1. **Program Number:**   *See,* Reporting Policy at III (C) (1).

14120114D

2. **Project Title:**   *See,* Reporting Policy at III (C) (2).

Gulf Watch Alaska Program - Data Management

3. **Principal Investigator(s) Names:**   *See,* Reporting Policy at III (C) (3).

Rob Bochenek

4. **Time Period Covered by the Report:**   *See,* Reporting Policy at III (C) (4).

February 1, 2014-January 31, 2015

5. **Date of Report:**   *See,* Reporting Policy at III (C) (5).

March 1, 2015

6. **Project Website (if applicable):**   *See,* Reporting Policy at III (C) (6).

www.gulfwatchalaska.org

https://workspace.aoos.org/group/5186/projects

7. **Summary of Work Performed:**   *See,* Reporting Policy at III (C) (7).

Project investigators continue to provide core data management support and services to Gulf Watch Alaska (GWA), an EVOSTC-funded long-term monitoring (LTM) program. The focus continues to be on refining protocols for data and metadata transfer, data formatting and metadata requirements, improving search and discovery services, and salvage of historic data, for both those data funded by the *Exxon Valdez* Oil Spill Trustee Council and ancillary historic data from other projects.

PIs have participated in regular GWA Program PI meetings, including the in-person meeting in November 2014 and the January 2015 meeting at the Alaska Marine Science Symposium, and are coordinating activities between the Herring Research and Monitoring (HRM)and GWA programs. PIs also worked with the Program Management Team to follow up on recommendations developed through the January 29-30, 2014 Data Review Workshop.

The AOOS Ocean Research Workspace, rolled out to PIs in Year 1, continues to be used as the internal staging area for PI data and work products, with individual PI user and group profiles created and further refined. Axiom staff continue to provide training via webinars and support through email and in person meetings. GWA Program PIs continue to use the system to organize and consolidate their project level data. Software engineers at Axiom provide support for the Workspace, resolve bugs and implement new functionality in response to user feedback.

The GWA project data is a key component of the Alaska Ocean Observing System's Gulf of Alaska Data Portal. The portal showcases GWA project data once it becomes public alongside other environmental data sets ingested by the project team.

Table 1. Project Milestone Status

Objective/Deliverable/Milestone	Status
Objective 1: Provide data management oversight & services, including data structure optimization, metadata generation & data transfer. Audit data and restructure and reorganize for public access.	Ongoing All 2012 and 2013 data posted on Workspace.
Objective 2: Consolidate, standardize, and provide access to study area data sets that are critical for retrospective analysis, synthesis, and model development.	Ongoing
Objective 3: Develop tools for user groups to access, analyze, and visualize information produced by the GWA effort.	File level metadata exposed in the GWA portal. Portal search is currently being rebuilt. 4D visualization system in the planning stages. Initial GWA data sets being converted to netCDF for early 4D demo and long term preservation.
Objective 4: Integrate all data & metadata into AOOS data system and Gulf of Alaska Data Portal for long term storage and public use.	Ongoing Historical data portal is a DataONE member node. With new additional funding in 2015, will work to make full AOOS data portal a DataONE node.

Objective 1

The primary results produced by this project include the acquisition and documentation of GWA PI-produced data sets and the aggregation of ancillary environmental data sets for integration into the AOOS Gulf of Alaska Data Portal. Investigators continue to improve the Ocean Workspace in response to user feedback. The increase in use by PIs is represented in the following figures. All 2013 data are now posted on the Workspace, per the Program Management data sharing protocols.

The Ocean Workspace is used by Gulf Watch Alaska program managers and investigators to facilitate many of the logistical, curatorial, and preservation-oriented aspects of data collection and management. Improvements to the Workspace, while not explicitly funded as a part of the GWA Data Management project, will continue to be made based on feedback from users. In 2014, the Workspace metadata editors were expanded in response to a request from GWA program management to include new tool to provide detailed definitions of attributes used in tabular data files. This year, Axiom engineers developed and released a new tool to automatically read in tabular CSV data files, recognize column headings in the file, and provide metadata fields for defining those headers in standards compliant elements. The tool is shown below in Figure 4. Immediately below, figures 1-3 summarize Ocean Workspace use by GWA PIs in FY2014, followed by a description of the Ocean Workspace.

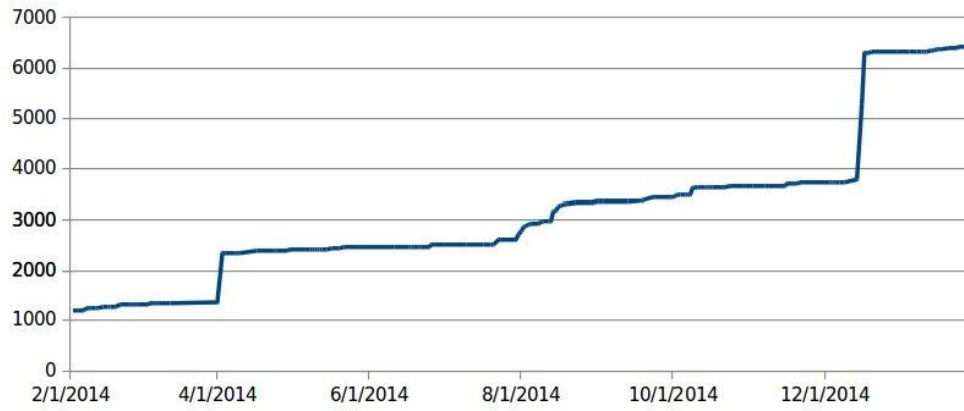*Figure 1. The number of files uploaded by GWA team members in FY 2014.*
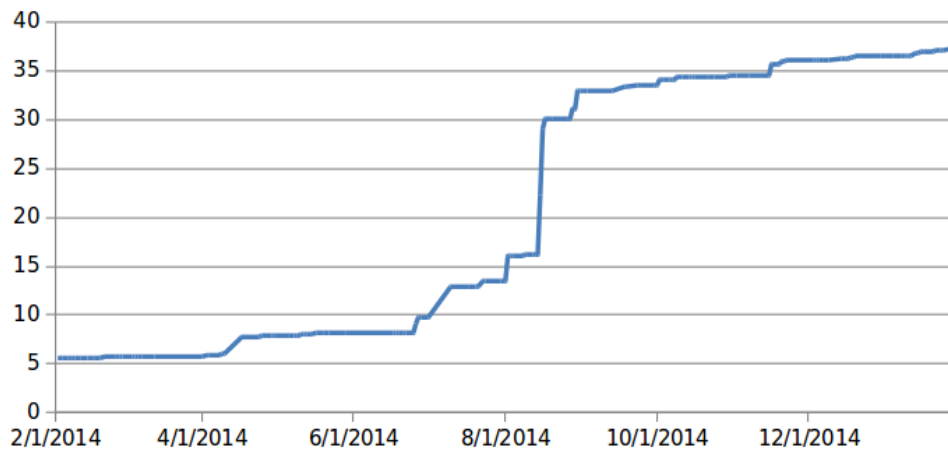
Figure 2. The amount of total storage in Gb used by GWA team members in FY 2014.
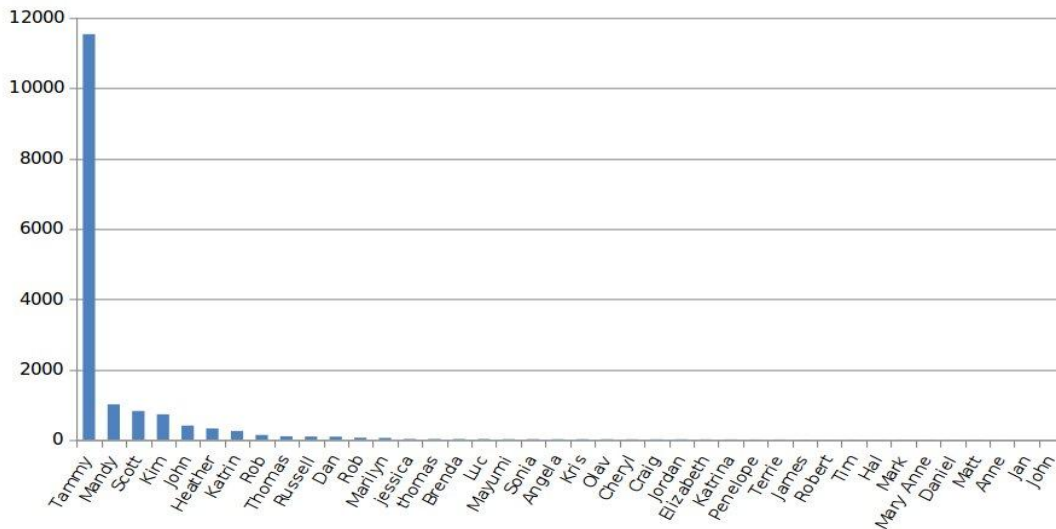
Figure 3. The distributions of file upload effort across individual GWA users.

The Ocean Workspace

The Ocean Workspace is a web-based data management application built specifically for storing and sharing data among members of scientific communities as an internal staging area prior to public release of data on a completely public portal. In addition to the Gulf Watch Alaska program, more than twenty regional, national, and private research groups currently use the Workspace, which has over 350 active individuals sharing thousands of digital files. The Workspace provides users with an intuitive, web-based interface that allows scientists to create projects, which may represent scientific studies or particular focuses of research within a larger effort. Within each project, users create topical groupings of data using folders and upload data and 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 into their web-browser. Standard, ISO 19115-2 compliant metadata can be generated for both projects and individual files. Users of the Workspace are organized into campaigns, and everyone within a campaign can view the projects, folders and files accessible to that campaign. 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 is shared with the public. It also gives program managers, research coordinators and others 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. The interface allows users to navigate between groups in which they are involved through a simple drop down control. Transfer of data and information occur over Secure Socket Layer (SSL) encryption for all interactions with the Workspace. The Workspace supports authentication through Google accounts, so if users are already logged into their Google account (e.g., Gmail, Google Docs, etc.), they can use the Workspace without creating a separate username and password.

Metadata authoring — Metadata elements currently available to researchers in the Workspace are common to the Federal Geographic Data Committee (FGDC) designed Content Standard for Digital Geospatial Metadata (CSDGM) and the ISO 19115 standards for geospatial metadata, extended with the biological profiles of those standards. Axiom also developed an integrated FGDC biological profile extension editor that allows users to search the ~625,000 taxonomic entities of the Integrated Taxonomic Information System (ITIS) and rapidly generate taxonomic 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.
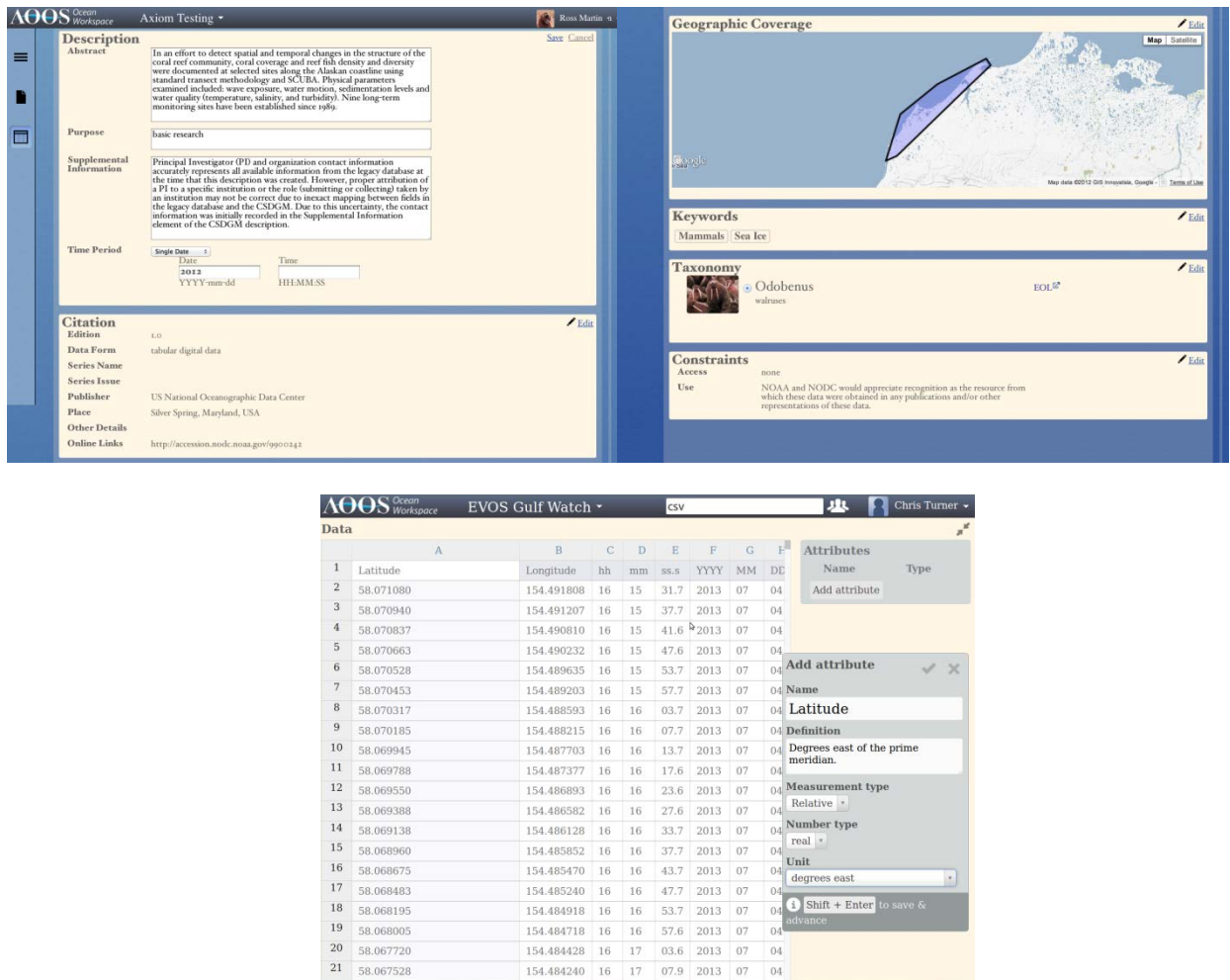

Figure 4. Screenshots of the Workspace metadata interface, clockwise from top-left: the interface to author basic descriptive and citation metadata fields, a tool which allows researchers to describe the geographic extent of the project, keywords, taxonomic information and data constraints, the new tool for editing attributes in tabular data files.

Advanced and secure file management — A core functionality of the Workspace is the ability to securely manage and share project-level digital resources in real-time with version control among researchers and study teams. Users of the Workspace are provided with tools that allow them to bulk upload files, organize those documents into folders or collections, create projects with predefined and user-created context tags, and control read and write permissions on files within projects. The

Workspace also has the ability to track file versions: if a user re-uploads a file of the same name, the most current version of the file is displayed, but access is provided to past versions as well.

Figure 5. Screenshots of project and file management in the Workspace. The first screenshot shows a list of projects to which the example user has access rights. The second screenshot displays the interface a researcher would use to organize independent files into folders, and the way two versions of the same file are tracked by the Workspace.

Objective 2.

Consolidating, standardizing, and providing simple access to relevant study area datasets that are not part of the GWA effort adds value now and beyond the life of the GWA effort. By leveraging the work done by other research, modeling, and monitoring efforts in the Gulf of Alaska, the GWA project contributes to a deeper understanding of the Gulf of Alaska ecosystem. In 2014, Axiom data analysts added more than 160 additional data layers to the Gulf of Alaska portal. Many of these new data layers expand on or update existing datasets in the portal, such as the new Shorezone and Environmental Sensitivity Index layers describing the concentrations of sensitive resources along sections of the Cook Inlet shoreline, and the new Aquarius satellite Version 3.0 measurements of sea surface salinity and sea surface winds,

Objectives 3 & 4.

In 2013, the data management team released the Alaska Ocean Observing System's Gulf of Alaska data portal, which integrates data and project information produced by GWA researchers with 260 additional GIS, numerical modeling and remote sensing data resources specifically for the Gulf of Alaska region. The team leveraged the AOOS portal, which was developed using other funding and had these additional features: an integrated search catalog which allows users to search by category or key word, ability to preview data before downloading files, and advanced visualization tools. Once the program's monitoring data has been ingested into the Ocean Workspace, quality controlled, and approved as final, then it is ingested into the Gulf of Alaska Data Portal for full public access.

During 2014, a number of updates were made to the AOOS data system, the benefits of which are shared by the EVOS GWA program and the other research groups supported by or working with

AOOS. These improvements are separated below into work completed in 2014, and work begun in 2014 and still underway.

Table 2. Gulf Watch Alaska and Herring Research and Monitoring programs: Data life cycle.

	PIs	Program Mgmt Team	AXIOM	NCEAS	EVOSTC & Trustee agencies
Data collection & any telemetry	PI/agency responsibility; established sampling protocols for each component.	Review & maintain sampling Standard Operating Procedures (SOPs). Coordinate, with Science Coordinating Committee, consistency in sampling across the program.	Store current Standard Operating Procedures within Ocean Research Workspace.		Fund data collection projects and programs. Establish basic requirements: quality data, well documented, publicly accessible, archived.
QA/QC	PI responsibility based on agency or entity requirements. Documentation of instrument calibration & data QA/QC procedures to be included in sampling SOPs & project metadata.	Review QA/QC documentation before accepting data. Limited QA/QC performed on metadata to ensure it has required information (e.g. date, time, location, etc.) and data fields are appropriately documented (e.g. units in column headers).	Working with GULFWATCH program coordinator (Tammy Neher), specific datasets are aggregated together and reviewed for problems to prepare for synthesis efforts. Mostly rely upon PI for QA/QC.	For historical data, limited QA/QC (e.g., columns, domain, units) is performed and provided in metadata documentation to ensure it has required information. If original PIs are unavailable then any issues are simply noted in metadata.	Establish clear requirements for program and coordinate on agency data standards.
Metadata	PI responsibility to provide metadata according to agency and team standards.	Works w/PIs & data team to develop requirements. Assists PIs & reviews project level and file level metadata files.	Metadata can be created through the Workspace on the project level or file level using the ISO suite of protocols with taxonomic extensions (ITIS). Other metadata formats can be incorporated as well.	For historical data projects, NCEAS researches data and provides metadata as available to reconstruct the data set. Metadata are extracted from reports, papers, and other available materials. Metadata are provided in EML format using tools developed at NCEAS (web entry, and Morpho entry).	Coordinate on agency metadata requirements and standards.
Internal data access and staging	Post data on Ocean Research Workspace as soon as possible, but no later than 1 year after collection.	Keeps records of data availability. Assists PIs in posting data on Ocean Research Workspace. Coordinates with Axiom/AOOS and NCEAS on user requirements for Workspace.	Provide Workspace as internal staging area for use by team. Work w/team to develop additional functionality for team use. Workspace is highly leveraged tool that is password protected.	Use Redmine ticket system to track the lengthy process of finding, acquiring, and processing historical data. As data are processed, they are inserted as private objects into the GoA Member Node, and then made public as the documentation is completed.	
Data security			Data are archived on AOOS server in Anchorage & at mirror site in Portland OR.	Historical data are archived on the NCEAS GoA Member Node, replicated to DataONE, and a copy is made on the AOOS data servers. DataONE checks	Provide requirements, if any, for agency data archive.

				validity of content through rolling audit.	
Data analysis, synthesis & visualization	Produce data analyses, synthesis documents and data visualizations from project data.	Coordinates with PIs, AOOS, Axiom and NCEAS to produce synthesis and visualization products and reports.	Provides team with full access to all data for potential applications. Provide team access to all ancillary AOOS data & tools. Provide time series animations & syntheses on request from science team & outreach team.	Historical data are made publicly available via the GoA Member Node, and can be accessed from the web, analytical environments like R, and workflow systems like Kepler and VisTrails.	
Data discovery (search function)	Ensures that data are complete, QA/QCd & have complete metadata records.	Determines when data & metadata are ready to be published to public AOOS portal.	Incorporates data & metadata into AOOS GoA data search catalog w/additional GWA & historical EVOSTC tags. Setting up process for connecting to DataONE.	Historical data are listed on the AOOS GoA data portal, and are searchable on the DataONE portal as well as the KNB.	
Public data delivery	Reviews published data on data portal for accuracy.	Reviews published data on data portal for accuracy. Keeps track of program data delivery status.	When data meet all above requirements, publish data & metadata into the AOOS Gulf of Alaska portal for broader public access & use.	Historical data and metadata can be downloaded from AOOS GoA Data Portal, the GoA DataONE member node, and DataONE replica servers.	Public data access is required.
Long-term archive			AOOS data system is being used for long-term storage. With other funding, now developing methods for automated delivery to national archives (e.g., NODC) and to DataONE nodes.	Provide linkages to DataONE to replicate data across diverse institutions to protect against funding and policy failures. Historical data have 3 replicas nationally, working with Axiom on replication processes for current data streams.	Long-term archiving required by trustee agencies.

Work Completed

Axiom software engineers redesigned the display in the gulf of Alaska Data Portal of metadata created in the Ocean Workspace and imported into the portal. Upon initial release of the portal, project metadata created in the Workspace was visible as an HTML webpage and file-level metadata from the Workspace was available in the portal as raw, unstyled JSON documents. In the time since the launch of the portal, the metadata editors in the Ocean Workspace have been harmonized to provide the same interface and fields for project and file metadata, and have expanded to provide new metadata fields. This year, Axiom's interface designer created a new stylesheet to display the both the project and file level metadata from the Workspace in a much more human-readable form. The design of the metadata pages in the portal underwent several design iterations based on user feedback before settling into their current form.
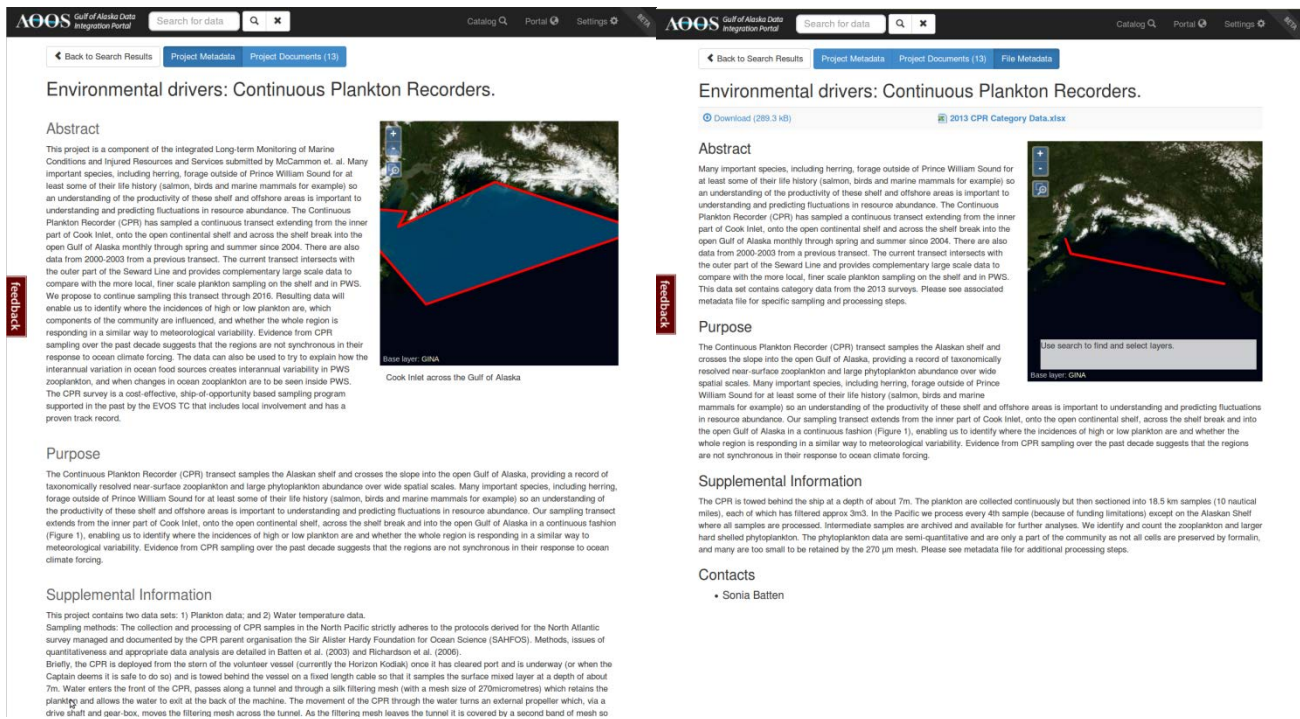
Figure 6. Screenshots of metadata imported from the Ocean Workspace into the public Gulf of Alaska Data Portal. On the left: project metadata for the Continuous Plankton Recorder (CPR) project; on the right: metadata for a single data file within the CPR project.

Work Underway

Axiom software architects and engineers have begun work to improve the Gulf of Alaska Data Portal's data catalog user interface and portal visualization capabilities. Improvements currently underway to the user interface include rebuilding the search tool to improve the precision and relevancy of search results, and indexing datasets' spatial and temporal metadata to allow advanced catalog searches. These upgrades to the data system are motivated by feedback received from GWA managers as well as external sources. Improvements to the catalog search tool will expand the range of material indexed for search to include file-level metadata and the text content of files imported into the Gulf of Alaska data portal from the Workspace. It will also suggest synonymous terms for users to search based on their search queries, e.g. - the new search tool would suggest 'sea surface temperature' when a user searched 'water temperature'. Indexing spatial and temporal metadata will allow users to limit the results of their searches to show only the data in the area selected during the time span indicated. Users will be able to set these limits by drawing a polygon on a map, inputting a spatial bounding box, and/or using a time slider to set a time range.

Data visualization is limited by the underlying data structures used by the data collectors. Axiom and AOOS are at work on a next-generation data portal based on a 4-dimensional data model enabled by the netCDF data format. This system is in the very early stages of development by Axiom software architects, but data analysts have already begun converting targeted datasets into the netCDF format. NetCDF is a well documented, open, and self-describing format that was designed with the needs of long term preservation in mind. Once these conversions are complete, the datasets can be more robustly visualized along standardized parameters while being ready for archiving in a long term preservation

environment. From the GWA project, Axiom analysts have worked with program management to convert three seasons of CTD data into netCDF files that will be used to create rich, 4D visualisations once the conversion is complete. An example of a preliminary visualization of netCDF data for another program, is below in Figure 7.
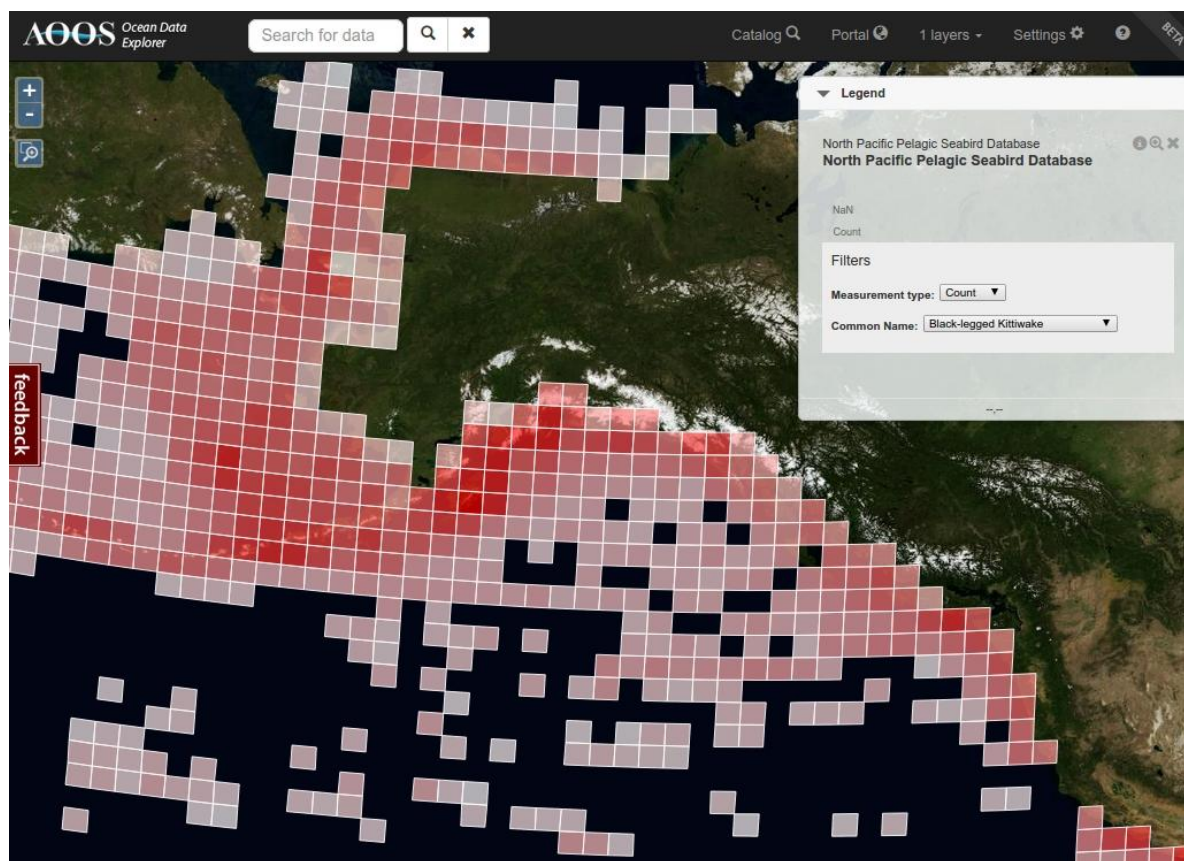

Figure 7. Screenshot of AOOS Gulf of Alaska Data Portal with the North Pacific Pelagic Seabird Database layer loaded. Color represents raw counts of Black-legged Kittiwake in the waters surrounding Alaska.

Responsibilities over the Data Lifecycle

The following table was created by the project team and GWA program management to make explicit and summarize the responsibilities of the various parties involved in planning or implementing data collection, management, and publication tasks during the data lifecycle.

8. Coordination/Collaboration: *See,* Reporting Policy at III (C) (8).

A. Collaboration and coordination both within your program and between the two programs:

The data management tools and services provided to the EVOSTC LTM and Herring programs are coordinated and collaborative by their very nature. As users of a central data management system, both programs provide useful feedback that informs the features Axiom develops and implements for the Ocean Workspace and the Gulf of Alaska Data Portal. Through ingesting, synthesizing, and prioritizing feedback and feature requests from both programs, the project team coordinates the needs of each program into a set of tools useful to both. Similarly, by making data from each program available in the Gulf of Alaska Data Portal, the project team helps the two programs collaborate to provide a comprehensive, holistic portrait of the conditions monitored in the Gulf of Alaska by both programs.

B. Coordination with other EVOSTC funded projects: None
C. Coordination with our trust agencies:

The project team provides data management visualization, and preservation services, including providing access to and facilitating the use of the Ocean Workspace, to a number of other programs that receive funding from or are administered or are overseen by representatives from the trustee agencies. Some of these programs and their associated trustee agencies are given on the table below.

Table 3. Collaborating projects and trust agencies

Arctic Marine Biological Observation Network (AMBON)	BOEM
Arctic Ecosystem Integrated Synthesis (Arctic EIS)	BOEM
Marine Arctic Ecosystem Study (MARES)	BOEM
IOOS Systems Integration	NOAA
Beluga Sightings Database Visualization	NMFS
Alaska Ocean Observing System (AOOS) Data Management	NOAA
Central and Northern California Ocean Observing System (CeNCOOS) Data Management	NOAA
Gulf of Alaska Integrated Ecological Research Program (GOAIERP)	NMFS
Russian-American Long-term Census of the Arctic (RUSALCA)	NOAA

Spatial Tools for Arctic Mapping and Planning (STAMP)	NOAA
Alaska Data Integration working group (ADIwg)	USGS

9. Information and Data Transfer: *See,* Reporting Policy at III (C) (9).

Publications produced during the reporting period: None completed.

Conference and workshop presentations and attendance during the reporting period:

The AOOS data team at Axiom Data Science attended the GWA PI meeting in November 2014, and the team meeting in January 2015 at the Alaska Marine Science Symposium (AMSS). The team participated in the webinar held in September 2014 to educate agency managers. Throughout the year, the project team keeps in contact with the GWA program management team with regular email and phone calls. Beyond the scope of just the GWA effort, Axiom Data Science held a user feedback meeting in the summer of 2014 to better understand how users browse and search in the portal. Tammy Neher, GWA Science Coordinator, called in to this meeting.

Demonstrations of the Ocean Workspace have been given to a wide variety of users including GWA PIs. Demos have also been given to PIs with the North Pacific Research Board's Gulf of Alaska Integrated Ecosystem Research Program, the BOEM-funded Arctic Ecosystem Integrated Survey, the Distributed Biological Observatory, and many other related research programs for which AOOS or Axiom also provides data management or visualization services. The AOOS Gulf of Alaska Data Portal, featuring GWA data sets, was demonstrated at AMSS during several workshops and was on display at the AOOS booth during the AMSS poster session.

10. Response to EVOSTC Review, Recommendations and Comments: *See,* Reporting Policy at III (C) (10).

Science Panel 2014 Comments

It was encouraging for the Science Panel to hear via a conference call with Kris Holderied, Tammy Neher, and Scott Pegau that the standardized forms for metadata submission had been recently modified, and that a more refined version is now available to investigators. The Panel is hopeful that this will facilitate all investigators' compliance on submission of both metadata and data in a timely manner (within one year of collection) as agreed upon when accepting funding from EVOSTC.

Data Management Team Response

In 2015, the project team will work with the GWA program management team to continue to track what data has been delivered, which PI is responsible for the dataset, and the status of data preparation, processing and metadata development.

11. Budget: *See,* Reporting Policy at III (C) (11).

Axiom's budget for data management support originally included 9.2K for equipment to purchase a storage array for Gulf Watch data. Axiom was able to leverage the AOOS data system at no cost for storage of Gulf Watch data. The surplus funds were utilized by personnel budget line instead resulting in improved services and support for investigators.