NANOOS Data Management Plan

Table of Contents

Version Tracking

Date	Persons	Change Description
2020-11-20	Craig Risien	Updated for inclusion in the NANOOS 2021-2026 Proposal
2020-08-16	Emilio Mayorga	Certification Approved
2018-2-22	Emilio Mayorga	Response to IOOS Request for Additional Information: - Added Version Tracking section and Table of Contents - Enhanced Archiving and Quality Control descriptions, including timelines - Clarified roles of NANOOS “internal” providers vs NANOOS DAC - Added citation (Seaton et al., 2018)
2017-07-11	Emilio Mayorga	Initial version, submitted with Certification Application

A. Background

The NANOOS Data Management And Cyberinfrastructure (DMAC) effort involves a collaboration of several NANOOS partners who are responsible for developing and maintaining a regional Data Assembly Center (DAC) for robust, operational data aggregation, management, quality control, distribution and archiving; and informative, user-friendly products encompassing the web portal and interactive, data access and visualization applications online and on mobile platforms. NANOOS DMAC integrates and manages data from a variety of sources and types of assets that include in-situ observations, remote sensing observations and products, processed data products (such as climatologies), and numerical model nowcasts and forecasts.

NANOOS organizes the broad DMAC effort into two overlapping committees or teams, with overlapping members and regular cross-communication: DMAC and User Products (UPC) Committees. The DMAC Committee oversees DAC operations, including compliance with IOOS DMAC functional roles, while the UPC oversees web portal and user product development, prioritizing and usability assessment for new features, and user engagement. The UPC strives to create innovative and transformative user-defined products and services for Pacific Northwest stakeholders, sustaining innovations with the NANOOS Visualization System (NVS) to succeed in this vital translation: meaningful and informative data products that connect with user applications and serve society. NVS, including its supporting data and metadata stores and integrated visualization products, play a central role in NANOOS DMAC and UPC efforts. Weekly "tag-up" calls of these two committees plus a third, the Outreach, Engagement, and Education (OEE), comprise a team, the NANOOS Tri-Com, who collectively produce and review enhancements and development of our data products and services to meet user needs. Annual Tri-Com meetings are used to review progress and set priorities for the coming year.

The goals of the DAC system are to provide the functional capabilities described in the "Guidance for Implementation of the Integrated Ocean Observing System (IOOS) Data Management and Cyberinfrastructure Subsystem" document (https://ioos.noaa.gov/data/contribute-data/), and to support the data and metadata needs of NANOOS user applications that serve a range of partners and consumers including government (local, state, federal and tribal), industry, education sectors, non-profits and the general public. These goals are carried out as a distributed collaboration involving primarily the University of Washington (UW), Oregon State University (OSU) and the Columbia River Inter-Tribal Fish Commission (CRITFC), which in 2020 assumed stewardship of observing and modelling assets previously operated and maintained by Oregon Health & Science University's (OHSU) Center for Coastal Margin Observation and Prediction (CMOP).

Data for in-situ observations and metadata for all asset types are integrated, managed and centralized on UW servers to support user access by the NVS application, programmatic access to asset inventories via open standard services, and programmatic access to regional in-situ, fixed location data via ERDDAP and the IOOS implementation of a 52North Open Geospatial Consortium (OGC) Sensor Observation Services (SOS) instance, all of which are UW hosted. This integration, also supported by OSU and CRITFC, encompasses "internal" data from NANOOS-funded assets as well as "external" data from other regional assets and federal and Canadian assets. NANOOS submits glider and High Frequency (HF) Radar data to the corresponding national DACs for national integration and distribution, while also consuming those data streams in our user applications.

In addition to direct observation data, NANOOS also integrates and manages data from other types of assets, including gridded data from models and processed data products. These are currently distributed via THREDDS and consumed into our user applications.

The NANOOS DMAC team keeps abreast of IOOS and U.S. IOOS Program Office data management activities via regular communication with Program Office personnel, other Regional Association (RA) DMAC teams, and community-based discussions. It maintains communication particularly with West Coast RA's through common participation in the West Coast Ocean Data Portal (http://portal.westcoastoceans.org, a component of the West Coast Ocean Alliance), the IOOS Pacific Region Ocean Acidification (IPACOA) Data Portal (http://www.ipacoa.org), and other cross-regional initiatives. It also maintains international engagement via collaborations with Canadian partners (particularly in British Columbia, but also nationally) and technical leadership in the Global Ocean Acidification Observation Network (GOA-ON) Data Portal (http://portal.goa-on.org). (NANOOS DMAC regional, cross-regional, and international collaborations on ocean acidification are discussed in Mayorga et al., 2016, and Newton et al., 2012).

B. Roles and Responsibilities

Both NANOOS DMAC (DAC) and UPC teams include personnel from UW, OSU, CRITFC and DOGAMI (Oregon Department of Geology and Mineral Industries). The DMAC Committee oversees DAC operations, including compliance with IOOS DMAC functional roles, while UPC oversees web portal and user product development, prioritizing and usability assessment for new features, and user engagement. NANOOS OEE members provide oversight on usability. OSU and UW co-lead DAC efforts as well as implementation and operational activities involving centralized NANOOS resources including the web portal and NVS; however, prioritization and definition of enhancements and new capabilities are arrived at collaboratively among these NANOOS partners. In addition, data and visualization product development, multi-source data assembly, and IOOS protocol or functional capability implementation are all activities carried out jointly by these partners in support of NANOOS DMAC efforts.

All data flows from instrument platforms to shore-side servers are managed by platform operators, both for "internal" (NANOOS) and "external" data streams. NANOOS DAC ingest of data streams begins with data already found on a shore-side server. See Section E for discussion of the distinction between those two types of data streams, and data flows and transformations.

Craig Risien (OSU), NANOOS DMAC Committee Chair. Leads the NANOOS data management (DAC) team, ensuring that all data collected by the program are timely, properly preserved, and made available via IOOS standard services. Coordinates and leads the implementation of IOOS DMAC functional capabilities involving data integration, management, quality control, distribution, and archiving. Coordinates DAC activities among NANOOS DAC partners and serves as primary point of contact between NANOOS, the DAC team, data providers, peer RA DMAC teams, and the IOOS Program Office DMAC team. Also provides coordination with other relevant cyberinfrastructure and data initiatives regionally, nationally, and internationally.

Troy Tanner (UW), leads NANOOS web portal and user application development, including mobile applications. Lead developer for NVS, including plot and map rendering capabilities such as map tile generation from gridded data. Coordinates integration of distributed data products into cohesive and user-friendly user applications. Also coordinates NVS metadata and data store development and maintenance. Supervises staff who perform system administration for all UW NANOOS servers, including those supporting DAC capabilities.

Jonathan Allan (DOGAMI), NANOOS UPC Chair. Coordinates UPC activities to seek user feedback and prioritizing of user product development and enhancements. Also provides and oversees direct data product development in his areas of expertise, including shoreline observations, near-shore bathymetry, tsunami hazards, and climatology.

NANOOS selected these individuals based on their achievements, qualifications, and regional knowledge. All are known experts in their respective fields. Each of these NANOOS leads provides performance reports to the Executive Director and are part of NANOOS reviews. Additionally, all are subjected to the annual review processes of their home institutions. UW and DOGAMI each have a process in place for personnel evaluation. All personnel listed have received excellent evaluations that are on file with their respective Human Resources departments.

C. Implementation of Data Management Protocols

(Certification requirements Section f5) NANOOS DMAC personnel maintain regular communication with the U.S. IOOS Program Office through a variety of mechanisms including in-person meetings, phone calls & webinars, email conversations, and GitHub repositories. This communication ensures that the DMAC team is aware of all new practices and protocols, as promulgated by the IOOC and the IOOS Program Office and understands how and when to implement them. NANOOS DMAC personnel also play active roles in IOOS DMAC projects, contributing to software development, implementation of standards, and system documentation.

The NANOOS servers and data management software are maintained by the NANOOS DMAC team so data management protocols can be applied as soon as practicable, limited only by resource restrictions (personnel time, budget, or server capabilities). The NANOOS DMAC team plans to implement all new protocols as soon as possible and within one year of adoption by IOOS. Once the new protocol is received by the DMAC lead, an implementation plan, including a more detailed schedule, will be outlined, and executed by the DMAC team.

D. Computing Infrastructure

DAC operations and user applications are supported by servers primarily at UW (Applied Physics Laboratory, APL), with complementary servers at OSU and CRITFC. **UW servers** run on Ubuntu 14.04 and 16.04 and include four physical servers and two Virtual Machines (VM's) on one of those servers. These servers host MySQL and PostgreSQL relational database management systems, Nginx and Tomcat web servers, Tomcat web apps (GeoServer, IOOS 52North SOS, ERDDAP, a UW-APL-developed map tile server supporting NVS), the NVS server-side application framework (in PHP), a Python data harvesting and monitoring library for NVS (Python 3, with heavy reliance on *conda* virtual environments), and numerous automation scripts. The NVS MySQL asset database is backed up nightly to an external hard drive. The servers are in a dedicated server room with power protection. Important files from the two VM's serving data ingest, storage and distribution roles (including applications providing web services: IOOS 52North SOS, GeoServer, and ERDDAP) are backed up nightly and stored offsite at the UW IT Service Center at two separate data centers each in a different seismic zone. Other files are backed up to nightly to external servers. Master, versioned repositories of all in-house code is maintained in a host git server at UW-APL using *Phabricator*.

Three **OSU servers** (Dell PowerEdge R520, R310, R330) process and serve plots and data files for the NANOOS web portal and NVS. They are situated in the CEOAS (College of Earth, Ocean, and Atmospheric Sciences) Environmental Computing Center (ECC) that is protected by a 24x7 uninterruptible power generator. These servers run on CentOS 6.10 and are configured using *RAID10*. The servers are backed up using *rsnapshot* and 2 x 2TB external hard drives. Baseline backups of critical directories such as the web server and home directories, which contain the shell and Matlab scripts that generate plots and data files, occur monthly to one of the external hard drives; and to the second hard drive using a 3-day rotation backup scheme. All external hard drives are in the ECC.

A **CRITFC server** processes and serves plots and data files for the NANOOS web portal NVS. It is situated in the Oregon Health & Science University (OHSU) Advanced Computing Center (ACC) that is protected by a 24x7 uninterruptible power generator. The server runs on CentOS 6.3 and its hard drives are configured using *RAID6*. Backups of critical scripts that generate plots and data files for NVS occur as needed to a separate subversion server. CRITFC observation data are stored in a PostgreSQL 9.1 database on a separate server (CentOS 5.11) also situated at the ACC. The database is backed up using *WAL* archiving.

E. Data Streams

NANOOS DMAC integrates and manages data from a variety of sources and types of assets that include in-situ observations, remote sensing observations and products, processed data products (such as climatologies and anomalies), and numerical model nowcasts and forecasts. This document focuses on data obtained directly from observations. **Variables** handled encompass meteorology; physical, biogeochemical, and biological oceanographic properties; and nearshore geomorphology. For NANOOS data streams (E.1), the specific variables and data types produced by each observational asset are described in Section I of the **individual provider Data Management Plans (DMP's)** (see DMP links in Table 1); variables for all integrated assets are provided by the asset inventory; descriptions of **NANOOS asset inventories**, including access information, are available at http://www.nanoos.org/documents/certification/DMP/NANOOSAssetInventory.pdf.

Data for in-situ observations and metadata for all asset types are integrated, managed and centralized on UW servers to support user access by the NVS application, programmatic access to asset inventories via open standard services, and programmatic access to regional in-situ, fixed location data via ERDDAP and an IOOS 52North SOS instance, all of which are UW hosted. This integration encompasses **"internal" data** from NANOOS-funded assets (managed by operators – NANOOS Principal Investigators – who are an integral part of NANOOS) as well as **"external" data** from other regional assets or larger scale observation systems, and federal and Canadian assets. Federal sources of in-situ observation data (NOAA NDBC, NOAA CO-OPS, NOAA PMEL, USGS, CDIP, etc.) are managed by federal partners and are not discussed further in this document. NANOOS submits glider and High Frequency (HF) Radar data to the corresponding national DACs for national integration and distribution using previously defined DAC standards, while also consuming those data streams in our user applications.

NVS (http://nvs.nanoos.org) serves DAC functions as a central integrator and distributor of data, asset metadata, graphical products (time series plots, map overlays, section plots, etc) and related information available to users in a consistent fashion (Mayorga et al., 2010). Its data flows and components are illustrated in Figure 1, though emphasizing fixed-location in-situ assets. Hosted on UW servers, the NVS asset metadata is maintained in a MySQL relational database linked to additional, external information stored on UW and other NANOOS DMAC and UPC partner servers. It relies on consistent categories of platform types, measured variables (labeled as "measurements") and units, data types, etc, and mapping and translation between heterogeneous source metadata and this consistent metadata. In situ observational data, particularly for – but not limited to – fixed-location assets (including depth profilers), are also integrated into this database and leverages this asset metadata as well as a coherent, harmonized representation of time, location, depth/altitude, and units; transformations from original source formats, encodings and conventions are performed using Python data harvesters sharing a common library. These transformations and checks **represent a base level of**

Quality Control applied by the DAC to all in situ observational data streams. Gridded data for map overlays (such as gridded maps of currents from HF Radar) are handled via a parallel system that converts data files into time-stamped Google Map tiles, or consumes external tiles or map services, all controlled by NVS asset metadata.

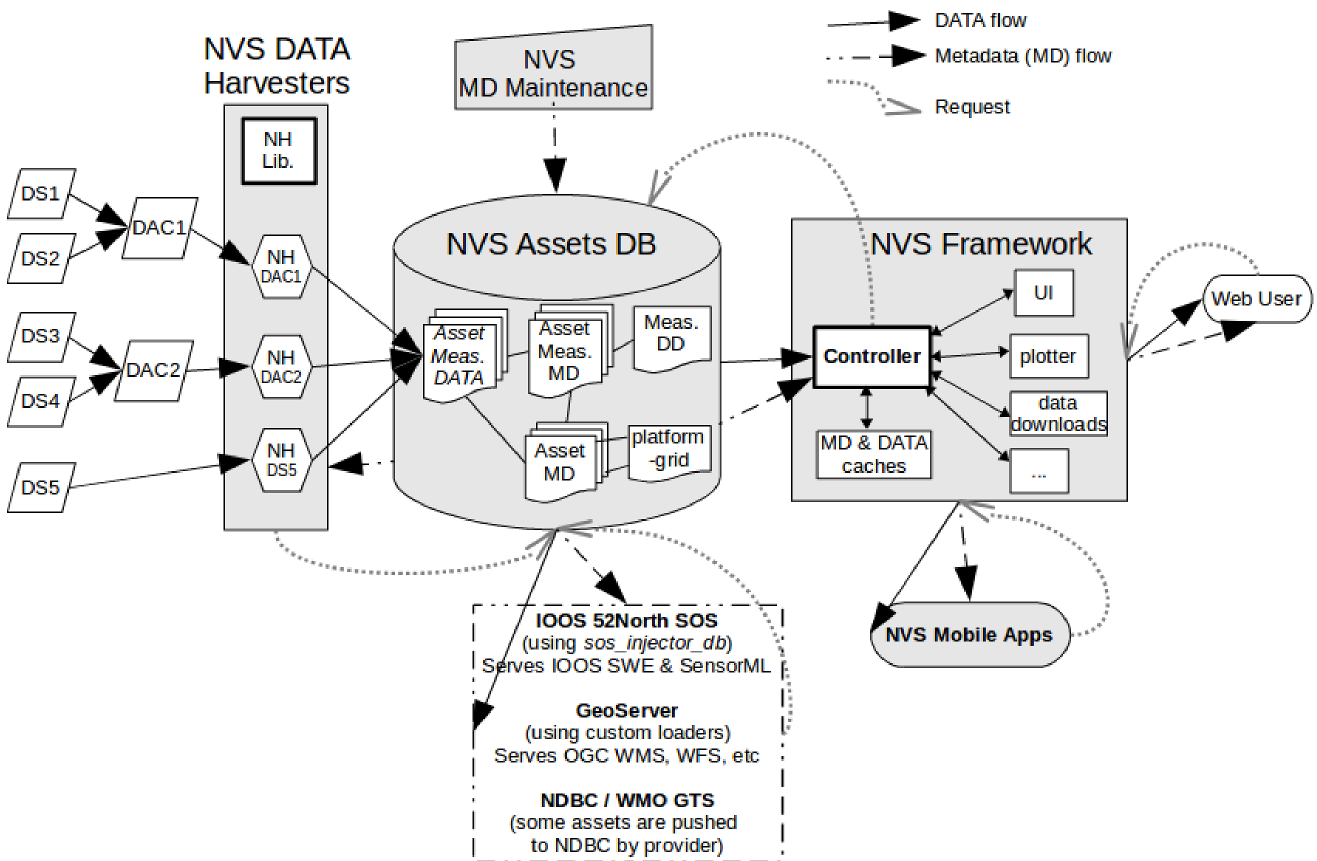

Figure 1. NVS data flows and components, emphasizing fixed-location in-situ assets (platforms). The NVS Assets Database contains the NANOOS DAC metadata store for all asset types and time series data store for fixed-location and some mobile in-situ assets. Shaded blocks represent broad NVS components. DAC1 and DAC2 illustrate complementary functionality served by NANOOS DMAC partners (e.g., OSU) in aggregating data from a few sources and redistributing them in a common form for easier access by the centralized DAC. Data flows from instrument platforms to operator shore-side servers (which in turn serve as NVS data sources) are not shown here. DAC (Data Assembly Center), DD (Data Dictionary), DS (Data Source), Lib. (Code Library), MD (Metadata), Meas. (Measurements), NH (NVS Harvester), UI (User Interface). Adapted from Mayorga et al. (2010).

The NVS metadata and data store in turn support **web services** for programmatic access, including **GeoServer** (http://data.nanoos.org/geoserver/) for access to inventory metadata and small data subsets via widely used OGC web services, including Web Mapping Service (WMS) and Web Feature Service (WFS); IOOS 52North **SOS** (http://data.nanoos.org/52nsos/), including **SensorML** metadata, for time series data from fixed-location in-situ assets, with a scope focused on regional assets supported by NANOOS and external, non-federal sources not otherwise available via IOOS compliant web services and which have provided their consent for web service redistribution; and a script to push fixed-location in-situ time series data to **NDBC**, which in turn pushes data to **WMO / GTS** (NDBC pushes are currently enabled for a subset of assets; in addition, some NANOOS providers push their data directly to NDBC). NVS itself provides a custom, lightweight set of HTTP GET web services that return readily parsed JSON responses.

E.1 NANOOS (Internal) Data Streams

Table 1 provides a summary of NANOOS-supported ("internal") observational data streams, including important data management and data distribution services or capabilities in place as they pertain to each data stream. Each of the 13 data streams is grouped into four categories that share many characteristics:

1. Surface Currents and Waves
2. Fixed-location Sensor Platforms
3. Gliders and Ferries
4. Beach and Shoreline Observations

These categories generally correspond to those presented in the NANOOS Strategic Operational Plan.

For each data stream category, this section will describe these functional capabilities:

1. Data Ingestion and Management *(certification requirements Section f2; individual DMP Sections I.1, I.2, I.3, III.1, V.1)*
2. Quality Control *(certification requirements Section f3; individual DMP Section I.3)*
3. Data Access and Sharing *(certification requirements Sections f2 and f4; individual DMP Sections II and V)*
4. Archiving *(certification requirements Section f6; individual DMP Section IV.2-8)*

Individual assets within the data stream categories will be described as appropriate but pointing to the individual provider DMP's for more complete details on the specific asset and the actions performed by providers (Table 1). Variables measured by each asset are listed in the individual DMP, Sections I.1-2. Each provider (operator) manages its own data flow from sensors to shore-side servers, computer infrastructure, and instrument logs for its data files before integration into the NANOOS DAC; these are described in the individual DMP's. The descriptions below include brief summaries of actions performed by NANOOS providers for assets under their responsibility (gleaned from the individual DMP's), as well as broader, across-the-board actions performed by the NANOOS DAC. The latter are indicated explicitly

Table 1. Data Management summary for NANOOS-supported ("internal") assets; for additional, detailed information on each asset, see the corresponding Data Management Plan (DMP) file(s).

	Operator	Asset Count	NVS Metadata Store	NVS Access*	52N SOS	GeoServer	NDBC / WMO GTS	National DAC	Archiving	DMP File
Surface Currents and Waves										
HF Radar (currents)	OSU	11#	X	All			X	X	NCEI	1.DMP.HFRadar.pdf
Port X-Band Radar (waves)	OSU	2#	X	Plots						2.DMP.PortsXBandRadar.pdf
Fixed-location Sensor Platforms										
Washington Shelf Buoys	UW	2	X	All	X	X	X			3.DMP.WAShelfBuoys.pdf
Oregon Shelf Buoy	OSU	1	X	All	X	X	X			4.DMP.ORShelfBuoy.pdf, 5.DMP.ORShelfBuoy_OA.pdf
Puget Sound, ORCA Buoy Program	UW	6	X	All	X	X	X			6.DMP.PugetSoundORCABuoys.pdf
Columbia River estuary & plume SATURN network	OHSU	14+	X	All	X	X	X		NCEI	7.DMP.ColumbiaSATURNNetwork.pdf
South Slough Estuary Obs.	SSNERR	6	X	All	X	X	X			8.DMP.SouthSloughNERR.pdf
Gliders and Ferries										
N. California Shelf Glider	OSU	1	X	Plots			X	X	NCEI	9.DMP.NorthernCAGlider.pdf
SW WA Glider	OHSU	1x	X				X	X	NCEI	7.DMP.ColumbiaSATURNNetwork.pdf
Victoria Clipper Ferry	WDOE	1	X	All						10.DMP.VictoriaClipperFerry.pdf
Beach and Shoreline Observations										
Oregon Shoreline Obs.	DOGAMI	-	X	Plots					State Agency	11.DMP.ORBeachShorelineObs.pdf
Washington Shoreline Obs.	WDOE	-	X	Plots					State Agency	12.DMP.WABeachShorelineObs.pdf
Nearshore Bathymetry	OSU	-	X							13.DMP.NearshoreBathymetry.pdf

* For NVS Access, "All" represents both data download and graphic presentation, and "Plot" only includes graphic presentation; # Number of radar sites;
+ Several stations are currently inactive but may be redeployed as resources allow; x Not currently deployed, pending additional funding and servicing.

**** Surface Currents and Waves***

HF Radar and *Port X-Band Radar* provide complementary but distinct capabilities for observing surface currents and waves, respectively. As the *HF Radar* technology is more mature, procedures for data management, quality control, data access and sharing, and archiving are more standardized and agreed upon nationally compared to *Port X-Band Radar*. NANOOS follows national HF Radar best practices, as referenced in the "National Surface Current Mapping Plan".

Data Ingestion and Management. Details for each asset type are described in its individual DMP (Table 1). *HF Radar* data are submitted to the National HF Radar Network (DAC) according to DAC specifications. *Port X Band Radar* data, representing radar backscatter intensity as a function of range, azimuth, and time, are recorded on-site and processed on a workstation either on-site or at OSU. Raw data are stored in NetCDF files that include all necessary metadata for making the data meaningful as well as other useful information pertaining to the recordings: Radar antenna elevation, Radar zero heading, Radar antenna location (latitude/longitude or UTM coordinates), Recording year, Recording UTC offset, and the "donut", or number of range bins that should be removed from each radar pulse in order to correctly map the data. The backscatter intensity data are rectified to geographic coordinates and used to create mean images (time exposures) and bulk wave parameters (peak directions and frequencies). Processed data is generally stored as MATLAB MAT files. Imagery is stored as PNG files that are generally available on the website for use by NVS within one hour of the end of the recording.

Quality Control (QC). Details for each asset type are described in its individual DMP (Table 1, section I.3). *HF Radar* QC procedures adhere to the National HF Radar Network (DAC) according to DAC specifications. There is no established standardized protocol for *Port X-Band Radar* image data. However, times when data is not available are flagged.

Data Access and Sharing. *HF Radar:* As it is distributed to the National HF Radar Network, these data are openly and readily available via web services and download capabilities provided by the National DAC. The DAC also makes the data available to the WMO GTS for use in national forecast capabilities. In addition, NANOOS produces its own gridded product from subsetting and reprocessing National DAC products; these are accessible for browsing and visualization as map overlays on NVS. *Port X-Band Radar:* Custom plots are created by the provider and integrated into NVS for the two port sites. Raw data are available upon request from the provider; community data standards for encoding and delivery of this data type are not available.

Archiving. *HF Radar* data are archived on NCEI by the National DAC. *Port X-Band Radar raw* data (with metadata) are currently not archived nationally but are preserved by the provider (OSU) in either raw or NetCDF format using local redundant HDD storage. Discussions between NCEI, the provider and NANOOS will be required to lay out appropriate options for long-term, national archiving.

**** Fixed-location Sensor Platforms***

Data for this asset type is generated from 5 datasets (*Washington Shelf Buoys, Oregon Shelf Buoy, ORCA Buoy Program, SATURN network, and South Slough Estuary Observations*) managed by 4 providers (Table 1). Platform types span shelf and estuarine moorings, profiling moorings, and sensors attached to piers or other coastal fixed structures. Variables handled encompass meteorology as well as physical and biogeochemical oceanographic properties. *South Slough Estuary Observations* are also part of the National Estuarine Research Reserves (NERR) network and follow NERR System-Wide Monitoring Program (SWMP) data management, processing, QA/QC and distribution procedures established by the NERR Centralized Data Management Office (CDMO). These procedures are described or pointed to in the individual DMP for this data stream (Table 1), taken from the SWMP network DMP; therefore, specific procedures will not be discussed in detail below, except as noted.

Data Ingestion and Management. Details for each asset type are described in its individual DMP (Table 1). Initial raw and pre-processed data are assembled by the data providers and made available for near real-time integration (harvesting) by the NANOOS DAC via data pulls, typically from a provider web server, using a variety of access protocols and data encodings. Procedures for harvesting and DAC management generally follow the workflow described in the introduction to Section E, including Figure 1, and include substantial data and metadata harmonization, including conversion to common metric units. The NVS data store currently maintains the most recent 60 days of data, though the lifting of this limitation is being incrementally explored and will be removed once QARTOD QC flagging and flag rendering is implemented. The IOOS 52North SOS ingests these data from the NVS data store hourly using the IOOS/Axiom *sos_injector_db* tool. Currently the SOS data store is retaining all ingested raw data since the initial asset ingest; however, this arrangement will change in the near future to the provision of near real-time data, including QARTOD test results, via ERDDAP.

Quality Control (QC). All NANOOS providers follow best practices for QA/QC procedures, as documented in the individual DMP's (section I.3 of individual DMP's; see Table 1). The current NANOOS DAC QC process for all fixed-station assets, in addition to what's described in Section E include syntax checks to reject malformed data sources; location (latitude & longitude) validation after each deployment; and feed/continuity monitoring to alert the DAC team when data are lagging well behind the expected schedule so that appropriate action(s) can be taken. NANOOS DMAC has participated actively in QARTOD implementation discussions, including the IOOS DMAC QARTOD Working Group and contributions to several QARTOD manuals. We initiated a QARTOD flagging pilot on the NVS data store in Winter 2017 using the *ioos_qartod* Python package. In addition, OHSU has taken steps to implement QARTOD flagging on the *SATURN network* data in 2018 (Seaton et al., 2018) in coordination with the NANOOS DAC (see the individual DMP), and King County – an external regional partner – is already implementing QARTOD-based flagging on their data streams (see Section E.2). NANOOS is also building regional climatologies and historical records that will help establish QARTOD ranges. For variables with QARTOD manuals, we anticipate operational implementation of QARTOD flagging at the NANOOS DAC by mid-2021, for all Fixed-location Sensor Platforms integrated by the DAC.

Data Access and Sharing. As described above in the introduction to Section E, data are available for browsing and visualization on NVS, as well as download or programmatic access via NVS, SOS, NDBC / GTS, and ERDDAP. In addition, two THREDDS servers are currently maintained by CRITFC (*SATURN network*) and OSU (*Oregon Shelf Buoy*) with partial to complete historical data access: http://data.stccmop.org:8080/thredds/archive.html and https://wilson.coas.oregonstate.edu/thredds/catalog/NANOOS/NH10/catalog.html, respectively. A centralized NANOOS THREDDS catalog is planned for deployment on UW servers in 2021.

Archiving. NANOOS started archiving data with NCEI in Spring 2017. In this initial stage, the *SATURN network* data (Table 1) from OHSU-CMOP (now CRITFC) was used as a pilot to develop metadata conventions, file segmentation and archival procedures that can be readily adapted to fixed-location in-situ data streams from other NANOOS sources. Nearly all historical data have been submitted, and all current data is being submitted monthly starting on July 15 (a search on NANOOS and "Buoy / Station" at https://www.nodc.noaa.gov/ioos/ yields 32 results as of 2020-11-19). Archival files are in the NCEI NetCDF Templates v2.0 format and follow ACDD and CF conventions. The NANOOS DAC automatically stages monthly incremental archive updates on the 10th of the month at http://data.nanoos.org/ncei/ohsucmop/, and these are automatically downloaded (pulled) by NCEI by the 15th of the month. The NCEI Submission Agreement (SA) is provided as the Appendix document *NANOOS-NCEI-Submission_Agreement_2017-03-13T19-51-12.pdf* (click on the file name for direct access). The SA has extensive technical details about the archiving procedures. Below are relevant fragments from the SA, with small additions:

Submission Information Packages (SIP) will be organized into 'bags'. Each 'bag' will contain data, metadata, and manifest files which fully document the files intended to be submitted. The 'bags' will be folders on http://data.nanoos.org/ncei/ohsucmop/ which correspond to the name of the platform. E.g. abpoa/, riverrad/, saturn01/, etc. Within the station folder (or 'bag') there will be four standard files with the following names: bag-into.txt, bagit.txt, manifest-sha256.txt, and tagmanifest-sha256.txt as well as a data/ directory which will contain folders for all of the netCDF files to be submitted. Each of the folders within the data/ directory represent an instrument/instrument deployment.... NCEI will organize the Archival Information Packages (AIP) by station. Each time a new station arrives, a new AIP will be generated.

NANOOS DMAC will work with other NANOOS-funded providers to develop plans for NCEI archival following the conventions and procedures developed for the NANOOS *SATURN network* data set. Some providers may choose to archive only delayed-mode, post-processed data within a couple of months after mooring or sensor recovery after each deployment (typically every 6 months in the case of *Washington* and *Oregon Shelf Buoys*). Data from *South Slough Estuary Observations* are expected to be archived with NCEI via separate arrangements between NERR CDMO and NCEI.

*** Gliders and Ferries**

Of the two NANOOS gliders, *Northern California Shelf Glider* (a Seaglider) and *SW WA Glider* (a Slocum glider), only the former (a collaboration between NANOOS and CeNCOOS) is currently deployed. However, all past transect deployments from both gliders have been submitted to the national Glider DAC. *Victoria Clipper Ferry* collects near real-time geo-referenced environmental data during daily transits of the private passenger ferry vessel, Victoria Clipper IV, as it travels between Seattle, WA and Victoria, B.C. (Canada). WDOE collaborates in the management of this data stream with UW-APL and with Integral Consulting, Inc.

Data Ingestion and Management. Details for each asset type are described in its individual DMP (Table 1). *Northern California Shelf Glider* data are received and automatically processed by OSU using standard procedures described in the individual DMP. The processed deployment file is then shared with CeNCOOS, where it is automatically transformed into the Glider DAC NetCDF template and pushed to the Glider DAC. *Victoria Clipper Ferry:* Data are transmitted daily from the ferry vessel to a Digital Ocean cloud server for data processing and dissemination.

Quality Control (QC). Details for each asset type are described in its individual DMP (Table 1, section I.3). All QC procedures are currently applied by the providers. *Northern California Shelf Glider:* The data processing and quality-control procedures for profile data are defined in the "Seaglider Quality Control Manual" document, Appendix D. *Victoria Clipper Ferry:* Data from the sensors are run through a set of automated and manual (staff-driven) quality control procedures based on criteria developed from QARTOD guidance, sensor manufacturer recommendations, in conjunction with location specific and climatology tests to ensure that measurements meet and pass these tests.

Data Access and Sharing. *Glider data:* Archived and near real-time *Northern California Shelf Glider* and archived *SW WA Glider* are available via the Glider DAC, making these data readily available via web services and download capabilities provided by the National DAC. The DAC also makes the data available to the WMO GTS for use in national forecast capabilities. In addition, NANOOS serves glider data files submitted to the Glider DAC for direct distribution from the NANOOS ERDDAP server. NANOOS has developed user-friendly NVS glider applications for data visualization and access to *SW WA Glider* (http://nvs.nanoos.org/GliderLaPush) and *Northern California Shelf Glider* (http://nvs.nanoos.org/GliderTrinidadHead) observations. *Victoria Clipper Ferry* data are openly accessible from the Cloud, as Raw and NetCDF files, and from a THREDDS server for "Level 1" processed data; details are found in the individual DMP (Table 1). A subset of sensors is also available on NVS.

Archiving. Glider data from active deployments (*Northern California Shelf Glider*) is submitted to the Glider DAC in near-real-time by CeNCOOS, and from there it is archived at NCEI. *SW WA Glider* have been submitted to the Glider DAC, which in turn archived it at NCEI. *Victoria Clipper Ferry* data are currently not archived nationally but are preserved by the provider on a cloud service. Discussions between NCEI, the provider and NANOOS will be required to lay out appropriate options for long-term, national archiving of this trajectory data. We anticipate developing automated, monthly submission of NetCDF trajectory data adhering to the NCEI 2.0 Template.

* Beach and Shoreline Observations

NANOOS collects information on the morphodynamics of beaches along the coasts of OR and WA in order to document the seasonal, interannual, and long-term changes taking place at multiple beach study sites and at a range of spatial scales, in support of hazard mitigation. Data collected include beach profiles, and topographic, shoreline and bathymetric mapping, generated from three programs: *Oregon (OR) Shoreline Observations*, *Washington (WA) Shoreline Observations*, and *Nearshore Bathymetry*.

Data Ingestion and Management. Details for each asset type are described in its individual DMP (Table 1). For all data programs, field survey data are collected using Real-Time Kinematic Differential Global Positioning System (RTK-DGPS) and downloaded back in the program lead's office, where post-processing and QA/QC procedures are applied and the data files are managed as described in the DMP's. For *OR & WA Shoreline Observations a c*onsistent set of beach profile plots are created and distributed for wider use.

Quality Control (QC). Details for each asset type are described in its individual DMP (Table 1, section I.3). *OR & WA Shoreline Observations* use nearly identical QA/QC procedures, developed collaboratively, to ensure locational and elevation accuracy, remove spurious elevation values, and characterize uncertainties. Similar approaches are used for the *Nearshore Bathymetry* data. QARTOD manuals do not currently exist for these data types.

Data Access and Sharing. *OR & WA Shoreline Observations*: Typically, data are made available after they have been collected with a lag of approximately 1-2 weeks for OR and 1-2 months for WA. The data are disseminated as plots via the NVS Beach and Shoreline Changes App, see http://nvs.nanoos.org/BeachMapping. Data files are available upon request as Excel and matlab files for OR, and ASCII files for WA. *Nearshore Bathymetry* data are available upon request. Efforts are underway to distribute these data sets via geospatial data formats and OGC web services provided by the NANOOS GeoServer.

Archiving. *OR & WA Shoreline Observations* data are archived at their respective state agencies following state data preservation directives. *Nearshore Bathymetry* data are currently stored with backups at OSU servers. The providers and NANOOS are currently working with the IOOS program office and NCEI on the final details related to archiving Oregon shoreline data at NCEI. It is anticipated that the initial archive submission of these data will take place during the first quarter of 2021 and that additional data will be archived after each collection campaign. Once a final process is in place for archiving Oregon shoreline data at NCEI, NANOOS will leverage this work to develop the process for archiving Washington shoreline data at NCEI.

E.2 External Data Streams

The NANOOS DAC also integrates external data streams (not funded by NANOOS) from other regional assets or large-scale observation systems, and federal and Canadian assets. Data from US federal or Canadian federal (currently Environment Canada) sources are not discussed further, as they are managed by their respective agencies. Remaining observational data streams are **primarily fixed-location in-situ assets** similar to the NANOOS Fixed-location Sensor Platforms described in Section E.1. They include shelf and estuarine moorings, benthic cabled platforms, sensors attached to piers or other coastal fixed structures, sensors on pumped water intakes at shellfish hatcheries and other facilities. Currently active providers are described in Table 2. The range of variables measured is similar to those from NANOOS Fixed-location Sensor Platforms.

Data Ingestion and Management. These providers range from small groups managing a single platform to ONC and OOI, which are large, long-term observation systems with highly established and documented data and asset management procedures and highly capable data distribution systems. Metadata and near real-time data streams from these sources are integrated into the NANOOS DAC using the same procedures as for NANOOS Fixed-location Sensor Platforms as described in Section E.1. As with those assets, data are provided in heterogeneous formats, data access mechanisms, and encodings. Data from these external sources are currently integrated only in a "pass through" basis where the NVS data store retains the most recent 60 days. This limitation may be lifted on a case-by-case basis in the future based on data assessments, partnership strengthening, and increasing NANOOS capabilities.

Several providers, particularly those in the shellfish aquaculture community and ocean acidification (OA) concerns, maintain their sensors and data in collaboration with NANOOS partners (Newton et al., 2012); that community in particular has developed common, shared approaches for data management around the Burke-o-lator OA sensor and related sensors, based on collaborations between industry and academic and federal scientists. In addition, NANOOS DMAC provides data management guidance to external partners, based on requests, and needs.

Quality Control (QC). Several providers already apply robust and well documented QA and QC procedures. King County is implementing QC adapted from QARTOD. ONC and OOI have their own robust, published QC procedures. Providers using Burke-o-lator sensors share and continue to develop common best practices developed by that collaborative community. Public QC documentation for a substantial subset of providers is available as follows (refer to Table 2 for provider codes):

- *Hakai*: https://data.hakai.org/sensor-network
- *King County*: http://www.kingcounty.gov/services/environment/water-and-land/puget-sound-marine/marine-mooring/data%20quality%20control.aspx
- *ONC*: http://www.oceannetworks.ca/data-tools/data-quality
- *OOI*: http://oceanobservatories.org/quality-control/

- *WADOH*: https://www.doh.wa.gov/CommunityandEnvironment/Shellfish/CommercialShellfish/VibrioControlPlan/TemperatureData

In addition, the NANOOS DAC applies to external-provider data streams the same set of base-level harmonization and QC checks described in the Quality Control sub-section above under E.1, *Fixed-location Sensor Platforms.* Once NANOOS DMAC has developed and operationalized QARTOD checking and flagging capabilities (see Section E.1, *Fixed-location Sensor Platforms*), we will apply QARTOD flagging for variables with published QARTOD manuals to all near real-time data streams, both internal and external.

Table 2. External, fixed-location in-situ data streams other than those from federal or Canadian federal agencies. These assets are integrated into the NANOOS DAC and NVS.

Provider Information						
Code	Name & URL	Type	Contact Name & Email	Asset Count	52N SOS	Notes
Hakai Institute	Hakai Institute	Academic	Wiley Evans, wiley.evans@hakai.org	2	X	Canadian. Burke-o-lator (OA)
HMSC	Hatfield Marine Science Center	Non-Profit	Dann Cutter, Dann.Cutter@oregonstate.edu	1	X	Currently Offline
King County	King County	County	Stephanie Jaeger, Stephanie.Jaeger@kingcounty.gov	4	X	Implementing QARTOD-based QC flagging.
NWIC	Northwest Indian College	Academic	Beth Curry, beth4cu@uw.edu	1	X	Close partnership with NANOOS, UW
ONC	Ocean Networks Canada	Academic	Mike Morley, mmorley@uvic.ca	4		Canadian. Large, long-term observation system.
OOI	Ocean Observatories Initiative	Academic	Edward Dever, edward.dever@oregonstate.edu	6		Large, long-term observation system. Using Endurance Array platforms.
PennCoveShellfish	Penn Cove Shellfish	Industry	Jim Nagel, jim@penncoveshellfish.com	2	X	
PSI	Pacific Shellfish Institute	Non-Profit	Andy Suhrbier, suhrbier@pacshell.org	2	X	Includes one Burke-o-lator (OA). Close partnership with NANOOS.
TaylorShellfish	Taylor Shellfish	Industry	Benoit Eudeline, BenoitE@taylorshellfish.com	1	X	Burke-o-lator (OA). Close partnership with NANOOS.
WADOH	Washington Department of Health	State	Clara Hard, clara.hard@doh.wa.gov	18	X	Seasonal network (late Spring to early Fall).
WhiskeyCrShelfish	Whiskey Creek Shellfish Hatchery	Industry	Alan Barton, alan_barton22@yahoo.com	1	X	Burke-o-lator (OA). Close partnership with NANOOS.

All assets are in the NVS Metadata Store and all provide data+graphic access on NVS. OA: Ocean Acidification, where the Burke-o-lator is a specialized OA sensor.

Data Access and Sharing. All data streams listed in Table 2 are publicly available via NVS for graphical browsing and data downloads, and most are also ingested into the NANOOS 52N SOS service for distribution and national integration via standard IOOS web services. NWIC data is also redistributed to NDBC / WMO GTS.

In addition, several providers maintain their own robust, open data distribution mechanisms, including Hakai Institute, King County, ONC and OOI.

Archiving. ONC and OOI have their own published, long-term data preservation plans and mechanisms. The other external providers are not currently archiving data with NCEI or a national archive. NANOOS DMAC plans to engage these partners to discuss such archiving, once the mechanisms for regular NCEI archiving of all NANOOS internal assets have been put in place.

F. Web Portal and User Applications

The **NANOOS web portal** (http://www.nanoos.org) provides access to a wide range of resources about NANOOS, its members, IOOS and other organizations. The Products page (http://www.nanoos.org/products/products.php) provides user-friendly access to relevant web resources, user applications and products from both NANOOS and other organizations.

Figure 2. NVS (V6.3) Apps at http://nvs.nanoos.org.

NVS (http://nvs.nanoos.org) serves as the primary data browsing, visualization and access application provided by NANOOS to its user community (Risien et al., 2009; Risien et al., 2019). In order to better serve a wide range of users, NVS provides customized "Apps" targeted to specific user communities, but sharing a common user interface for enhanced usability and typically drawing from the same common data stores. Custom Apps include the *Tsunami Evacuation Zones App* (see Martin et al., 2011), *Boaters App*, *Tuna Fishers App*, *Climatology App*, and others (Figure 2). The NVS infrastructure has been generalized into a software framework labeled "vizer" that has been adapted to other applications such as the thematic ocean acidification IPACOA and GOA-ON data portals described earlier (Section A).

The NVS *Data Explorer App* (Figure 3) provides NVS' richest and most comprehensive – but often the most complex – access to data assets integrated by NANOOS, including site time series, depth profile section plots, glider section visualizations, site comparisons between observed and modeled variables, map overlays (rendered as map tiles) from model forecasts and gridded data products such as remote sensing, and climatological, cyclical map overlays; these are accessible from an interactive, central "timeline" selector tool.

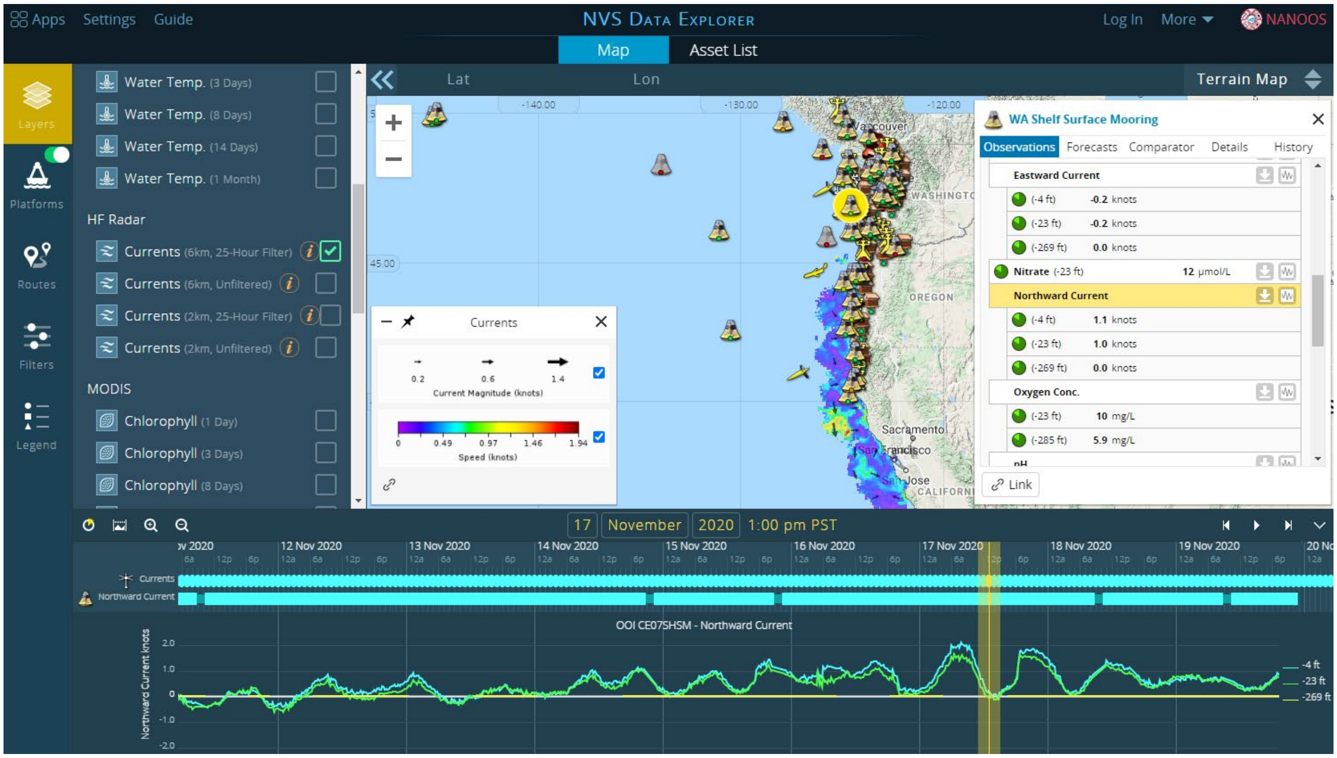

Figure 3. NVS *Data Explorer App* (http://nvs.nanoos.org/Explorer) showing multi-depth salinity time series at an OOI asset that also includes meteorological observations; and a surface currents map overlay from HF Radar.

The **CMOP Data Explorer with NANOOS Data** enables advanced users to generate and combine different kinds of plots using multiple sites and variables (e.g. winds, waves, temperature) obtained via web service access from the harmonized NVS data and metadata store for fixed-location site time series. It is a collaborative effort between OHSU-CMOP and UW-APL that adapted the pre-existing "CMOP Data Explorer" tool (Baptista et al., 2015) to Pacific NW data distributed via the JSON-based, REST, light-weight web services provided by NVS. It's hosted by OHSU-CMOP at:
http://www.stccmop.org/datamart/observation_network/dataexplorer_nanoos
Figure 4 shows a sample plot from this application combining to sites and two variables over a common time period.

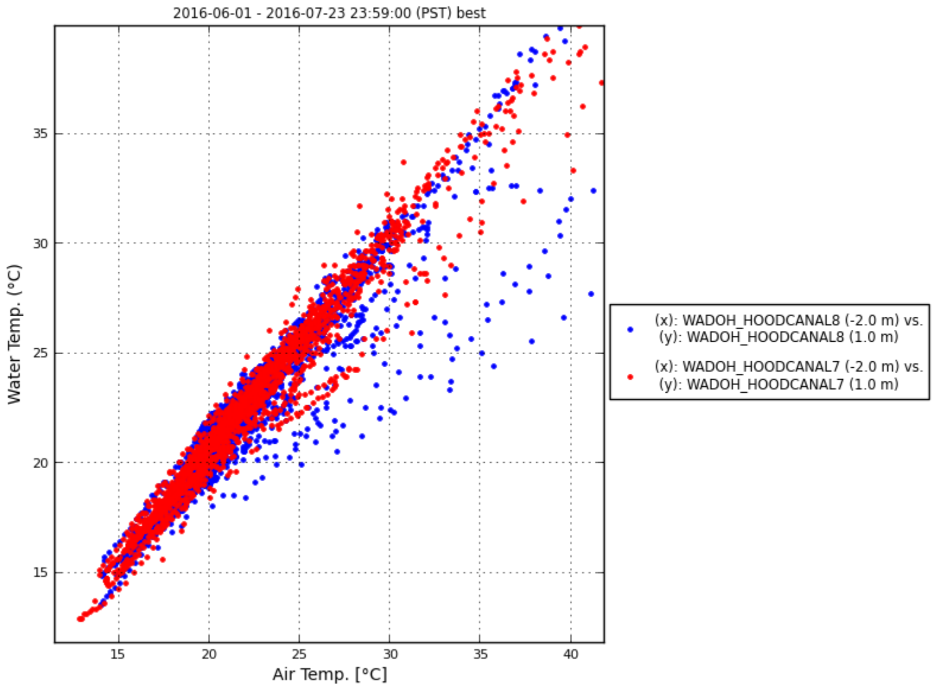

Figure 4. Sample plot from the "CMOP Data Explorer with NANOOS Data" application combining two sites (Hood Canal 7 [red] and Hood Canal 8 [blue] from WA Department of Health) and two variables (air and near-surface water temperature) over a common time period (2016-06-01 to 2016-07-23).

Finally, NANOOS hosts other online user products that leverage NVS components or other NANOOS data integration activities. Prominent examples include **J-SCOPE**, the JISAO Seasonal Coastal Ocean Prediction of the Ecosystem (http://www.nanoos.org/products/j-scope/); and **Real-Time HABs**, Real-time Information About Harmful Algal Blooms (http://www.nanoos.org/products/real-time_habs/).

G. References

Baptista, A.M., C. Seaton, M.P. Wilkin, S.F. Riseman, J.A. Needoba, D. Maier, P.J. Turner, T. Karna, J.E. Lopez, L. Herfort, V.M. Megler, C. Mcneil, B.C. Crump, T.D. Peterson, Y.H. Spitz & H.M. Simon. 2015. Infrastructure for collaborative science and societal applications in the Columbia River estuary. *Front. Earth Sci.* 9(4):659–682, doi:10.1007/s11707-015-0540-5

Martin, D.L., J.C. Allan, J. Newton, D.W. Jones, S. Mikulak, E. Mayorga, T. Tanner, N. Lederer, A. Sprenger, R. Blair and S.A. Uczekaj. 2011. Using web-based and social networking technologies to disseminate coastal hazard mitigation information within the Pacific Northwest component of the Integrated Ocean Observing System (IOOS). *Proc. MTS/IEEE Oceans'11*, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6107278

Mayorga, E., J. Newton & T. Tanner. 2016. Ocean Acidification monitoring data collaborations, integration and dissemination: The US Pacific NW regional IOOS experience with local to global efforts. *2017 Ocean Sciences Meeting*, New Orleans, LA, 23 Feb.

Mayorga, E., T. Tanner, R. Blair, A.V. Jaramillo, N. Lederer, C.M. Risien and C. Seaton. 2010. The NANOOS Visualization System (NVS): Lessons learned in data aggregation, management and reuse, for a user application. *Proc. MTS/IEEE Oceans'10,* doi:10.1109/OCEANS.2010.5663792

Newton, J., D. Martin, E. Mayorga, A. Devol, R. Feely, S. Alin, B. Dewey, B. Eudeline, A. Barton, A. Suhrbier, A. Baptista and J. Needoba. 2012. NANOOS partnerships for assessing ocean acidification in the Pacific Northwest. *Proc. MTS/IEEE Oceans'12,* doi:10.1109/OCEANS.2012.6405086

Risien, C.M., J.C. Allan, R. Blair, A.V. Jaramillo, D. Jones, P.M. Kosro, D. Martin, E. Mayorga, J.A. Newton, T. Tanner, and S.A. Uczekaj, 2009. The NANOOS Visualization System: Aggregating, displaying and serving data. *Proc. MTS/IEEE Oceans'09,* doi: 10.23919/OCEANS.2009.5422325.

Risien, C.M., J.A. Newton, T. Tanner, P.M. Kosro, E. Mayorga, R. Wold, J.C. Allan, and C. Seaton, 2019. The NANOOS Visualization System (NVS): A Decade of Development and Progress Addressing Stakeholder Needs. *Proc. MTS/IEEE Oceans'19,* doi: 10.23919/OCEANS40490.2019.8962588.

Seaton, C.M., M. Wilkin and António M Baptista. 2018. OD24C-2735: Implementation of a Standardized Real-time Data Quality Assurance System for the Columbia River Estuary. *2018 Ocean Sciences Meeting*, Portland, OR, 13 Feb. https://agu.confex.com/agu/os18/meetingapp.cgi/Paper/322768