

Standard Operating Procedure (SOP) 16: Managing the Production Environment version 3

Glacier Bay National Park and Preserve Oceanographic Monitoring Protocol

Change History

Previous Version #	Revision Date	Revised By	Changes	Justification	New Version #
2	February 2020	B. Johnson	Complete rewrite.	Replaced core information architecture.	3
1	September 2014	B. Johnson	Minor corrections.	Revised monolithic protocol OC-2014.1	2
-		B. Johnson	Initial version	For original protocol	1

Please cite this Standard Operating Procedure as:

Johnson, W. 2020. Standard Operating Procedure (SOP) 16: Managing the Production Environment version 3 Glacier Bay National Park and Preserve Oceanographic Monitoring Protocol. National Park Service, Juneau, Alaska. Available online at <https://irma.nps.gov/DataStore/Reference/Profile/2272159>.

NOTE

This version of SOP 16 is considerably different from its predecessors. In 2018 the SEAN authoritative repository and its associated websites had to be abandoned. Most of the vital sign deliverables were moved to the NPS Data Store repository which now serves as the authoritative source. Some detail is gone into in order to clarify the new workflow required to support the new data architecture.

Summary

The output of SEAN vital sign monitoring programs is always in the form of a very tightly defined set of formal deliverables. For the formal deliverables to be disseminated both internally and to the public they must be installed in the NPS Data Store repository at NRSS. (A version of one of the Data Store deliverables is also installed in the NOAA NCEI/ NODC World Ocean Database repository). In order for this production content to be generated, various tasks must be performed in SEAN's data management environment. Once content is built, validated, and "accepted" (referred to as "certified" before 2018) in the management environment, it then gets copied to Data Store for permanent storage and dissemination.

Most of the detailed SOPs in the oceanographic monitoring protocol end with a reference to propagating the final deliverable into Production. This is an implicit reference to this SOP.

Schematic of the environment

Figure SOP 16.1 illustrates the major components used to collect, build, quality check, store and disseminate the program's formal deliverables. The physical location of components is included in order to make clear who is responsible for their maintenance. Each object is color-coded to clarify its position in SEAN's four-phase model specified in the network's Data Management Plan (<https://irma.nps.gov/DataStore/Reference/Profile/660529>).

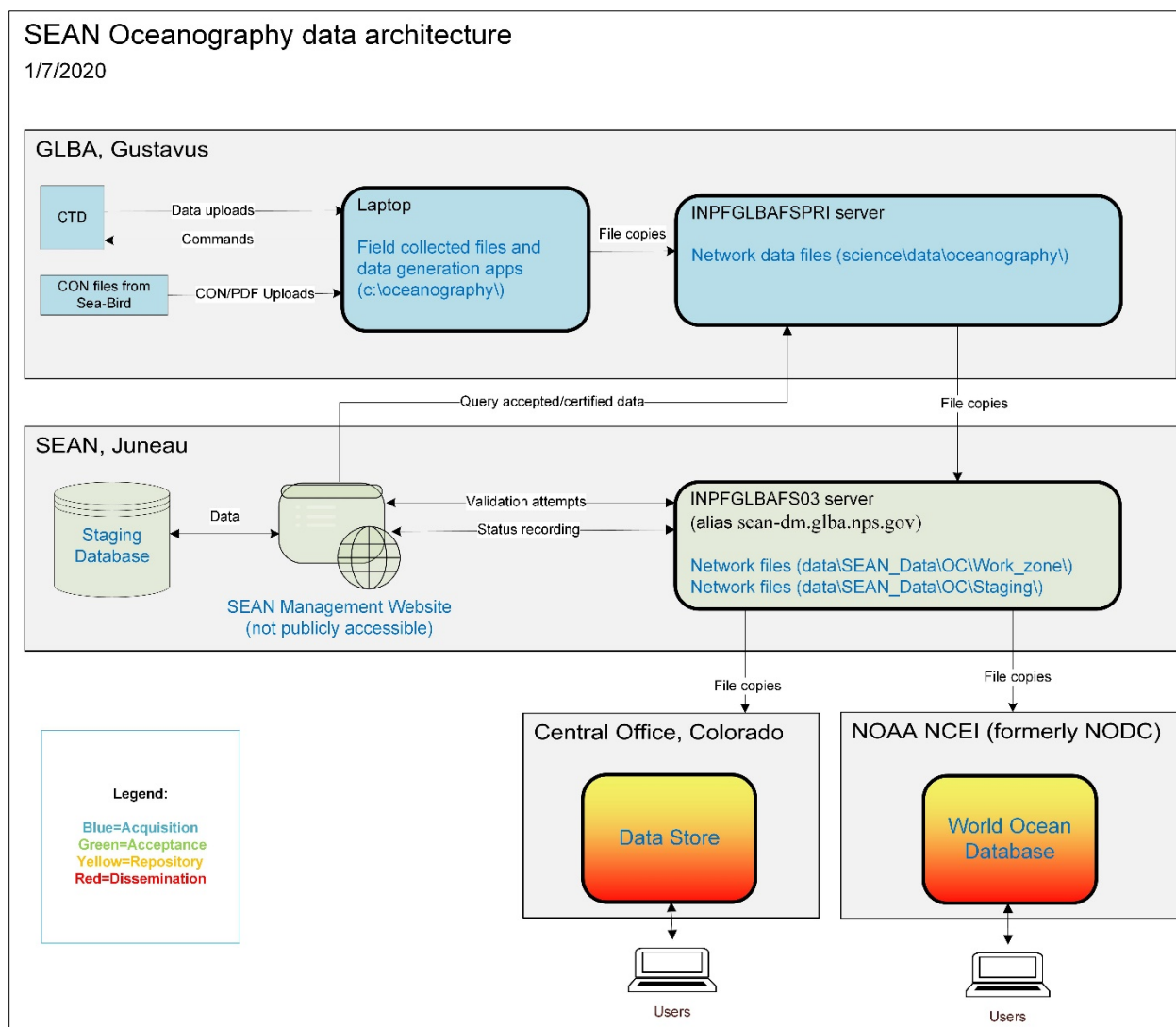


Figure SOP 16.1. Components of the data management environment are illustrated in schematic form. Gray rectangles represent distinct physical locations.

Individual components of the data management environment

Acquisition

All physical oceanography data are collected at Glacier Bay National Park. A designated Park Lead has ultimate responsibility for this. Data come from two sources. “CON” files are delivered to the park by Sea-Bird Scientific corporation and reflect the current CTD instrument configuration as well as their calibration parameters. CON files are typically generated biennially for each CTD. A CTD is typically put into service every other cruise-year. Along with the CON file, Sea-Bird also provides PDFs of calibration certificates for each instrument.

The second source of data are the water column profile data streams collected by the CTD itself. A field laptop computer periodically gets connected to the CTD. An operator issues commands to the CTD and it returns raw data consisting of strings of hexadecimal numbers to the laptop in the form of one “HEX” file for each cast. The laptop houses data processing applications that generate meaningful “CNV” files through signal-processing transformation of the CON and HEX files. Draft CNV files are used during cruises to verify proper functioning of instruments. At the end of the cruise year a complete set of CNVs is generated by the laptop and copied to a local file server at Glacier Bay for formal review. The underlying CON and HEX files are also preserved on the file server before being forwarded for acceptance, archiving and dissemination.

Acceptance

The data manager is responsible for all quality control tasks up to final acceptance. Candidate deliverables are copied from GLBA to the “staging” area of a Juneau file server. They are then subjected to formal validation as described in the protocol document and associated deliverable-based SOPs. For a particular deliverable this may require either an automated web application housed on the SEAN Management Website or a manual review by the data manager. The process of submission by the project lead and validation by the data manager is typically iterative, terminating when the data manager finds no “mandatory errors” and the project lead determines any remaining “optional errors” cannot be resolved. The Park Lead then provides an acceptance via email to the Data Manager and the deliverable is moved on to permanent repository storage.

An audit trail of every submission of a deliverable is maintained in the staging directory. The submissions and validation results are tracked in the monolithic SEAN Staging database which holds a status table that records the outcome of each staged submission.

The Staging database also holds a cumulative table of all accepted oceanographic water column profile records. This serves to inform other SEAN deliverables, such as annual reports, and is also the source for the NOAA World Ocean Database submission.

The Staging database currently uses Microsoft SQL Server 2017 Standard Edition to hold the OC_D product and the submission status tables. It currently resides on INPGLBAFS03 and its connection string name is SEAN_Staging_2017. The oceanography objects reside in schema OC. The status objects reside in schema SEAN.

Repository

The authoritative repository is NPS Data Store. The root of the SEAN Data Store structure is currently at <https://irma.nps.gov/DataStore/Reference/Profile/2170061> and Oceanography is linked to it as a Data Store “project”.

Data Store already contains a structure of “reference” entries for all formal deliverables. A reference contains detailed metadata regarding the deliverable as well as any attached file content (called “holdings” in Data Store parlance). Adding a deliverable to Data Store for a new cruise year typically involves uploading a file covering the period as well as revising the reference metadata to reflect the new scope of the deliverable. But some deliverables may instead require creating a separate new reference through Data Store’s “versioning” facility and weaving it into the SEAN structure. Also, NRR and NRDS reports that are published at the central office have their own Data Store references defined by the publications section. To incorporate them into the SEAN structure, special “assemblage” references have been defined that accept links to the central office publications. As an example, see <https://irma.nps.gov/DataStore/Reference/Profile/2257810>.

The NOAA World Ocean Database is populated using special methods defined in SOP 17. It is cumulative and based on the accepted Staging database content. Since this repository is not under direct control of National Park Service, its content is not authoritative. Only Data Store is the authoritative source.

Dissemination

Both repositories involved maintain their own web dissemination facilities. It is no longer necessary for the SEAN to maintain separate web dissemination functions. We have it in writing from the GLBA Superintendent that oceanography deliverables are not sensitive, so all associated Data Store references and holdings have been made open to public and are also replicated on DATA.GOV.