DATA MANAGEMENT PLAN

Version 2-00
Document Control Number 1102-00000
2014-03-21

Consortium for Ocean Leadership
1201 New York Ave NW, 4th Floor, Washington DC 20005
www.OceanLeadership.org

in Cooperation with

University of California, San Diego
University of Washington
Woods Hole Oceanographic Institution
Oregon State University
Scripps Institution of Oceanography
Rutgers University

Document Control Sheet

Version	Date	Description	Originator
1-00	Nov 1, 2010	Initial Draft	B. Bergen
1-01	Nov 12, 2010	Post-San Diego Mtg-adoption/modification of TOC; rearrangement of content; correction of some terms	F. Vernon/team
1-02	Nov 12, 2010	Cost section added	J. Graybeal
1-04	Nov 26, 2010	Completion of first draft from DMP committee/working group	DMP working group
1-05	Nov 30, 2010	Cleanup	B. Bergen
1-06	Dec 1, 2010	Reviewed for Creative Commons and Science Commons usage; Incorporated WWP's comments	B. Bergen
1-07	Dec 6, 2010	Formatting, copy edits	E. Griffin
1-08	Dec 13, 2010	Incorporation of Pre-CCB review comments; clarification of roles and responsibilities; addition of TBD/TBR table; addition of "Candidate preamble"	Bill Bergen, Frank Vernon
1-09	Dec 16, 2010	Modified/updated sections on physical samples and sampling strategy	Frank Vernon, Bill Bergen
1-11	Jan 3, 2011	Formatting, copy edits	E. Griffin
1-12	Jan 17, 2011	Committee updates, addition of 3 TBRs	B. Bergen
1-13	Feb 28, 2011	Additional committee updates; changed archive to storage; changed "must" to "should" in acknowledgement section of A-2; changed Observation strategy working group to Sample Strategy Working Group; restructured TBD/TBR table per committee direction	B. Bergen
1-14	Mar 2, 2011	Formatting for ECR submission	E. Chapman
1-15	Mar 11, 2011	Liens from ECR #1300-00136. Initial Release.	E. Chapman
1-16	Aug 10, 2011	Revision of section 4.5 and section 4.1	L. Heilman
1-17	Aug 30, 2011	Revised sections 4.1 and 4.5, revised data/product to data/data product throughout entire document, replaced Data/Product Working Group with Algorithm Working Group throughout document based on liens from ECR submitted to and accepted at CCB on 8/25/2011.	L. Heilman

Formatted: Font color: Black

1-18	Aug 31, 2011	Revised sections 2.3.2, 3.5, 3.10, 4.1, 4.2, 4.4, 4.9, 4.10.1, 4.10.3, 4.13 (new numbering) as a result of open tasks completed. Removed section 1 "Additional DMP Effort Remaining" and document disclaimer.	R. DelCoco
1-19	Sept 28, 2011	Revisions based on comments to ECR 1300-00199	
1-20	Nov 7, 2011	Edits throughout. Changed mode definitions, changed level definitions	L. Heilman
1-21	Dec 22, 2011	Removed appendices, changed mode and level definitions, removed definitions of QA/QC and pointed to Ref module; removed working groups, defined child documents	B. Bergen
1-22	Feb 12, 2012	Incorporated review changes from CI	B. Bergen
1-23	March 5, 2012	Incorporated additional comments from CG and CI	B. Bergen
1-24	March 26, 2012	Incorporated changes S. Ackleson, removed senior project scientist position; typos. Addressed comments from CCB reviewers	B. Bergen, ECR #1300-00228
1-25	March 30, 2012	Formatting, copy edits	E. Griffin
1-26	Aug 5, 2013	Updated Data Product Algorithm definition iaw ECR 1300-00368	E. Chapman
1-27	March 20, 2014	Major update per initial deployments and changes to CI scope... Picked up and included unapproved changes (from NSF) for versions -27 and -28. Included input/reviews from ML, MV, AJP	B. Bergen, S. Banahan, E. Chapman, J. Thomas, K. Stocks, E Griffin, T Cowles
2-00	Mar 21, 2014	Resolution of liens from 1300-00422.	B. Bergen

Formatted: Font color: Black

Table of Contents:

Formatted: Font color: Black

Formatted: Font color: Black

1 Executive Summary

Effective management and storage of data are fundamental requirements for successful scientific research. Oceanographic research depends on the availability and clarity of existing data. The Consortium for Ocean Leadership (COL) through the Ocean Observatories Initiative (OOI) and partner institutions plan, design, deploy, and operate an integrated, sustained scientific ocean observatory as NSF's contribution to the U.S. Integrated Ocean Observing System (IOOS). Internationally, IOOS is the U.S. contribution to the Global Ocean Observing System (GOOS) and the Global Earth Observation System of Systems (GEOSS). In fact, much of this Data Management Plan (DMP) is based on the work contained in the IOOS Data Management and Communications (DMAC) documents. OOI is a network of regional, coastal, and global arrays that rapidly and systematically acquires and disseminates data/data products to serve scientific/societal needs.

This DMP is divided into two main parts. Part I addresses the management aspects of data/data products. Specifically, it addresses:

- Governance
  - Organization
  - Interfaces
  - Training Needs and Schedule
  - Data/Data Product Management Tools
  - Required Facilities
- Procedures/Policies
  - Data/Data Product Retention
  - Data Policy
  - Algorithms
  - Compliance and Review
  - Data/Data Product Release, Use, Rights, and Rules
  - Cost Awareness
  - Security
  - Transition at End of Program
  - O&M Aspects of Data

Part II presents important aspects of the data and how this document, other plans, and other related OOI documents approach the technical details associated with data/data products. Specifically, Part II is organized to discuss:

- Basic OOI data/data products
  - Data Product Levels
  - Data Modes
  - Flow through the OOI System
  - OOI Data Product List
  - Metadata (Including Engineering Data, photographs and associated products)
  - Standards
  - Volume of data
  - Acquisition/Collection
  - Data/Data Product and Information Tools
  - Processing
  - Storage
  - Access/Delivery
- Variants of above data products for independent Principal Investigator (PI) data products

This DMP is the pinnacle in a series of documents that addresses OOI data/data product management. The DMP does NOT specify the implementation details for data/data product generation and management but rather provides the guidance upon which implementation is

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

based. Implementation is via a series of data/data product specifications, procedures, and work instructions.

This Plan presents a short overview of OOI and provides both an organization/management then technical focus on the issues of data collection, transmission, quality control processing, storage, discovery, and access. As one of many OOI Program documents, the DMP is developed, implemented and updated according to the planning and governance procedures described in OOI governing documents such as the Project Execution Plan (PEP), System Engineering Management Plan (SEMP), and the OOI Final Network Design (FND). It is important to recognize that the DMP is a "living document." It is changed and adapted as needed throughout the lifetime of the OOI.

The DMP describes the approach to detailed procedures and implementation steps for system-wide management of OOI data, so that users of the OOI may receive and/or retrieve OOI data in a uniform manner. Underlying the OOI Data Transport mechanism is the unifying vision of services. Through services and various types of client applications—for example, tools for end-users, modelers, and planners and value-added marine information web sites— users can access data from the OOI distribution points (servers). OOI provides the means to connect the OOI network to data management systems operated by marine data partners and by partners in other disciplines such as meteorology. The DMP calls for the use of standards such as OPeNDAP.

OOI data management consists of dedicated hardware-based facilities, software, and human data specialists. The DMP provides the foundation for the OOI observatory to utilize DMP standards and protocols and conform to OOI Data Policy.

As used in this Data Management Plan, data in the OOI are digital in nature and capable of electronic transmission. Besides the digital, temporal, and spatial data streams characterizing observational measurements, which are the commonly assumed forms of "data," the complete definition of data must allow for computed and imaged results, such as integrated or modified data streams, hand-written documents, technical reports, bathymetry, images, and metadata, including the analytical results and metadata from physical items such as water, sediment, rocks, and biological samples. In OOI, these "other forms of non-digital data" are converted digital representations and associated with instruments and/or their specific data products for inclusion within OOI.

2 Introduction

2.1 Overview

The OOI comprises observing nodes spanning Global, Regional, and Coastal scales and a computational and communication cyberinfrastructure:

- Global Scale Nodes - The Global observing nodes address mesoscale phenomena via a network of moored buoys and mobile platforms linked to shore via satellite for command and control plus limited near real-time data access.
- Coastal Nodes – The Coastal observing nodes comprise moored and mobile platforms that provide extended opportunities to characterize the effects of high frequency forcing on the coastal environments
- Regional Scale Nodes - The Regional cabled observing nodes 'wires' a region of the Juan de Fuca tectonic plate in the Northeast Pacific Ocean providing high-speed optical data transfer and high power to instrumented platforms.
- Cyberinfrastructure (CI) - The CI integrates the marine observing nodes; associated sensors; and land data processing, storage, and user distribution services into a coherent system of systems.

- The OOI collects ocean surface, water column, and seafloor data at specified and adjustable sampling rates over years to decades. Researchers can evaluate interdisciplinary measurements to investigate a spectrum of phenomena including:

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

episodic, short-lived events (tectonic, volcanic, biological, severe storms), and more subtle, longer-term changes or emergent phenomena in ocean systems (e.g., circulation patterns, climate change, ocean acidity, ecosystem trends). The availability of ample power and data rate to select parts of the ocean by the OOI provides the ocean science community, educators, policy makers, and the public with access to detailed data on multiple spatial and temporal scales. This enables studying the coastal, regional, and global-scale ocean, and using mobile assets (autonomous underwater vehicles, gliders, and vertical profilers) to complement fixed-point observations. Additional information regarding the OOI design and deployment is found in the Final Network Design (FND) document (DCN 1101-00000 Final Network Design).

2.1.1 Data Management Vision, Philosophy, Objectives, and Challenges

Central to the success of OOI and other regional, national, and international ocean and coastal observatories in delivering data to the user is the presence of comprehensive, integrated data/data product management.

The OOI focuses on the needs of end users in the implementation and operation of data management in order to maximize the application of OOI data to critical scientific questions. At a minimum, the OOI makes data/data products readily accessible, allows users to locate data and information products, and provides essential metadata (descriptive information about the data) along with the data/data products.

OOI is an integrated system of regional, coastal, and global observing networks that systematically acquires and disseminates data/data products to serve the needs of scientists, government agencies, educators, non-governmental organizations, and the public. The OOI vision is one of collaborative and integration of these various spatial scale arrays.

The design requirements of the OOI lead to three essential objectives for data management:

- Sustain, for decades, the delivery of multiple streams of ocean data, that address specific science themes including resolving ocean processes and properties across a range of temporal and spatial scales;
- Facilitate the use of the designed capabilities of the OOI for specific at-sea experiments or extension of the temporal and spatial reach of ocean research;
- Sustain the expansion capabilities of the observatory as technology evolves and in response to new science questions.

These overarching objectives are extended into a set of guiding principles that address the OOI vision:

- **Open and easy access and discovery**: OOI enables users from all over the globe to locate, access, and use the diverse forms of marine data and their associated metadata and documentation in a variety of computer applications (e.g., Geographic Information Systems- GIS, scientific analysis applications). Users are unencumbered, as much as practical, by traditional barriers such as specific/unique data formats.
- **Reliable, sustained, efficient operations**: OOI provides 24/7 delivery of select near-real-time data/data products from instrumentation to users with time-sensitive requirements. It provides high reliability in the delivery of delayed-mode and near-real-time data to users and sufficient capacity (e.g., computational power and communication links) for distribution of raw data and data products among high-volume users.
- **User feedback**: OOI provides an outreach process for marine data/data products, emphasizing the benefits of participation in OOI, and helping to identify and remedy difficulties encountered by those who are participating.

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

- **Open design and standards**: OOI is built on an open software design to the greatest extent possible. The design standards and protocols for external facing interaction (e.g., APIs) are available to the user community. Existing information technology and scientific standards are used in preference to development of new solutions, whenever suitable standards exist.
- **Preservation of data/products:** Observations, raw data, and select data products and associated metadata are stored for posterity in an automated manner. Data of long-term interest to specific domains are archived to domain-focused repositories wherever possible. All other data and associated metadata are maintained for the longest possible time given funding constraints. (See Cost Impacts and Awareness below.)

The OOI design faces a variety of challenges:

1. **Maintaining the continuous delivery of data/data products**: This challenge has two major components: a) a time variable (continuous or nearly continuous) and b) a quality variable (high). Traditional systems of hardware and software (HW/SW) commonly address the first variable. Attention to detail, scientific expertise, and vigorous production quality processes are all required to achieve a high quality data product. Production of high quality data products commonly requires considerable time. OOI strives to accomplish both components simultaneously.
2. **Algorithm maintenance for long-term time series analysis**: Over the course of the OOI, pre-processing and product generation algorithms mature and change. It is critical to capture this "induced" fluctuation in the data and data products so that long-term analyses do not misinterpret these changes. Creation and capture of detailed metadata is essential.
3. **Ensuring system adaptability and flexibility**: This supports long-term evolution to meet changing user needs and observing opportunities.
4. **Properly balancing opposing observational needs**: For example, a typically stable, long-term temporal and spatial observation strategy must be balanced with dynamically responding to unanticipated events of interest. It is paramount for OOI to support both of these needs so that data product sets can serve the largest possible user community. The OOI must be capable of emphasizing different observational strategies under different circumstances, as determined by the ocean science community.
5. **Achieving acceptance and recognition by marine data providers and data user communities**: Challenges for initial OOI acceptance and subsequent growth in its usage include OOI community outreach and organizational behavior. Challenges exist at both management and technical levels. Occasionally, short-term inconveniences to organizations may occur. The developed OOI observing arrays and components must conform to the expectations and standards of the user community.

2.2 Purpose and Scope of this Plan

2.2.1 Purpose

Effective management, storage, and delivery of data are fundamental requirements for successful scientific research endeavors. The data obtained through the OOI must meet Federal data policy requirements as well as be consistent with needs of users, comparable to other major programs in ocean science (e.g., ARGO, IOOS, WOCE). This document provides guidelines to implement Federal data policy by assuring timely submission of quality controlled oceanographic data to users, including distribution to national data centers. This Plan describes the overall approach for the production and management of the OOI data/data products and the processes invoked to accomplish the Plan.

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

2.2.2 Scope

This document is applicable to all OOI observing nodes, systems, subsystems, components, and policies and procedures directly or indirectly associated with the production and delivery of specified oceanographic data/data products. It includes the processes used by OOI and all of the Implementing Organizations (IOs) and as a result, the IOs do not maintain separate or subordinate DMPs.

This document serves as the 'parent' for the data management procedures the detail level work instructions that provide more specific implementation actions and day-to-day operations.

2.3 Documents

The following documents are maintained in the OOI Document Management System (DMS) and include those that are dependent (children) on this DMP as well as other additional documents for guidance and reference.

2.3.1 Dependent Documents

- 1102-00003 Pre-Commission Data Procedure
- 1102-00010_Data_Use_Policy_OOI
- 1102-00020_Data_User_Terms_Conditions_OOI
- 1102-00200 Observation and Sampling Approach
- 1102-00300_Protocols_Procedures_Data_Products_QAQC_Cal_Physical_Samples_OOI
- 1131- series CI-RSN Interface Control Documents
- 1132- series CI-CG Interface Control Documents
- 1133- series CG-RSN Interface Control Documents
- 1134- series CI-EPE Interface Control Documents
- 1321-00000 Metadata Implementation
- 1341-series Data Product Specifications Documents
- 1342-series Data Processing Flow Documents
- 1350-series Instrument Operational Specifications
- 1351-series Node Operational Specifications

2.3.2 Guidance Documents

- 1001-00000 Project Execution Plan (PEP)
- 1100-00000 Systems Engineering Management Plan (SEMP)
- 1101-00000 Final Network Design (FND)
- NSF 04-004 Division of Ocean Sciences Data and Sample Policy
- NSF Award & Administration Guide (AAG) Chapter VI.D.4.

2.3.3 Reference Documents

The following documents may provide useful information regarding data management.

- 1000-00000 Configuration Management Plan (CMP)
- 1003-00000 Quality Assurance and Quality Control (QA/QC) Plan
- 1007-00000 Risk Management Plan (RMP)
- 1010-00000 Operations and Maintenance (O&M) Plan
- 1150-00000 Test and Evaluation Strategy
- 1012-00000 Security Management Plan (aka Cyber Security Plan)
- 2130-00003 Architecture Specification
- OOI Data Products List (this list is currently maintained on the Instrument Application on the OOI Software Application Framework tool)

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

2.4 Precedence

In the event of conflict between/among documents/plans, policies, and guidelines relating to data/data product issues, the following precedence shall exist when not otherwise specifically identified. First in precedence is the OOI PEP, then the OOI SEMP, followed by this DMP. The remaining OOI pertinent documents are of subsidiary, equal importance.

The OOI Program Director, or designee, shall have authority to resolve any conflict between/among guidelines and directions.

2.5 Requirements

The OOI has been designed and built to an extensive set of requirements to assure alignment with the scientific goals of the OOI. As specified in Program governing documents (e.g., SEMP), these OOI requirements are maintained in a requirements software management tool Dynamic Object Oriented Requirements System (DOORS). A synchronized Excel export of the requirements is maintained in the OOI document control system (Alfresco), enabling easy access by all Program personnel.

Oceanographic measurement requirements, as referred to in this DMP, may be categorized in many ways including: science/scientific, common characteristics, or performance. The responsibility of the OOI with respect to data/data products may also be described as that of "facilitation"; this is generally referred to as functional requirements. The major functional requirements of OOI include:

- **Data collection/transport**: The OOI provides capability for the collection/transmission of data from sensors to network entry points where the data become available to assembly centers, users, and external data centers for research and product generation applications.
- **Quality control (QC)**: The OOI documents the steps taken to maximize the quality of the data obtained through the sensor network. QC operations are performed in advance of deployment of sensors, during the at-sea deployment and recovery activities, and following receipt of the signals by the data acquisition systems.
- **Data assembly and selection:** The OOI provides mechanisms for aggregation and buffering of data over useful spans of time and space. Data assembly allows users to more easily exploit near-real-time data, especially data from distributed sensor arrays or sensor arrays that utilize a data transport mechanism that has intermittent or "burst" transmissions. Users can specify time periods, geospatial areas, and/or variables of interest for display and download. For example, a user can specify a time period during the passage of a strong winter nor-easter and examine all those data captured by the Pioneer Array.
- **Product generation**: The OOI provides a specified set of generated products. The list of OOI produced products is contained in the Data Products List. The list of produced products may vary over time as a result of policy directives and suggested improvements or additions from the user community. Additional product generation is the responsibility of end-user groups.
- **Metadata management:** The OOI provides guidelines and utilizes extensible standards (e.g., ISO 19115) for metadata; ensures that the linkages between data and metadata are maintained and provides for communication of metadata among the OOI observing nodes and components.
- **Data/data product storage and persistence**: The OOI provides for the medium-term (months to years) storage of operational and science data, as well as the long-term (permanent) archival and stewardship of OOI data/data product sets that conform to national archive standards, as well as OOI standards and established user requirements. The Data Policy (1102-00010_Data_Use_Policy_OOI) specifies when data are shared with or transferred to national archives and how long the data are kept immediately available from the OOI. Agreements with national archives will be established, documented, and published before end of the OOI construction period.

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

- **Discovery**: The OOI provides a means for determining what data/data products are available within the OOI based upon queries that can be issued by users or directly by machines.
- **Administrative functions**: The OOI provides oversight mechanisms to ensure proper functioning and evolution of OOI data/data product and related activities. These mechanisms include: fault detection and correction, security, monitoring and evaluation of component performance, providing for extensibility, establishing and publicizing policies for data availability, and soliciting and responding to user feedback.

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

3 Management

3.1 Governance

OOI data products are managed by the OOI Program Management Office (PMO) in cooperation with the IOs. A variety of data transmission, distribution, and dissemination methods (including dedicated fiber, satellite links, and national fiber infrastructure) are used to transfer data from the measurement platforms through various procedures for data verification, processing, storage, and distribution to users. In general, some form of data assembly, data verification, and data quality assessment are required before ocean measurements can be used.

This DMP provides the roadmap and guidance for the requirements and design to achieve the following functionality:

1. Collection, processing, and storage of measurements from a wide variety of sensors
2. Clear descriptions of data sets (via metadata);
3. The ability to search for and find data sets, products, associated data items;
4. The ability to access measurements and data products, via services, from computer applications across the Internet

The data/data products aspects of OOI operate within the context of the overall OOI governance mechanisms described in the OOI PEP, FND, and SEMP. Some of these governance mechanisms include:

- Community-based user outreach programs to provide information about and solicit input of future needs and functionality with respect to data/data products.
- OOI program management and system engineering to facilitate proper execution of plans, designs, and requirements. They address key information technology (IT) issues including security, performance, availability, and cost.
- Periodic NSF detailed reviews of all aspects of the OOI program, particularly those of the data/data products streams.

Section 4.3 (Instrument List and Data Product List) provides further details about where and how data are generated and products produced.

During the construction period, the OOI Senior Systems Engineer (SSE) and Senior Program Scientist assigned roles and responsibilities to OOI staff regarding data/data products. These roles and responsibilities have four distinct foci:

- Data Product Specifications– the selection; creation, if necessary; and maintenance of all algorithms (data product, calibration, and quality control) to create OOI data/data products. The Program Scientists are responsible for this work.
- Metadata– the creation of all metadata associated with OOI data/data products. CI is responsible for implementing and maintaining all metadata capabilities.
- Calibration/Verification/Quality Control– direct the implementation of the instrument, time, and data calibrations necessary for collection of best possible OOI scientific data. This includes generation and maintenance of the QC algorithms, calibration and QC look-up tables, and all data field verification techniques for or applied to OOI data.
- Instrument measurements– define initial sampling strategy, may provide limited alternate choices and any available options for adaptive sampling capabilities

3.2 Roles and Responsibilities

The DMP defines the high-level data/data product guidelines for the program. The DMP requires that the management and engineering processes include the necessary discipline and participation to effectively perform the complete, integrated data management functions. During the construction phase, the OOI Senior Systems Engineer has the responsibility to oversee, approve, and conduct the policies and activities described herein. During O&M, the Program

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

Director (or their designee) assumes these roles and responsibilities. Through the management and engineering processes, the approval of the OOI Senior Systems Engineer or designee is required on all documents and decisions that may impact the ability of the observing nodes and components, including instruments, to meet their allocated data/data product requirements.

The Marine Implementing Organizations (MIOs) have the responsibility to design and implement the deployed marine observing nodes that implement the data collection and transmission capabilities.

The OOI cyberinfrastructure group has the responsibility to design and implement the hardware, software components, and networks that provide instrument command and control; data management; and, data processing, storage and dissemination.

3.3 Procedures/Policies

3.3.1 Data Policy- General

The OOI exists to enable new and important research. For this research to be of maximum benefit to society, data/data products must be carefully processed, stored, and released in a timely fashion as constrained by available funding and other resources. Rapid release, utilization and publication of results from OOI data/data products are encouraged. Data users from government and academic and public sectors should be able to use OOI data and information.

OOI data/data product policy is based on and consistent with NSF, U.S. Federal and international policies and principles. The OOI data policy is described in the Data Policy document (1102-00010). Changes to the data policy follow the standard OOI change control process as defined in the CMP.

All data/data products collected, prepared or supported with OOI funding are subject to the OOI data policy. The policy calls for full and open sharing of data and metadata, processing algorithms, products, code, and related information. It also calls for adherence to data, metadata, and data products standards promulgated by OOI.

3.3.2 Data/Data Product Retention

The driving goal of the OOI is to provide not only the data/data products to users as quickly as possible, but also the indefinite stewardship of these data. This goal also applies, where practical, to data and other important artifacts produced by the OOI as a result of individual PI activities (i.e., not OOI provided instruments) that stem from externally-funded research proposals. Their data, as deemed practical by the OOI, are stored and provided to users in the same fashion as basic OOI data/data products.

The retention of data/data products requires a balancing of constraints. The constraint driving the retention procedure is mainly the cost to persist total observatory data volume, which in turn is affected by number of instruments, observation interval, data volume per observation, and number of data/data products. Direct retention of OOI data is determined on a product-by-product basis Current plans are discussed below in Section 3.9- *Operations and Maintenance (O&M)/ Long Term Management (LTM).*

3.3.3 Observation Strategy and Planning

The OOI Senior Systems Engineer and Program Scientists, with feedback from external advisory groups: define, develop, and monitor the default and pivotal observation strategies for all OOI instruments on each platform (fixed or mobile). (Default sampling strategies are the sampling intervals and procedures used during operations that maximize the utility of the data received. Pivotal sampling strategies are the data sampling sequences chosen to ensure that the observations obtained meet key OOI science objectives. They are considered the 'minimal' level of sampling required to meet the long-term science goals of the sensors.) These strategies are

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

publically available on the OOI web site and openly communicated after review and approval by the OOI. During the construction phase, the OOI SSE is the final decision authority regarding default and pivotal sampling strategies. During the O&M phase, the Observatory Director is the decision authority. The strategies are implemented by the IOs during activation and operation of the OOI network.

The term "sampling (or observation) strategies" include:

- temporal resolution of the observations– timing of individual samples as well as timing of profiles or bursts of samples
- spatial resolution of the observations– details of instrument locations as well as flight paths of mobile platforms

Changes to any pivotal strategy follow the normal OOI change control process. (Pivotal Sampling Strategy is a data sampling sequence chosen to ensure that the observations obtained meet key OOI science objectives, as practicable.) Procedures and justification for the approval of -potential changes to the observation strategies, including, for example, pre-approved interruption of the default strategies to respond to unusual events (e.g., volcanic eruption; coastal hypoxia; severe atmospheric storm), are developed by the OOI Program Advisory Committee in collaboration with the Senior Project Scientist.

Temporary changes to exploit scientific opportunities due to sudden events including for example, interruption of the default strategies to respond to unusual events (e.g., volcanic eruption; coastal hypoxia), are delegated to the O&M manager. External investigators may also request a non-emergency deviation of the default observation strategy plan via submission of a proposal to NSF.

Obviously, events will occur that require near immediate operational decisions and actions (e.g., power constraints, hardware problems, communication bandwidth changes). The O&M lead will establish and publish a policy to deal with such occurrences. Sampling changes are implemented by the observing node managers and communicated to the OOI webmaster for posting on the OOI website (currently planned for the metadata section.) Every change must be logged within both the OOINet and the marine IO OMC/OMS. Users with event subscriptions or operator access will get a notification when changes are made.

3.3.4 Changes

As the OOI evolves, the DMP must also evolve. Any proposed changes to the OOI DMP follow the established OOI change control process.

3.4 Data Processing Techniques/Algorithms

Data processing algorithms range from simple calibration corrections (e.g., gain and offset) to complicated numerical methods for deriving products from multiple interdependent measurements. Automated Quality Control (QC) and secondary calibrations (based on information obtained at deployment and recovery) are also implemented via algorithms. The following subsections briefly discuss the OOI approach to data product algorithms. There are expected to be at least three different basic algorithm types necessary in order to create and maintain OOI data/data products. These algorithm types are:

- Data product algorithms – Software that takes one or more data inputs, processes the input based on the steps documented in the Data Product Specification (DPS), and outputs a Data Product. (See section 3.11.1 for definitions of OOI data levels.)
- Calibration algorithms – algorithms used to adjust data to account for the result of primary or secondary calibration procedures carried out on the instruments and sensors. Calibration algorithms result in the changing of values of a data product but do NOT change the data product level. Primary calibrations may occur onboard the instruments and/or applied during the data product algorithm.

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

- Quality control algorithms – algorithms used to apply various quality control procedures to data/data products, usually resulting in setting data quality flags that are visible to users. These algorithms do not change the data.

3.4.1 Algorithm Management

During the construction phase, the SSE is ultimately responsible for delivery of integrated, operational algorithms to O&M. This includes the definition and documentation of the algorithms plus implementation and maintenance procedures. CI is responsible for actual implementation. The SSE ensures the complete definition and documentation of all issues and activities directly related to the production of the scientific and engineering algorithms necessary to produce data products from OOI infrastructure and instrumentation. The SSE ensures that the algorithm development group has all necessary information needed to design, code, test, and integrate all product algorithms. This information includes specifications; instrument manuals; test data/results; and, if necessary, actual access to specific instruments.

The SSE:

- Ensures correct requirements directly associated with data product production.
- Approves algorithms for all OOI data products.
- Ensures calibration processes and algorithms are captured in clear, concise, useable specifications.
- Works closely with CI to assure metadata fields are captured in data product, calibration, and quality control algorithms as necessary and appropriate.
- Develops procedures for maintaining accurate current and historical descriptions of all algorithms.
- Develops procedures to verify all algorithms to be implemented within the OOI.

The SSE does not:

- Perform algorithm coding or integration. As algorithms are identified and documented, the information is used by cyberinfrastructure for development and integration within the OOI. This includes maintenance of the algorithm code. This is referred to as the operational code (or software code) in this document.
- Provide advice to the user community. After algorithms are integrated within the OOI baseline, O&M procedures are devised and implemented to disseminate algorithm descriptions to the user community.

3.4.2 Data Product Specifications (DPS) & Software Code

Prior to approval of algorithms, the SSE conducts a formal algorithm definition, documentation, and two step review process. The exact manner of creating, revising, and approving DPSs is captured in the DPS process document in the OOI process library. The following summarizes this process.

Step one is the algorithm definition and collection of all materials that will be needed to implement the algorithm. The SSE, in collaboration with the IO PMs and the Program Scientists assign responsibility for creation of a Data Product Specification (DPS) to a particular person. This person is provided a DPS template for completion and a schedule for delivery. Some of the sensor-specific algorithms may be provided by the sensor vendors while multiple-sensor products may require reaching out to existing user communities for algorithm advice.

Step two occurs upon delivery of the initial draft DPS to the SSE. The SSE designated DPS “scheduler” and “gate keeper” reviews the draft DPS submittal, makes any easy corrections, goes back to the submitting Program Scientist (PS) if needed and when satisfied that the DPS is complete in terms of “all sections filled in”, submits the draft DPS to a small group (typically three or four subject matter experts [SMEs]) for a five day Focused Review.

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

Step three begins with the return of the reviews from the focused SMEs. These reviews are provided to the DPS author for adjudication. When these comments have been adjudicated, the DPS scheduler posts the candidate DPS for anyone in OOI to review. This is called the Formal Review and lasts for five days. During this second, program-wide review process, reviewers assess: algorithm correctness and completeness, operational viability, and potential trade-offs between product accuracy versus factors such as life cycle cost, computational requirements, and community acceptance. Close review by the implementing and integration groups is critical to this step; they must assure that all required information is available and properly explained.

The final step is for the SSE to confirm with the Marine Integration Lead (via signature) that the DPS is complete and acceptable. The SSE then signs the DPS; it is entered into the OOI DMS; and it is delivered to CI for implementation.

All data product, calibration, and quality control algorithms used by OOI to produce data products are captured in DPSs and/or Data Processing Flow Diagrams (DCN 1342-xxxxx) for the specific instrument classes. DPSs enable technical staff ("coders") to write the software needed for implementation, testing, and deployment. DPSs are also documentation vital to data users so that they can better understand the data products and judge whether specific products are appropriate for their needs. Each DPS describes the data product(s) and the algorithm (including primary calibrations) used to create the products). Each DPS is assigned document control and version numbers based on the OOI CMP and maintained in the OOI DMS. Test data sets (inputs and results/outputs) are generated and provided as part of the DPS where feasible. DPSs also contain only essential discussion of:

- Dependencies
- Input and output products
- Equations, in finite difference forms as required
- Data flows (also in the Data Processing Flow Diagrams)
- Exception handling
- Graceful degradation (if applicable)
- Processing requirements
- Any external/ancillary data needed
- Metadata fields directly related to the algorithm
- Pseudocode and code examples
- Measurement error quantification

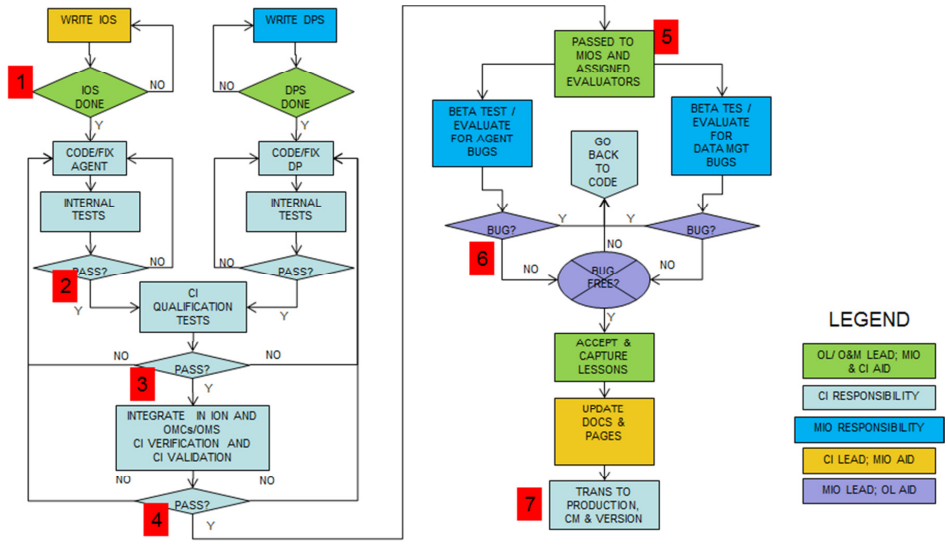
Technical and operational references are included in DPSs to allow traceability to algorithm origin and examples of previous use. A DPS template has been developed and is found in the OOI DMS under OOI document 1341-00000_DPS_Template_OOI.

3.4.3 End-to-End Algorithm Development Sequence

Algorithm development, test, integration and transfer to operations follow the steps in the Marine Integration Procedure (MIP) as shown below.

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial



DPSs containing algorithms, pseudocode, code examples, and test data sets are delivered by the PMO systems engineering organization to the algorithm development group for development. Once developed and unit tested, the algorithm code is delivered to CI for integration. Generating and maintaining all aspects of algorithms is the responsibility of the PMO at Ocean Leadership. Maintaining the algorithms includes making sure the DPSs are properly approved, communicated, and correctly placed in the OOI DMS. See section 3.4.6 below for a description of algorithm change processes.

3.4.4 Algorithm Storage

DPSs consisting of algorithm test data sets and pseudocode or code examples are retained in the OOI DMS under configuration control. DPSs are numbered 1341-XXXXX in accordance with the CMP. Test data set results are archived in the DMS as test reports (see 1150-00000 OOI *Test and Evaluation* Strategy document section 7.2.3.)

3.4.5 Algorithm Access

Read only access via electronic means to DPSs is provided on a free and open basis, within the constraints of national security, and proprietary restrictions, to all users of OOI data. Algorithms and supporting documentation (for example DPSs) are posted on our website (currently planned for the 'metadata' section).

3.4.6 Algorithm Changes

Changes to DPSs or code follow the established change control process for specifications, which includes final approval by the Senior Systems Engineer or designee of the Program Director. When changes are made, the change documentation must describe what was done, and why the action was performed in the Algorithm History section of the DPS. The OOI Senior Systems Engineer will have the OOI web page report these changes to communicate to the community.

3.5 Data/Data Product Release, Use, Rights, Rules

All OOI data/data products/metadata and supporting information developed by OOI are made available as quickly as possible. Some measurements and observations are released in near-real time if communication capabilities are available. All recipients of OOI data/data products may use

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

them freely for research and/or publication. OOI requires the user properly credit the data source. The OOI *Data Policy and User Terms and Conditions document* provides additional detail on the philosophy, approach and implementation of OOI and user responsibilities.

3.6 Cost Impacts and Awareness

The data management capabilities of OOI support the delivery of data products, as listed in the Data Product List.

The OOI data management function supports multiple levels of data assurance activity (i.e., no QC performed, automated QC, human-in-the-loop QC) and allows the OOI community and leadership to tailor the resources applied to generation of data/data products, according to user requirements and budgetary constraints.

Decisions to achieve the OOI community's priorities take into account the cost impacts and alternatives. Trade studies are used to evaluate the scientific and operational impacts of an action versus costs and available resources. The more extensive the issue, the more formal the trade study. For example, simple, minimal cost impacts may be evaluated at a subsystem or IO level, while expensive and/or complicated issues involve evaluation by all IOs and OOI leadership. The results of those decisions are reflected in observatory or component configurations, and implemented through the O&M Annual Work Plan.

3.7 Security

The security aspects for OOI data management have multiple facets. These include the need for:

- Data management governance and policies
- Prevention of external aggression (e.g., “hacking”)
- Backups of data and physical components to prevent data loss
- Minimizing single points of failure
- Providing physical protection to the components of the OOI
- Sequestering data with potential national security concerns

It is recognized that a significant risk for data management is the accidental corruption of data and metadata, but it is also recognized that it is possible to have deliberate falsification of data. Tracking data provenance helps mitigate this. The *OOI Cybersecurity Plan* (DCN 1012-00000) is the governing document for the security aspects of data management.

The other security aspect of data management is the acquisition of data that have national security implications. Certain data that may include (but are not limited to) seismic and hydro-acoustic data (depending on observation strategy plan, frequency, data rate and dynamic range) may be delayed by diversion to a secure facility, and may only be released to the OOI after review. It is possible that some data may not be released to the OOI at all. The NSF, US Navy and OOI have appointed representatives for cyber-security issues and are charged with developing procedures as needed.

3.8 Transition of Data/Data Products Post OOI

The Consortium for Ocean Leadership, a non-profit 501-C3 corporation under contract with the NSF, is implementing the OOI. It is assumed that this program, as currently structured, has an end of life; currently the OOI is intended to remain in service for a period of 25 years. There must be advance consideration for all facets of the OOI post this contract period. Areas to consider are: saved data/data products, physical samples, physical facilities, data communications, maintenance, and operations. All of these are either directly or somewhat impacted by the approach and philosophy specified in this DMP.

A final data transition plan (FDTP) is required and is highly influenced by the manner of the transition (e.g., end of mission, continuation by new contractor, transfer to an operational government organization such as National Oceanic and Atmospheric Administration). This FDTP

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

should be created under the direction of O&M after the manner of transition is selected, with the goals of the transition defined and with sufficient lead time to both develop and execute the plan.

Specific transition with respect to data/data products may vary by item/infrastructure. The current plans are as follows:

3.8.1 Algorithms

All calibration and processing algorithms are readily available throughout the operation of OOI. Any algorithms produced or used by the OOI and not already present in national repositories are transferred to NSF-specified agencies or groups for safe keeping and stewardship. The formal mechanisms needed to accomplish this transfer are developed as part of the FDTP.

3.8.2 Data/Data Products

Any data/data products (and their associated metadata) produced by the OOI and not already present in national repositories are transferred to NSF-specified agencies or groups for safekeeping and stewardship. The formal mechanisms needed to accomplish this transfer are developed as part of the FDTP. It is intended that open and unfettered public access continues. Data/data products already in national and thematic repositories continue to be managed by those repositories.

3.9 Operations and Maintenance (O&M)/ Long Term Management (LTM)

The Operations and Maintenance (aka Long Term Management) of the OOI is described in the O&M Plan. (As of this DMP version/date, the O&M Plan does not contain sufficient procedures and work instructions to address data management.) OOI intends for the data/data products produced by the Program to be available for the life of the program, and indefinitely if feasible. Data are eventually moved from OOI into long-term storage after a period of time. This period of time depends on the type of data and is further described in the *Data Policy and User Terms and Conditions* documents (1102-00010_Data_Use_Policy_OOI; 1102-00020_Data_User_Terms_Conditions_OOI).

.

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

4 Data and Data Products

The primary objective of the OOI is to provide data/data products to the user community. These data/data products are of scientific value and made available to users via standard internet-based communication paths. A second, vital, objective of the OOI is to provide a sustainable hardware and software infrastructure by which observations are obtained and transferred to land-based systems for ultimate consumption by the scientific community. The observing nodes operate and collect data that are transmitted to land-based processing centers. Availability by system and subsystem is specified and maintained in the OOI requirements repository. Requirements on the accuracy and precision of the measurements within each data set or product are also specified and maintained in the OOI requirements management tool. These data/data product performance requirements are traceable to the fundamental OOI science requirements as captured in the OOI top-level science themes and science requirements in the OOI requirements repository.

4.1 OOI Data products

4.1.1 Data/Data Product Levels

A variety of schemes exist for the storage and access of environmental data; all have advantages and limitations. Interactive, menu driven web pages can facilitate the search and provision of data/data products at different levels of processing. These web pages are typically site, phenomena, and instrument-centric although broader and complex searches are possible. In order to facilitate flexible yet maintainable and manageable data sets, the OOI has adopted commonly used numbered data/data product levels to describe various degrees of processing. Each numbered level has a simple descriptive synonym. The italicized text is the formal definition maintained in the OOI requirements repository. They are listed below for reader convenience:

- **Instrument raw data**: *The data as they are received from the instrument. These data are in instrument-specific format and may, depending on the instrument, contain data for multiple sensors (interleaved), be in native sensor units (e.g., counts, volts) or have processing steps already performed within the instrument (e.g., primary calibration.)* They are always persisted and archived by the OOI. Storage of instrument raw data enables reprocessing and/or reanalysis. Instrument raw data may be packed/interleaved and/or at any of the following levels and should not be confused with formal data products. An example of instrument raw data would be a format 0 binary file from an SBE-37IM on a Global Flanking Mooring.

- **Level 0 (L0)**: *Unprocessed data: Data that are in instrument/sensor units and at instrument/sensor resolution. They are sensor by sensor (unpacked and/or de-interleaved) and are available in OOI supported formats (e.g., NetCDF).* These data are sometimes mistakenly called “raw” data. They are always persisted and archived by the OOI. Storage of Level 0 data enables reprocessing or reanalysis. It also enables researchers to perform their own fundamental quality control and calibration.

- **Level 1 (L1)**: *Basic Data: L0 data that have been calibrated, are in well-defined scientific units and may have some QC applied.* The amount and type of applied QC in the delivered data set may be selectable by the end user (e.g., user may request only good data or all data). QC may utilize simple automated techniques, complex algorithms, or human inspection. Level 1 data are sometimes called “processed” data, though this is not encouraged because it can be confusing. Actions to transform Level 0 to Level 1 data are captured and presented in the metadata of the Level 1 data.

- **Level 2 (L2):** *Derived Data Products: Data that have been manipulated and/or combined and include new values not present in original data*. The OOI provides only a limited number of derived products.

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

4.1.2 Data/Data Product Time Stamping

Data and Products are assigned 'time stamps' in UTC. The time stamp for each data element or resulting product is selected based on the 'first' point in the data path where a reliable time is available (i.e., 'as far upstream and as close to the sensor as possible'.) GPS is the master time source.

4.1.3 Data/Data Product Geolocation

Data and products are geolocated. The highly mobile platforms (i.e., gliders and AUVs) will contain geolocation (either directly measured or linearly interpolated) as part of each acquired data point. All other OOI data collection platforms (including wire following profilers) will contain a single geolocation point included in the product metadata. This geolocation point will be the deployment location determined during each deployment.

4.1.4 Temporal decimation

Temporal decimation of instrument data is a consequence of telemetry constraints (e.g., CGSN global arrays). (All instrument raw data that were NOT received from deployed assets due to transmission constraints are provided after instrument recovery from the full, recovered data set.) The metadata contain information describing any telemetry-driven decimation. In some cases, if decimation is necessary, the conversion to L1 from the instrument data may occur prior to telemetry, either through the application of the appropriate data product algorithm on the platform or onboard the instrument itself. All temporal changes to the data are captured in the metadata. Decimation does not affect the modes or the levels of the data products.

4.1.5 Data/Data Product Acquisition Modes

Data Modes describe the manner/mechanism/method of data acquisition. There are two data modes for OOI data- transmitted and recovered. Note that the mode of data acquisition is not an indication of data level or any temporal averaging/decimation. It is planned that recovered mode data products normally be at full temporal resolution and transmitted data products be at full resolution whenever allowed by data rate constraints.

1. **Transmitted:** describes data received through a transmission media over distance. Examples are: Fiber optic cable, satellite, glider to satellite, acoustic modem. These data may be transmitted via streaming or in periodic "batch."
   a. Transmitted streaming data are provided at full temporal resolution and near-real-time by fiber optic cable.
   b. Data received through satellite relay or other mechanisms results in "batch" receipt and may be interleaved or decimated in time. These data have greater latencies than the transmitted streaming data.

2. **Recovered:** Describes data offloaded directly from an instrument or data logger; usually by connecting the instrument to a computer after the instrument has been recovered and writing to files, often on the recovery vessel.

4.1.6 Data/Data Product Calibration

The OOI recognizes and implements two levels of data calibration. The document- 1102-00300_Protocols_Procedures_Data_Products_QAQC_Cal_Physical_Samples_OOI, defines the calibration steps and procedures. The two basic calibration equations (which represent steps in data processing) are defined below for reader convenience.

Primary Calibration equation:
The primary calibration equation describes the sensor characteristics and uses the sensor and time or frequency (e.g., response poles and zeroes or finite impulse response filter coefficients) specific coefficients for conversion of the measured values to physical units. Typically, this

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

equation and the coefficients from a recent calibration are provided by the vendor with a new instrument. These can sometimes be modified, if applicable, by shore-side calibrations performed prior to deployment. Generally, the OOI does not modify the primary or shore side calibration coefficients while that instrument is deployed with the possible exception of those instruments which can be calibrated in situ.

Secondary Calibration equation:
Any equation that adjusts the data, performed in addition to the primary calibration. Typically, the OOI attempts to keep this computationally simple (e.g., changes in gain and offset). The coefficients used in the secondary calibration equation may change with time (e.g., interpolating between a pre-deployment calibration and a post-recovery calibration.) Care must be taken in how the pre-deployment and post-recovery calibrations are interpolated over the recording period including estimates of calibration errors as a function of deployment time.

OOI may perform secondary calibration. This secondary calibration is completely recorded and explained in the metadata accompanying all products. If performed, this calibration must be captured and detailed in the metadata. Data that has undergone secondary calibration will be indicated via unique product names.

4.1.7 Quality Assurance (QA) and Quality Control (QC)

QA and QC are formally defined in the OOI requirements repository- reference module. QA is performed by the O&M staff.

OOI provides automated QC software (currently 7 tests, e.g., spike test) that is used by the different L1 and L2 product production algorithms to set quality flags on a data point-by-data point basis. Data Quality Control Levels (QCL) describe the quality control steps that have been applied to the data.

Data quality control steps can be applied to any L1 or L2 data product level of either acquisition mode. Note that quality flags indicate if a value has exceeded a set limit (aka threshold) and therefore may (or may not) have an issue. The Data Processing Flow Diagrams (DCN 1342-xxxxx) specify which automated QC algorithms have been utilized. Speed of data/product availability is influenced by the degree of quality control that has been applied to the data/products.

In addition to being collected by a few specific OOI instruments, physical samples are also utilized as part of the OOI data calibration/verification/quality control efforts. Any digital documentation accompanying the physical samples (e.g., analytical results and metadata) is stored in its original form. Other non-electronic artifacts can be converted to digital form by the provider, submitted to OOI, and appropriately cataloged in OOI for future search and use.

Sampling strategies/ protocols for the collection and analyses of physical samples are developed by the responsible science lead for the relevant parameters. Based on these protocols, a physical sample management procedure is developed to detail the storage of both the physical samples and the derived data. The stored data include shipboard recovery and processing procedures, analytical methods, and resultant analyses of the physical samples (e.g., seawater samples from a ship CTD/rosette system for analysis of dissolved oxygen, in order to cal/val a DOSTA instrument; fluid and rock chemistry; 16S rRNA). These protocols also detail the process for release of physical samples to interested users and the required agreements.

A complete discussion of the collection and processing of physical samples from OOI instruments and for field verification/secondary calibration and the overall QA/QC approach and procedures is found in:

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

1102-00300_Protocols_Procedures_Data_Products_QAQC_Cal_Physical_Samples_OO. Users can access the QA/QC documentation (e.g., algorithms, threshold tables, procedures) via the OOI website under the 'metadata' section.

4.2 Flow through the OOI

Data arrive at the edge of the OOI integrated observatory network through various communication channels and according to basically two transmission types and two recovery approaches. In the first transmission type, data are received one record (or block of data) at a time, typically as the records are generated by the sensor or platform originating them. This is referred to as *streaming*. Initial metadata are usually attached to the generated records by the local platform or the sensor. Not all metadata need to be transferred with each record (e.g., sensor manufacturer). The metadata contain all the key knowledge about the data record (e.g., time of collection; unique source; and record description identifier, including version, platform identification), to enable it to be understood by the system and its users.

The second transmission type results from possible communication channel data rate limitations. Records may accumulate in queues. The records are then downloaded in rapid succession; this is called *batch transmission*. While batch does not affect the way the data are processed, often the handling up until the data are received by OOI cyberinfrastructure is quite different. For example, an autonomous underwater vehicle (AUV) may send all its records upon docking; or an instrument may collect an extended data set on storage media internal to the instrument and then upload the records upon recovery of the instrument.

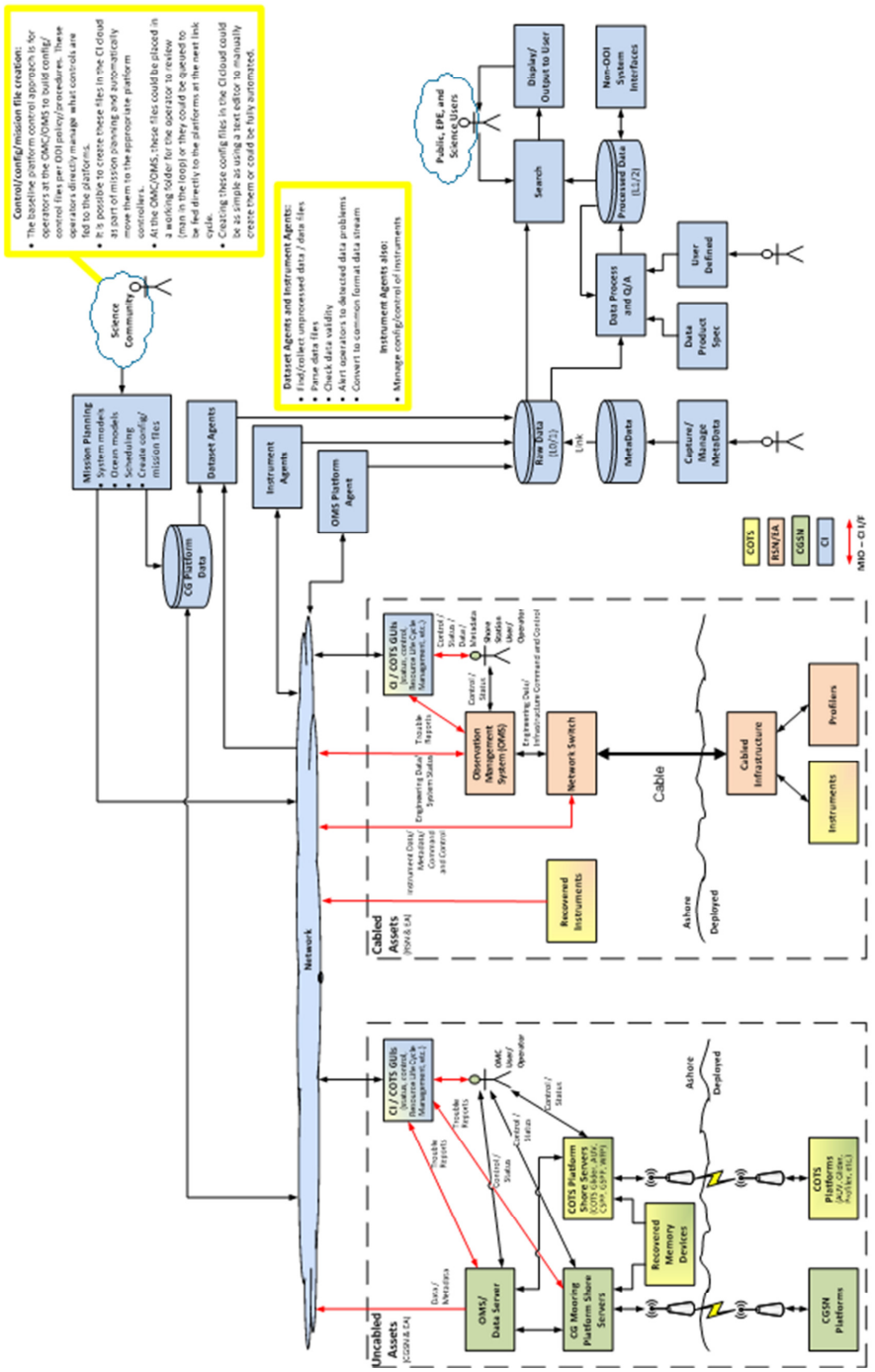
Recovered data enters the OOI networks via one of two methods: direct connection to the instrument or removal of a digital storage type device (e.g., flash drive). In some cases, the data received through different communication channels may contain duplicate records. The handling of such data is described under "Acquisition/Collection—Policies for handling duplicate data," section below.

Once "inside" the OOI networks, data moves from the Marine network to the Observatory Network where processing, cataloging, and persistence occur. They are then published (exposed) on a User network via Human Machine Interface (HMI) web pages.

Data flow from sensor to user is depicted in the following diagram

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial



Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

4.3 Instrument List and Data Product List

The list of instruments deployed on the OOI network is currently maintained in the Instrument Application housed within the Software Architecture Framework (SAF). Confluence pages contain additional vendor references, instrument manuals, and information about the instruments. A published version of this material is available on the public OOI website.

The OOI data products are also currently listed in the Instrument Application maintained in the OOI SAF. Each Data Product is described in a separate OOI Data Product Specification document (DCN 1340-XXXXX). The Data Product Specification includes the information needed by developers and users. DPSs contain the specific processing steps necessary to create calibrated, quality controlled level 1 and level 2 data products from the source instrument.

A Data Processing Flow Diagram (DCN 1342-XXXXX) for each instrument class illustrates the processing and QC steps to obtain the specific data products produced by that instrument.

Each level and mode of data products is registered. All registered data products are made available according to OOI data policy.

Ancillary data (data from non-OOI sources- e.g., shipboard cruise data or data from other government agencies) can be attached to the OOI homepage under the “data tab”. These data may be used as a basis for data quality control (e.g., HITL) and by researchers to use ancillary data to enhance analysis of OOI data. Responsibility for attachment and exposure mechanisms for ancillary data will be designated by the Observatory Director. The OOI also has the capability to link ancillary data to a specific data product.

Non-digital data (e.g., log sheets or other printed matter) are not accepted for storage by the OOI. Instead, digitized versions are eligible for inclusion as ancillary data as described above.

4.4 Metadata (Including Engineering Data)

Metadata (‘data about data’) take many forms, all of them critical for understanding the provenance (history) of the observational data, and therefore are critical for appropriate analysis and use of the data. Metadata include both human-readable and computer-parseable information about the data and their collection. Metadata include descriptions of data, sensors, platforms, and processes, including any engineering data that are generated during OOI activities (observatory operations or processes). Metadata can also include pointers to other data, such as auxiliary or coincident environmental observations, that are important for understanding the data (e.g., to support calibration of sensors at sea).

Traditional uses of metadata (i.e., to provide catalog browse/ search capabilities, and to describe observatory/ cyberinfrastructure assets) are supported as capabilities available to the end user. As described in the Acquisition/Collection—Metadata section below, appropriate metadata are obtained for all data acquired or collected by OOI.

4.5 Standards

OOI is standards-oriented, standards-compliant and supports community specifications. Choice of standards is driven by requirements of functionality, community acceptance, and potential interoperability. Standards adoption and specifications analyses are performed by the OOI Project Managers, Program Scientists, Engineering team and other OOI IO partners.

The standards include:

- data and metadata formats and services
- supported metadata content (required and optional) (based on ISO 19115 and NetCDF-CF data format standards)
- naming conventions

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

- transmission protocols.

4.6 Naming Conventions

OOI uses controlled vocabularies to define metadata content, and particularly acceptable naming conventions. The Uniform Resource Locator, or URL, is expected to be the unique identifier of choice due to its resolvability. OOI works with the community to develop mappings necessary to translate between its internal names and other terms and vocabularies supported by the community.

CI increases the level of specific definitions as the scheduled OOI releases are developed. For example, initial naming conventions were established for the model data included in Release 1 and additional metadata content and naming schemes are matured during Releases 2 and 3. All mappings and content are complete by Release 3 acceptance.

4.7 Data Volumes

With all the OOI sensors deployed and operating, and assuming that all sensors are producing continuous data, the OOI could potentially produce over three petabytes/year. This data volume is dominated by the high definition (HD) cameras on RSN. The next largest data producers are the broadband hydrophones which produce ~10% of the HD camera data volume. The third largest producers are the still cameras, followed by seismic and low frequency acoustics. The expected data volume estimates exceed OOI's ability to persist all the data in user ready storage.

The approach to provide both affordable permanent stewardship and user access to these data is described in: *1102-00010 Data Use Policy* and *1102-00020 Data User Terms and Conditions*.

4.8 Acquisition/Collection

The OOI supports data collected from sensors, OOI observing nodes/ components, physical samples, data processing software, and people (operators). Each of these data sources adhere to OOI data management policies, procedures and interfaces.

4.8.1 Metadata

Any data that OOI collects are associated with appropriate and at least minimally complete metadata. Specific metadata to be applied are defined as part of the OOI observatory and component development process. Examples of metadata include:

- time of data collection
- location of data collection
- instrument transfer function
- data source information, including details of data origination (provenance) as appropriate to the data source type (e.g., instrument ID)
- data description information

The form in which these metadata is provided is specified for each data type within the DPS document for that data product. OOI metadata follows the ISO19115 standard for geographic metadata information and NetCDF-CF with additional metadata types and fields specific to OOI as necessary. Additional information is contained in the document: *1321-00000 Metadata Implementation*.

4.8.2 Duplicate data

When OOI acquires data, it is possible that the data are full or partial duplicate of data previously acquired (e.g., repeats from telemetry sources or external sources). Some data sources, particularly remote ocean sensors, transmit data subsets initially via satellite telemetry, followed by a full data set when higher data rate transmission is available.

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

In the cases where summary data were initially transmitted and more detailed data become available later, the summary data are not deleted, and the more detailed data are added. This constitutes two similar, but unique data sets with content overlap.

4.8.3 Procedures for collecting operations and support data

An extremely wide variety of operational and support data are collected by OOI to support the analysis of science/sensor data. Typically the associated data sets support the understanding of the provenance of the data—the history of processes, components, people, and activities that explain how the scientific data came into existence. Generally the availability of this information, or lack thereof, determines whether a particular analysis can be performed.

On occasion, the support data themselves have scientific value under direct analysis. For example, operational weather data from a ship's cruise can prove significant to analysis of air-sea heat exchange; or internal measurements of a sensor's operating environment may carry unexpected clues about external environmental conditions.

Thus, it is not possible to confidently know all data to be used for specific scientific applications. However, we do know that specific data are collected for specific purposes (e.g., calibration and/or verification for surface buoy meteorological instrument inter-comparisons, CTD secondary calibration). The listing and accounting of operations and support data are developed as the instruments and algorithms to process/ calibrate their data are defined. These listings reside in DPSs, Instrument Operational Specifications (IOSs), Interface Control Documents (ICDs), and/or 1102-00300_Protocols_Procedures_Data_Products_QAQC_Cal_Physical_Samples_OOI.

4.8.4 Human provided data

Many types of data are collected from personnel participating in OOI activities. These people occupy a wide range of roles including: marine operators (e.g., mooring logs, cruise reports), cyberinfrastructure operators, managers, science users, and educational and public participants. The procedures to include data from the affiliated personnel take into account all the related issues. Files can be attached to products (e.g.,.xls, Word, or PDF/ JPG for notes and photographs relevant to the data collection, Perhaps the most significant human input to the science data is any human performed QC. The OOI system is designed to permit human annotation of data sets reflecting their QC efforts.

4.9 Right to Use licensing

The OOI has a policy for use of all the data that it serves. The *Data Policy and User Terms and Conditions* presents this right to use in detail.

4.10 Data/Data Product and Information Tools

The OOI provides multiple access options to the OOI data. OOI Release 2 contains many user facing web interfaces and displays. These interfaces/ displays continue to evolve and mature during Release 3 development and are made publically available at Release 3 acceptance. At that point, all data, products, and tools are openly available to users.

Capabilities of the OOI ultimately include:

- browser-based data access for human end users, including visual interfaces and basic metadata search functionality by time, location, and type of measurement
- application programming interfaces (APIs) to enable computerized data access and exporting OOI data/data products into external software applications
- ability to evaluate data/data products based on the use of “quality flags”

The OOI Internet web pages contain links to these data access points as well as documentation of how to use them.

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

4.11 Processing

There are three data product levels produced by the OOI as well as two Acquisition Modes. Instrument raw data are stored in a lossless manner. Instrument agent drivers create Level 0 data products from the instrument raw data with no calibration or any manipulation other than separating of the data products from the instrument raw data and rendering them into OOI data formats from instrument specific, possibly proprietary formats. Initial calibrations and quality checks are then applied to the data automatically to create Level 1 data products. When improved calibration information (e.g., secondary calibration after deployment or recovery) or quality annotated data become available, improved Level 1 data products are available.

The OOI approach to data quality assurance and quality control is captured in a separate document and will be available via the OOI website.

The goal of the OOI data QA/QC processing effort is twofold:

- Maximize the amount of "good" data from the OOI instrumentation, where "good" means that the reported data may reasonably be expected to be close to the true values within the nominal accuracies of the instrumentation
- Provide an assessment for each measurement indicating whether it is "good" or not

The OOI Data QA/QC Plan covers the following topics:

- Data QA/QC approach by sensor/instrument/platform
- Algorithm definitions for data QA/QC (not to be confused with the algorithms to process or produce the data/data products)
- Procedural definitions for data QA/QC

One fundamental underlying OOI data procedure is the persistence and archive of all original, unprocessed "raw" data. Improved data sets (e.g., those with improved secondary calibrations) may be made available with metadata information such that improved data can always be traced back to the raw data (data provenance).

The OOI performs the maximum amount of QC allowable within the operations budget. The OOI automated data QA/QC procedures utilize the common features of community-supported, techniques (e.g., as used by Argo, OceanSITES, QARTOD).

Data quality flags are added through the QA/QC process. The capability is provided to flag data problems down to the individual observation. All QA/QC lookup tables, thresholds, algorithms are stored and available from the OOI website.

4.12 Data Storage

4.12.1 OOI Approach

The OOI includes data storage capabilities, and utilizes external archives to extend its own resources. It is the intent of OOI to provide all data of interest to national archives at the earliest suitable time.

OOI stores data to support the access needs of the OOI system and its users. Not all data are kept in OOI disk storage indefinitely. Instead they may be moved to external (delayed access) storage, archived to external repositories, or deleted if the volume exceeds OOI's capability, consistent with: *1102-00010_Data_Use_Policy_OOI and* 1102-00020 Data User Terms and Conditions. The timing and nature of the transition is driven by policy, costs and system capability/performance.

The data storage policy considers factors such as the data's:

- origin (science data and closely related metadata are the highest priority),

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

- size (very large data sets can be stored within OOI for briefer periods; most OOI data sets are relatively small and can be stored indefinitely)
- domain (seismic data are quickly delivered to external repositories)
- adequacy of metadata
- overlap with redundant data (i.e., more or less processed versions of the same data)
- nature of ownership (OOI, non OOI (e.g., ancillary data required for product production), or individual investigator/Principle Investigator produced).

4.12.2 Back-up Guidelines

The OOI incorporates appropriate reliability measures to meet the defined requirements, including backups of operational and persisted information.

The procedures used for performing backups are designed and implemented by the OOI CI, with oversight by the OOI Operations Manager. The policies and procedures are reviewed by the OOI Program, and in particular the OOI Senior Systems Engineer. System designs may be reviewed at any time in the OOI System Architecture and Design sections of the DMS.

4.12.3 Local Storage and Caching

If the instruments have the capability, data are stored locally on the instrument. This local storage provides an important data backup role in case there are problems with the communications infrastructure or the infrastructure lacks sufficient data rate to transmit the entire data set. These data can then be recovered when circumstances allow. Marine IOs (either on a platform or at a shore-station “cyber PoP”) are responsible to provide local caching of data at their OMC/OMS, to similarly guard against data loss. Once the data are completely transferred to the CI, the MIO can delete or offload the OOI storage space as needed.

4.12.4 Guidelines for Managing Digital Collections

Guidelines and procedures for managing Digital Collections in OOI are developed by the OOI, with the cognizance and review of the OOI Science Advisory Panel.

4.12.5 Guidelines for Managing Document Collections

Guidelines and procedures for managing data/data product centric Document Collections in OOI, are developed by the OOI with the cognizance and review of the OOI Leadership and its advisory panels.

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial

5 PI Data Products

In addition to OOI produced data/data products, individual Principal Investigators (PI) are encouraged to utilize the OOI infrastructure. Individual PIs submit proposals to NSF in a fashion similar to current NSF procedures. As part of the NSF/OOI RFP guidelines, PIs are requested to follow the established OOI data/data products procedures, formats, metadata, etc. This includes use of the same definitions of data/data product levels/modes and provision of common metadata fields. Any variances in the PI proposal from this request are evaluated and adjudicated at that time. The intent is to use the same data management approach for PI data sets as those of OOI basic data sets. This enables broader community use of deployed assets.

The document- *1102-00020 DATA USER TERMS AND CONDITIONS* refers to these data as **Project Specific** that derive from instruments that are attached to or use the OOI infrastructure. They are owned and operated by persons/organizations other than the OOI Program Office. Regardless of source, all data produced as a consequence of the OOI infrastructure are considered OOI data and are governed by that data policy.

The intent and desire is for the OOI to collect, store, curate, and serve these PI data products when feasible. Implementation of this intent is predicated on adherence to the RFP guidelines mentioned above.

Essentially, PIs wishing to utilize the OOI infrastructure (hardware or software- either marine environments or land environments) have two options:

- Their data are in the OOI system; including OOI storage and OOI web pages to serve users. This may require development of additional DPSs, other documentation or software to deal with non-supported instrument types. This could add cost to PI proposals, but also facilitates usefulness of the data over time by a larger community. The closer the data and instruments adhere to the identical format of OOI data (e.g., from same instrument types), the better
- Their data are never in the OOI system other than utilization of the physical infrastructure (e.g., attached to moorings, power usage, transmission circuits)

Formatted: Font: (Default) Arial

Formatted: Font: (Default) Arial