

Axiom Data Science
SECOORA Base DMAC Services and Special Project Support

Statement of Work – Year 1

Performance Period: July 1, 2021 to June 30, 2022

PROJECT GOAL/PURPOSE:

The following document details specific milestones/deliverables, schedule, responsibilities, and staff roles for SECOORA core DMAC activities and two special projects. This effort will focus on maintaining and further cultivating the SECOORA cyber infrastructure, managing and ingesting new SECOORA data assets, extending the capabilities of existing data management tools, and developing new tools and functionality for the SECOORA web-based data portal.

Base DMAC Services

Task 1. Core Cyberinfrastructure Support (On-going)

1.1 - Support existing hardware (servers) & software that support data ingestion, search, aggregation, access/transport, & storage/archive systems in compliance with IOOS requirements

- Provide storage for data streams on Axiom hardware/infrastructure.
- Coordinate with DataONE and/or NOAA NODC/NDBC/NCEI as appropriate to provide the long-term archive.
- Backup data on physical storage media in preservation-oriented data formats, such as but not limited to, netCDF4, flat IEEE binary, ASCII, HDF, GRIB, and GIS formats.
- Purchase new and upgrade existing hardware (storage and computational) as needed to support ongoing dataset ingestion, visualization and data access services.
- Continue to increase the robustness of the data system to maintain high uptime performance and improve fault tolerance.

1.2 - Meeting new and existing requirements from IOOS and other affiliated organizations

- Maintain relationships and automated upload of observation data to NCEI for archival.
- Maintain data access services as required by IOOS.
 - Direct data access will be provided through a variety of services including OPeNDAP, WCS, WFS, WMS and in a variety of common formats including CSV, MATLAB, and JSON. These services will be provided by a combination of THREDDS, ERDDAP, and GeoServer. Axiom will continue to implement changes to SECOORA metadata profile (following IOOS Metadata Profile v1.2) to serve data to NDBC and the Environmental Sensor Map via ERDDAP - migrating from the existing solution that uses SOS and a custom NDBC XML format- relative to IOOS guidelines.
- Maintain relationships and transmissions of data to NDBC as needed, including migration to providing observations to NDBC via ERDDAP.

1.3 - Maintain and refresh existing data sources and metadata

- Maintain relationships with PIs and individual DMAC groups to pull in any new SECOORA funded data.
- Include data and metadata in the SECOORA portal as well as the IOOS catalogs, registries and inventories. Further, employ search utilities via a web browser to browse the data archives, and metadata and web service endpoints to the IOOS catalog on demand.

- Serve datasets through the portal with metadata conforming to the Federal Geographic Data Committee (FGDC) and/or ISO standards. Apply additional IOOS standards as they become fully developed, such as Darwin Core and Biological Data Services.
- Maintain a web catalog service to provide access to the metadata, as well as a SECOORA-specific instance(s) of the Research Workspace for metadata documentation and dataset creation, storage, and internal file sharing.

1.4 - User engagement -responding to feedback and data requests

- Respond to 100% of feedback button requests from the website and portal.
- Respond to 100% of data requests.
- Discuss feedback trends through all portals (SECOORA, AOOS, CeNCOOS, etc.) on biweekly DMAC calls between Axiom and SECOORA.
- Continue developing and maintaining the portal documentation at <http://portal.secoora.org/help> with common workflows and tasks.
- Expand documentation to include submission guidelines for various data types to the SECOORA data system.

1.5 - Website maintenance, support, and backup

- Working with SECOORA to maintain and help determine future website strategy at <https://secoora.org>.
- Implement revised website content based on SECOORA content mock-ups.
- Improve portal metrics available to SECOORA based on data portal metrics of interest.

Task 2. Cyberinfrastructure Enhancements

2.1 - Data ingestion (numerical models, GIS, in-situ, etc.)

- Prioritize new datasets already documented in Trello for ingestion, including new stations, gliders, HFR, water level, and other prioritized new data.
 - Each new dataset gets the Template assigned to it defining metadata and level of effort.
 - All existing New Dataset cards will be edited to match the Template.
 - Along with budget, choose datasets for ingestion and periodically revisit the remaining budget and cards to act on.
- Prioritize previously funded data not available in portal or never submitted to SECOORA
 - All Glider missions ([spreadsheet for reference](#)) not currently ingested - realtime and delayed mode- both in the portal and organized in the Research Workspace.
 - HFR
 - Work with SECOORA funded HFRadar providers to send their radial data to SECOORA DMAC (in addition to the DAC) so combined processing can be done outside of the DAC algorithms.
 - Process backlog of Radial data from NCEI when it is made available
 - Recovered observational data (via Research Workspace), as available from researchers in the SECOORA region.
- Collect, manage, quality control, and make available high impact and/or highly requested data in the SECOORA region.
- Work with data providers to curate metadata through the Research Workspace, as needed, for post-processed datasets lacking standards-compliant documentation.

2.2 - New capabilities for presenting, visualizing and processing data holdings

- New capabilities for managing and serving live media streams via RTSP protocol.

- New system development for gridded data display and virtual sensors. The current ncWMS1 software we are using is old, outdated and overloaded. An updated approach will allow us to better visualize unstructured grid data (USF FVCOM) and allow animations to be created within the SECOORA portal.
- Enable arbitrary, cross-data type groupings (i.e. ‘data packages’) in the SECOORA data portal to associate related data products originating from different components of the data system (sensor system, models, platforms, processed data and metadata from the Research Workspace, etc.).

2.3 -Implement QARTOD and QA/QC of data feeds through SECOORA data portal

- Maintain the automated pipeline for applying basic QARTOD checks and visualizations to real-time data feeds streaming into the SECOORA data portal, following v2.0 version [QARTOD GitHub library](#). Continue to make test results available for download within the data and metadata, and visually within the portal for roll up and individual flagged results. and visualizing test results in the portal.
- A basic test configuration management application for users to store and manage test configurations across parameters and geographies will be developed. Enhancements to this version will be isolated for future development of this tool, based on feedback from SECOORA, researchers, and the DMAC community.
- Other functionality may be included, including: adding libraries for calculating additional “advanced” QARTOD tests; being able to define test parameters per source, station, or region; or early prototype tool for providing alerts for failed QARTOD test results to Axiom, data providers, and/or external groups.

2.4 - Release updates and enhancements to the portal

- The SECOORA data portal will continue to be cultivated to integrate new technologies and functionalities as listed in the [FY21 Portal Release Schedule](#) for versions 2.14-2.16.
- Further, new features and improvements will be implemented as requested by SECOORA staff and as informed by end user feedback.
- Implement administrative log-on for SECOORA staff to the [sensor dashboard](#) to manage and update station alerts and narrative.

Task 3. DMAC Support to Existing Programs

3.1 - Progress reporting, workplan development, communication, SECOORA meeting attendance

- Attend SECOORA annual meeting.
- Produce written progress reports to SECOORA every 6 months.
 - Includes usage metrics on website and portal visits as well as service access where possible.
- Maintain the Jira tasking board to track all work.
- Participate as a member of the SECOORA DMCC Team.
- Participate in biweekly SECOORA DMAC calls between Axiom and SECOORA staff
 - Discuss any feedback received from portal system
 - Discuss portal analytics from previous week

3.2 - Participation in IOOS related workgroups and meetings

- Attend IOOS DMAC annual meeting (Spring 2022).
- Attend IOOS Code Sprint, as applicable.
- Attend monthly IOOS DMAC Webinars.

- Participate in ‘tri-RA’ DMAC meeting with AOOS, CeNCOOS, and SECOORA to discuss cross-regional data management objectives and products.
- Prepare for and engage in the yearly DMAC review from IOOS, as needed.

3.3 - Provide technical assistance to SECOORA PIs for Research Workspace

- Provide guidance on the use of Research Workspace.
- Establish new projects and organizational structures in the RW for research groups, as needed.
- Provide technical assistance for metadata authoring using RW metadata editor.
- Provide metadata quality reviews for datasets to be ingested into the data portal, as needed.

3.4 - Support the SECOORA Glider group

- Maintain the [spreadsheet of known SECOORA glider deployments](#).
- Maintain, document, and improve upon the existing SECOORA Glider System with feedback from the glider group.
- Process any SECOORA funded real-time data to the IOOS National Glider DAC in real-time.
- Work closely with the Glider group to obtain and process any delayed mode data into the Research Workspace and the IOOS National Glider DAC.

3.5 - Participate and assist in any DMAC related proposals

- Participate in funding proposals to support data management activities and/or the development of data products that uphold the SECOORA mission.

3.6 - Provide user support, technical documentation, and trainings/workshop for SECOORA community

- Provide one in-person or virtual training or workshop to the end user community
 - Consider holding a workshop in concert with a regional workshop or meeting, or in association with the SECOORA annual meeting.
- Host up to 3 webinars targeting Research Scientists to inform them of SECOORA data portal and/or products.
 - Webinar topics may cover, but not be limited to: hurricane tracking, using the portal in the classroom (high ed), physical/biological data integration, HFR data package, etc.

Task 4. Develop New Products and Capabilities

4.1 - Develop new data products- Based on input from SECOORA and its end-user community, this task will support development of new products as determined by SECOORA, with technical feedback provided by Axiom. These could include mobile phone apps; specialized data portals, products or tools for SECOORA subregions or for unique stakeholder groups (i.e. recreational fishers, commercial fisheries, and emergency planners), or for topical issues such as ocean acidification, web cameras, and community-based monitoring. Axiom will work closely with SECOORA, other data management awardees and appropriate advisory committees to implement identified user products, tools and their web interfaces, which includes developing product requirements; beta testing and refining products in order to increase their utility; providing reports as requested; developing detailed work plans with measurable timelines, deliverables, and performance metrics; and assisting with the development of funding proposals. At the onset of FY20 (July 2020), Axiom will work with SECOORA staff to identify new data products of interest to users in the region. This effort will involve isolating the desired functionality and technical requirements to ensure data products align with user needs and are implemented within scope of the allocated resources. Examples of prior data products developed under this effort include: [Eyes on the Hurricane](#), [HF Radar Surface Currents data product](#), [Text a Buoy](#), etc.

Special Projects:

Task 5. Enhancing the Capabilities of the SEAMAP-SA Biological Surveys Integrated into the SECOORA

Data Portal: The Southeast Area Monitoring and Assessment Program, South Atlantic (SEAMAP-SA, lead PI Tracey Smart, SCDNR) fully or partially supports a variety of long-term living marine resource surveys in waters of the Atlantic coast off the Southeast United States (generally referred to as the South Atlantic). These biological surveys provide much needed data for a variety of state and federally managed species in this region, including finfish, sharks, turtles, and invertebrates. Core data elements provided by all surveys include abundance, distribution, and community composition of teleosts, elasmobranchs, and invertebrates. SEAMAP-SA supported surveys include the Pamlico Sound Survey managed by North Carolina Department of Marine Fisheries (NCDMF), the Southeast Reef Fish Survey managed by South Carolina Department of Natural Resources (SCDNR) and Southeast Fisheries Independent Survey at the Southeast Fisheries Science Center, the Coastal Trawl Survey managed by SCDNR, and the Coastal Longline Survey managed by NCDMF, SCDNR, and the Georgia Department of Natural Resources (GADNR). In addition, SEAMAP-SA supports data visualization development for the South Atlantic Fisheries Management Council through the Florida Fish and Wildlife Research Institute (FWRI).

Beginning in 2018, SEAMAP-SA and SECOORA supported the migration of the combined SEAMAP-SA surveys' online database into the SECOORA data portal managed by Axiom Data Science. This began as a Pilot system with a single year and limited number of species from the SEAMAP-SA surveys (Reef Fish Survey) to provide proof of concept for incorporating biological surveys into the existing SECOORA framework and converting the survey data to Darwin Core standards. Development expanded to the full Reef Fish Survey time series and the other three SEAMAP-SA surveys, allowing load testing for the large number of records and needed adjustments for multiple surveys. The initial phases of this migration included reproducing most capabilities of the SEAMAP-SA system, including administrative tools for SEAMAP-SA data management staff to manage their own data within the system and end-user accounts for tracking purposes, and additional development of summarization and visualization tools not available in the SEAMAP-SA system. The migration of all SEAMAP-SA surveys is anticipated to be complete before the end of 2020. Although the current capabilities of the SECOORA data portal are well beyond those in the SEAMAP-SA system, not all survey elements were incorporated into the data portal and users are limited in their abilities to access or synthesize community data and to create advanced products that incorporate the extensive oceanographic data available in the SECOORA system.

In light of these limitations, the goals of this project are to:

1. Support the final integration of long-term living marine resources survey data types from SEAMAP-SA into the SECOORA data portal,
2. Support the development of new data access, data exploration, and data analysis or visualization tools within the SECOORA data portal,
3. Support expansion of long-term living marine resource survey data availability in the SECOORA data portal, and
4. Assess methods for inclusion of SECOORA oceanographic data into stock assessment data inputs for managed fisheries species.

Project Activities:

1. Expand the data tables and code tables currently in the system to add a Tagging table (which houses tagging and recapture information for a variety of fish and 2 elasmobranchs), additional

- life history data (e.g., mercury, genetics, and diet), and a Turtle table (which houses life history and tagging information for sea turtles encountered by the Coastal Trawl Survey). Year 1
2. Based on user feedback, develop tools or create training for users for pre-existing tools as appropriate. Year 1
 3. Create a multispecies data access product to produce the zero records for more than one species at a time. Year 1-2
 4. Develop a reporting system on top of EAV database (developed in previous years of the project) to support user metrics, user report queries/requests and exporting or querying of data for inclusion in the visualization system. Year 1-2
 5. Produce reports for inclusion into the SECOORA data system and change current visualization to use the new data source of the data. Year 2

Task 6. Augmenting Ocean Observing through Artificial Intelligence: Annotation, Data Standards, and

Applications: This project, led by Luke McEachron with FWRI, will build and demonstrate through two use cases, a SECOORA AI Annotation Data Portal (AI Portal) to support regional AI marine applications. Annotating refers to the process of assigning labels (e.g., species names) to data (e.g., images) to train machines to recognize patterns. We propose the following measurable objectives in support of this goal:

Objective 1: Establish Data Standards for AI Applications. Building upon the substantial progress made by Co-PI institutions Axiom Data Science and the University of South Florida (USF)-IOOS supported Marine Biodiversity Observation Network (MBON) to establish standards; defining metadata requirements; resolving file storage and access issues; and articulating the costs and benefits of different AI algorithm workflows that SECOORA, and other Regional Associations (RAs) can easily follow.

Objective 2: Construct an AI Portal to Serve as a New Digital Asset that Augments Multidisciplinary Ocean Observing Assets. Axiom Data Science, in consultation with team members, domain experts, and end-users, will construct an interactive data portal based on existing SECOORA data management and communications (DMAC) infrastructure to support regional AI applications and establish SECOORA as a leader in promoting marine AI among RAs.

Objective 3: Demonstrate Use Cases and Provide Training. In coordination with project PIs (McEachron, Muller-Karger) we will demonstrate the utility of the data portal and instruct users on the portal's use by initially focusing on two AI workflows as example use cases. These workflows will demonstrate two of the most common problems AI algorithms can resolve: image classification and anomaly alerts from video.

Project Activities:

1. Establish data and metadata standards for incorporation of AI data into the SECOORA data portal. Year 1-2
2. Resolve file storage and access issues; and articulate the costs and benefits of different AI algorithm workflows that SECOORA, and other Regional Associations (RAs) can easily follow. Year 1-2
3. Construct an interactive data portal based on existing SECOORA data management and communications (DMAC) infrastructure to support regional AI applications. Year 2-4
4. Develop training for end users on portal use using the two developed AI workflows. Year 5