

Southeast Coast Network

National Park Service
Inventory and Monitoring Division



Standard Operating Procedure 2.1.08

Dataset and Metadata Publication—Version 2.0

Please cite this as:

Southeast Coast Network. 2016. Dataset and Metadata Publication—Version 2.0. Southeast Coast Network Standard Operating Procedure NPS/SECN/SOP—2.1.08. Southeast Coast Network, Athens, Georgia.

Summary

This standard operating procedure (SOP) presents the procedure for publishing data and the accompanying metadata for Southeast Coast Network (SECN) monitoring protocols and projects.

Revision Log

Previous Version #	Revision Date	Author	Changes Made	Reason for Change	New Version #
1.0	12/23/2016	Editor	Editorial review	Protocol publication	2.0

Overview

The information obtained through the SECN monitoring program has a wealth of applications, including management decision-making, research, education, and promotion of public understanding of SECN park resources. Park managers are the primary audience for the results of vital signs monitoring. The Southeast Coast Network seeks to provide superintendents and resource managers with the data they need to make and defend management decisions and to collaborate with others for the benefit of park resources. Other key audiences for monitoring data include park planners, interpreters, researchers and other scientific collaborators, and the general public. To be effective, monitoring data must be analyzed and provided at regular intervals to end-users in a usable format.

This SOP presents the procedure for publishing data and metadata for SECN monitoring protocols and projects.

Scope

This document pertains to the publication of tabular data and spatial (i.e., geographic information systems [GIS]) data collected as part of standard, long-term monitoring or short-term studies. This document also details the publication of data photos (e.g., photopoints collected as part of the vegetation community monitoring and stream habitat protocols).

This document provides information regarding the standards, workflow and tools used to generate metadata for the publication of data in IRMA (National Park Service Integrated Resource Management Application— <https://irma.nps.gov>). Data availability in IRMA varies with the level of sensitivity assigned to the dataset.

Data Sensitivity

Network staff, park staff, and any cooperating (non-NPS) staff will classify sensitive data. Network staff work closely with park resource management staff to identify potentially sensitive park resources and ensure both tracking of information about these resources throughout the data / product lifecycle, and removal of potentially sensitive information from documents and products that will be released outside the network.

For some contractors (e.g., state employees), state sunshine laws may apply and should be handled explicitly in the cooperative agreement. Language about protecting all data from release is a good idea for all cooperators and contractors.

Network staff manage access to sensitive data handled by the network. All potentially sensitive park resources are identified and cooperators working on behalf of the Southeast Coast Network are informed that:

- Network staff must review all data and associated information prior to release in any format.

- Any information classified as protected should not be released in any form, except as approved in advance by the National Park Service.

Sensitive park resources potentially relevant to each protocol are identified through a cooperative effort by network and park staff. When preparing reports and associated data sets, SECN staff ensure that all protected information is properly identified and marked. All references to protected information are removed or obscured in any reports, publications, maps, or other public forum. Specific approaches to protecting data should be worked out with the park(s) involved.

The main distribution applications and repositories developed by the I&M Program are maintained on both secure and public servers and all records marked “sensitive” during uploading are only available on the secure servers. Thus, access to data on sensitive park resources can be limited to only the owner(s) of record.

Documentation (Metadata) Standards

Data documentation ensures that data sets are usable for their intended purposes into the future. This involves the development of metadata, including structured information about the content, quality, condition, and characteristics of a data set. In addition to spatial information, metadata include information about data format, collection and analysis methods, time of collection, originator, access/use constraints, and distribution. Metadata also provide the means to catalog data sets within information-management systems, making their associated data sets available to a broad range of potential users. While most frequently developed for geospatial data, metadata describing non-geospatial (e.g., tabular) data sets are also needed.

The following procedures and specifications were developed for the SECN Inventory & Monitoring (I&M) Program to support the creation of metadata documenting data sets generated by the I&M division. These procedures and specifications should also be used for data sets not collected by the I&M division but used for analysis or provided to parks (e.g., weather, surface water, eBird) solely as reference information. These guidelines outline the general approach for generating metadata that complies with the Federal Geographic Data Committee (FGDC) standards. Metadata that meet these standards accompanies all datasets published by the SECN:

Federal Geographic Data Committee Standards

- Content Standard for Digital Geospatial Metadata (FGDC-CSDGM):
 - http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf
 - <https://www.fgdc.gov/metadata/geospatial-metadata-standards>
- Biological data profile:
 - <http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/biometadata/biodatap.pdf>

National Park Service Standards

- NPS Digital Photo Metadata Standard (2008 version)
https://science.nature.nps.gov/im/datamgmt/meetings/dm_mtg.../PhotoMetadata.ppt
- NPS Geospatial Metadata Standards
 - See Federal Geographic Data Committee standards above.
- Metadata and IRMA
 - See Section 9.5 in this document:
<https://irma.nps.gov/content/help/datastore/docs/ReferenceHelp.pdf>

Documentation (Metadata) Tools

Datasets (Tabular Data)

The Southeast Coast Network developed an integrated data management system, which encompasses a database back end in Microsoft SQL server and several front-end applications to enhance end-user interactions with the database. All long-term monitoring data are entered into the SECN data management system, and undergo rigorous, protocol-specific, quality assurance procedures, which are described in the data management sections of the individual protocols. These quality-assurance procedures ensure that data are complete and accurate prior to their use in analysis or distribution.

Following data certification, data are exported to a Microsoft Excel spreadsheet and an identical ASCII text file, .csv file or Microsoft Access database, as a stand-alone snapshot of data at one park for one field season or all parks for one field season. A dataset-specific metadata file documents each data file. The SECN Metadata Editor (Figure 1), a stand-alone tool linked to the SECN data management system, produces FGDC CSDGM formatted metadata as an .xml file that contains all mandatory metadata elements from the FGDC CSDGM standard. The data and metadata file are then zipped together for upload into the IRMA record for the appropriate annual data summary report. Refer to Appendix 1 for an example .xml metadata file published in IRMA.

Note:

- The Southeast Coast Network posts the dataset in the same record as the annual data summary report based on the posted data.
- The Southeast Coast Network decides on a case-by-case basis whether to publish the location data for sampling locations in a usable form (e.g., coordinate data or GIS files) in order to protect the site, permanent plots, and equipment.
- The Southeast Coast Network does not publish data obtained from other agencies (e.g., groundwater, stream flow, tides, air quality and weather). These data are already available online from the data collection agency. The Southeast Coast Network provides direct links to these datasets (in published reports and IRMA) whenever possible.

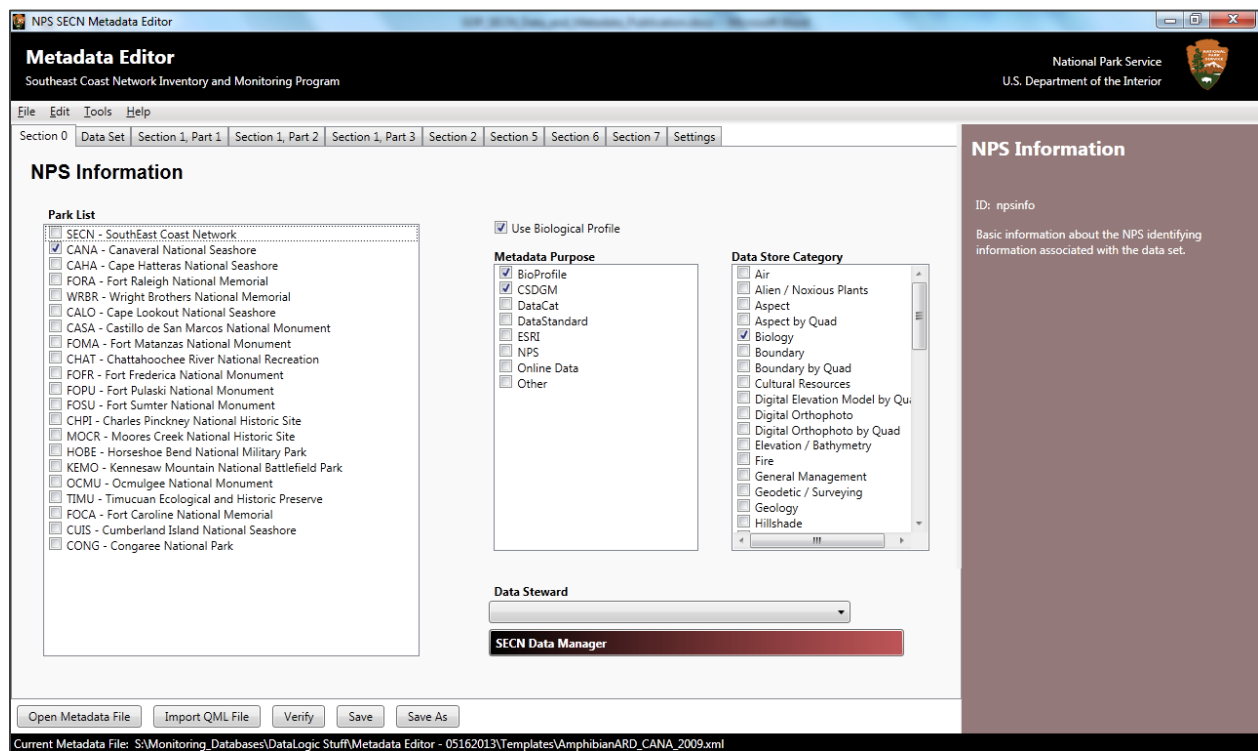


Figure 1. Screenshot of the SECN Metadata Editor.

GIS data

All GIS data (e.g. shapefiles, coverages, imagery) posted to IRMA and other online repositories are accompanied by FGDC compliant metadata generated with ESRI's metadata tools in ArcGIS. Extensive guidance exists regarding the generation of metadata for GIS data. The following documents provide detailed instructions for generating compliant metadata which meets Federal and NPS standards:

National Park Service. 2012. How to create metadata in ArcGIS 10.0. Intermountain Region GIS Program, National Park Service. Denver, Colorado. (Available at <http://imgis.nps.gov/documents/Metadata%20in%20ArcGIS%2010.0%20Ver1%2003222012.pdf>)

National Park Service. 2012. How to create metadata in ArcGIS 10.0. Appendices. Intermountain Region GIS Program, National Park Service. Denver, Colorado. (Available at <http://imgis.nps.gov/documents/Metadata%20in%20ArcGIS%2010.0%20Appendices%20Ver1%2003222012.pdf>)

GIS files are zipped together with the accompanying metadata file to ensure that the metadata travels with the GIS file when downloaded from IRMA or other online repositories. All mandatory fields are populated when generating metadata for GIS data. GIS files are posted as individual records in IRMA (i.e., Geospatial Dataset), with links provided in the corresponding data summary report or as a collection in IRMA to tie the records together for easier discovery.

Photographs

Southeast Coast Network protocols generate a large number of photographs, and documenting these photos can be quite onerous. To streamline the documentation of digital photographs, the National Park Service acquired an NPS-specific version of GPS-Photo Link to support the creation of metadata that meets the NPS photo metadata standard. NPS Photo Link supports embedding the full NPS Digital Photo Metadata Standard element set in the photo metadata header through the manual entry of metadata elements or through importing information from a comma-delimited spreadsheet (see Figure 2).

The National Park Service digital photo metadata standards are addressed in SOP 2.1.07 (SECN 2016).

All photos published to IRMA, NPS Focus, or other online repositories have the full NPS Digital Photo Metadata Standard element set embedded in the photo metadata header. End-users can view Photo Exif headers with GPS-Photo Link, a wide variety of free Exif viewers, or some commercial software programs that display the entire Exif header. Data photographs are published as a collection that includes all of the data photos for one field season. Data reports in IRMA link to corresponding photos and metadata for easier discovery.

Water Quality Data

The water quality component of the Natural Resource Challenge (NRC) requires that Vital Signs Monitoring networks archive all physical, chemical and biological water quality data collected with NRC funds in the National Park Service's STORET database, maintained by the NPS Water Resources Division (WRD). To facilitate archiving in STORET, the Water Resources Division developed a series of Microsoft Access-based templates (NPSTORET), in which networks enter their water quality data in a STORET compatible format. Vital Signs networks send data from NPSTORET or their own data system (following NPS developed specifications) to WRD annually for inclusion in EPA STORET.

NPSTORET includes a module for importing station information and results from other user databases, labs, or other sources. NPSTORET also provides an extensive metadata template to support the station information and monitoring data uploaded into NPSTORET. The metadata template describes in detail the protocols used to collect, store, transport, and analyze data. The metadata template also describes the parameters for which data are collected and analyzed—matching the characteristic definitions used by EPA STORET. This information should be readily available in the network's long-term water quality monitoring protocol. For more information on using NPStoret to documenting water quality data, please refer to the following standard operating procedures:

Wright, C. J. 2013. Exporting coastal assessment data to STORET. Southeast Coast Network Standard Operating Procedure NPS/SECN/SOP—2.2.02. Southeast Coast Network, Athens, Georgia.

Wright, C. J. 2012. Publication of fixed-station water quality monitoring data to STORET. Southeast Coast Network Standard Operating Procedure NPS/SECN/SOP—2.2.06. Southeast Coast Network, Athens, Georgia.

Data Sets for Publication

Table 1 provides an overview of the typical data sets published as a result of each long-term monitoring protocol. Table 1 also lists the methodology used for data collection and the repository for the data. Repositories for SECN data include the following:

- IRMA: <https://irma.nps.gov>
- STORET: <http://www.epa.gov/storet/>
- Aquarius: <https://irma.nps.gov/aqwebportal/>

Table 1. SECN datasets requiring metadata generation and publication, listed by long-term monitoring protocol, including the standard online repository for each dataset.

Protocol	Method	Data Type(s)	Repository
Amphibian	Automated Recording Device	Dataset	IRMA
	Visual Encounter Survey	Dataset	IRMA
Landbird	Variable Circular Plot	Dataset	IRMA
	Automated Recording Device	Dataset	IRMA
Coastal Water Quality	Fixed-station	Dataset	STORET, Aquarius
	Coastal Assessment	Dataset	STORET
Saltmarsh Elevation	Accretion Measurements	Dataset	IRMA
	Rod Surface Elevation	Dataset	IRMA
	Soil Porewater Salinity	Dataset	IRMA
Shoreline Change	Shoreline	GIS data	IRMA
	Beach Topography	GIS data	IRMA
Stream Habitat	Reach Assessment	Dataset, Photos	IRMA
	Transect Assessment	Dataset, Photos	IRMA
Vegetation Communities	Groundcover Cover & Condition	Dataset, Photos	IRMA
	Shrub Cover	Dataset, Photos	IRMA
	Canopy Cover	Dataset, Photos	IRMA
	Tree DBH	Dataset	IRMA
	Dead, Downed Materials	Dataset	IRMA
Water Level	Fixed Station	Dataset	Aquarius

Data Dissemination

The SECN data dissemination strategy ensures that:

- Data are easily discoverable and obtainable.
- Only certified data (data subject to complete quality assurance procedures) are released unless necessary to respond to limited/specific data requests (e.g., within-network requests and FOIA).
- Distributed data include appropriate documentation (metadata).
- Sensitive data are identified and protected from unauthorized access and inappropriate use.

SECN data and information products (e.g., final deliverables, periodic milestones) are “packaged,” to the maximum degree possible, and made available for distribution as a complete set. This “package” contains, as applicable: a snapshot of data; metadata and/or documentation; and reports, synthesis documents, and/or protocol documents. In general, information is available to two primary audiences: public and NPS employees, as determined by data sensitivity and development status. Only fully documented, certified, non-sensitive data and data products are uploaded to public distribution repositories (<https://irma.nps.gov>) or otherwise released to the public.

Literature Cited

Federal Geographic Data Committee. FGDC-STD-001-1998. Content standard for digital geospatial metadata (revised June 1998). Federal Geographic Data Committee. Washington, D.C. Available at: <http://www.fgdc.gov/metadata/csdgm/> (last accessed December 2016).

National Park Service. 2012. How to create metadata in ArcGIS 10.0. Appendices. Intermountain Region GIS Program, National Park Service. Denver, Colorado. Available at <http://imgis.nps.gov/documents/Metadata%20in%20ArcGIS%2010.0%20Appendices%20Version1%2003222012.pdf> (last accessed 14 December 2015).

Wright, C. J. 2013. Exporting coastal assessment data to STORET. Southeast Coast Network Standard Operating Procedure NPS/SECN/SOP—2.2.02. Southeast Coast Network, Athens, Georgia.

Wright, C. J. 2012. Publication of fixed-station water quality monitoring data to STORET. Southeast Coast Network Standard Operating Procedure NPS/SECN/SOP—2.2.06. Southeast Coast Network, Athens, Georgia.

Wright, C. J., J. C. DeVivo, and K. Funk. 2008. Information management and archiving plan, Southeast Coast Inventory and Monitoring Network. Natural Resource Report NPS/SECN/NRR—2008/062. National Park Service, Fort Collins, Colorado.