual Merit: This grant supports core data management of NSF OCE funded project data through the BCO-DMO repository containing over 9,000 diverse oceanographic datasets, and the implementation of augmented technologies and architecture. BCO-DMO also actively

contributes to informatics and data management science and is broadly considered a leader in this effort.
Broader Impacts: Data from BCO-DMO is widely used throughout the world for research and education, and the repository engages in training and cross-pollination of informatics with domain science students and early career individuals. BCO-DMO's website has been accessed from 216 countries worldwide and is has been used as a teaching tool for graduate classes in various programs.
Product(s) and Data Archive: BCO-DMO's database and website (https://www.bco-dmo.org)

NSF OCE-1435578 - PIs: P.Wiebe, A. Shepherd, M. Saito, D. Kinkade; 9/1/14 to 8/31/20 (with NCE); $9,600,001. *Biological and Chemical Oceanography Data Management Office (BCO-DMO): A System for Access to Ecological and Biogeochemical Ocean Data.*
Intellectual Merit: This grant supported the BCO-DMO office core mission to manage output from NSF OCE funded projects, in addition to contributing to development of new information infrastructure.
Broader Impacts: Data from BCO-DMO is widely used throughout the world for research and education. BCO-DMO's website has been accessed from 216 countries worldwide and is has been used as a teaching tool for graduate classes at URI.
Product(s) and Data Archive: BCO-DMO's database and website; Wiebe et al., 2015; Wiebe and Allison, 2015.

NSF ICER-1639714 - PIs: M. Saito, D. Kinkade; 9/1/16 to 8/31/19. $499,595 *EarthCube Data Infrastructure: Laying the Groundwork for an Ocean Protein Portal.*

Intellectual Merit: This collaborative project developed the Ocean Protein Portal for sharing ocean metaproteomic data launched in February of 2019. This grant and development also served as a testbed for the adaptation of new software infrastructure used in BCO-DMO architecture upgrades improvements.
Broader Impacts: This grant funded an Ocean Metaproteomics Workshop that brought together domain scientists and informaticists and led to the best practices publication, and the successful development of website for the sharing of ocean metaproteomic data.
Product(s) and Data Archive: Ocean Protein Portal and Metatryp websites; Held et al., 2018; Saito et al., 2019. Publications from this grant include the four indicated with a ‡ in the References.