1. Introduction

The U.S. Integrated Ocean Observing System (IOOS) is a nationwide effort to provide access to a wide variety of coastal oceanographic and environmental data. The Southeast Coast Ocean Observing Regional System (SECOORA) was established in 2007 by a consortium of partner agencies and research institutions and serves as the regional association for integrating coastal and ocean observing activities in the southeastern Atlantic and eastern Gulf of Mexico. SECOORA is one of 11 regional associations (RAs) acting as Regional Coastal Observing System (RCOS) under the authority of the Integrated Coastal and Ocean Observation System Act of 2009 (ICOOS Act).

As a member of IOOS, SECOORA has a mandate to collect, organize, and provide access to regional oceanographic data. These data need to be quality controlled (QC), easily understandable, electronically accessible, and well organized to allow researchers, policy makers, industry, and the public to make well-informed decisions. To satisfy this mandate, SECOORA supports a web-based data portal for the entire region providing ocean, coastal, and relevant interior environmental data and information products.

Typical data management activities for oceanographic information occur in isolated, physically distributed agencies, leading to low cross-agency utilization of data. Technical barriers, complex data formats, a lack of standardization, and missing metadata have limited the access to data in the past and made using available scientific information a cumbersome and daunting task. Therefore, existing data are often underutilized and typically have not undergone quality assurance.

SECOORA is implementing recommended and standard practices as defined by the IOOS Data Management and Cyberinfrastructure (DMAC) committee. This ensures data collected by SECOORA and member entities is distributed on the SECOORA web portal and are managed according to best practices identified by NOAA/US IOOS. This also ensures that appropriate metadata and QA/QC practices are followed, and that the data are of a known quality to the end user. 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 DMAC White Paper (v1.0), and data management and communications DMAC requirements for IOOS Regional Associations and other IOOS grant recipients who are providing data to IOOS.

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

- Section 2 describes the SECOORA DMAC System and its functions, goals, and objectives; the data management structure; and, information on the SECOORA data management team.
- Section 3 briefly describes the SECOORA data resources, defines data categories and asset types, and describes how the data categories are handled in the plan.
- Section 4 presents the SECOORA DMAC System statement of work and includes descriptions of the system computing infrastructure including details about the processes related to data ingestion, standards for format and content, metadata and data discovery, QC procedures (including procedures for data that cannot undergo QC tests) and flagging protocols. Additionally, this section

covers policies for stewardship, public access and dissemination, data archival and preservation, and data system performance and security measures.
- Appendix A provides information on efforts to comply with the 6 IOOS core capability requirements.

This document, unless superseded, pertains to a SECOORA Cooperative Agreement period of five years from July 1, 2021 through June 30, 2026. The document and associated appendices will be reviewed annually and updated, if needed.

2. SECOORA DMAC System

The mission of SECOORA DMAC System is to acquire, archive, and share coastal and marine data and information products to meet the needs of SECOORA stakeholders and the national US IOOS program. SECOORA uses a data management system that allows a complex array of oceanographic and environmental data types to be well organized, discoverable, accessible, and understandable. The SECOORA DMAC System employs a distributed data management approach, which allows data to seamlessly interchange between participating agencies. The system is composed of an internal master node coupled with external data provider nodes. External data providers include stakeholders, partners, and SECOORA funded projects which produce, manage, and share data. This distributed configuration increases capacity and technical knowledge within individual groups, 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 SECOORA users with minimal effort.

Integrating available interoperable data feeds into data access applications and data management systems adds a variety of 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 SECOORA region are openly available for use through interoperability protocols. Real-time and historical sensor data feeds for the SECOORA region are available for hundreds of sensors via the National Data Buoy Center (NDBC), the Center for Operational Oceanographic Products and Services (CO-OPS), National Estuarine Research Reserve System (NERRS) 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. SECOORA integrates all these data feeds and makes them available through the SECOORA data portal.

DMAC System Management Goals and Objectives

The SECOORA Deputy Director and Axiom Data Science, a Tetra Tech Company, (referred to hereafter as Axiom), comprise the DMAC System Management team and are tasked with fulfilling the primary goals and objectives within the SECOORA Data Management Work Plan.

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

1. Provide Technical Support for SECOORA cyberinfrastructure.
2. Data Portal maintenance, development, and upgrade.
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 SECOORA

users.
5. Implement appropriate QARTOD QC tests for SECOORA real-time sensor data feeds.
6. Archive SECOORA funded partner data at NOAA NCEI.
7. Provide system performance and security measures.

Goal 2: Provide DMAC support to the SECOORA program

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

Goal 3: Web Portal Hosting and Support

1. Host and maintain the SECOORA content website at https://secoora.org
2. Host and maintain the SECOORA data catalog and portal at https://portal.secoora.org
3. 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.
4. Work with SECOORA staff to update the website periodically, to improve clarity, ease of use, and the overall "look and feel."
5. Work with SECOORA staff, PIs, and member organizations to update the data portal periodically, to improve access to data, ingest new data, develop new tools, improve clarity and ease of use, and the overall "look and feel."

SECOORA Data Sharing Policy

As a rule, SECOORA does not embargo data. Providers of real-time data must submit their data to SECOORA as it is collected. For non real-time data, in order to be compliant with the IOOS DMAC requirements and FAIR (find, access, interoperate, and reuse) data principles, all data must be made discoverable and accessible to the general public in a timely fashion (no more than two years after data collection). Data should be made available in at least one machine-readable format, preferably a widely used or open-standard format, and should be accompanied by machine-readable documentation (metadata), preferably based on widely used or international standards. For more information on best-practices for data and metadata formats please refer to the Axiom published Best Practices for Scientific Data Management website. There are levels of data probation that SECOORA respects:

1. Immediately available (near real time) - examples of this type of data are most physical

oceanographic and environmental data, models, and satellite data.

2. Probation - any species sensitive data that may influence "bad behavior" (e.g., resource depletion) can be probated 60 days in the case where real-time data is an issue. An example of this type of data is animal borne sensors that collect physical data.
3. Embargo - for fisheries data where the PI has justified embargo *in consultation with SECOORA*, data can be embargoed for 2 years after the data is collected, or until the PI publishes the data, whichever comes first.

DMAC System Management Structure

The SECOORA DMAC System Management Structure is composed of the SECOORA Deputy Director, Axiom DMAC Technical Lead, Axiom DMAC Coordination Lead, Data Team, and Web Team. Both the Data Team and the Web Team are led by the Axiom DMAC Technical and Coordination Leads. The SECOORA Deputy Director is responsible for the RCOOS and DMAC Operations. The Deputy Director provides overall project management expertise and oversees all aspects of the SECOORA DMAC System and data management reporting requirements. The Deputy Director reports to the SECOORA Executive Director.

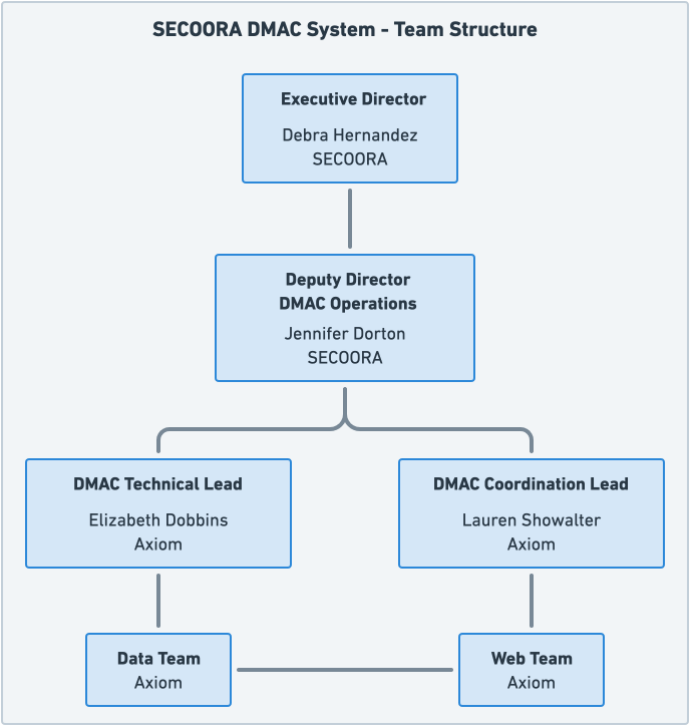

Figure 1: Team Structure

The DMAC Technical Lead (Elizabeth Dobbins, Axiom) and the DMAC Coordination Lead (Lauren Showalter) oversee the Axiom Data Team and Web Team staff, and the SECOORA DMAC system. Additionally, the DMAC Technical and DMAC Coordination Leads contribute to proposal development and general SECOORA data management reporting requirements. The DMAC Technical Lead is the main point of contact for all technical data-related questions and is an expert in managing large scale datasets related to the SECOORA mission. The DMAC Coordination Lead submits bi-annual reports to SECOORA Deputy

Director for RCOOS and DMAC Operations and website services. Workplan reviews are conducted quarterly (see *Figure 1 – Team Structure*).

Data Team

The Data Team is composed of staff from Axiom and are involved with all aspects of the SECOORA data flow, including data ingestion, creation of metadata, conversion, discovery, maintenance of data feeds, storage, and any necessary archival services. Its primary goal is to gather and serve data important to SECOORA end users via standard services as recommended by the IOOC and the IOOS Program Office (e.g., OPeNDAP, ERDDAP, THREDDS). It is also tasked with managing and archiving any SECOORA-funded and non-funded data generated by oceanographic models, buoys, or other devices to enable generation of data products.

The Data Team is responsible for the design and deployment of a DMAC System to meet the needs of the SECOORA user-base. This system must provide the functional components required by IOOS RCOS as described in this plan. The team offers comprehensive technical solutions to data management needs, underpinned by a scalable, open-source system that uses existing and emerging software, high performance computer clusters, and interoperability services. This data management system provides an environment that increases the access and use of data by all user groups and allows data management staff to rapidly develop new capabilities and tools to meet emerging user needs.

The Data Team provides data management and informatics support for SECOORA and provides development capabilities 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 continually researches and employs new technologies to extend the capabilities of digital information and computer analysis systems.

Web Team

The Web Team is involved with all aspects of the maintenance and upgrade of the SECOORA content website that includes implementation and maintenance of content management software and functionalities, linkages to the data portal, and other information products. Its primary goal is to implement functionalities that will ensure the SECOORA web site is user friendly and easy to navigate and serves content and documents pertaining to SECOORA governance and constituents (Staff, Board, PIs, members, and end users). The team is responsible for archiving the content web site.

The DMAC Coordination Lead (Lauren Showalter, Axiom) oversees all aspects of the Web Team. The Axiom Web Team coordinates activities and tasks with SECOORA staff for content development and the Data Team when including new products and services. Additionally, the Web Team contributes to content development, website layout, and design. The Web Team submits quarterly web site analytics reports to the SECOORA staff. CVs for Axiom personnel are found in Appendix B.

3. SECOORA Data Resources and Asset Types

The SECOORA DMAC System provides data to the public from multiple sources including SECOORA-funded projects and data from numerous and diverse external federal and non-federal organizations.

Observational Data Types

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

- *Real-time data* are ingested, served, and displayed by the SECOORA DMAC System at the same frequency the data are collected (and sometimes reported) by the originator with little to no delay. Real-time assets primarily include shore stations, HF Radar, gliders, oceanographic buoys, and numerical model data. Axiom also serves all buoy and coastal data through the ERDDAP system to NDBC for incorporation into their data system.
- *Near real-time data* are ingested by the SECOORA DMAC 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. Historical data are sometimes collected in real-time and then archived, and sometimes ingested from local or national archives on request. Examples of historical data include data recovered from a glider post-deployment and a one-time model run for a specific domain or time period that covers a weather event.

Data Categories

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

1. Federally Sourced Data
2. Model Products
3. Static Data Products
4. Funded Data Streams (SECOORA funded)
5. Regional Partner Data Streams (not SECOORA funded)

Federally Sourced Data

Federally sourced data incorporated into the SECOORA data portal are quality controlled following rigid data management and archival processes by the federal agency collecting the data. These data only require generic documentation by SECOORA on how these data are ingested and made available to the public via the SECOORA data portal (see Section 4 of this document). Federally sourced data served by the SECOORA DMAC System are listed in Appendix C. These data have already undergone QA/QC and have appropriate metadata available. Examples of federally sourced data include 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), Department of Agriculture, and U.S. Fish and Wildlife Service (USFWS).

Model Data Products

Model outputs and products served by the SECOORA DMAC System may incorporate or assimilate observational data (e.g., salinity from North Carolina State University (NCSU) Coupled Northwest Atlantic Prediction System (CNAPS)). These models are considered a product that falls outside the realm of "true" observations therefore they are exempt from detailed data stream documentation. See Appendix D for a list of federal and non-federal model products inventory.

Static Data Products

SECOORA 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 E contains the static products inventory.

Funded Data Streams

Data funded by SECOORA undergo data management by the data provider and/or by Axiom. Axiom data management includes data ingestion, standards and formatting, metadata and discovery, QC, stewardship and preservation, access and dissemination, archival and security. Descriptions of the processes that consistently apply to all SECOORA funded data streams are provided in Section 4. Additional data management documentation unique to individual data streams are provided through a systematic Data Stream Plan template that closely follows the RCOS Certification Guidance DMAC requirements. Use of this custom Regional Data Stream Plan template facilitates consistent documentation and streamlines future additions and edits to existing data stream protocol.

The Funded Data Stream Plans use a consistent and comprehensive template 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 dissolved 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., University of South Florida (USF) Coastal Ocean Monitoring and Prediction System (COMPS)).

Quality Control descriptions included in the Data Stream Plans 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 conducted 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.

See Appendix F for a complete list of SECOORA funded data streams including links to individual Data Stream Plans.

Regional Partner 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 and are not funded by SECOORA. These include regional data provided by local or state agencies, private companies supporting maritime activities in coastal waters, university projects, and research studies funded and conducted by local entities. None of these data streams served by the SECOORA DMAC System originate directly from SECOORA funding. Most regional data originate 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

NSIDC-binary format into netCDF). Data streams may be of any data type: real-time, near real-time, historical, citizen science. Leveraged projects in which SECOORA helps support but does not fund may also fit into this category. Due to the external nature of these data streams, they are usually exempt from detailed data stream documentation. On occasion, however, a data stream that would normally be considered exempt will require documentation in a Data Stream Plan. This includes:

- 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 are further documented in a Data Stream Plan.
- A federal data source that is translated or transformed in some way between the source at ingestion to the SECOORA access point of delivery (e.g., smoothing, block averaging).

See Appendix G for a complete list of Regional Partner Data Streams.

4. SECOORA DMAC System Architecture and Work Plan

Axiom has developed a framework for managing a variety of ocean data types (in-situ and remotely sensed data streams, multidimensional grids, GIS, and other structured formats). This framework exposes managed data through interoperability systems and uses several user interface tools that allow the data to be discovered and explored by the broader community. Use of this framework to power the SECOORA DMAC system will enable the SECOORA data team to rapidly ingest or connect to data sources relevant to SECOORA and develop advanced user tools and data products efficiently.

The SECOORA data system is divided into four logical tiers, which separate the suite of technologies composing the system. See *Figure 2* for a diagram of the system.

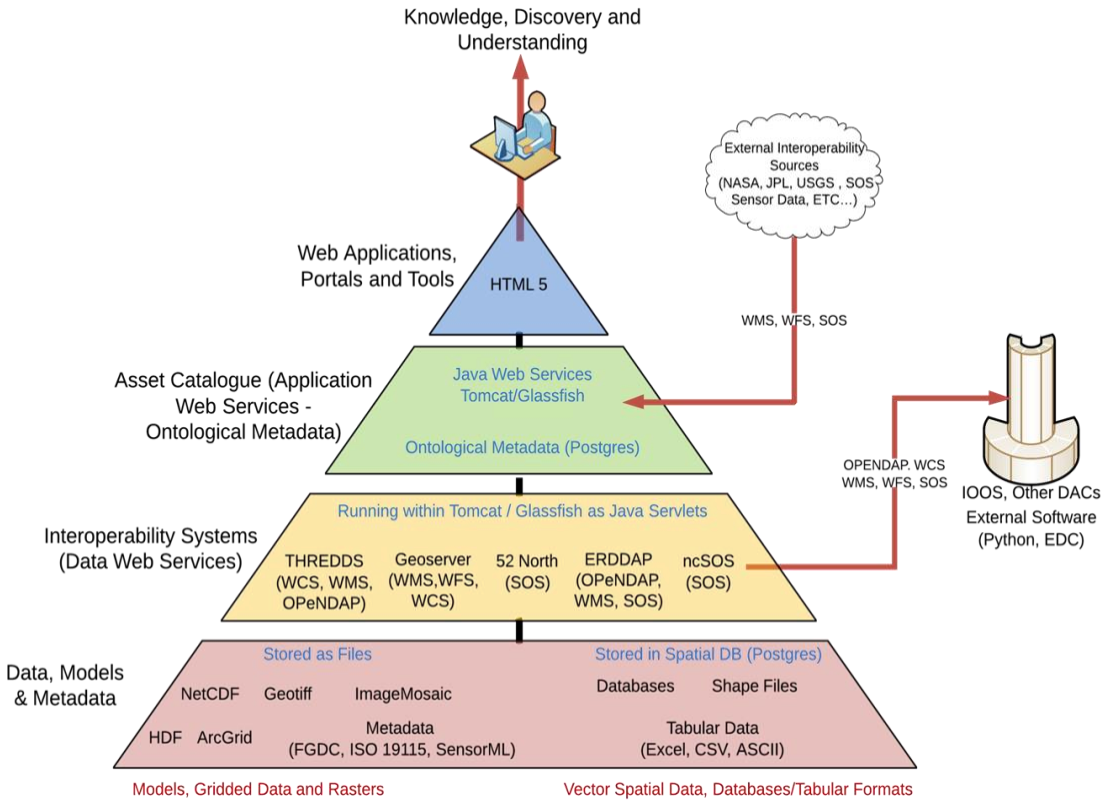

Figure 2. SECOORA DMAC system framework showing the flow of data through logical technology tiers, enabling discovery of data that enables understanding the ocean and coastal environments.

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 others 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 OGC WFS and OGC WMS protocols. Imagery, raster, and model data will be stored in a file server in their native file formats. THREDDS is 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 or other two- and three-dimensional imagery/raster data.

Tier 2 (Interoperability Systems) includes systems such as OGC Web Services (WMS, WFS, WCS) and other services such as ERDDAP to connect to the underlying data sources in Tier 1. Various interoperability servers (GeoServer, THREDDS, WMTS, ERDDAP 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.

Tier 3 (Asset Catalog) includes an Asset Catalog, which provides (1) ontological metadata and (2) connections to externally-hosted data via web services. The ontological metadata in the catalog 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 SECOORA 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 catalog 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 SECOORA.

Tier 4 (Web Applications, Portals and Tools) 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 Python and Java connect to the asset catalog and provide applications with access to the underlying descriptions of SECOORA data assets and sources. The asset catalog 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. 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 Southeast's marine and coastal environments.

User interfaces are web-based applications and tools that 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 the Southeast. 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 SECOORA DMAC System.

Data Ingestion

Observations and information are ingested into the SECOORA 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 SECOORA DMAC System. Each data asset ingested into the DMAC System has its own level of data processing maturity and quality with respect to the metadata available.

Data can be ingested into the SECOORA system using one of several pathways:

1. Contribution by the originator
2. Direct access or harvest from the originator website
3. Auto submission pathway from the Workspace

SECOORA funded partners provide data to SECOORA in a timely manner, as stipulated in the SECOORA's US IOOS Cooperative Agreement and in the annual descope proposals. Many data types are served in real-time. In cases where projects do not produce real-time data, the project PIs are responsible for making sure data becomes available to SECOORA as soon as possible.

Standards for Format and Content

Shared Data File Formats

SECOORA provides nearly all data in four open and 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. 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.

Data Access Points

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

1. Thematic Real Time 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. SECOORA provides THREDDS access points for raster (gridded) data and discrete time-series observations 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. SECOORA 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. SECOORA provides WMS access points for point, vector, and polygon information, as well as raster data through the open-source software packages ncWMS, ncWMS2, sci-wms and GeoServer.

4. Web Feature Service (WFS) - this service provides machine access to the vector elements of static layers. SECOORA provides WFS access points for point, vector, and polygon information, as well as time-series and raster data through the open-source software applications GeoServer.
5. Environmental Research Division's Data Access Program (ERDDAP) - ERDDAP is a common data server that provides access to subsetting and downloading data. SECOORA provides ERDDAP access to all time-series data in the region, a subset of gridded data, and some table-based GIS data products.
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.

The flow of data from the source to the SECOORA data portal follows the same general path for all sources as illustrated in the following flow diagram (Figure 3). For cases where the data are transformed or modified in any way, an explanation is provided in the individual Regional Data Stream Plan for that instance. This includes format translations or aggregations of component data streams into an integrated product.

Though SECOORA relies on local investigators to provide best practices for QA on their activities related to data submitted to SECOORA, 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.

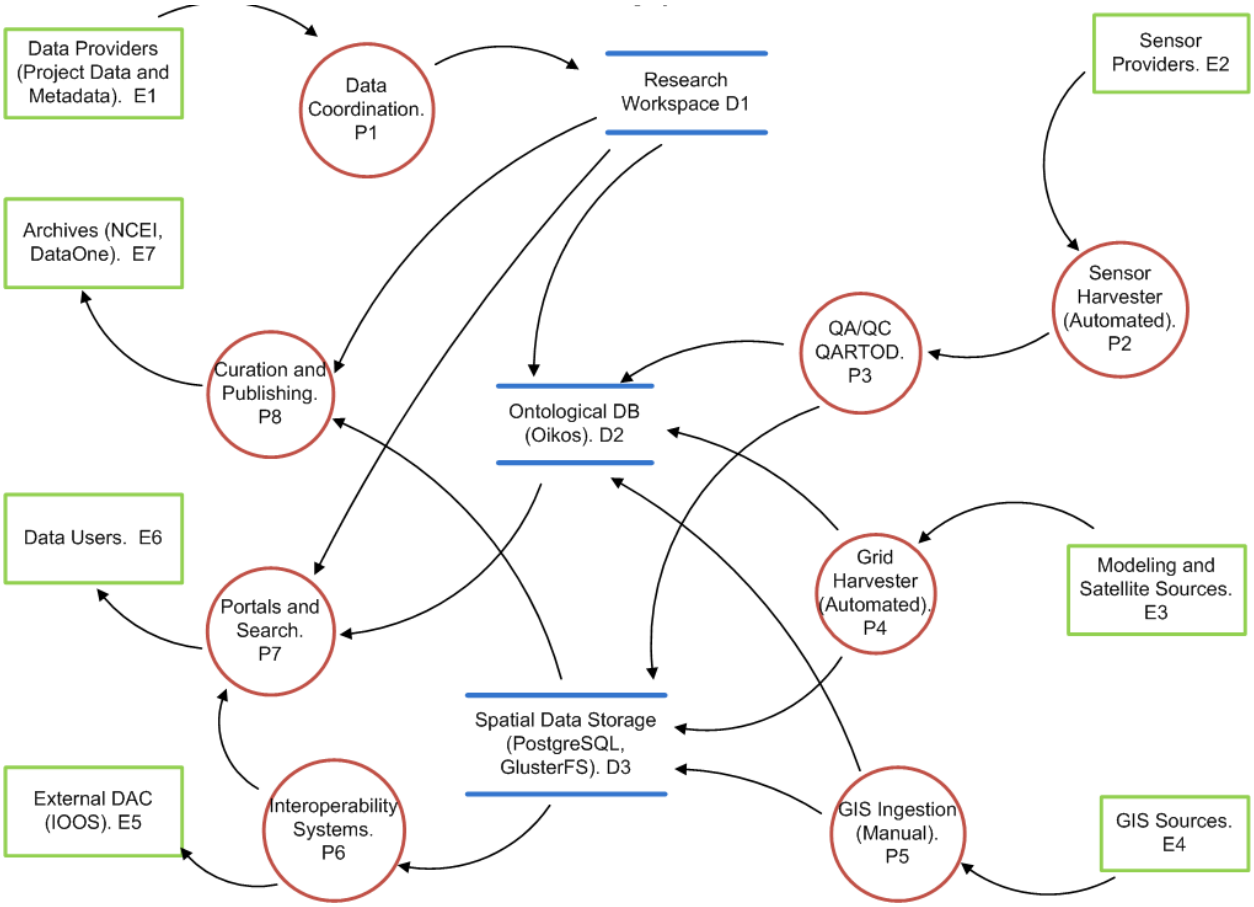

Figure 3. Flow diagram of data: From data, source, model, and GIS providers to users, archives, and external DACs.

Metadata and Data Discovery

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. However, many older data sets 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, SECOORA works towards 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 SECOORA data catalog. Details and availability of metadata are discussed in individual Data Stream Plans.

Quality Control Procedures

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

QC processes implemented by SECOORA 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.

SECOORA Implemented QC Protocols

SECOORA does not receive or serve any raw data transmitted directly from stations, so any applied QC procedures administered by SECOORA are in addition to those applied by the data provider. For many project-based and historical datasets, SECOORA 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 SECOORA by the source provider are included in the QC section of the individual Data Stream Plans (Appendix F).

SECOORA QC Protocols

Effective 2018, SECOORA began implementing policies outlined in the US IOOS Quality Assurance of Real-Time Oceanographic Data (QARTOD) manuals. As new data are ingested into the SECOORA DMAC 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 SECOORA. 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 SECOORA have different processes depending on the data type -- real-time data, historical data, citizen data, and federal data.

Real-time data

The SECOORA DMAC system ensures that QC standards are implemented, and QC flags made available, for all real-time data that are not received from a federal source. SECOORA currently serves various non-federal data streams that require QARTOD QC test implementations. QC procedures differ depending on whether they are implemented by the data provider or SECOORA.

QC by Data Provider

SECOORA funded data providers (e.g., UNCW CORMP, USF COMPS) implement and document the implementation of QARTOD recommended QC tests performed for their data. The University of Georgia maintained Ocean Acidification sensors on the NOAA NDBC Gray's Reef mooring are a part of the US National Ocean Acidification program, and NOAA Pacific Marine Environmental Lab (PMEL) reviews the data and implements quality control measures. In these cases, links to these procedures, or a summary of the QC performed is provided in the organization's individual Data Stream Plans (See Appendix F). The NWS, NERRS, the IOOS HFR DAC, and the IOOS Glider DAC all ingest and perform extensive QC on the raw data collected by these platforms prior to making them available to the public. In these cases, parameters and/or configuration for the quality test are defined by the data provider. SECOORA ingests these data and quality flags from the programs for display in the DMAC System and are not required to perform additional QC on these assets. Roll-up summary flags and individual test flags are shown visually in the SECOORA data portal with links to the QC documentation made available by the data provider. Flags are also stored alongside the data for download in CSV and netCDF downloads, as well as via

THREDDS and ERDDAP servers. See examples of roll-up summary and individual test flags for a UNCW CORMP sensor.

QC by SECOORA

For sources that do not provide quality flags, the SECOORA DMAC system runs QARTOD tests after ingesting observation data. Tests are run using the open source ioos_qc library 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 SECOORA data portal. Links to the ioos_qc methods used are available both within data charts and on sensor pages within the SECOORA 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 SECOORA DMAC 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 SECOORA data portal both visually and for download, similar to source-provided flags described above. 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 for Florida Department of Environmental Protection (FLDEP)’s Vilano Beach station run by the SECOORA system 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 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 for FLDEP's Vilano Beach wind sensor run by the SECOORA system can be seen here.

In addition to being viewable, quality test results are available for download in the SECOORA 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 FLDEP's Vilano Beach wind sensor dataset and embedded quality test results available in ERDDAP is here.

Historical Data

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 SECOORA Data System, follow the same QC protocols as the real-time data. In these cases, the applicable quality tests are run retroactively for all legacy time series data available within the SECOORA data portal. The quality test results are available visually along the time series continuum within the SECOORA data portal and for download following the same procedures as the real-time data.

Stewardship and Preservation

SECOORA stores ingested data in a secure, professionally managed external facility. SECOORA currently has total storage space for over 1.8 petabytes of data, supported by multiple redundant backups to Amazon Glacier, a cold storage backup and archival service offered by Amazon Web Services, and BackBlaze, an offsite backup service provider.

SECOORA 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 SECOORA to grant users rapid web-based access to all sensor data (federal and nonfederal) since SECOORA began aggregating feeds. The only assets that are not kept indefinitely in storage are webcam images, NEERS data (as it is strictly prohibited in their terms of service) and forecast products that have been replaced with a more accurate forecast.

Providing Public Access and Dissemination

All data served on the SECOORA DMAC System are fully available to the public and have no data restrictions or embargo periods placed on them. New datasets from either new or current data providers received by SECOORA are immediately available to the public after data ingestion and documentation is complete; however, they are not added to the searchable data catalog. Datasets are added to the searchable, public catalog only after the data provider is brought into a feedback loop to comment on the metadata, usage notes and citation information regarding their dataset. Once published in the catalog, datasets are promoted via the SECOORA website, social media accounts, outreach campaigns, and an email newsletter.

The SECOORA DMAC System provides a variety of environmental and socioeconomic data resources in a

one stop data portal, free to the public, with data originating from SECOORA funded data providers, federal and state agencies, local municipalities, academic institutions, research organizations, private companies, non-profit organizations, and community observers. Any data served by SECOORA data portals 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. Occasionally, a data sharing agreement between SECOORA and a data provider will identify the existence of intellectual property rights (IPR) to the data and this is noted in the applicable Data Stream plan. However, IPRs do not restrict access to any of the data that is freely served through SECOORA data portal. IPR information is provided out of courtesy to the data provider, and it is an unwritten expectation for all data used by someone other than the originator. It is a best practice to always clearly give credit to the data source (the originator) and data provider (in this case SECOORA) in any work or publications that emanate from using data accessed via the SECOORA data portals.

Data Archival

As a federally-funded program, SECOORA is required to submit data it generates to a national archive center. SECOORA works with the NOAA NCEI to assist with the preservation of appropriate data types. The SECOORA Data Management Team has worked and consulted with several NCEI staff members on automating the submission of SECOORA generated and managed data assets and SECOORA managed non-federal real-time assets to the NCEI. NCEI will advise SECOORA on the data submission forms and all necessary procedures. However, NCEI does not currently have the infrastructure nor the desire to directly ingest the many types of data that are aggregated and curated by SECOORA. To facilitate future interest, SECOORA will continue to make all data served available to NCEI and will push any data asset to NCEI if NCEI expresses interest in archiving. See Appendix H for the current NCEI Request to Archive submitted in 2017.

SECOORA also archives data through two national IOOS DACs: the IOOS HFRadar DAC and the IOOS Glider DAC. Both national DACs archive all submitted data, including that submitted by SECOORA, with NCEI.

Performance and Security

The SECOORA Data Team operates a High Performance Computing (HPC) cluster located in Portland, OR. This 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. SECOORA 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 SECOORA Data Team provides the following enterprise-level infrastructure capabilities:

Security and Redundancy

The SECOORA Data Team's data center in Portland, OR implements several levels of redundancy and backup. Our main storage cluster is distributed over several layers of physical hardware and we make use of Amazon Glacier, a cold storage backup and archival service offered by Amazon Web Services, for long

term archival and disaster recovery. Enterprise level firewalls and system monitoring software are also in place to provide hardened cyber security.

Capacity and Performance

High Performance Computing (HPC) has been a component of the SECOORA Data Team's technical strategy since early 2011. The SECOORA Data Team 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. SECOORA Data Team engineers 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.