1. Open Data Sharing

The SECOORA Data System provides data resources in a one stop data portal, free to the public, with data assets originating from federal and state agencies, local municipalities, academic institutions, research organizations, private companies, non-profit organizations, and community observers. Real-time and near real-time data are served as soon as practical as the data become available.

SECOORA works with data providers and its data partner, Axiom Data Science, to establish and maintain freely available data streams that allow for timely ingestion, processing, and serving of data. When possible, SECOORA aims to provide real-time or near real-time (as defined in RICE IOOS Guidelines) quality-assured and quality-controlled data. SECOORA adheres to data and metadata standards established by IOOS and leverages the experience and expertise of the community of data providers to improve data quality.

Status: All data currently served by the SECOORA data 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 the SECOORA data portal.

Challenges: Metadata for some data packages are sparse, often due to a lack of quality metadata from upstream data providers or the historic nature of the dataset; efforts are underway to enhance metadata records and develop tools to ease and democratize metadata curation using the SECOORA Research Workspace.

2. Data management planning and coordination

Data management is an increasingly important aspect of IOOS activities. Data management plans and the coordination of activities between Regions and the IOOS Program Office ensure that data are maintained in easily accessible formats that are archived for long-term storage.

The SECOORA Data Management Plan provides the approach to the necessary implementation, describing how data are ingested, managed and distributed from the source to public dissemination.

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. SECOORA and its data management partner, Axiom Data Science, serve data to users in common machine-readable data formats and provide the feeds to the GTS with their service-oriented architecture. SECOORA works with Axiom and data providers to ensure that IOOS standard ontologies and vocabularies are being used. SECOORA strives to maintain standards-compliant metadata and provide information to the IOOS catalog. All data received and made available through the SECOORA data portal is stored in standardized community driven formats on an infrastructure developed by Axiom.

Axiom maintains onsite storage at their facilities, as well as at a redundant offsite storage location.

Axiom also makes available open-source resources of software developed through the Axiom Data Science (@axiom-data-science) and SECOORA (@SECOORA) public GitHub organizations.

SECOORA officially became RICE certified by NOAA in 2017. As part of this process, the SECOORA Data Management plan was completed (April 2017) and the plan will be updated routinely (minimum 5 years) as needed to meet new requirements from the IOOS DMAC.

3. Provision of data to the Global Telecommunication System (GTS)

SECOORA has maintained their commitment to provide data to the GTS though NDBC. In some instances, the data is flowing from the SECOORA funded data provider (i.e. UNCW, USF COMPS buoys). The Big Carlos Pass station will be submitted to NDBC directly by SECOORA. Finally, CDIP moorings within the region are reported to the GTS by the CDIP program.

Update 2019: As of December 2019, the National IOOS Office has been working in collaboration with NDBC to set up data ingestion from IOOS RAs through standardized ERDDAP instances. This work includes updating the IOOS metadata 1.2 profile, updating the IOOS Compliance Checker to handle 1.2 profile, and developing pathways to serve data from the IOOS Catalog to IOOS RA ERDDAP instances. NDBC is actively pulling test data through this pipeline process.

4. Data access services

All data and products are registered in the IOOS Catalog. SECOORA offers six access points:
1. *Thematic Realtime Environmental Distributed Data Services (THREDDS)* - SECOORA provides THREDDS access points for raster (gridded) data stored in NetCDF format. THREDDS 4.6.10 - http://thredds.secoora.org and SECOORA ISO WAF - https://thredds.secoora.org/iso.
2. *Open-source Project for a Network Data Access Protocol (OPeNDAP)* - SECOORA provides OPeNDAP access points for raster (gridded) and time-series data.
3. Web Map Service (WMS) - SECOORA provides WMS access points for point, vector, and polygon information, as well as raster (gridded) data.
4. Web Feature Service (WFS) - SECOORA provides WFS access points for point, vector, and polygon information, as well as time-series and raster (gridded) data.
5. Environmental Research Division's Data Access Program (ERDDAP) - SECOORA primarily uses this service to facilitate device-level downloads (e.g., tabular data). ERDDAP 1.84 - http://erddap.secoora.org
6. File Downloads - SECOORA often provides data as downloadable files. These files are mostly served in the standard shared data file formats above, or in the case of project-specific data, in their native file formats.

Challenges: Large datasets and heavy usage can strain data access servers and negatively impact user experiences; Axiom and SECOORA are continually tuning and enhancing data

service software and developing deployment techniques to maximize performance and stability of these services. As new data types and variables come on-line routine coordination between Axiom, SECOORA, and IOOS will be required to make them available. Currently the IOOS Glider DAC only makes available a subset of data types transmitted by the profiling gliders.

5. Catalog registration

SECOORA maintains a WAF (https://thredds.secoora.org/iso), which is harvested by the IOOS Catalog. All applicable data sets and data products are registered in the IOOS catalog.

6. Common data formats

SECOORA offers data in IOOS compliant formats through the use of ncSOS, THREDDS and ERDDAP.

SECOORA provides nearly all data in four open, standardized forms:
1. *Network Common Data Form (NetCDF)* - a self-describing, machine-independent data format that SECOORA 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. SECOORA 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. SECOORA 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 SECOORA catalog. PNGs are limited in use as they are pre-projected, pre-scaled, and pre-sized images of data layers. However, SECOORA 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.

7. Metadata standards

All IOOS data providers are expected to ensure relevant metadata is produced, accessible and compliant with IOOS conventions, and to participate as appropriate in the development of such conventions. Descriptive information about datasets, sensors, platforms, models, analysis methods, quality-control procedures is essential for the long-term usability and reuse of information.

SECOORA requires standards-compliant metadata for project-level data (SECOORA or IOOS-funded projects). Though SECOORA does not require specific metadata standards for ingesting other types of data, most modern data submittals are accompanied by standard ISO/FGDC metadata records.

RW is the SECOORA web-based data management application. RW is being phased in within SECOORA and it will be used to assemble, store, and share data by researchers or SECOORA

partners. RW provides users with a web-based interface that allows researchers to create *projects* to represent particular scientific studies or focuses of research within a larger effort. Standard, discovery-level ISO 19115-2 and 19115-10 compliant metadata can be generated for both projects and individual datasets.

Many historical datasets come with informal metadata documentation that is variable in terms of completion and detail required by modern standards. Some data sets are only accompanied with narrative information. In these cases, SECOORA plans to work with the data provider to create more up-to-date metadata records and share the data within RW so that it can be ingested into the SECOORA data portal.

8. Storage and archiving

SECOORA ingested data is stored in a secure, professionally managed external facility and currently has total storage space for over 1.8 petabytes of data. Those resources are geo-replicated between Portland, Oregon and Providence, Rhode Island. All aggregated data is stored indefinitely beyond the life of each individual project. Real-time sensor feeds will become historical sensor feeds one-month after collection. The only assets that are not kept indefinitely in storage are webcam images.

As a federally funded program, SECOORA is required to submit data it generates to a national archive center. SECOORA is working with the National Centers for Environmental Information (NCEI) to assist with the archival of appropriate data types accepted by NCEI. SECOORA maintains an NCEI archive WAF at https://ncei.axiomdatascience.com/secoora/ which is regularly harvested by NCEI. The bulk of the data assets managed by SECOORA are non-real-time, nonfederal assets, sometimes from small data originators, and often from distinct research projects or large, integrated ecological research programs. These data may not fall under the purview of the NCEI. Accordingly, SECOORA plans to archive these data in the DataONE network through RW.

9. Ontologies, vocabularies, common identifiers

SECOORA makes use of IOOS ontologies, vocabularies and common identifiers as needed.

10. Consideration for Long-term Operations

The SECOORA Data System hosts several integrated data management tools to ease data access, storage, and sharing by its users including the RW and its metadata editor, and the SECOORA Portal and catalog system. The SECOORA RW, the web-based data management application, will be used to assemble, store, and share data by researchers or SECOORA partners.

RW includes an integrated metadata editor to support the documentation of data and facilitate its accuracy and reuse. Content collected in the RW 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. Standard,

discovery-level ISO 19115-2 and 19115-10 compliant metadata can be generated for both projects and individual datasets.

Axiom Data Science has made significant progress redesigning the back end SECOORA Data System to implement and support QARTOD checks for real-time data. Axiom has implemented the display of QARTOD flags to be available through the SECOORA data portal. The basic/required quality tests for IOOS RA assets are now being run on observation data with documentation of the test code and thresholds available through open-source QARTOD libraries accessible through the portal. Flags from quality tests run by data provider are viewable and available for download in the data portal and/or ERDDAP data servers for the SECOORA region. This work will be ongoing as QARTOD manuals are further developed for additional parameters to provide the correct level of guidance for implementing appropriate QC at the regional level. SECOORA will continue working with regional data providers and experts to develop user established ranges tests that are specific to each sensor and its location.