ALASKA OCEAN OBSERVING SYSTEM
DATA ASSEMBLY CENTER AND DATA MANAGEMENT SUBSYSTEM PLAN

EFFECTIVE: JUNE 1, 2016 – MAY 31, 2021

1.0 INTRODUCTION

The Alaska Ocean Observing System (AOOS) was established in 2004 by a consortium of partner agencies and research institutions and serves as the regional association for integrating coastal and ocean observing activities in Alaska's three regions: Arctic, Bering Sea/Aleutians and Gulf of Alaska. AOOS formally serves as the Alaska regional component of the U.S. Integrated Ocean Observing System (IOOS) and is considered to be a Regional Information Coordination Entity (RICE) under the authority of the Integrated Coastal and Ocean Observation System Act of 2009 (ICOOS Act).

As a member of IOOS, AOOS has a mandate to collect, organize, and provide access to Alaska oceanographic data. These data need to be easily understandable, electronically accessible and well organized to allow policy makers, industry and the general public to make well-informed decisions. To satisfy this mandate, AOOS supports a regional Data Assembly Center (DAC) and web-based data portal (called the Ocean Data Explorer) for the entire state of Alaska providing ocean, coastal and relevant watershed environmental data and information products.

To ensure data collected by AOOS and other regional entities and distributed through the AOOS DAC and on the AOOS web portal are managed according to best practices identified by NOAA, and that the data are of a known quality to the end user, AOOS is implementing recommended and standard practices as defined by the U.S. Integrated Ocean Observing System (IOOS) Data Management and Communications (DMAC) committee. These practices apply to data standards, metadata and data, transport and access, archival, information technology (IT) security, quality control and quality assurance, described in the NOAA IOOS Program Office whitepaper on the data management and communications subsystem (2010).

The AOOS Data Management Subsystem (referred to hereafter as the AOOS Data System) must adhere to these practices, and the AOOS Data System Plan (referred to hereafter as the Data Plan) provides the approach to the necessary implementation, describing how data are ingested, managed and distributed from the source to public dissemination. The Data Plan is organized as follows:

- Section 2 provides an overview of the AOOS Data Management Subsystem, describing the function and goals and objectives of the AOOS Data System management, the data management structure, and details related to the AOOS data management team.
- Section 3 briefly describes the AOOS data resources, defines data categories and asset types, and describes how the data categories are handled in the Plan.
- Section 4 presents the AOOS Data System statement of work and includes descriptions of the system computing infrastructure including details about the processes related to data flow including data ingestion, standards for format and content, metadata, quality control procedures and flagging protocols, data stewardship, preservation, public access and dissemination, data archival and preservation, and data system performance and security measures.
- Section 5 provides summaries of primary AOOS Data System applications.

This document, unless superseded, pertains to a period of five years from June 1, 2016 through May 2021.

2. AOOS DATA MANAGEMENT SUBSYSTEM

The mission of the AOOS Data Management Subsystem (AOOS Data System) is to acquire, archive and share marine data and information products to meet the needs of Alaska stakeholders and the national IOOS program. AOOS uses a data management system that allows a complex array of oceanographic and environmental data types to be well organized, accessible, and understandable. The AOOS Data System uses a distributed data management approach, which allows data to seamlessly interchange between participating data providers, which are primarily government agencies and research entities. The system is composed of an internal master node coupled with external data provider nodes existing within the agencies producing and managing source data. This distributed configuration increases capacity and technical knowledge within agencies, allowing them to better meet their own internal data management goals. The distributed architecture leverages hardware, bandwidth and staff resources across multiple systems and groups. Utilization of currently available external data feeds for sensor, remote sensing and other data sources improves access to data for AOOS users with minimal effort.

Integrating available sources of interoperable data feeds into data access applications and data management systems adds a variety of data resources at a low cost. Large quantities of real-time and historical sensor information, remote sensing satellite information and marine habitat and biological data for the Alaska region are openly available on the AOOS Data System for use through interoperability protocols. For example, NASA Earth Observations (NEO) provides an expansive array of long term oceanographic, climate and atmospheric remote sensing datasets. Real-time and historical sensor data feeds for the Alaska region are available for hundreds of sensors via SNOTEL, the National Data Buoy Center (NDBC), the Center for Operational Oceanographic Products and Services (CO-OPS) and other NOAA programs. Additional sources of interoperable data include those hosted at NASA's Jet Propulsion Laboratory (JPL), U.S. Geological Survey (USGS) TerraServer and other research organizations.

As referred to in the AOOS Strategic Operations Plan, the purpose of the AOOS Data Management Plan is to document the following:

1. The overall management objectives and protocols for the data served on the AOOS website;
2. The individuals responsible for coordination and management of observational data across the region and the procedures for soliciting and evaluating the data management team's capability;
3. The data resources including descriptions of data categories and types served by the AOOS Data System; and
4. Regional Data Stream (Management) Plans, which are data management implementation protocols for aggregate sets of non-federal data, either by source or platform, that AOOS provides access to on its data system.

2.1 AOOS Data System Management Goals and Objectives

The Data Manager and Data System Management Team for AOOS are tasked with overseeing fulfillment of four primary goals and numerous objectives within the AOOS Data Management Work Plan.

Goal 1: Provide Core Data Management Support to the AOOS Program

1. Provide technical support for AOOS cyber infrastructure;
2. Develop and maintain web-based data portal;
3. Deliver real-time, delayed-mode and historical data for in-situ and remotely-sensed physical, chemical and biological observations;
4. Deliver model-generated outputs, including both nowcasts/forecasts and reanalysis, to AOOS (and IOOS) users;
5. Implement Quality Assurance of Real-Time Oceanographic Data (QARTOD) quality control (QC) checks for AOOS real-time data feeds if available;
6. Develop and implement process for archiving into federal archives; and
7. Provide system performance and security measures.

Goal 2: Provide Data Management support to the AOOS program

1. Provide overall data management project management and oversight;
2. Engage with data providers to access, understand, and appropriately document data (metadata and QA/QC) that is ingested through the AOOS infrastructure;
3. Participate in Alaska committees and teams (including teams as determined by the Executive Director, and the joint State-Federal Data Integration Initiative) in order to facilitate data integration and interoperability within Alaska;
4. Participate in national and cross-regional committees, workshops and teams in order to further the development of a coordinated approach to IOOS data management;
5. Participate in international development of Arctic Observing Network (AON) and Sustained Arctic Observing Network (SAON) data management;
6. Work closely with the AOOS office, other data management awardees if selected, and appropriate advisory committees to implement identified user products, tools and their web interfaces;
7. Develop product requirements;
8. Beta test and refine products in order to increase their utility;
9. Provide reports as requested;
10. Develop detailed work plans with measurable timelines, deliverables, and performance metrics; and
11. Assist with development of funding proposals.

Goal 3: Develop and maintain special data products

1. Support existing products;
2. Ingest new data according to priorities developed by AOOS staff; and
3. Develop new data and information products.

Goal 4: Web Portal Hosting and Support

1. Host and maintain the AOOS web portal at www.aoos.org ;
2. Provide access to the user interface and visualization tools, data products, data query and access tools, decision-support tools, agency project tracking systems and databases, as well as IOOS Registry tools; and

3. Work with AOOS staff to update the website periodically in order to improve clarity, ease of use, and the overall "look and feel."

2.2 Data System Management Structure

The AOOS Data System Team(s) reports to the AOOS Executive Director and works directly with AOOS staff as an integral component of the overall AOOS program. An external Data Management Advisory Committee (http://www.aoos.org/about/committees.html) provides advice to the Executive Director, AOOS Board and the Data Team (Figure 1). Other external committees are used as needed to provide additional advice on implementation and user needs.

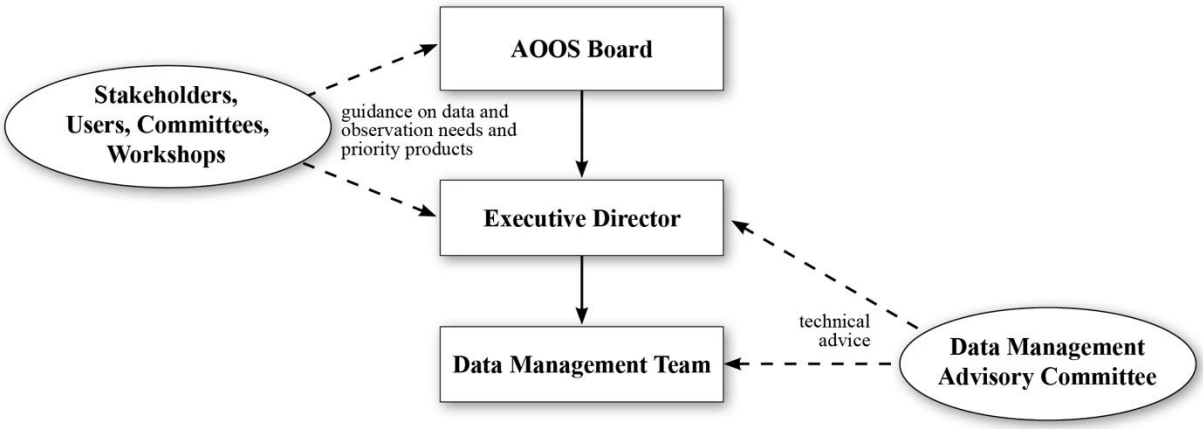

Figure 1. AOOS Data Management System.

2.3 AOOS Data System Management Team

The Data Team is involved with all aspects of the AOOS data flow, including data ingestion, conversion, discovery, maintenance of data feeds, storage, and any necessary archival services. Its primary function is to gather and serve data important to the AOOS region to end-users via standard services as recommended by the Integrated Ocean Observing Committee (IOOC) and the IOOS Program Office (e.g., OPeNDAP, SOS, etc). It is also tasked with managing and archiving any AOOS-funded data assets including oceanographic models, buoys, or other devices.

AOOS Data Manager: The AOOS Data Manager is the Data System management lead and provides project management expertise while overseeing all aspects of the AOOS Data System. The Data Manager supervises all staff on the AOOS data management team, referred to as the Data Team. Additionally, he/she contributes to proposal development and general AOOS data management reporting requirements. For the purposes of certification under the 2009 Integrated Coastal Ocean Observing System Act, the Data Manager is considered a RICE employee (see RICE Certification Requirements, Section 997.26(c)) and is one of the AOOS individuals responsible for collection, management or dissemination of observing data, and is responsive to federal government control.

AOOS Data Team: The AOOS Data Team is responsible for all AOOS deliverables in the annual AOOS data management workplan (See Section 4.0 of this Plan). Working under the direction of the AOOS Executive Director and with the AOOS staff, the Data Team designs and deploys a data management subsystem to meet the needs of the AOOS user base and that provides functional components required by IOOS as described in this Data Plan. The resulting management system will increase the access and use of data by all user groups and allow data management staff to rapidly develop new capabilities and tools to meet emerging user needs. The Data Team also provides data management and informatics support for AOOS and development capacity for map-based data portals, spatial planning tools and data management frameworks, which transfer and ingest data from external systems via interoperability protocols. The team ensures transparency and communication between client and contractor about design requirements and development progress, and remains current with, researches and employs new technologies to extend the capabilities of digital information and computer analysis systems.

2.4 AOOS Procedures for Evaluating the Data Subsystem Management Team

In 2010 AOOS chose to contract for professional data management services. Data Manager and Team selection followed a process of broadly soliciting competitive proposals to provide web portal, data management, communication and user product services for AOOS for up to 5 years. AOOS solicited proposals for two sets of services, and proposers were encouraged to bid on one or both of these components, separately or combined.

> *I. Data Management and Communication Services:* Services include providing data management support (data ingestion, metadata, relational database development and maintenance) and communication services (web portal, data clearinghouse, coordination, and communication), building upon the hardware, software, query tools and products developed over the previous years, and following the national IOOS Program guidance.
>
> *II. User Product Development Services*: Develop user-driven products and associated interface and visualization tools that will be maintained by and interact with the data system developed under component #1 above.

AOOS procedures followed during the solicitation, evaluation and selection of contractual data management support are fully described in Appendix A.

2.5 Statement of AOOS Data System Ownership and Intellectual Property

All equipment, hardware, software code and products purchased and/or developed as a result of the data management system proposal award(s) belong to the Alaska Ocean Observing System.

3. AOOS DATA RESOURCES AND ASSET TYPES

The AOOS Data System provides data to the public from multiple sources and diverse external organizations in addition to AOOS funded projects. AOOS data inventories, organized by category and

accessible through the AOOS Data System as of June 30, 2016 are provided in Appendices B, C, D and E. These appendices are updated annually, or more often as needed, to reflect changes in the data asset inventory on the AOOS Data System.

3.1 Observational Data Types

The AOOS data inventories include multiple types of data, including real-time data, near real-time data, historical data, and citizen science data. AOOS defines each data type in a consistent manner with IOOS RICE Guidelines as follows:

- *Real-time data* are ingested, served, and displayed by the AOOS Data System at the same frequency the data are collected (and sometimes reported) by the originator with little to no delay. Examples of real-time assets include weather stations, oceanographic buoys, and webcams.
- *Near real-time data* are ingested by the AOOS Data System at the same frequency that the data are made available; however, there is some delay (hours to days) between data collection and when the data provider makes it available. Examples of near real-time assets include satellite images and derived satellite products.
- *Historical data* are data that are one month old or older. AOOS historical data were sometimes collected in real-time and subsequently archived; other historical data are ingested from local or national archives upon stakeholder request.
- *Citizen science data* are collected by members of the general public who are not necessarily trained as scientists. They might be affiliated with a non-profit, education or local community organization (e.g., marine science center), or private business or enterprise. As these data are collected by volunteers and typically have limited quality assurance and quality control (QA/QC), these data require quality flags to alert the end user that the data may not be “science” quality.

3.2 Data Categories

AOOS data types are divided into four major categories that determine the level of documentation and quality control (QC) that is required for the data assets within each category:

Federally Sourced Data: Federally sourced data are assumed to be quality controlled following rigorous and documented data management and archival processes by the provider, and thus only require generic documentation by AOOS on how these data are ingested and made available to the public (Section 4.0). Federally sourced data served by the AOOS Data System are exempt from detailed data stream documentation. The AOOS Data System provides access to data from the National Oceanographic and Atmospheric Administration (NOAA), the U.S. Geological Survey (USGS), the U.S. Forest Service (USFS), the Federal Aviation Administration (FAA), the National Science Foundation (NSF), the U.S. Department of Agriculture (USDA), and U.S. Fish and Wildlife Service (USFWS) (Appendix B).

Model Products: Model products served by the AOOS Data System may incorporate or assimilate observational data (e.g., all bathymetric charts served by AOOS are from gridded models derived using “true” observations). However, they are considered a product that falls outside the realm of “true” observations, therefore are exempt from detailed data stream documentation (Appendix C).

Static Data Products: AOOS static data products are typically derived from observed data, but are displayed in a way that the original data are no longer reproducible and cannot be used to assemble a numerical observational dataset in time or space. Other types of static data products are merely representations of fixed political or legal boundary information. These products fall outside the realm of "true" observations, therefore are exempt from detailed data stream documentation (Appendix D).

Regional Data Streams: Regional Data Streams are defined here as any data resource that does not fit into the exempt categories already discussed: federally sourced data, model product, and static data product. These include not only AOOS funded observational assets, but also regional data provided by local or state agencies, private companies supporting maritime activities in coastal waters, university projects, research studies funded and conducted by other organizations. Many regional data streams originate from leveraged projects AOOS helps support, but also come from sole source providers affiliated with other entities (research, private, NGO, etc.). Occasionally, a federally sourced data asset is manipulated in some fashion prior to display and, therefore, requires documentation (e.g., federal satellite data that is transformed from a National Sea Ice Data Center (NSIDC)-binary format into netCDF). Data streams may be of any data type: real-time, near real-time, historical, citizen science (Appendix E).

The primary processes involved with data management and flow include data ingestion, standards and format, metadata and discovery, quality control, stewardship and preservation, access and dissemination, archival and security. Descriptions of the processes that consistently apply to all data are provided throughout Section 4.0. Additional data management documentation unique to individual regional data streams are provided through a systematic data stream plan template that closely follows the RICE Certification Guidance Data Management and Communications (DMAC) requirements (section 997.23 f (1-6)) and the NOAA Data Sharing Template. AOOS implements this customized Regional Data Stream Plan template to facilitate consistent documentation and to streamline any future modifications to existing data stream protocols.

The Regional Data Stream Plans use a consistent and comprehensive set of questions designed to describe how data streams with similar procedural controls are handled and managed end to end. Grouped parameters may originate from a single platform type (e.g., a mooring that provides temperature, salinity, and oxygen data, all of which are treated in a standard way); a data type that is handled similarly across all platforms (e.g., webcam imagery); or originate from a single data source (e.g., Alaska State Department of Transportation weather observations).

Exempt Data Requiring a Data Stream Plan: On occasion, a data stream that would normally be considered exempt will require documentation:

- Data products that include representations that can be used to reproduce numerical data in time or space are considered observing data, are treated as a Regional Data Stream and documented as such;
- A federal data source that is translated or transformed in some way between the source of ingestion to the AOOS access point of delivery (e.g., smoothing, block averaging) is treated as a Regional Data Stream and documented as such.

Quality Control Documentation: IOOS DMAC guidelines require that no raw data will be served on the regional association websites, and no raw data are served by the AOOS Data System. Generic quality control (QC) applied to all data ingested and provided by the AOOS Data System is described in Section 4.4.4 of this plan. QC descriptions unique to individual regional data streams are included in the Regional Data Stream Plans, and may follow one of four paths for a given data stream:

1. Follows prescribed QARTOD guidelines (required for real-time data only if a QARTOD Manual exists for the parameters in the data stream).
2. When QARTOD guidelines do not exist, some other suitable form of QC implementation is completed and described;
3. A description of the QC completed by the data provider (e.g., brief description or link to QC protocols performed at the source).
4. Data are considered exempt from QC documentation or requirements if federally sourced or designated as citizen science.

Specific QC and QARTOD applications applicable to AOOS Regional Data Streams are summarized in Appendix F. Regional Data Stream Plans are available in Appendix G, with individual data stream plans as sub-documents numbered G-1 through G-40.

4. AOOS DATA SYSTEM WORK PLAN

Details of specific work plan activities and projects are determined in agreement with the AOOS Executive Director and the Data Manager. The AOOS Board and the AOOS Data Management Advisory Committee (DMAC) contribute to the data management work plan on an annual basis. The plan is reviewed monthly by the Data Manager and AOOS Staff together, and updated as necessary on a quarterly basis. The status of the work plan is reviewed by the AOOS Data Management Advisory Committee semi-annually.

4.1 Computing Cyberinfrastructure

The AOOS Data System is the backbone of the cyberinfrastructure that is leveraged to acquire, archive and share marine data and information products. This infrastructure has been developed to meet the guidelines and specifications recommended by the NOAA-funded Integrated Ocean Observing System (IOOS) and endorsed by the federal Interagency Ocean Observation Committee (IOOC) and Global Earth Observation (GEO) Program. The data system is built using several mature, open-source interoperability and data stewardship systems to provide full-lifecycle data management services, including: data ingestion, metadata, data aggregation and assembly, data catalogue and discovery, data access and transport, data storage, and end user input and feedback.

The AOOS data system is divided into four logical tiers, which separate the suite of technologies composing the system summarized here (Figure 2).

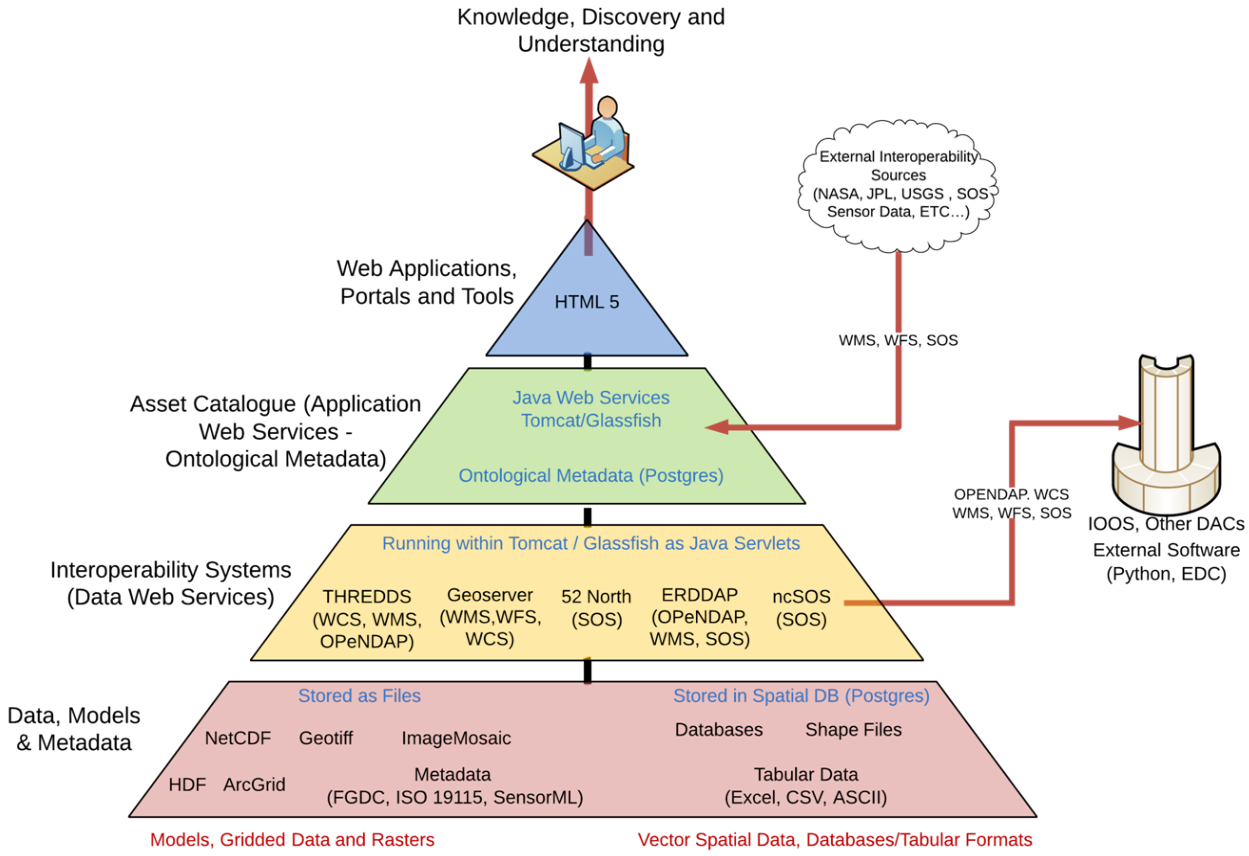

Figure 2. A schematic of the AOOS Data System that details the flow of data through logical technology tiers so that it can be consumed by users to enable discovery and understanding of marine data and products.

Tier 1 (Data, Models and Metadata) represents the source data, which are stored as files or loaded within geospatial databases. Tier 1 is the base of the data system framework where the source datasets produced by researchers, instruments and remote sensing platforms, metadata, and model outputs, enter the data system to provide the foundation for applications and user tools. These resources can be stored either in native formats or spatially enabled databases. The decision to choose one method over the other is dictated by the requirements of the interoperability system that will be serving the data. Data which has a tabular or vector form (Shapefiles, databases, Excel spreadsheets, comma separated values (CSV) text files, etc.) will be converted into netCDF files when appropriate, and will be loaded into a PostgreSQL database and spatially indexed. When possible, GeoServer, an open source geospatial data server, will then connect to the database and serve the data via WFS and WMS protocols. Imagery, raster, and model data will be stored in a file server in their native file formats. THREDDS and/or ncWMS will be used to serve NetCDF and HDF files which may contain two, three, four or higher dimensional gridded datasets. GeoServer or other OGC compliant mapping servers will be utilized to serve GeoTIFF, ArcGrid, or other two dimensional imagery/raster data.

Tier 2 (Interoperability Systems) includes systems such as Web Map Services (WMS) and Web Coverage Services (WCS), that are then implemented and connected to the underlying data sources in Tier 1.

Various interoperability servers (GeoServer, THREDDS, ncWMS, 52 North SOS, etc.) are implemented on top of source data to expose a powerful set of interfaces for other computing systems and humans to extract, query, and visualize the underlying source data. These systems will facilitate all aspects of data delivery to users in addition to providing the muscle for the machine-to-machine data transfer to national data assembly systems as required. These systems have been developed using the Java programming language and run within Tomcat servlet containers.

Tier 3 (Asset Catalogue) includes an Asset Catalogue, which provides (1) ontological metadata and (2) connections to externally-hosted data via web services. The ontological metadata in the catalogue describes the characteristics including geographic locations, spatial and temporal resolution, units, source location and CF parameter, taxonomy, date of last update, etc. of each data resource. Dynamic fields in the ontological metadata (e.g., coverage dates, which change when new data arrive) are updated automatically by the system as new data are ingested; static metadata fields (e.g., narrative descriptions of the data) are updated manually. Storing the metadata outside of the files themselves is critical to providing a responsive, up-to-date public-facing catalog. It also allows AOOS to optimize data discovery tools such as advanced searching by parameter or geographic location and build tools such as on-the-fly unit conversions for gridded datasets. External web services in Tier 3 provide the catalogue access to external (web-based) sources of information. This is commonly used to display data and basemaps from reliable data providers so data do not have to be stored and maintained by AOOS. CF Standards are provided in Appendix I.

Tier 4 (User Applications) is the final technical level and is composed of the web-based applications and tools that allow users to discover and explore the data resources in the system. Web services written in Java connect to the asset catalogue and provide applications with access to the underlying descriptions of AOOS data assets and sources. The asset catalogue contains relationally-structured maps between data types, sources, and a controlled set of definitions so that user interface applications can connect users to vast arrays of data through simple but powerful interfaces. These interfaces are available at both statewide and regional scales, including the Arctic, Cook Inlet, and Gulf of Alaska, to provide users with access to public data. These interfaces include the following:

- Public-facing data catalogs showing data assets that are updated automatically when new data are published into the system;
- A powerful, prioritized, Google-like search interface that allows users to search by geography, time, access method, or words contained in metadata descriptions;
- A secure method to share project-and file level metadata and data files with the public;
- Interactive maps that allow users to explore other, related datasets relevant to the Alaska marine and coastal environments.

User interfaces comprised of web-based applications and tools provide users access to all the data and products within the data management system. These applications make it easy to discover and explore data that have been published throughout Alaska. Finally, at the top of the pyramid atop Tier 4, users have a powerful and intuitive experience of the underlying systems working together to facilitate the discovery, accessibility and comprehension of data served by the AOOS Data System.

4.2 Data Ingestion

Observations and information are ingested into the AOOS Data System from a variety of sources, including both historical and real-time observations, forecast, nowcast and hindcast model outputs, GIS

information, and synthesized products that can be useful for layering with other data in the AOOS system. Each data asset ingested into the DMAC has its own level of data processing maturity and quality with respect to the metadata available.

Data are ingested into the AOOS system using one of several pathways:

1. Auto submission pathway from the Workspace- not visualized data;
2. Contribution by the originator;
3. Direct access or harvest from the originator website (real-time sensors, models and satellite imagery).

Project data supported by AOOS using IOOS funding provide data to AOOS in a timely manner, as stipulated in the agreed contractual statement of work. When possible, data are served in real-time; however, in many areas of the AOOS region, real-time data are not possible due to weather, ice interference with surface buoys used to transmit data, and lost power or data telemetry. On projects that do not produce real-time data, the project investigators are responsible for making sure data become accessible by AOOS as soon as possible.

4.3 Contributing Data to AOOS

The general process for data submission is outlined below:

1. Data are organized in the AOOS Research Workspace (Section 5.0) and accompanied by robust, descriptive metadata.
2. When a project is complete or its embargo period ends, the Principal Investigator (PI) selects the 'Published' option for their project in the Research Workspace.
3. The PI selects the files or folders of content to be published.
4. The Research Workspace creates a resource map of the content to be published.
5. Automated checks look for and prompt for and capture metadata content required for publishing but found to be absent from the metadata record, including:
    a. Parameters or observations;
    b. Collection methods and quality controls procedures;
    c. File formats;
    d. Update and submission frequencies;
    e. Responsible parties and affiliations;
6. A metadata rubric is returned to user to show completeness, and the PI provides additional metadata content until they're satisfied with rubric results.
7. The PI pushes the archived package on for review by a program manager or AOOS staff.
    a. The reviewer checks data files to ensure they match program data management plan requirements for formatting, quality, naming conventions, etc.;
    b. The reviewer verifies metadata quality, and if necessary, improves or sends the metadata back to the PI for additional details, depending on the assignment of responsibility within the program.
8. Reviewed data, metadata, and resource maps are finally pushed into the member node for long term preservation.

4.4 Data Flow Processes

The flow of data from the source to AOOS data portals follows the same general path for all sources as illustrated in the following flow diagram (Figure 3).

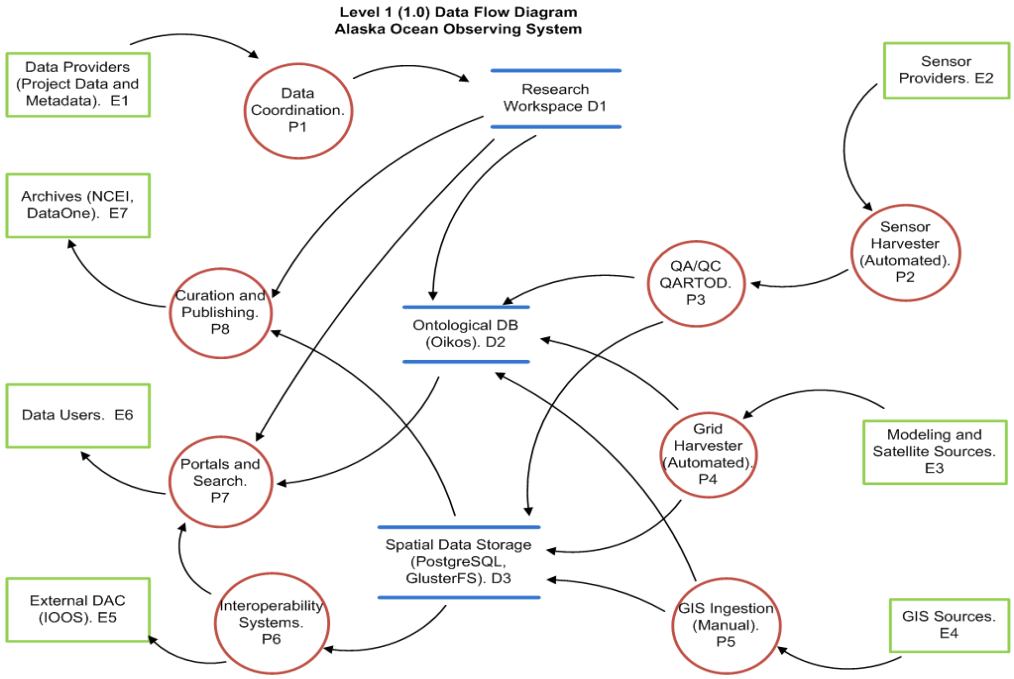

Figure 3. Flow diagram of data: From data, source, model, and GIS providers to users, archives, and external Data Assembly Centers.

4.4.1 Shared Data File Formats

AOOS provides data through six data access services (listed in Section 4.4.2), the last of which is "direct file downloads" of any of the original shared data file formats:

1. *Network Common Data Form (NetCDF)* - a self-describing, machine-independent data format that AOOS uses primarily for raster (gridded) data. Some data stored as unstructured grids use this format as well.
2. *Comma Separated Values (CSV)* - a human-readable ASCII format that is nearly universally accepted by spreadsheet and programming languages. AOOS uses CSV formats to allow users to download (1) time-series extractions from raster data, and (2) GIS vector and polygon information (e.g., boundaries).
3. *Shapefile* - an open geographic information system format for point, vector, and polygon data. AOOS allows users to download shapefiles of static GIS layers such as boundaries, biologic distributions, etc.
4. *Portable Network Graphics (PNG)* - PNG is a lossless, image format provided as an alternative to shapefiles in the AOOS catalog. PNGs are limited in use as they are pre-projected, pre-scaled, and pre-sized images of data layers. However, AOOS provides PNG files as example WMS requests, which are useful to users who cannot access GIS services and who do not understand how to manipulate WMS requests.

Data available for digital download allows users to download files with usable numbers on which they can do further analysis or visualization of their own. These fundamentally differ from images (provided

by WMS services and PNG files), which are pre-rendered with a fixed scale and applied color ramp. These types of downloads are differentiated under the “download” button in the catalog interface.

4.4.2 Data Access Points

Access points provide standardized, documented services that allow users to download what they need from AOOS without having to make person-to-person data requests. AOOS offers six access points:

1. *Thematic Realtime Environmental Distributed Data Services (THREDDS)* - THREDDS is a set of services that allows for machine and human access to raster data stored in NetCDF formats. THREDDS provides spatial, vertical, and temporal subsetting as well as the ability to select individual dimension or data variables to reduce file transfer sizes. AOOS provides THREDDS access points for raster (gridded) data stored in NetCDF format.
2. *Open-source Project for a Network Data Access Protocol (OPeNDAP)* - OPeNDAP is a protocol that can transfer binary or ascii data over the web. Like THREDDS, it provides spatial, vertical, and temporal subsetting and the ability to select individual variables to reduce file transfer sizes. Unlike THREDDS, requested data are provided as non-NetCDF, structured output. OPeNDAP output can be imported directly into graphical programs such as GrADS, Ferret, or R. AOOS provides OPeNDAP access points for raster (gridded) and time-series data.
3. Web Map Service (WMS) - WMS provides machine access to images, which can be used by individuals or programs (e.g., tiling services). Accessing programs use GetCapabilities requests to ask for image data in whatever format they require, which allows them to gather image tiles over specific areas with the projections, styles, scales and formats (PNG, JPG, etc.) that fits their needs. AOOS provides WMS access points for point, vector, and polygon information, as well as rasterwell raster data through open source software packages ncWMS, ncWMS2, sci-wms and GeoServer.
4. Web Feature Service (WFS) - WFS provides machine access to the vector elements of static layers. AOOS provides WFS access points for point, vector, and polygon information, as well as time-series and raster data data through the open source software applications (i.e. GeoServer).
5. Environmental Research Division's Data Access Program (ERDDAP) - ERDDAP is a common data server that provides access to subsetting and downloading data. AOOS provides ERDDAP access to all time-series data in the region, a subset of gridded data, and some tabled based GIS-data based products.
6. Direct File Downloads - AOOS often provides data as downloadable files. These files are mostly served in the original standard shared data file formats above, or in the case of project-specific data, in their native file formats.

4.4.3 Metadata

AOOS relies on local investigators to provide best practices for Quality Assurance (QA) on their activities related to data submitted to AOOS. Part of the data ingestion process is to establish adequate metadata and provide metadata links that provide the necessary background information to establish the purpose of the data and expected quality. AOOS requires standards-compliant metadata for project-level data (AOOS or IOOS-funded projects). Dynamic fields in the ontological metadata (e.g., coverage dates, which change when new data arrive) are updated automatically by the system as new data are ingested; static metadata fields (e.g., narrative descriptions of the data) are updated manually. Though AOOS does not require specific metadata standards for ingesting other types of data, most modern data submittals are

accompanied by standard ISO/FGDC metadata records. However, many older datasets come with informal metadata documentation that is variable in terms of completion and detail required by modern standards, and some are only accompanied with narrative information. In these cases, AOOS works to making the source information easily accessible to the end-user by providing links to source data or data providers, and making all available metadata information that came with the data available in the data catalogue.

Details and availability of metadata are discussed in individual Regional Data Stream Plans.

4.4.4 Quality Control Procedures

A primary mission of AOOS is to serve as the Alaska Regional Data Assembly Center (DAC), aggregating data from local and federal sources and making them available, accessible, and understandable to the public. Quality Assurance (QA) are procedures undertaken during the experiment and/or instrument design phases of data collection, ensuring that all the data collected is as accurate and precise as possible. Providing very few data collection devices itself, AOOS is reliant on individual data providers to provide adequate QA procedures, and they will not be discussed in this document.

Quality control (QC) processes implemented by AOOS are used to identify and flag or remove bad data after data collection. Sharing these protocols and quality flags are an important component of publicly serving data, and QC documentation is provided here and in the individual Regional Data Stream Plans.

AOOS Implemented QC Protocols

AOOS does not receive or serve any raw data transmitted directly from stations, so any applied QC procedures administered by AOOS are in addition to those applied by the data provider. For many project-based and historical datasets, AOOS provides the same data (though sometimes in converted formats) that are available from the source provider. Any QC procedures that are documented and made available to AOOS by the providers are included in the QC section of the individual Regional Data Stream Plans (Appendix G: G1-G40).

Effective 2018, AOOS began implementing policies outlined in the US IOOS Quality Assurance of Real-Time Oceanographic Data (QARTOD) manuals. As new data are ingested into the AOOS Data System, they are assessed and classified accordingly, to allow for full documentation as described in this plan, including Data Stream Plans for new assets that do not come from federal sources and that will be archived by AOOS. As new QARTOD protocols are updated and new parameter manuals developed over time, Data Stream Plans will be updated accordingly to include newly required QARTOD implementations. When QARTOD guidelines do not exist for a variable, other suitable form of QC implementation is conducted and described.

Implementation of QARTOD tests by AOOS have different processes depending on the data type -- real-time data, historical data, federal data, data without available or clear QC, and citizen science data.

1. Real-time data: AOOS ensures that quality control (QC) standards are implemented and QC flags made available for all real-time data that are not received from a federal source. AOOS currently serves various non-federal data streams that require QARTOD QC implementation. Details on AOOS planned QARTOD QC implementation are described in Appendices F and G. QC procedures differ depending on whether they are implemented by the data provider or AOOS.

QC by Data Provider

Highly leveraged programs manage some of the non-federal, real-time regional data streams from Alaska including observations from CDIP wave buoys, water level data collected by the Alaska Department of Natural Resources (ADNR) and served by NWS, and the High Frequency Radar (HFR) derived surface current data. CDIP, the NWS and the HFR Network ingest and perform extensive QC on the raw data collected by these platforms prior to making them available to the public. AOOS sources these data from these programs for display on the AOOS Data System, and therefore, is not required to perform additional QC. Please refer to the Regional Data Stream Plans for Cook Inlet CDIP Wave Buoy (Appendix G-1), ADNR Water Level (Appendix G-2) and HFR Surface Currents (Appendix G-3).

As some other data providers document existing QC performed at the source by the data originator (e.g., Canada Water Office stream gauge data), links to these procedures, or a brief summary of the QC performed are provided in the Regional Data Stream Plans. In these cases, parameters and/or configuration for the quality test are defined by the data provider. In cases where the data provider makes quality flags available within their data, those flags are ingested for display in the AOOS Data System and additional QC are not required on those assets. Roll-up summary flags and individual test flags are shown visually in the AOOS data portal with links to the QC documentation made available by the data provider (refer to QC by AOOS section below). Flags are also stored alongside the data for download in CSV and netCDF downloads, as well as via THREDDS and ERDDAP servers.

QC by AOOS

For sources that do not provide quality flags, the AOOS Data System runs QARTOD tests after ingesting observation data. Tests are run using the open-source ioos_qc library (https://github.com/ioos/ioos_qc) which implements a suite of QARTOD tests as well as other quality control algorithms. The quality test code and test thresholds are documented and publicly available through the AOOS data portal. Links to the ioos_qc methods used are available both within data charts and on sensor pages within the AOOS data portal. Thresholds used for each test are also viewable on sensor pages and users are linked to the test code in GitHub.

Within one hour after observations are ingested to the AOOS Data System, a process is run to calculate flags for the following QARTOD tests, depending on the parameter:

- Gap Test- checks that the times supplied are in monotonically increasing chronological order, and optionally that time intervals between measurements do not exceed a value.
- Syntax Test- checks for parity errors by testing if data can be extracted from the downloaded or scraped data.
- Location Test- checks that a location is which reasonable bounds.
- Gross Range Test- Checks that values are within reasonable range bounds.

- Climatology Test- Checks that values are within reasonable range bounds for a given location and depth
- Spike Test- checks if the difference in values between a data point and its neighbors exceeds a threshold.
- Rate of Change Test- checks if the first order difference of values exceeds a threshold.
- Flat Line Test- checks for consecutively repeated values within a tolerance.

Tests are run for all sensor data that do not already have QC tests applied to it and are applied continuously as new data enter the system. The quality test thresholds can be defined per sensor parameter, when input from the sensor operator or subject matter expert has been obtained. When a specific sensor QC configuration has not been defined it will fall back to a default set of thresholds for each test. For example, the Gross Range thresholds for Air Temp might be (-90C, +60C), and Barometric Pressure might be (800, 1090) mbar.

The quality flag results are made available in the AOOS data portal both visually and for download. By default data is provided with a "rollup" (e.g. summary) flag applied to it. This rollup flag is the worst case of all individual tests (see "Primary Flag" in the QARTOD Data Flags Manual). If any of the flags fail, that data point is not shown in portal visualizations, such as time series plots or anomaly charts, but the data is still available when the dataset is downloaded. An example of the rollup flags run by the AOOS system for AOOS data assets can be seen here.

For each individual quality test, the individual test flags are shown visually alongside the data. Within a timeseries chart the "flag statistics" are shown as a stacked bar plot showing the test results (i.e. pass, fail, suspect, missing data) at each data point alongside the observation data. Users can interact with flags by hovering over them to view a breakdown of individual test results. Additionally, users can turn on and off quality filtered tests entirely or by test results type using a checkbox. The quality flags are responsive to the time binning represented within the chart thereby allowing a user to view summary flags or narrow down to raw data points. An example of the individual test flags run by the AOOS system for the Marine Exchange Alaska's wind sensor from the Anchorage Port station can be seen here.

In addition to being viewable, quality test results are available for download in the AOOS data portal. The single rollup flag variable is served alongside the data in CSV downloads, as well as THREDDS and ERDDAP servers. For serving individual quality flags, a second flag variable is also available within the downloaded data for each measured parameter that describes all individual quality flags in one value. Quality tests are described in a standard way, as described by the IOOS Metadata Profile v1.2, in which QARTOD flag variables are associated with data variables using the CF "Ancillary Variables" approach. An example of the Marine Exchange Alaska's wind sensor dataset from the Anchorage Port station and embedded quality test results available in ERDDAP is here.

3. Historical Data: The need for quality metadata, including documented QC procedures, that accompanies historical and project data being received and made available through AOOS is currently being addressed at data inception. All modern (current day) incoming project data are managed and

curated by the project researchers through a tool called the Research Workspace. To facilitate the entry of more consistent and complete metadata, including documentation of QA/QC performed on the data from project researchers, a comprehensive metadata editor used within the Workspace allows the AOOS Data Management Team to directly verify the existence of appropriate and required QA/QC documentation before data are pushed to the AOOS catalog.

QA/QC procedures for historical (older than one month) data can vary widely among data providers. In the past, prior to the use of the Research Workspace, QC procedures may not have been consistently available or provided with the data. Therefore, not all of the historical project data ingested by AOOS has been accompanied with current day standard documentation in the metadata record, and might refer back to the data provider. This is often the case where AOOS re-serves and repackages (via products) historical data assets that have limited documentation and QC information. AOOS will continue to provide access to such data assets, balancing the desire to make available valuable historical data resources (that might otherwise not be discoverable anywhere else) with the need for current-day-standards quality QC documentation. As part of that effort, AOOS always provides links to the data provider, regardless of the metadata status, and all AOOS data resources provide originator information that is available with the metadata links. When documentation of methods and procedures are available, that information and the associated links to documents are provided as part of the metadata and will have associated links or bibliographies provided in the individual Data Stream Plans.

When QARTOD applies, data assets that were previously reporting real-time data and that have had their historical data stored and made available in the AOOS Data System, follow the same QC protocols as the real-time data (e.g., King Island Wave Buoy; non-federal weather data). In these cases, the applicable quality tests are run retroactively for all legacy time series data available within the AOOS data portal. The quality test results are available visually along the time series continuum within the AOOS data portal and for download following the same procedures as the real-time data.

4. Citizen Science Data: Any data that are collected by members of the general public (including private businesses) who are not trained scientists, though sometimes working in cooperation with a scientific program, will be documented as citizen or community observer sourced. These data do not require documented QC procedures. However, they do require a description of the process of how data are displayed and clearly marked as “citizen science” on the data portals. Citizen science data assets have individual Regional Data Stream Plans (Appendix G). Citizen science data served by AOOS will be flagged accordingly (e.g., Primary Level flagging standard (UNESCO 2013) Value 2, “Not evaluated, not available or unknown.”). If QC has been performed, this will be documented in the Regional Data Stream Plan for that asset. A disclaimer will be associated with any citizen science data transmissions. Flagging of data will occur as needed on existing Citizen Science Data within 12 months of certification.

5. Federal Data: AOOS serves many federal datasets that already have been through a rigorous QC process, and therefore, do not require additional QC implementation by AOOS. The National Data Buoy Center (NDBC), a part of the National Oceanographic and Atmospheric Administration’s (NOAA) National Weather Service (NWS), operates and quality controls data from moored buoys, Coastal-Marine Automated Network (C-MAN) stations, and Deep-Ocean Reporting and Assessment of Tsunamis (DART®) tsunameter stations. NDBC also quality controls and distributes environmental data from partner

program stations, including the Integrated Ocean Observing System (IOOS) (approximately 300 stations nationwide -- See the NDBC webpage at http://www.ndbc.noaa.gov/ioos.shtml for a description of the IOOS program at NDBC). All of these stations acquire environmental data used primarily for preparing weather warnings, analyses, and forecasts. The quality control procedures used by NDBC are either completely automated, or may include a manual (human intervention) component in addition to the automated QC. The completely automated procedures are performed at the National Weather Service Telecommunications Gateway (NWSTG) for real-time messages used in operational forecasts and warnings. The other procedures are performed at NDBC for data submitted for archival. The generic real-time automated procedures performed at NWSTG check to eliminate gross errors (spike test), transmission parity errors (syntax test similar to what AOOS applies), range limit exceedances (gross range) and time continuity checks. Relational checks, such as examining a wind gust to wind speed ratio, are also performed to check the quality of both measurements. Other checks assess instrumentation functionality to ensure measurements are not impacted by issues like a low battery voltage. When sensor or system degradation is detected, the affected data are removed before posting on the NDBC Web site or archival. The real-time processing procedures followed by NWSTG will not allow release of data from a degraded sensor. For more information, please refer to the NDBC Handbook of Automated Data Quality Control Checks and Procedures.

6. Procedure for data that cannot be QC'd as Directed: There may be situations that arise where AOOS or a stakeholder or client requests ingestion of a data stream that comes with informal metadata documentation and has limited QC documentation (if any). In these cases, AOOS works towards researching and documenting to the best of their ability, the expected QC for these limited data sets, making the source information easily accessible to the end-user by providing links to source data or data providers, and making all available metadata information that comes with the data available in the data catalogue. Often, these data are used to develop a product (e.g., a geospatial parameter distribution). If no QC can be documented or post-applied to a given data stream, perhaps because it is in the form of a data product, such data streams will have a low confidence score for QC, and this will be clearly defined and spelled out in the metadata file for that data stream. These data will be flagged consistently with the Primary Level flagging standard (UNESCO 2013) as value 2, "Not evaluated, not available or unknown."

7. Procedure for Citizen Science Data QC and Flagging: Citizen Science data do not have established QARTOD QC or flagging protocols. In some cases, Citizen Science data will not be accompanied by rigorous QA or QC, and data may not be qualified. Citizen Science data that have no known implemented QC tests, or projects that provide community based information where the information quality is not available will have their data flagged consistently with the Primary Level flagging standard (UNESCO 2013) as value 2, "Not evaluated, not available or unknown." If QC has been performed, this will be documented in the Regional Data Stream Plan for that asset. All Citizen Science data will be defined as such in the description and metadata associated with these data streams, and associated disclaimers will be provided.

4.4.5 Stewardship and Preservation

AOOS stores ingested data in a secure, professionally managed external facility. AOOS currently has total storage space for over 1.8 petabytes of data, and those resources are geo-replicated between Portland,

Oregon and Providence, Rhode Island. Local data storage in Anchorage is limited to temporary files only that are checked in to the main servers on a sub-daily basis.

AOOS stores all aggregated data, be it real-time sensors, forecasts results, static GIS layers, etc., indefinitely beyond the life of each individual project. This means that real-time sensor feeds will become historical sensor feeds one-month after collection, and it allows AOOS to grant users extremely rapid web-based access to all sensor data (federal and non-federal) since AOOS began aggregating feeds. The only assets that are not kept indefinitely in storage are webcam images.

4.4.6 Public Access and Dissemination

The AOOS Data System provides a variety of environmental and socioeconomic data resources in a one stop data portal, free to the public, with data originating from federal and state agencies, local municipalities, academic institutions, research organizations, private companies, non-profit organizations, and community observers. Any data served by the AOOS portal carries with it the permission to view and access, and carries no privacy or ethical restrictions. Data access is defined here as being permitted to download data through an AOOS data portal.

Occasionally, a data sharing agreement between AOOS and a data provider will identify the existence of intellectual property rights (IPR) to the data and this is noted in the applicable Regional Data Stream Plans. However, IPRs do not restrict access to any of the data that is freely served through AOOS data portals. IPR information is merely provided out of courtesy for the data provider, and it is an unwritten expectation as well as a best practice, that as with all data used by someone other than the originator, clear credit is given to the data source (the originator) and data provider (in this case AOOS) in any work or publications that emanate from using data accessed via the AOOS portals.

New datasets received by AOOS are immediately available to the public after data ingestion and documentation is complete; however, they are added to the searchable, public catalog only after the data provider is contacted and any feedback (if given) is taken into account. Once published in the catalog, datasets are promoted via the AOOS website, social media accounts, and email newsletter.

4.4.7 Data Archival

As a federally funded program, AOOS is required to submit data it generates to a national archive center. AOOS is working with the National Centers for Environmental Information (NCEI) to assist with the preservation of data generated or made publicly available by AOOS. The bulk of the data assets managed by AOOS are non-real-time, nonfederal assets, sometimes from small data originators, and often from distinct research projects or large, integrated ecological research programs. AOOS will continue to make all data served on the AOOS portal available to NCEI, and will work with NCEI to make available any data asset they would like to curate. The AOOS Data Management Team is consulting with several NCEI staff members (Matthew Biddle, John Relph, and James Partain) on automating the submission of AOOS-generated and managed data assets and AOOS-managed non-federal real-time assets to the NCEI.

AOOS Generated and Managed Assets: AOOS serves many datasets that already have archival mechanisms in place, including CDIP wave buoy data, real-time sensor streams from federal sources (e.g., NSF Circum-Arctic Lakes Observing Network, NOAA CO-OPS, NOAA NDBC, NOAA PMEL, USGS NWIS, etc.), and marine mammal telemetry data from the BOEM-funded MARES program. NCEI is working with the AOOS Data Management Team on drafting a Submission Information Form (SIF) for AOOS as well as IOOS sensor data. The AOOS Data Management Team completed what is required for the SIF, and NCEI will determine the final scope of the SIF for sensor-based data, projected to be completed by Spring 2017.

AOOS generated and managed data stream submissions will be completed using the sensor SIF, describing elements similar to the information provided by the Regional Data Stream Plans (Appendix G). The King Island Buoy (Data Stream Plan G23) is currently the only AOOS generated asset which will utilize this SIF, and is used here to illustrate the kinds of information to be included on the sensor SIF:

1. List of parameters collected: sea water temperature, significant wave height, maximum wave height, wave from direction, dominant wave period, air temperature.
2. Process steps/quality control including final format: When buoy is redeployed, data will be sent from the buoy via modem to the AOOS Data System, converted to ASCII format, tested according to applicable QARTOD procedures, and aggregated into daily netCDF files.
3. Timing of data submissions: Historic data will be submitted once, as a single netCDF file. When the buoy is redeployed, submission timing will be determined during SIF revision in coordination with NCEI and based upon the projected size of the netCDF files over several durations (e.g. annually, semi-annually, quarterly, etc.).
4. Development of data documentation: The submitted netCDF files will contain attributes describing buoy location, deployment and retrieval dates, instrumentation, points of contact, and other information captured by the Attribute Convention for Data Discovery, as applicable.
5. Data disposition: While developing the SIF, the AOOS Data Team will consult with NCEI staff to determine mechanisms acceptable to both parties to transfer the data, with an initial intent to explore the appropriateness of having the NCEI implement FTP pull processes from a specified FTP location and using an agreed-upon netCDF format and manifest procedure.
6. Data affiliations:
   - Submitting institution: AOOS;
   - AOOS Operations Director point of contact: Dr. Carol Janzen (janzen@aoos.org);
   - AOOS Data Manager point of contact: Rob Bochenek (rbochenek@axiomdatascience.com).

Other real-time sensor Data Streams that will utilize the sensor SIF include Alyeska Resort Weather Stations (Data Stream G7), Marine Exchange of Alaska (AIS) Weather Stations (G10), and Cook Inletkeeper Stream Temperature Monitoring (G17).

NCEI does not currently support an automated way to generate SIFs for non-sensor data (e.g., historical), which comprises a significant portion of AOOS data offerings. It is our understanding that NCEI can generate a SIF per each non-sensor dataset for certain datasets they've identified as of interest

for NCEI archival, probably as one-offs through the Send2NCEI application. An example of non-sensor data is the Cook Inlet Beluga Whale Prey Availability in Winter Habitats (Data Stream G15).

AOOS Managed Non-federal Assets

The bulk of the data assets managed by AOOS are non-real-time, nonfederal assets, sometimes from small data originators, and often from distinct research projects or large, integrated ecological research programs (e.g., *Exxon Valdez* Oil Spill Trustee Council, Gulf Watch Alaska Long-term Monitoring Program). Currently, 12 (including King Island Buoy) of the 40 AOOS Data Streams are of interest to NCEI, three of which may already be archived. Data stream plans provided in Appendix G outline any plans for archival of each data stream asset. Though much of these data may fall outside the immediate purview of the NCEI, all data will be offered to NCEI for archival and will remain available for future archival there.

To facilitate archival of the valuable assets not accepted by NCEI, AOOS plans to also archive selected project data (primarily integrated ecological research projects) in the DataONE network. DataONE is a community driven project funded by the National Science Foundation, providing access to data across multiple member repositories, supporting enhanced search and discovery of Earth and environmental data. DataONE promotes best practices in data management through responsive educational resources and materials. More information about DataONE can be found at this link: https://www.dataone.org/

DataONE is one of the only available options for archiving the data that is regionally very valuable, but not of interest to NCEI in the near term. A major advantage of DataONE is the fact that NCEI is a DataONE member node itself. As a member node, NCEI has access to all data archived in the DataONE network and, should it choose, can expose those data through its own catalog and search interfaces at anytime.

The AOOS Data Management Team has begun work to stand up a DataONE Tier 3 Generic Member Node (GMN) to be attached to an updated version of the Research Workspace. This Tier 3 member-node will serve as the primary archive for AOOS-managed data assets. DataONE Tier 3 status is the second highest tier, and it encompasses the abilities of Tiers 1-3: reads all public objects in the data store including all science data, science metadata, and resource maps (Tier 1); controls access to objects using authorization and certificate-based authentication (Tier 2); and has ability to write to the member nodes allowing us to create and maintain data and metadata in the network (Tier 3). Tier 4, which AOOS will not be providing, allows the DataONE infrastructure to geo-replicate copies of data objects in the network onto member node servers, so it doesn't add anything new to AOOS's existing archive objectives.

To date, developers on the AOOS Data Team (Axiom Data Science) have consulted with DataONE developers employed by the National Center for Ecological Analysis and Synthesis (NCEAS) to understand best practices and implementation options for standing up a member node. To date, the local deployment of the DataONE GMN has been installed in a virtual machine on AOOS cyberinfrastructure resources. Upon completion of the updated metadata editor (Section 5.2) the member node will be ready for local testing, and is planned to enter a staging environment for more

robust testing to comprehensively evaluate content and member node to coordinating node interactions. This member node is planned to be online and operational in 2017.

After deployment and launch, data and metadata submitted will be submitted to the new DataONE GMN through the Research Workspace. Documentation of specific submission procedures will be created to ease the organization, evaluation, and submission of data to the member node after testing internally and with select user-groups.

4.4.8 Performance and Security

The AOOS operates two High Performance Compute (HPC) clusters located in data centers in Portland, Oregon and Providence, Rhode Island. These HPC resources are composed of approximately 2500 processing cores staged in a series of interconnected blade arrays as well as 1.8 petabytes of storage. Compute nodes and storage nodes are connected over a low latency, converging network fabric (40 Gb/Sec Infiniband). GlusterFS is employed as a storage software abstraction layer that enables clients and storage servers to exploit data transfer over Remote Direct Memory Access (RDMA) protocols. This configuration enables data throughput from the storage cluster to the compute cluster to reach speeds greater than 160 Gb/Sec in high-concurrency situations. The AOOS Data Team also has a dedicated multi-braided 1 Gb/Sec high speed internet connection for large file transfers between external data centers and for high-bandwidth demands of centralized web based applications. The current AOOS Data Team (Axiom Data Science) provides the following enterprise-level infrastructure capabilities:

- *Security and Redundancy* — The AOOS Data Management Team operates two physically distributed, mirrored data centers (Portland, OR and Providence, RI). Each independent cluster also implements several levels of redundancy and backup. The two physical locations ensure that multiple redundant copies of data exist in addition to web application servers. Several layers of physical hardware (enterprise level firewalls) and system monitoring software (NAGIOS) are also in place to provide hardened cyber security.
- *Capacity and Performance* — High Performance Computing (HPC) has been a component of the AOOS technical strategy since early 2011. The current AOOS Data Team (Axiom) operates its own private “cloud” of compute and storage resources that data managers can provision to specific tasks and roles. The current number of processing cores and storage is scalable to allow additional resources to be added as necessary. AOOS Data Team engineers at Axiom have demonstrated that large GIS, model, and remote sensing datasets require HPC environments to be visualized and queried over web-based interfaces. Because HPC is achieved through load balancing and parallelization, these types of systems also provide the added bonus of high availability and redundancy.

4.4.9 System Change Management Processes

AOOS has a system change management process that is designed to control the implementation of all changes made to any device or application within the AOOS production environment. The operational requirements of the AOOS data assembly center demand highly available and functional data services. The system change management process exists to ensure AOOS can provide a high level of availability and integrity in the delivery of technology services. The scope of the process includes all infrastructure, applications, and services used by the AOOS client community for research or data access purposes. The

objectives of the process are to ensure that all changes are properly analyzed, documented, and communicated to AOOS staff and to all functional groups and clients potentially affected by, or involved in, their execution. Procedures required before, during, and after any change execution and the respective areas of responsibility are documented. The proper analysis and testing is performed a priori to assess the need for a change versus the potential impact (good/bad) of the change. No change is executed without first being properly planned, documented, peer reviewed, tested, and approved. Specifically, AOOS utilizes a separate staging environment for isolating and testing new software systems and evaluating how they affect the AOOS data system before they are deployed into production systems. User communities are notified prior to release of any major upgrades or system changes. Each system change is tracked via several mechanisms including task tracking systems and an internal Gitlab software repository, which enables the team to document what was changed, how it was changed, when it was changed and by whom. The AOOS data management staff at Axiom can roll back system changes to previous checkpoints if for some reason the new system change causes unforeseen problems in addition to running ad hoc reports to describe changes over time.

5. AOOS DATA SYSTEM APPLICATIONS

The AOOS Data System hosts several integrated data management tools to ease data access, storage, and sharing by its users including the Research Workspace and its metadata editor, and the AOOS Ocean Data Explorer, the statewide data portals. These are further described below.

5.1 Research Workspace

The AOOS web-based data management application, named the Research Workspace ('Workspace'), is used to assemble, store, and share data by researchers or AOOS partners. Since its release in April 2012, the Workspace user-base has grown to more than 500 individuals from a number of large-scale scientific research programs, including the *Exxon Valdez* Oil Spill Trustee Council programs, the North Pacific Research Board's Gulf of Alaska Integrated Ecosystem Research Program and Annual Research Programs, the Distributed Biological Observatory, the Marine Biological Observation Network (MBON, which includes the Arctic MBON), the Arctic Animal Telemetry Network, the Marine Arctic Ecosystem Study, and several other integrated multidisciplinary programs. Users have uploaded over 18 terabytes of data spread across more than 800,000 files.

The Workspace provides users with an intuitive, web-based interface that allows scientists to create *projects* to represent particular scientific studies or focuses of research within a larger effort. Within each project, users may create topical groupings of data using folders and upload data and add contextual resources (e.g., documents, images and any other type of digital resource) to their project by simply dragging and dropping files from their desktop (Figure 4). Standard, discovery-level ISO 19115-2 compliant metadata can be generated for both projects and individual datasets. Users of the Workspace are organized into groups, and everyone within a group can view the projects, folders, and files uploaded by other group members. This allows preliminary results and interpretations to be shared by geographically or scientifically diverse individuals working together on a project or program before the data are shared with the public. It also gives program managers and other stakeholders a transparent and front-row view of how users have structured and described projects, and how their programs are progressing through time. The Workspace has the following capabilities:

Secure group, user, and project profiles: Users of the Workspace have a password-protected user profile that is associated with one or more disciplinary groups or research programs.

Advanced and secure file management: A core functionality of the Workspace is the ability to securely manage and share any type of digital resource in real-time among researchers and study teams.

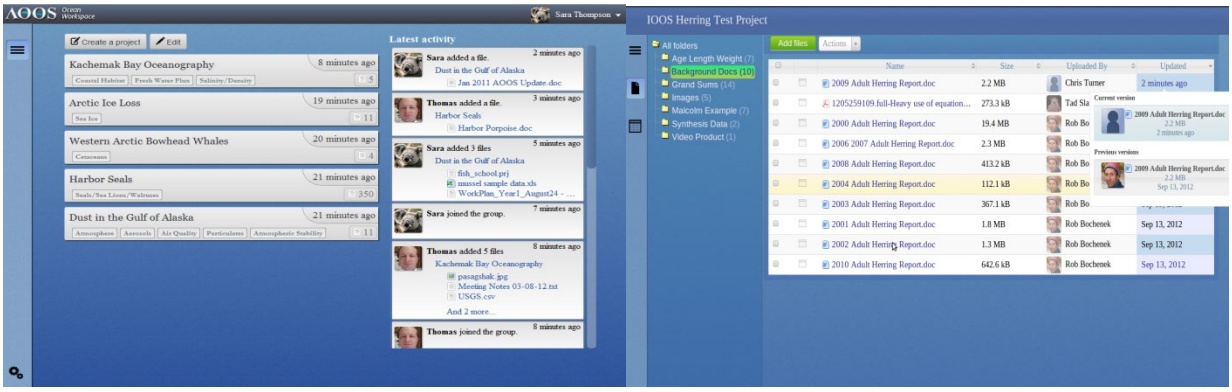

Figure 4. Screenshots of project and file management in the workspace. Left: A list of projects to which the example user has access rights. Right panel: The interface a researcher would use to organize independent files into folders and the versioning system in the workspace.

Specifically, the Workspace employs the following technological components:

- Database systems — PostgreSQL 9 is used for storage of tabular and relational data representations and is extended with PostGIS for spatial data. All data uploaded to the Workspace is replicated across multiple database servers to provide redundancy and ensure high availability.
- Object storage and schema-less data representations — MongoDB is used as a persistent NoSQL storage and query system for file objects, tabular data (flat structures) and hierarchically structured data (generally XML). MongoDB allows horizontal scaling through sharing across physical devices and provides redundancy and high availability through replication. The MongoDB instance consists of a three-node cluster, and each node maintains a complete replicate of the others. Data within each node is further redundant by virtue of RAIDed disk arrays.
- Web tier — The web services used by the Workspace are developed using Java and integrated into a web application framework called Play!, which provides a stateless architecture for Java and Scala development. The RESTful, stateless design allows services to be scaled across application nodes for load balancing, redundancy and horizontal scalability.
- Caching and pub/sub — Redis is used as an intermediary between the web and data tiers. It also serves as a pub/sub interface for managing communications between web tier nodes and serving real-time connections to browser clients in a scalable manner.
- User interface — The user interface of the Workspace is composed of several JavaScript and HTML5 libraries and integrates with server side modules wrapped into the Play! framework.

5.2 Metadata Editor

The Workspace includes an integrated metadata editor to support the documentation of data and facilitate its accuracy and reuse. Content collected in the Research Workspace metadata editor uses fields from the ISO 19115 suite of standards for geospatial metadata, which is the FGDC endorsed successor to the CSDGM, extended to describe taxonomic classification for biological datasets. To facilitate taxonomic description, the AOOS Data Team, Axiom, developed a tool that allows users to search the ~625,000 taxonomic entities of the Integrated Taxonomic Information System (ITIS) in order to rapidly add species information to metadata. Because the Workspace is a cloud-based service, researchers can move between computers during the metadata generation process in addition to allowing team members and administrators to simultaneously review and edit metadata in real time.

The metadata editor is used as a core component in the AOOS Data System. Metadata are generated at multiple points throughout the project lifecycle. At the onset of the project high-level overview information about the project is documented, which includes the location, project timeline, contacts, keywords, taxonomic species, and expected data. As the project progresses and data are collected and moved through the quality control, processing, and analysis phases, more descriptive metadata can be captured by researchers. While the workflow for creating metadata may vary project-by-project, annual metadata revisions help to keep pace with new data submissions and changes to the collection procedures. As the project data matures, the completeness, content, and quality of the metadata record should also mature to robustly describe the data and meet national format standards.

After metadata has been written that complies with content and quality requirements, the Workspace can be used as a gateway to publish data and associated metadata to AOOS's publicly-accessible data portals (Figure 5), described in more detail below. This feature simplifies the publishing of data and metadata for researchers and data managers.

Figure 5. Left: the current Workspace metadata interface. Right: a Workspace project as seen from the AOOS data portal.

An updated metadata editor is under development for the Workspace (Figure 6). The new editor includes more metadata fields, allowing more robust descriptions of datasets and their connections to other resources, and eases the metadata generation process with short, modular, and easily understood entry forms. The new editor will be released in Fall 2016 with templates being released periodically afterward.

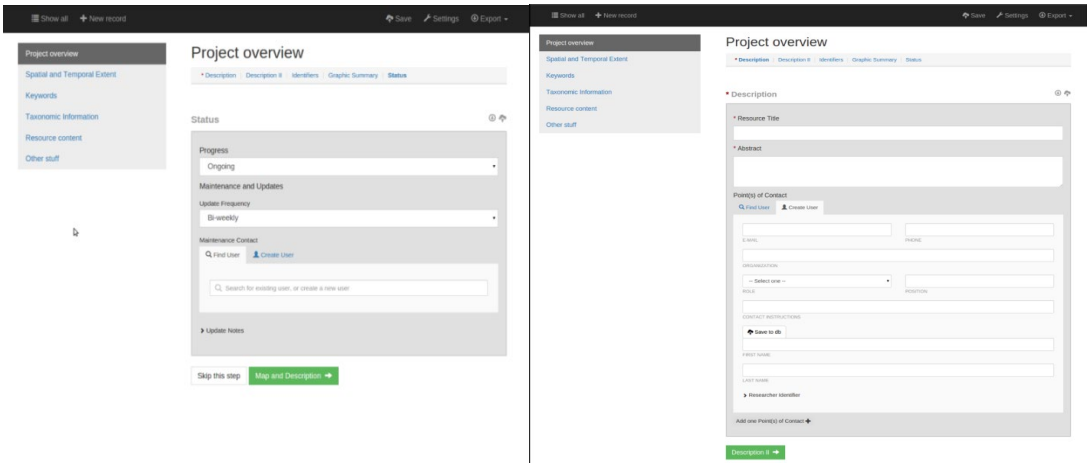

Figure 6. Screenshots of the in-development metadata editor. Left: the 'Status' tab on the 'Project Overview' metadata page, with fields to capture status, update frequency, and point of contact for project updates. Right: the basic 'Description' tab on the "Project Overview' page, showing title, abstract, and point of contact fields.

5.3 AOOS Data Portals

Through its Data System, AOOS maintains numerous data portals, which are customized public web interfaces that allow scientists, managers, and the general public to discover and have access to public data from many sources (http://www.aoos.org/aoos-data-resources/). Some of these portals include: the Ocean Data Explorer, a statewide portal providing access to all of AOOS' public data; regional subsets of the Ocean Data Explorer, such as the Arctic Portal and the Gulf of Alaska Portal; the Real-time Sensor Map that accesses 3,000 real-time sensors; the Model Explorer that displays satellite observations, model predictions, and numerical simulations, and more. Other portals are more thematically organized such as the Cook Inlet Response Tool and the Cook Inlet Beluga Whale Ecosystem Portal. The portals use the metadata and other contextual information to develop a series of search indexes using a highly scalable technology called Elastic Search. Elastic Search is a Java-based distributed indexing scheme that allows entire collections of documents, databases, and flat files to be indexed via several dimensions. When implemented, collections of information can be searched rapidly by spatial queries, time, text patterns, parameter and taxonomy. This technology facilitates data discovery and access to information, metadata, and data using a Google-like search interface.

Within the data portals, users can search or browse real-time conditions, operational and research forecasts, satellite observations, and other spatially referenced datasets that describe the biological,

chemical and physical characteristics of Alaska and its surrounding waters. Data in the portals can be interactively mapped by adding and removing layers, selecting base maps, and seeing changes over time with an interactive time slider. Using the portals, users can also access metadata and project contacts, as well as download data in a variety of formats.

6. REFERENCES

AOOS Strategic Operational Plan, URL

Public Law No. 111-11, Integrated Coastal and Ocean Observation System Act of 2009 (ICOOS Act).

https://www.ioos.noaa.gov/wp-content/uploads/2016/04/Public_LawNo111-11HR-146_-PassedSigned_033009.pdf

NOAA Integrated Ocean Observing System (IOOS) Program Office Certification Criteria, 2012. Interagency Ocean Observation Committee, Washington, D.C. 11p.

https://www.ioos.noaa.gov/wp-content/uploads/2016/04/IOOS-Certification-Criteria_4-25-12.pdf

NOAA Integrated Ocean Observing System (IOOS) Program Office, 2015. Regional Information Coordination Entities (RICE) Certification Requirements Guidance.

https://www.ioos.noaa.gov/wp-content/uploads/2016/04/cert_requirementsguidance_081015.pdf

NOAA Integrated Ocean Observing System (IOOS) Program Office White Paper (v1.0), March 12, 2010, titled “Guidance for Implementation of the Integrated Ocean Observing System (IOOS) Data Management and Communications (DMAC) Subsystem”

(http://ioos.gov/library/dmac_implementation_2010.pdf)

UNESCO, Intergovernmental Oceanographic Commission of UNESCO, 2013. Ocean Data Standards, Vol.3; Recommendation for a Quality Flag Scheme for the Exchange of Oceanographic and Marine meteorological Data (IOC Manuals and Guides, 54, Vol.3) 12 pp. (English.)(IOC/2013/MG/54-3).