WRITING METADATA

How to write FGDC compliant metadata using Metavist
Metadata = Data Documentation

- What data were collected
- Why data were collected
- How data were collected
- Reliability of data
- Issues that should be accounted for when working with data
- How to get data
- Tools needed to work with the data
Purpose of Metadata

• Primary
  – Guide for you to understand your data
    Helps you remember what you and your team did
  – Guide for others to understand your data
    Helps users today and 20 years from now be successful

• Secondary
  – Guide to other documentation in pub
    Documentation not limited to formal metadata
  – Marketing for your data
    Viewed by most Archive customers
Writing Quality Metadata

• Complete and correct
  – Data are useless without proper documentation
  – This will help ensure data are used correctly
  – Other files or publications can be referenced

• Comprehensive and comprehensible
  – Should fully describe data
  – Should be understandable by non-experts

• Focus on the data
  – Describe the data – not the results of analysis

• Start early
  – Don’t wait until paper is written to start writing the metadata – begin right away
Metadata Standards

• CSDGM*
  – Federal Geographic Data Committee (FGDC): Content Standard for Digital Geospatial Metadata
  – Designed for geospatial data
  – Mandatory for Federal use since 1994

• BDP*
  – Biological Data Profile: formally approved by FGDC as a superset of CSDGM
  – Additional elements: Taxonomy, Methodology, Analytical tools
  – Works for nearly all geospatial / non-geospatial data

• ISO 19115 family
  – Designed for geospatial data; supersedes CSDGM
  – More complex, but also more flexible than CSDGM
  – BDP is approved part of standard

* Use of “FGDC” from here on out refers to both CSDGM and BDP unless otherwise noted.
Metadata Standards cont.

• EML
  – Ecological Metadata Standard
  – Originally developed for the ecology discipline
  – Used by Long Term Ecological Network
  – Crosswalk between BDP and EML in progress

• Others
  – Dublin core
  – …
Tools Used to Write FGDC Metadata

• Metavist software (FS R&D: Dave Rugg)
  – Free! User friendly!
  – Requires some knowledge of FGDC standards
  – Works for geospatial and non-geospatial data
  – Works for CSDGM or BDP metadata
  – Generates XML file, can export as HTML

• ArcCatalog
  – For geospatial data only – doesn’t understand BDP
  – May not be complete form of metadata
  – Default standard is ISO 19115, but knows CSDGM

• Microsoft Word Form
  – Easy to fill out
  – Requires no prior knowledge of FGDC standards
  – Works for all data except geospatial data
  – Works for CSDGM (except spatial sections) or BDP metadata
FGDC Metadata - Main Sections

1. Identification
   What data were collected, why collected, where collected, tools need to work with data

2. Data Quality
   How data were collected, reliability of data, data omissions

3. Spatial Data Organization

4. Spatial Reference

5. Entity and Attribute
   Description of all files, list and description of variables within each file

6. Distribution
   How to get data, data formats

7. Metadata Reference
   Metadata currentness, responsible party
Metadata Example – using Metavist

1. Identification
What data were collected

Why data were collected

When data were collected
Description of Geographic Extent

The USDA Forest Service Marcell Experimental Forest (MEF) is an 1140-hectare tract of land located 40 kilometers north of Grand Rapids in Balsam Township, Itasca County, Minnesota, USA. The bounding coordinates are the maximum extents of western, eastern, northern, and southern corners of both the North and South Units of the Marcell Experimental Forest.

Watershed areas:
- S1 = 33.2 hectares (ha)
- S2 = 9.7 ha
- S4 = 34.0 ha
- S5 = 32.6 ha
- S8 = 8.9 ha

Where data were collected - description

Where data were collected - coordinates
Keyword categories we use:
1. ISO 19115
2. R&D taxonomy
3. Author choice
4. Place keywords
For BDP metadata only:
If applicable, include taxonomy for each species. We can auto-generate this using https://www.itis.gov
Access Constraints

(Data default is "None")

None

Use Constraints

(Data default is "None")

These data were collected using funding from the U.S. Government and can be used without additional permissions or fees. If you use these data in a publication, presentation, or other research product please use the following citation:


Point Of Contact (optional)

Security Information (optional)

Security Classification System

Security Classification

Security Handling Description
For BDP metadata only:
If applicable, include citation and details for tools needed to work with the data.

Tools, models, or statistical procedures that the data set is intrinsically bound to and are available for use in analyzing the data set. Examples include reconstructions of phylogenies, population viability analyses, community ordinations, most atmospheric and hydrologic transport analyses, and inferences on the effects of climate change on forest composition and productivity. Enough information should be included so that a potential data user can easily determine why they might wish to acquire the analytical tool, and the methodology to acquire it.
Who funded this work, and other data credits

Important image/graphic(s)

Publications that USE the data we are publishing, or are related
Metadata Example – using Metavist

2. Data Quality
Accuracy of data
Fidelity of the relationships in the data and tests used

Completeness of data: data omissions and why, how missing data are recorded, etc.
Both tabs for spatial data: Positional Accuracy and Cloud Cover information.
For BDP metadata only: How data were collected (thorough and understandable)

Citation(s) for data/info from other sources

Post-collection data processes
How data were collected, methodology citations, methodology keywords
Post-collection data process steps, sources used, sources produced

Process step date
Metadata Example – using Metavist

3. Spatial Data Organization
For Spatial Data Only:
Mechanism used to represent spatial information in data
(can obtain info directly from ArcGIS)
Metadata Example – using Metavist

4. Spatial Reference
For Spatial Data Only:
Reference frame for, and the means to encode, coordinates in data
(can obtain info directly from ArcGIS)
Metadata Example – using Metavist

5. Entity and Attributes
To keep things simple we typically do not use the Detail Description section unless data are spatial.
Depending on complexity, variable descriptions can be provided in a separate file.

Citation for publication associated with these data, and any other references noted in variable descriptions.
Metadata Example – using Metavist

6. Distribution
Information about data distributor, data formats, and data access
Distribution Information

Distributor
Organization: USDA Forest Service, Research and Development

Resource Description
RDS-2018-0009

Distribution Liability
Metadata documents have been reviewed for accuracy and completeness. Unless otherwise stated, all data and related materials are considered to satisfy the quality standards relative to the purpose for which the data were collected. However, neither the author, the Archive, nor any part of the federal government can assure the reliability or suitability of these data for a particular purpose. The act of distribution shall not constitute any such warranty, and no responsibility is assumed for a user’s application of these data or related materials.

Contact Information

The primary contact is a(n):
- Person
- Organization

Contact Person (Optional)

Contact Position (Optional)
Research Data Archivist

Contact Address
mailing and physical: Fort Collins

Contact Voice Telephone
see Contact Instructions

Contact TDD/TTY Telephone
(Optional)

Contact Fax Telephone
(Optional)

Contact E-Mail Address
(Optional)

Hours of Service
(Optional)

Contact Instructions (Optional)
This contact information was current as of April 2018. For current information see Contact Us page: https://doi.org/10.2737/RDS.
Metadata Example – using Metavist

7. Metadata Reference
Choose metadata standard

Who to contact if there are questions
Metavist – File Options

Metavist saves file as XML, but can export to HTML.
Marcell Experimental Forest daily streamflow data

Data publication access: https://doi.org/10.2737/RDS-2018-0009

Abstract:

Purpose:
The Marcell Experimental Forest was formally established in 1962. The MEF contains six watersheds (and other study sites), each consisting of an upland portion and a peatland that is the source of a stream leaving the watershed. The watersheds and environmental monitoring at the MEF are part of a long-term research program on the hydrology and biogeochemistry of watersheds with uplands and northern peatlands (Kolka et al. 2011).

Supplemental Information:
Additional information about the Marcell EF, its instrumentation, and data can be found at https://www.nrs.fs.fed.us/ef/marcell/ and in Sebestyen et al. 2011.

Original metadata date was 02/22/2018. On 02/26/2018 we added additional supplemental files containing details about instrumentation. On 04/30/2018 we updated the instrumentation supplemental files and updated the
Metadata Example – using Word Template

https://www.fs.usda.gov/rds/archive/metadata

(Word Template does not contain spatial sections)
1. IDENTIFICATION INFORMATION

Citation for Data Publication

Originators (author names, please include middle initial): 

Title: 

Data Presentation Form (tabular digital data, raster digital, database, document…): 

Publication Place: Fort Collins, CO

Publisher: Forest Service Research Data Archive

Description of Data Publication

Abstract (narrative summary of data): 

Purpose (why data were collected): 

Supplemental Information (other important info): 

Time Period of Content

Beginning Date: 

Ending Date: 

Other: 

Status

Progress (in progress, complete): Complete

Maintenance and Update Frequency (as needed, none planned, annually…): As Needed
Description of Geographic Extent (description of where data were collected):

Bounding Coordinates

West Bounding Coordinate: 
East Bounding Coordinate: 
North Bounding Coordinate: 
South Bounding Coordinate: 

Coordinates Unit: Longitude and Latitude in decimal degrees

Bounding Altitudes

Minimum Altitude: 
Maximum Altitude: 

Altitude Distance Units (feet, meters): 

Theme Keywords
(for more info see: https://www.fs.usda.gov/rds/archive/submittdata/Keywords_for_FS-RDA_archive.pdf)

Author’s choice Keywords: 
ISO 19115 Keywords: 
R&D Taxonomic Keywords: 

Place Keywords (include state(s) if applicable): 

Use Constraints (any constraints with sharing these data?): 

Point of Contact (for data)

Contact Organization: 

Contact Person: 

Contact Position: 

Contact Address: 

Contact Voice Telephone(s): 

Contact Email Address: 

Data Set Credit (who funded this work?): 

Native Data Set Environment (software, operating system, etc. - if important):

Cross-References (citations for publications that USE or are related to Data Publication, please include DOI/URL)

Complete Citation(s): 

2. DATA QUALITY INFORMATION

Attribute Accuracy

Attribute Accuracy Report (assessment of how “true” attributes values are):

Logical Consistency Report (methods used to check for inconsistencies):

Completeness Report (info about omissions, selection criteria …):

Lineage- Methodology (how data were collected or obtained, steps in field or laboratory work…)

Methodology Keywords:

Methodology Description:

Methodology Citations (publications that describe methods or are referenced in methods, please include DOI/URL)

Complete Citation(s):

Source Citations (if any data were obtained from another source please provide: source citation, description of data obtained, and where data were obtained)

Complete Citation(s) and Data Obtained:

Lineage- Process Steps (steps or data manipulations applied after data collection, or modifications made to source data)

Process Descriptions (include dates):
3. ENTITY AND ATTRIBUTE INFORMATION
   Overview description of variables in each data set (literally need a list and description of variables in each file and be sure to include units - this can also be done in a spreadsheet):

   Citation(s) that contain data summary or details about these variables:

4. DISTRIBUTION INFORMATION
   Data type (need a list and description of all file formats used)
   (Example: CSV = comma-delimited ASCII text file)
   Format Names and Description:

5. METADATA REFERENCE INFORMATION
   Metadata Contact (who to contact if there are questions about the metadata)
   Contact Organization:
   Contact Person:
   Contact Position:
   Contact Address:
   Contact Voice Telephone(s):
   Contact Email Address: