

Southeast Coast Network

National Park Service
Inventory & Monitoring Program



Standard Operating Procedure (SOP) #2.1.8.

Dataset and Metadata Publication

Please cite this as:

Southeast Coast Network (SECN). 2013. Dataset and Metadata Publication. Southeast Coast Network Standard Operating Procedure NPS/SECN/SOP-2.1.8. National Park Service, Athens, Georgia.

Summary

This Standard Operating Procedure (SOP) presents the procedure for publishing data and the accompanying metadata for Southeast Coast Network monitoring protocols and projects.

Revision Log

Previous Version #	Revision Date	Author	Changes Made	Reason for Change	New Version #

Overview

The information obtained through the Southeast Coast Network (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. Our goal is to provide superintendents and resource managers with the data they need to make and defend management decisions and to work with others for the benefit of park resources. Other key audiences for monitoring results include park planners, interpreters, researchers and other scientific collaborators, and the general public. To be effective, monitoring data must be analyzed, interpreted, and also provided at regular intervals to end-users in a format they can use.

This Standard Operating Procedure (SOP) presents the procedure for publishing data and the accompanying metadata for Southeast Coast Network monitoring protocols and projects.

Scope

This document pertains to the publication of tabular data and geographic information systems (GIS) data collected as part of a standard, long-term monitoring protocol or short-term study. 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 is not considered a ‘cheat sheet’ used to walk the users through the use of software designed to generate metadata for the data types discussed above; this document provides information regarding the standards, work flow and tools used to generate metadata for the publication of data in IRMA (National Park Service’s Integrated Resource Management Application – <https://irma.nps.gov>). Data in IRMA are available to NPS-only, or the Public – depending on the level of sensitivity assigned to the data.

Data Sensitivity

Classification of sensitive data will be the responsibility of network staff, park staff, and any cooperating (non-NPS) staff. Network staff will classify sensitive data working closely with park resource management staff to ensure that potentially sensitive park resources are identified, that information about these resources is tracked throughout the data / product lifecycle, and that potentially sensitive information is removed from documents and products that will be released outside the Network.

Network staff are responsible for managing access to sensitive data handled by the Program. All potentially sensitive park resources will be identified and cooperators working on behalf of the SECN will be informed that:

- All data and associated information must be made available for review by network staff prior to release in any format

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

Sensitive park resources will be identified through a cooperative effort. The network and park staff should identify all potentially sensitive park resources to the program ecologists for each protocol. Reciprocally, the program ecologists must identify any known references to potentially sensitive park resources. When preparing reports and associated data sets, SECN staff will 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.

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 will only become available on the secure servers. Thus, access to data on sensitive park resources can be limited to network staff or research partners. I&M Program applications such as IRMA provide a means to flag data as “NPS Only” or “Public.” Additional notes describing data quality should be added to the records in these applications, as resources permit.

Documentation (Metadata) Standards

Data documentation is a critical step toward ensuring that data sets are usable for their intended purposes well into the future. This involves the development of metadata, which is defined as structured information about the content, quality, condition, and the definition of 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 provides 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 (tabular) data sets are also needed.

The following procedures and specifications were developed for the Southeast Coast Network (SECN) of the National Park Service (NPS) Inventory & Monitoring (I&M) Program for the creation of metadata to document I&M-generated data sets. It is recommended that these procedures and specifications also be used for other SECN data sets that are not necessarily a part of the I&M Program. These guidelines outline the general approach to generating metadata compliant with the Federal Geographic Data Committee (FGDC) standards. All datasets published by the SECN will be accompanied by metadata which meets these standards:

Federal Geographic Data Committee Standards

Content Standard for Digital Geospatial Metadata:

http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf

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

<http://npsfocus.nps.gov/docs/guide/metadata/MetadataStandards.html>

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 has developed an integrated, decision support system which encompasses a database back end in 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 decision support system, as well as undergoing rigorous, protocol-specific, quality assurance procedures which are described in the data management sections of the individual protocols. The goal of these quality assurance procedures is to ensure that data are complete and accurate prior to being used for analysis and/or distribution.

Following data certification, all data will be exported to a MS Excel spreadsheet and an identical ASCII text file, or MS Access database, as a stand-alone snapshot of data consisting of one park, one field season (or all parks for one field season). Each data file produced will be documented by a dataset-specific metadata file. Metadata is produced by the SECN Metadata Exporter (Figure 1), a stand-alone tool linked to the SECN data management system, which produces FGDC CSDGM formatted metadata. The metadata file produced is an xml file and contains all mandatory and mandatory if applicable metadata elements from the FGDC CSDGM standard. The data and metadata file are then zipped together for upload into the IRMA record for the annual data summary report, based on the dataset being uploaded. Please refer to Appendix 1 for an example xml metadata file published in IRMA.

Please Note:

- SECN will post the dataset within the same record as the annual data summary report which utilized the data being posted
- SECN will decide on a case by case basis whether or not 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.

- SECN will 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. SECN will provide direct links to these datasets (in published reports and IRMA) whenever possible.

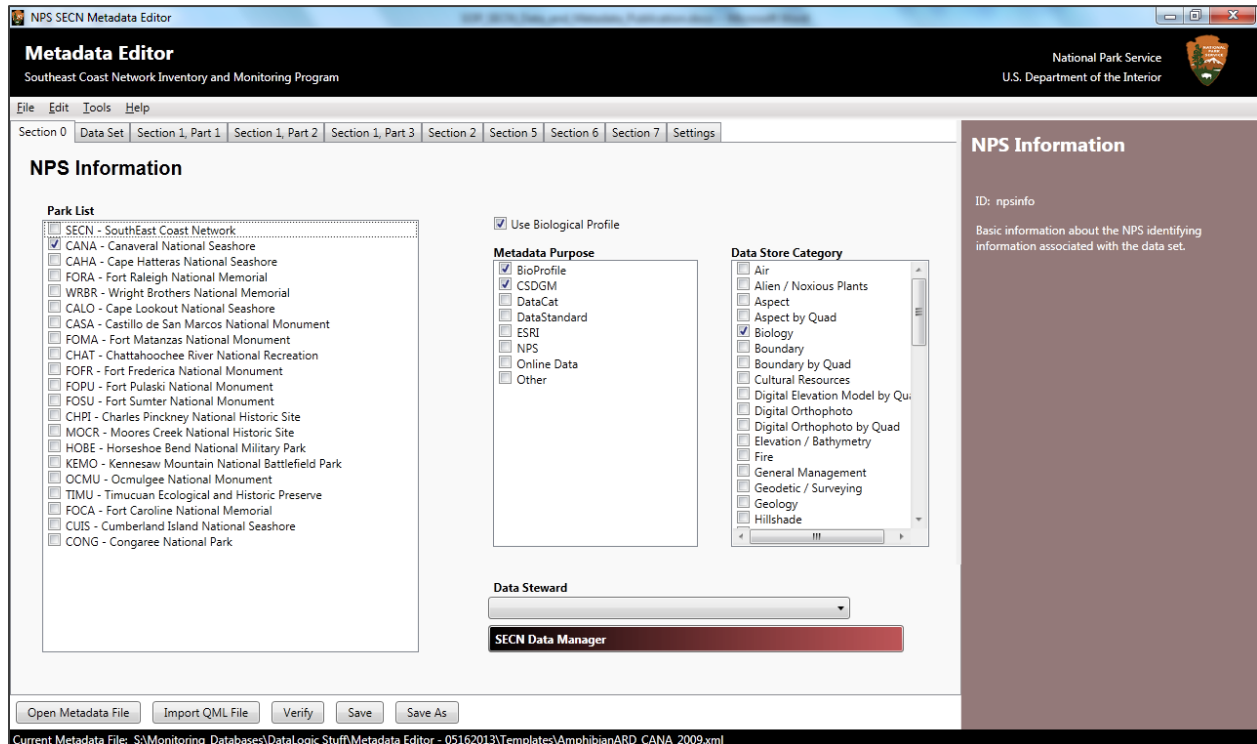


Figure 1. Screenshot of the SECN Metadata Editor Tool.

GIS data

All GIS data (e.g. shapefiles, coverages, imagery) posted to IRMA and other online repositories will be accompanied by FGDC compliant metadata generated with ESRI's metadata tools in ArcGIS. There is extensive guidance 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, CO. (Available at <http://imgis.nps.gov/#Resources>)

National Park Service. 2012. How to create metadata in ArcGIS 10.0. Appendices. Intermountain Region GIS Program, National Park Service. Denver, CO. (Available at <http://imgis.nps.gov/#Resources>)

Metadata templates are also available at <http://imgis.nps.gov/#Resources>.

GIS files will be zipped together with the accompanying metadata file in order to ensure that the metadata travels with the GIS file when downloaded from IRMA or other online repositories. All mandatory and mandatory if applicable fields will be filled in when generating metadata for GIS data. GIS files will be posted as an individual record in IRMA (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.

- SECN will decide on a case by case basis whether or not 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.

Photographs

Due to the number of photographs taken as part of ongoing, long-term monitoring efforts, the task of documenting these photos can become quite onerous. In order to streamline the documentation of digital photographs, the NPS has 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.

The National Park Service digital photo metadata standard specifies that seven metadata elements should be considered mandatory. These are the minimum elements required to discover, interpret and manage the image and are indicated below in **bold**. Other elements are required or recommended metadata components by NPS information management systems (e.g. NPSFocus and IRMA) in order to develop a more complete metadata record (M – mandatory, MA – mandatory if applicable)

- Abstract: Summary of content and significance of resource (**M**)
- Author/Contributor: Persons responsible for resource in current form (**M**)
- Citation: If item is part of a larger work, this gives the citation of that work (**MA**)
- Contents: Used when multiple images are described with one record (**MA**)
- **Coverage/Place:** Geographic area of resource. Coordinates are sub-elements (**M**)
- Coverage/Time: Time period of resource (e.g. begin and end dates) (**M**)
- **Date Created:** Date created (**M**)
- Format/Size: Quantity of images, electronic versus physical format, file size (**M**)
- Is Part Of: Name of larger collection; may be none (**M**)
- **Keywords:** Subject terms, either use Keywords or Subject (**M**)
- **Park Code:** NPS Unit Code (**M**)
- Park Name: Park Name (**M**)
- **Publisher:** Agency that supported the item being produced in its current form (**M**)
- Resource Type: Image, text, video, webpage... (**M**)
- **Rights/Access:** Restricted access or publicly available (**M**)

- **Source/Original:** Description of original source prior to scanning or alteration (**MA**)
- **Subject:** Subject term from thesaurus (use this or Keywords or both) (**M**)
- **Title:** Descriptive title (**M**)

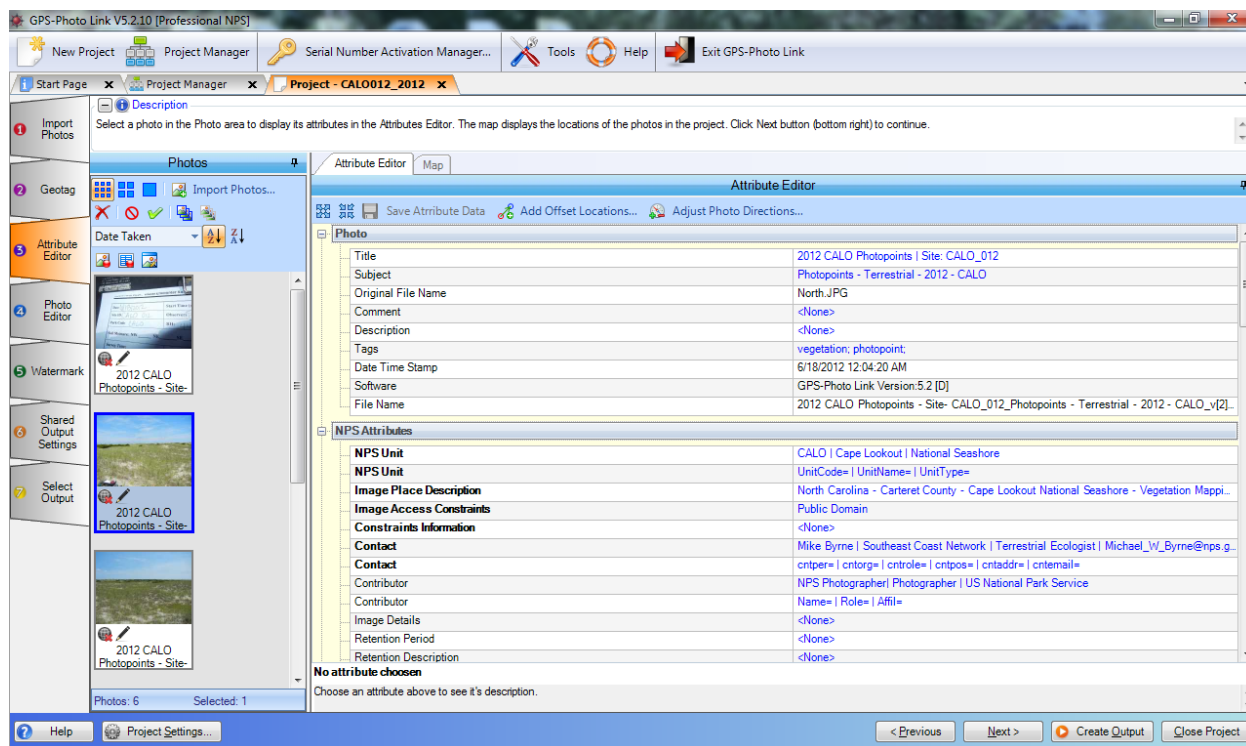


Figure 2. Screenshot of the attribute editor from GPS-Photo Link, using the NPS photo metadata standard.

All photos published to IRMA, NPS Focus or other online repositories will have the full NPS Digital Photo Metadata Standard element set embedded in the photo metadata header. Photo Exif headers may be viewed by end-users through the use of GPS-Photo Link, a wide variety of free Exif viewers and through some commercial software that may show the entire Exif header as well. Data photographs will be published as a collection that includes all of the data photos for one field season. The photos and metadata will be linked to the corresponding data reports in IRMA 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 WRD has developed a series of MS Access-based templates (NPSTORET), for Networks to use to enter their water quality data in a STORET compatible format. Vital Signs Networks will send their data from

NPSTORET or their own data system (following NPS developed specifications) to WRD on an annual basis for inclusion into EPA STORET.

NPSTORET includes a module for importing station information and results from other user databases, labs or other sources. In addition, NPSTORET 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. In addition, the metadata template describes the parameters collected/analyzed for — 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 for 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-2.2.2. National Park Service, 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-2.2.6. National Park Service, Athens, Georgia.

Data Sets for Publication

This table provides an overview of the typical data sets for publication based upon each long-term monitoring protocol. Data sets are published by monitoring protocol and methodology. Also included in the table is the name of the repository where the data are published.

- IRMA: <https://irma.nps.gov>
- STORET: <http://www.epa.gov/storet/>
- Aquarius is available to NPS staff only

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

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
Marine 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

Data Dissemination

The SECN data dissemination strategy aims to ensure 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 are accompanied by appropriate documentation (metadata).
- Sensitive data are identified and protected from unauthorized access and inappropriate use.

Access to Southeast Coast Network (SECN) data products is facilitated via a variety of means that allow users to browse, search and acquire Network data and supporting documents. SECN data and information products (final deliverables or 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 will be made 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 may be uploaded to public distribution repositories (<https://irma.nps.gov>) or otherwise released to the public.

Additional Information

Content Standard for Digital Geospatial Metadata:

http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf

Biological data profile:

<http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/biometadata/biodatap.pdf>

NPS Digital Photo Metadata Standard

<http://npsfocus.nps.gov/docs/guide/metadata/MetadataStandards.html>

Metadata and IRMA

See Section 9.5 in this document:

<https://irma.nps.gov/content/help/datastore/docs/ReferenceHelp.pdf>

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.

National Park Service. 2012. How to create metadata in ArcGIS 10.0. Intermountain Region GIS Program, National Park Service. Denver, CO. (Available at <http://imgis.nps.gov/#Resources>)

National Park Service. 2012. How to create metadata in ArcGIS 10.0. Appendices. Intermountain Region GIS Program, National Park Service. Denver, CO. (Available at <http://imgis.nps.gov/#Resources>)

Wright, C. J. 2013. Exporting Coastal Assessment Data to STORET. Southeast Coast Network Standard Operating Procedure NPS/SECN/SOP-2.2.2. National Park Service, 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-2.2.6. National Park Service, 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.