Standard Operating Procedure (SOP) #19

Metadata Development

Version 1.01 (July 29, 2021)

Change History

New Version #	Revision Date	Author	Changes Made	Reason for Change	Previous Version #
1.01	7/29/2021	Kelly Kozar, Kim Weisenborn	Updates to general instructions. Updated outdated SOP references. Added Metadata Interview form as appendix.	To accurately reflect the metadata procedures. To update outdated SOP references and external links. To make the interview form easily accessible.	1.0

Only changes in this specific SOP will be logged here. Version numbers increase incrementally by hundredths (e.g., version 1.01, version 1.02) for minor changes. Major revisions should be designated with the next whole number (e.g., version 2.0, 3.0, 4.0). Record the previous version number, date of revision, author of the revision, changes made, and reason for the change along with the new version number.

Purpose

This SOP document describes the guidelines for documenting data and how it should be accomplished for the Pacific Island Network (PACN) Focal Terrestrial Plant Communities (FTPC) Monitoring Protocol.

Metadata Documentation

Data documentation is a critical step toward ensuring that datasets are usable for their intended purposes well into the future. This involves the development of metadata, which can be defined as structured information about the content, quality, condition, and other characteristics of a given dataset. Additionally, metadata provide the means to catalog and search among datasets, thus making them available to a broad range of potential data users. Metadata for all PACN monitoring data will conform to Federal Geographic Data Committee (FGDC) guidelines and will contain all components of supporting information such that the data may be confidently manipulated, analyzed, and synthesized. Initial metadata development describing individual measurements is based on the definitions and procedures described in SOP #10 Conducting Community Vegetation Surveys.

Updated metadata is a required deliverable that should accompany each season's certified data. For long-term projects such as this one, metadata creation is most time consuming the first time it is developed – after which most information remains static from one year to the next. Metadata records in subsequent years only need to be updated to reflect changes in contact information, taxonomic

conventions, publication listings, data disposition, quality descriptions, collection methods, analysis approaches or quality assurance for the project.

Specific procedures for creating, parsing, and posting the metadata record are found in National Park Service (NPS) Metadata Recommendation and Implementation Plan for the Inventory and Monitoring Program[1].

General Procedures

1. After the annual data quality review has been performed and the data are ready for certification, the PACN Botanist (or a designee) updates the PACN Metadata Interview Form in Appendix SOP 19.a.

    a. The metadata interview form greatly facilitates metadata creation by structuring the required information into a logical arrangement of 15 main questions, many with additional sub-questions.

    b. The first year, a new copy of the metadata interview form should be used. Otherwise, the form from the previous year can be used as a starting point, in which case the Track Changes tool in MS Word should be activated in order to make edits obvious to the person who will be updating the XML record.

    c. Complete the metadata interview form and maintain it in the project workspace. Much of the interview form can be filled out by cutting and pasting material from other documents (e.g., reports, protocol narrative sections, and SOPs).

    d. The PACN Data Manager can help answer questions about the metadata interview form.

2. Deliver the completed interview form to the PACN Data Manager according to SOP #21 Product Delivery Specifications.

3. The PACN Data Manager (or PACN GIS Specialist for spatial data) will then extract the information from the interview form and use it to create and update an FGDC- and NPS-compliant metadata record in XML format. Specific guidance for creating the XML record is contained in NPS Metadata Implementation Plan for the Inventory and Monitoring Program.

4. The PACN Data Manager will post the record and the certified data to the Integrated Resource Management Applications Portal[2], and maintain a local copy of the XML file for subsequent updates.

[1] https://irma.nps.gov/DataStore/Reference/Profile/2265315 (last accessed 29 July 2021)
[2] https://irma.nps.gov/Portal (last accessed 29 July 2021)

5. The PACN Botanist should update the metadata interview content as changes to the protocol are made, and each year as additional data are accumulated.

Identifying Sensitive Information

Part of metadata development includes determining whether or not the data include any sensitive information, which is partly defined as the specific locations of rare, threatened or endangered species. Prior to completing the metadata interview form, the PACN Botanist and park biologists should work together to identify any sensitive information in the data after first consulting SOP #20 Sensitive Information Procedures. Their findings may be documented and communicated to the PACN Data Manager through the PACN Metadata Interview Form.

Appendix SOP 19.a. PACN Metadata Interview Form

National Park Service
U.S. Department of the Interior

Pacific Island Network

Instructions

Type responses directly into this document using a different font color or into a separate text or Word document file and send to the project metadata contact.

1. Title of project

2. Originator / PACN Botanist Contact information: Include address, email and telephone number
    a. Additional contacts for project (address, email and telephone number)
    b. Other individuals/organizations to receive credit for project support (e.g., funding, data collection and analysis)

3. Is the dataset published? If yes, provide the reference

4. Brief description of the dataset (just a few sentences)

5. Purpose of data collection

6. Time period represented by the dataset

7. Primary method of dataset development
    a. Field visits?
    b. Remote instrumentation (e.g., temperature recorders, remote cameras, etc.)?
    c. Existing data sources?

8. Status of the dataset
    a. Is it complete, in progress, or planned?
    b. Will the dataset be updated? If so, how frequently?

9. Location of data collection. Include place names/descriptions and longitude and latitude bounding coordinates (need values for the four cardinal directions - N, S, E, and W). If UTMs are known, provide those as well.

10. Keywords for searching on this dataset
    a. Thematic
    b. Place
    c. Temporal
    d. Taxonomic
    e. Strata

f. If a controlled vocabulary or thesaurus was used, provide the reference

11. List all related datasets for cross-reference. This includes any tabular and spatial (GIS) data as well as reports.

12. Biological data: the Federal Geographic Data Committee (FGDC) Biological Profile for metadata provides the means to document tabular datasets, taxonomy, field methods, and the use of analytical tools or models.
    a. Was dataset developed using a model or other analytical tool?
        i. If yes, reference
        ii. If the model or tool is available include a contact and/or URL
    b. Does the dataset contain biological information? If no, skip to item **13**.
        i. What species or communities were examined?
        ii. Was a taxonomic authority or field guide used for identification? If yes, reference
        iii. Briefly summarize field methods (copy and paste from other documents)
        iv. If existing protocols or methods were used, provide the references
        v. Were voucher specimens collected? If yes, list names of repositories and all contact information for each repository

13. Information needed for archiving all datasets (tabular, spatial and biological)
    a. Describe quality assurance and quality control measures taken
    b. Identify items excluded from data collection: (e.g., stems less that a certain diameter or streams without surface flow)
    c. Type/format(s) of dataset – list all that apply (spreadsheet, ascii text file, database, GIS attribute table, etc.)
    d. What is the filename (and directory path if available) for dataset?
    e. Field/attribute information:
       List the fields (columns) in the dataset for *each* file or table (if available, a data dictionary should contain this same information and can be referenced instead)

       For each field list:
        i. Definition of the field (brief description), field size, and type (e.g., numeric, date, text, etc.)
        ii. Domain - each attribute is described by a domain of allowable values:
            Enumerated – a finite list of codes or values - list the codes and definitions
            Range – measured or calculated values - identify the units as well as the minimum and maximum values ("0 - no limit" is acceptable)
            Codeset – codes that come from a published code set or from another file or table - list the reference
            Unrepresentable – anything that does not fit into the above three domains (e.g., open ended list of landowners, site IDs, comments fields) - briefly describe what is in the field
        iii. If in an Access database, indicate if the field is a primary key or link field

14. Is this a GIS dataset? If not, skip to item **15**
    a. Identify a pathway to where the data can be accessed over a network or send a copy of the ESRI File geodatabase (.gdb) spatial data, project specific ArcGIS Pro layers (.lyrx), and Excel (.xlsx) station coordinate table files or the shapefile and indicate the following:
        i. Projection and datum parameters (e.g., Projection: UTM Zone 10, Datum: NAD 1983)
        ii. Map units (e.g., meters)
    b. Brief description of what the dataset contains and how it is to be used
    c. List any source datasets used. For each source list:
        i. Source name, originator and publication date
        ii. Source time period and scale
        iii. Source presentation form and media type
        iv. Contribution of source to the analysis
    d. List the processing steps used to create the dataset, and for each include:
        i. Who (and contact information)
        ii. Approximate date of processing
    e. Data quality:
        i. Status - planned, in progress, complete
        ii. Logical consistency (e.g., are lines snapped, polygons closed, etc.)
        iii. Positional accuracy (estimate of distance from true location and reason for discrepancy - e.g., GPS (Global Positioning System) accuracy limitations)
        iv. Attribute accuracy

15. For all datasets: is the dataset available for distribution? Indicate to whom it can be distributed: Public, NPS Only, PACN Only or Park Only
    a. Is there sensitive information (specific locations of rare, threatened or endangered species)?
        i. Can the resolution of those specific locations be ‘degraded’ to allow distribution?
    b. Are there legal restrictions on who may use the data?
    c. Any advice/caveats for potential users of the dataset?
    d. What are the distribution instructions?

*Adapted from “Electronic Metadata Interview” (Lienkaemper, no date).