

\*Please refer to the Reporting Policy for all reporting due dates and requirements.

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

14120111-C

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

PWS Herring Program – Data Management Support

**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).

February 2015

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

<http://pwssc.org/research/fish/pacific-herring/>

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

From the beginning of the EVOS Herring project investigators have been focused on establishing protocols for data transfer, metadata requirements and initiating the data salvage effort. PIs have participated in several Herring Program PI meetings and coordination activities between the Herring and GWA programs. In addition, the AOOS Ocean Workspace has been in use by Herring Program PIs since 2013 for staging, organizing, and sharing their datasets. Software engineers at Axiom have worked throughout FY2014 to support the Workspace, resolving bugs and implementing new functionality in response to user feedback.

In coordination with Herring Program management, data management project PI determined the need for additional staff time beyond that currently funded for the effort the project that would be dedicated to Herring Program data management needs. Project team applied for and received funding for this new position in FY2015

**Table 1. Project milestones status**

Deliverable/Milestone	Status
GoA Data Portal Showcasing Herring data sets	Completed, 1 September 2013.  GoA Data Portal interface updates in 2014, larger updates begun.
Continue to support the transfer and documentation of Herring data sets. Auditing and restructuring/reorganizing	Ongoing  Applied for and received additional funding to support half of a data coordinator position to be dedicated to the EVOS Herring Program and

	GWA.
Continue to cultivate and support the functional capabilities of the AOOS Ocean Workspace to address Herring researcher needs	Ongoing
Improved Herring Portal project profile by exposing underlying file level metadata	Completed, 15 January 2014

## **Objective 1 & 4**

The Gulf of Alaska Data Portal was launched in September of 2013. At that time, all of the data that had been in the dedicated Herring Portal was migrated to the Gulf of Alaska portal where it benefits from the additional context of more than 400 other data layers describing other observed and modeled parameters in the Gulf of Alaska. The Gulf of Alaska Data Portal leverages the cyberinfrastructure behind the AOOS Ocean Data Explorer, which was developed using other funding and has 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 Herring Program's data has been ingested into the Ocean Workspace, quality controlled, documented, and approved as final, it can then be 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 available to be shared by the EVOS Herring 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.

### **Work Completed**

The display of metadata imported from the Workspace into the Gulf of Alaska data portal was redesigned. Upon initial release of the portal, project metadata created in the Workspace was visible as an HTML webpage, while file-level metadata was available in the portal as JSON documents for users to read and understand themselves. This year, Axiom's interface designer created a stylesheet to display the Workspace metadata in a much more human readable form. This file-level metadata page in the portal underwent several design iterations based on user feedback before settling into its current form.

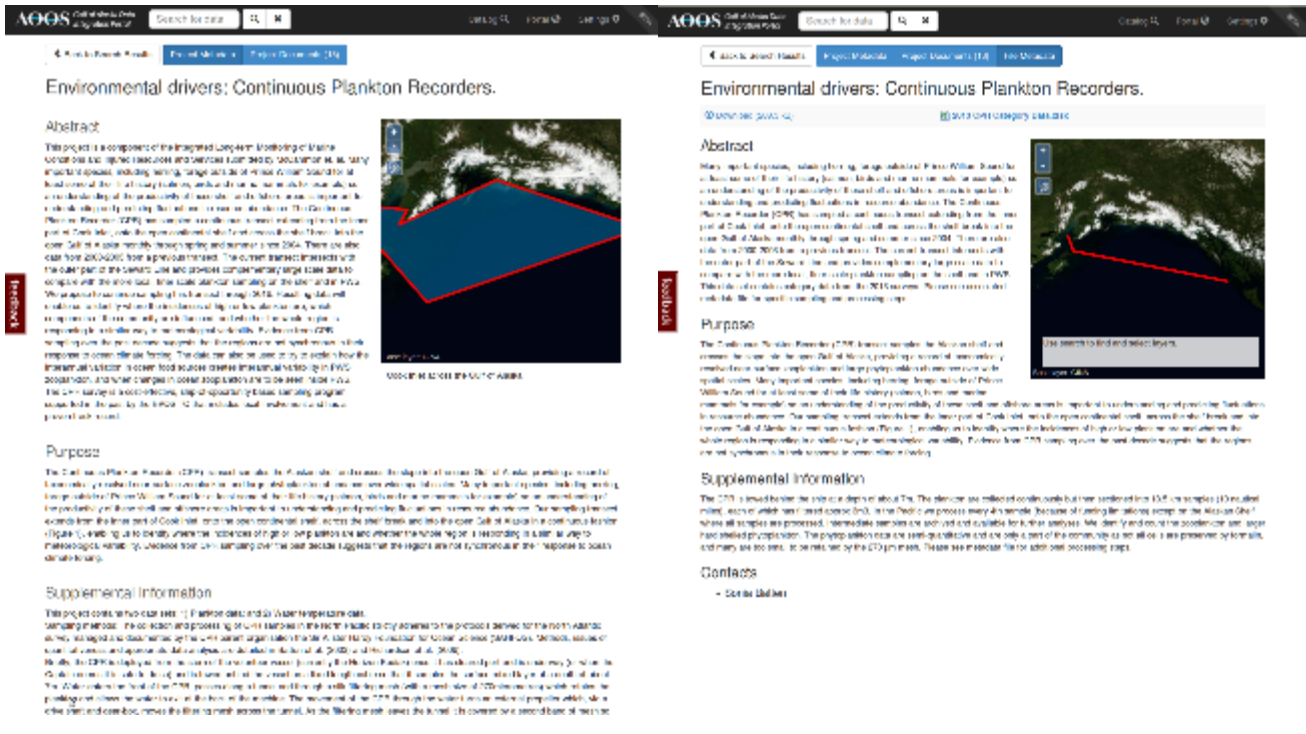


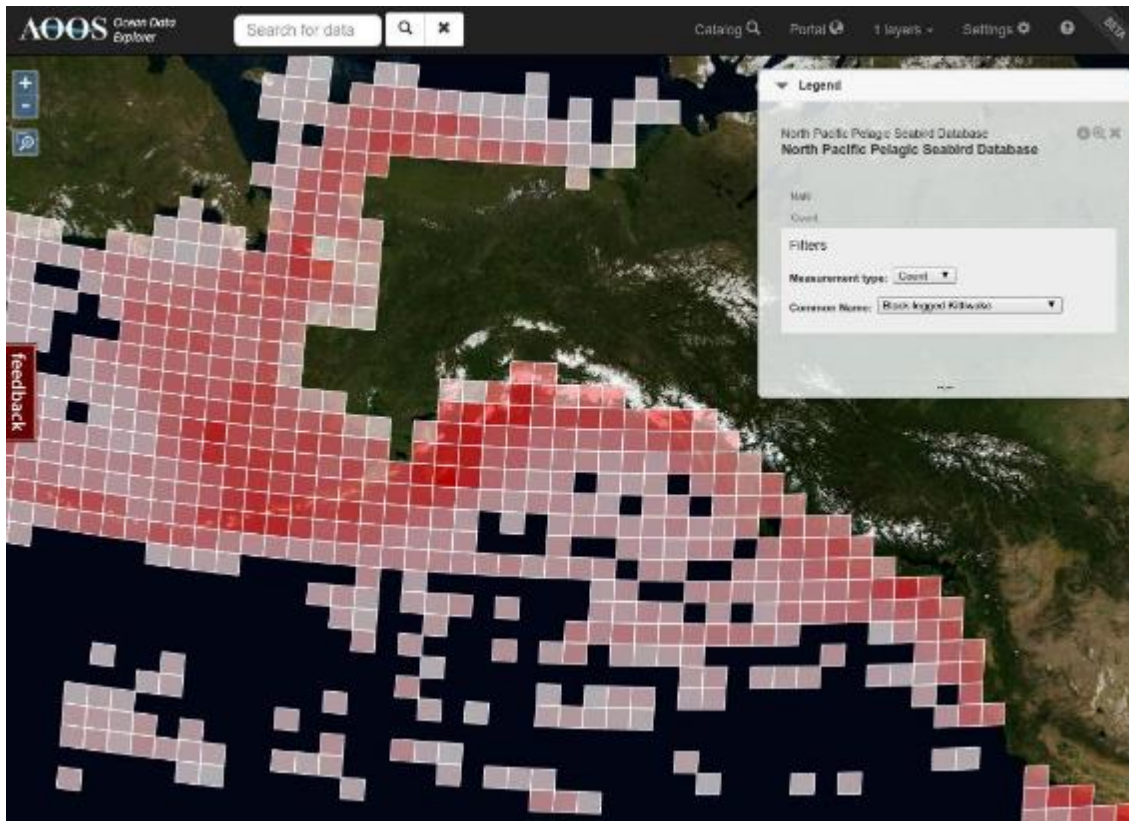
Figure 1. Screenshots of metadata imported from the Ocean Workspace into the Gulf of Alaska Data Portal. From left to right: project metadata for the Continuous Plankton Recorder (CPR) project, metadata for a single data file within the CPR project.

## Work Underway

Axiom software architects and engineers have begun work to improve the data catalog user interface and portal visualization capabilities. Underway improvements 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. In 2015, Axiom will begin working with Herring program management to identify Herring Program datasets that are ready for

conversion. Figure 2, an example of a preliminary visualization of netCDF data for another program, is below.



*Figure 2. 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.*

### **Objective 2 & 3**

The primary results produced by this project include the acquisition and documentation of EVOS Herring Program PI-produced data sets and the aggregation of ancillary environmental data sets for integration into the AOOS GOA Data Portal. To facilitate the acquisition and documentation of Herring Program data, the project team provides Herring Program PIs with access to the Ocean Workspace, a web-based collaborative data management environment. The project team has supported the use of the Ocean Workspace for data ingestion and documentation through webinars, email support, and by making functional improvements to the Ocean Workspace based on user feedback.

The project team and the Herring Program management decided that the Herring Program required more active data management facilitation than was possible using the Ocean Workspace and leveraging work done with other research groups to develop data lifecycle and management plans. In 2014, the project team applied for additional funding from the EVOSTC to partially a full time data coordinator position at the Axiom Data Science office to provide dedicated one-on-one work with EVOS Herring Program and GWA PIs. This new member of the project team will contact individual PIs about their datasets, work in person or over the internet to help PIs develop robust metadata, and assist Herring Program management with the essential inventorying and status-checking of data submitted to the Ocean Workspace. Through identifying the Herring Program's

need for this position, and in finding and being awarded the funds to pay it, the project team has increased the capacity to support Herring Program data management.

Investigators continue to improve the Ocean Workspace in response to user feedback. As a result of this, the Ocean Workspace has become more useful and easier to use. The increase in use by Herring Program PIs is represented in the figures below, followed by a description of the Ocean Workspace.

## Total Files, FY2014

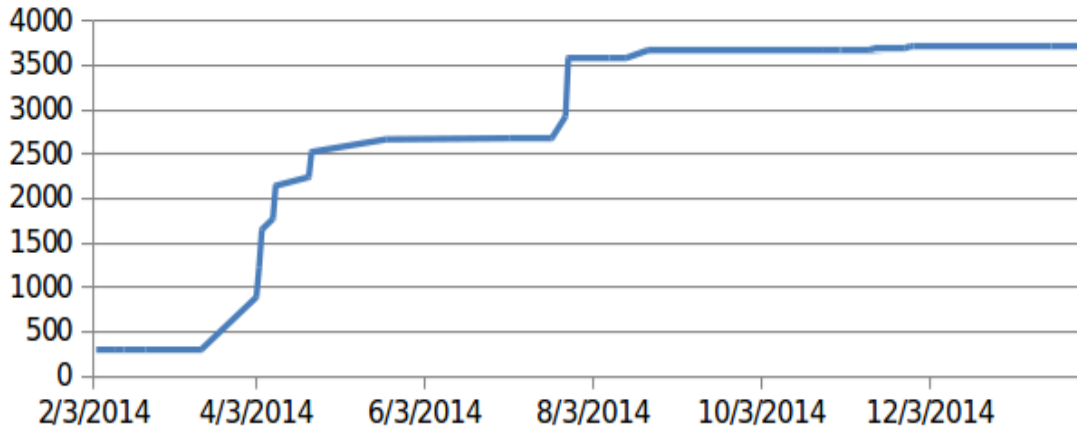


Figure 3. The number of files uploaded by Herring team members in FY2014.

## Total Gigabytes, FY14

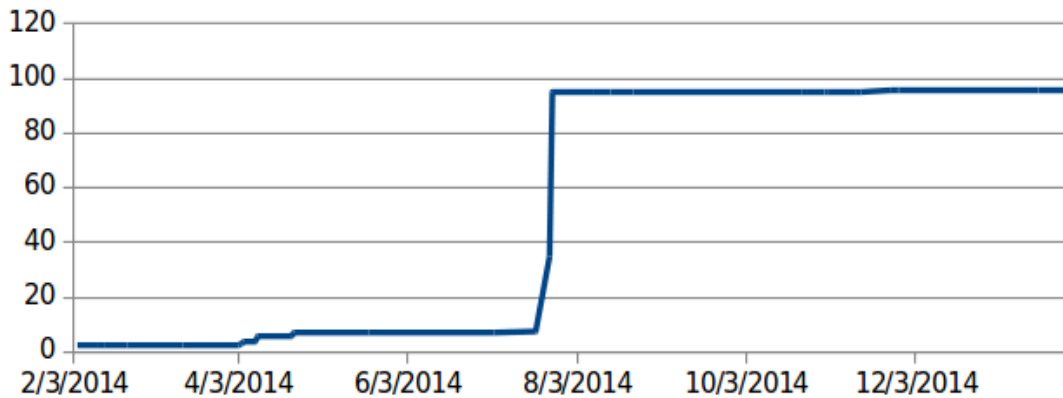


Figure 4. The amount of total storage used by Herring team members in FY2014.

## Files Uploaded per User

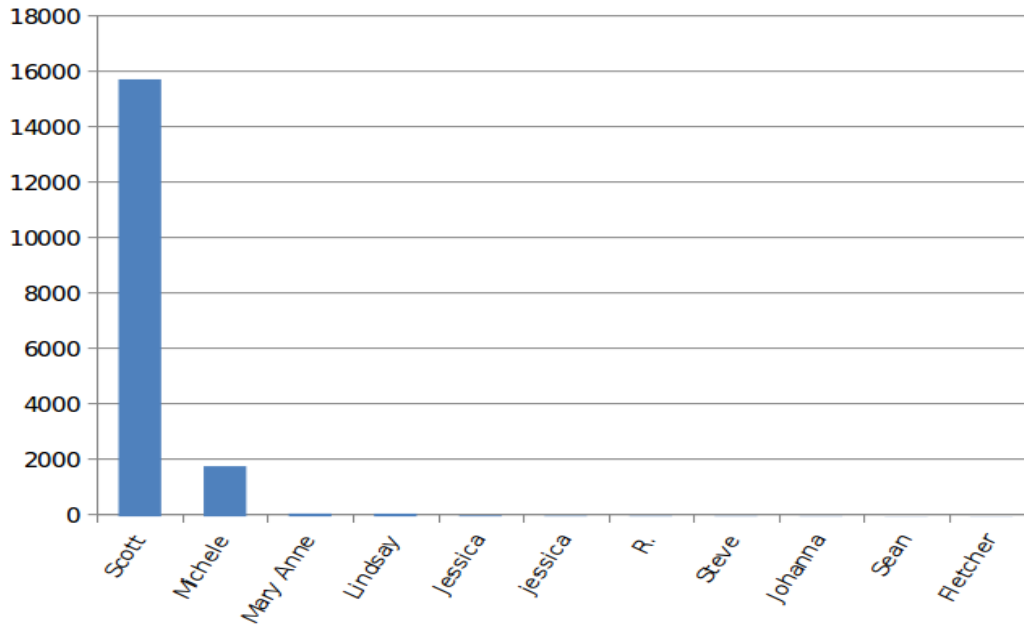
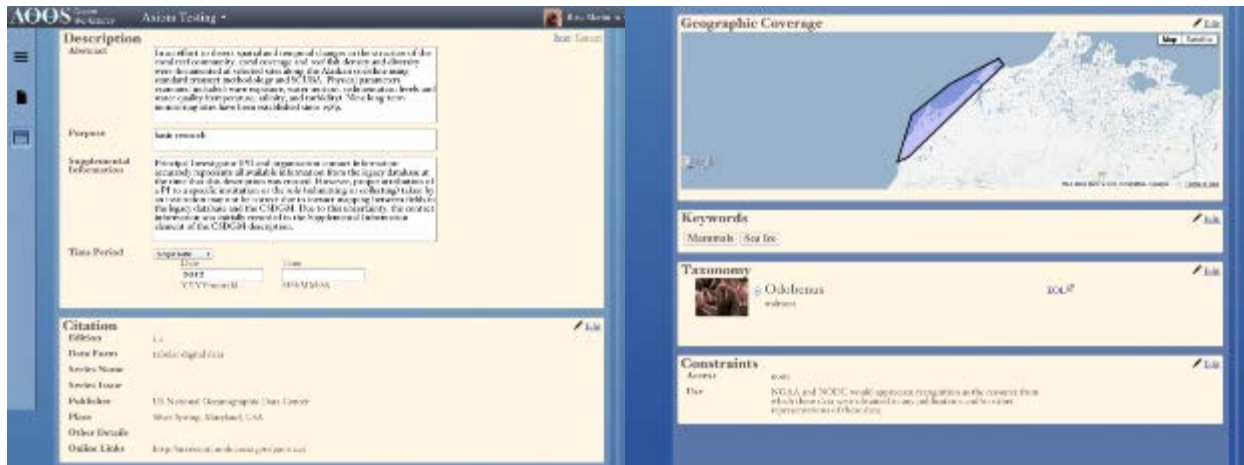


Figure 5. The distributions of file upload effort across Herring Program users.

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. 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. The Workspace metadata editors were expanded in 2014 to include a new tool to provide detailed definitions of attributes used in tabular data files. This tool automatically reads in tabular CSV data files, recognizes column headings in the file, and provides metadata fields for defining those headers in standards compliant elements.



The screenshot shows a table of data with columns A through G. An 'Add attribute' dialog box is overlaid on the right side of the table, showing fields for 'Name', 'Definition', 'Measurement type', and 'Number type'.

	A	B	C	D	E	F	G
1	24.021242	LongIsle	33	103	54.8	2013	07
2	32.021243	254.481093	12	15	31.7	2013	07
3	33.021244	154.481092	12	15	32.7	2013	07
4	33.021245	154.480013	12	15	-1.6	2013	07
5	34.021246	154.480012	12	15	-2.6	2013	07
6	34.021247	154.480011	12	15	33.7	2013	07
7	35.021248	254.480009	12	16	37.7	2013	07
8	32.021249	254.480008	12	16	37.7	2013	07
9	32.021250	254.480007	12	16	37.7	2013	07
10	32.021245	254.480006	12	16	37.7	2013	07
11	33.021246	154.480005	12	16	17.6	2013	07
12	33.021247	154.480004	12	16	23.6	2013	07
13	34.021248	154.480003	12	16	22.6	2013	07
14	34.021249	154.480002	12	16	34.7	2013	07
15	35.021250	254.480001	12	16	37.7	2013	07
16	35.021245	254.480000	12	16	43.7	2013	07
17	32.021245	254.480000	12	16	37.7	2013	07
18	33.021246	154.480000	12	16	31.7	2013	07
19	33.021247	154.480000	12	16	37.6	2013	07
20	34.021248	154.480000	12	17	33.6	2013	07
21	34.021249	154.480000	12	17	37.6	2013	07

Figure 6. 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 7. Screenshots of project and file management in the Workspace. From left to right: a list of projects to which the example user has access rights, 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.

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

- a) This project is responsible for working with all other projects in the Herring Research and Monitoring program to provide data management services. This service is the same as being provided to the Gulf Watch Alaska program and both programs benefit from shared services.
- b) There is no coordination with other EVOSTC funded projects.
- c) Most of the herring data provided in the Gulf of Alaska Portal is from Alaska Department of Fish and Game. We work with Steve Moffitt of the Cordova office to update data from their annual surveys.

**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 technology

Several demonstrations of the Workspace have been given to a wide variety of users including PIs from the EVOS Herring Program, Gulf Watch Alaska, the Gulf of Alaska Integrated Ecological Research Program (GOAIERP), the Distributed Biological Observatory (DBO), and the Arctic Ecosystem Integrated Survey (ArcticEIS).

The AOOS Gulf of Alaska Portal featuring Herring Program data sets was demonstrated at the Alaska Marine Science Symposium during several workshops and was on display at the AOOS booth during the AMSS poster sessions.



Project team members participated remotely in the Herring Program PI meetings in 2014.

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

It was encouraging for the Science Panel to hear via a conference call with Program Science Leads that the standardized forms for metadata submission had been recently modified, and a more refined version is now available to investigators. However, it was discouraging to learn that not all investigators were compliant on submission of both metadata and data in a timely manner (within one year of collection) as agreed upon when accepting funding from EVOSTC. In the future we see submission of required data and metadata as a condition of funding renewal.

**Project Team Response**

In 2015, the new data coordinator will work with the Herring Program management team help PIs achieve compliance with the EVOS requirements regarding submission of data and metadata. The data coordinator will also assess the need for a data inventory to track dataset status. This inventory could track delivery dates for Herring Program datasets, which datasets have 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).

Budget Category:	Proposed FY 12	Proposed FY 13	Proposed FY 14	Proposed FY 15	Proposed FY 16	TOTAL PROPOSED	ACTUAL CUMULATIVE
Personnel	\$94,400.0	\$93,700.0	\$16,700.0	\$17,300.0	\$17,900.0	\$240,000.0	\$ 210,247
Travel	\$0.0	\$0.0	\$0.0	\$0.0	\$0.0	\$0.0	
Contractual	\$0.0	\$0.0	\$0.0	\$0.0	\$0.0	\$0.0	
Commodities	\$0.0	\$0.0	\$0.0	\$0.0	\$0.0	\$0.0	
Equipment	\$3,900.0	\$4,800.0	\$0.0	\$0.0	\$0.0	\$8,700.0	
Indirect Costs (23%)	\$21,700	\$21,500	\$3,800	\$4,000	\$4,100	\$55,100.0	\$ 45,478
<b>SUBTOTAL</b>	<b>\$120,000.0</b>	<b>\$120,000.0</b>	<b>\$20,500.0</b>	<b>\$21,300.0</b>	<b>\$22,000.0</b>	<b>\$303,800.0</b>	<b>\$255,725.0</b>
General Administration (9% of	\$10,800.0	\$10,800.0	\$1,845.0	\$1,917.0	\$1,980.0	\$27,342.0	
<b>PROJECT TOTAL</b>	<b>\$130,800.0</b>	<b>\$130,800.0</b>	<b>\$22,345.0</b>	<b>\$23,217.0</b>	<b>\$23,980.0</b>	<b>\$331,142.0</b>	
Other Resources (Cost Share Funds)	\$0.0	\$0.0	\$0.0	\$0.0	\$0.0	\$0.0	

**COMMENTS:**  
 This summary page provides an five-year overview of proposed funding and actual cumulative spending. The column titled 'Actual Cumulative' should be updated each fiscal year to provide information on the total amount actually spent for all completed years of the project. On the Project Annual Report Form, if any line item exceeds a 10% deviation from the originally-proposed amount; provide detail regarding the reason for the deviation.

There were no deviations from the proposed budget.



*We appreciate your prompt submission  
 and thank you for your participation.*