

How to Prepare and Submit Electronic Research Data to the Forest Service Research Data Archive

<https://www.fs.usda.gov/rds/archive>

This document provides an overview of the steps needed to prepare both data and documentation for submission to the Forest Service Research Data Archive (FS-RDA), along with instructions on how to submit this package.

1. PREPARE FILES

- **Determine what to include**
 - Raw research data file(s)
 - This can be data used for a particular publication, observational or historical data related to a short or long-term study, or secondary data that has been substantially modified (with documented concurrence of originating sources)
 - Supplemental files (maps, photographs, unpublished reports, lab notes, study plan, etc.)
- **Ensure data files are complete and correct**
 - Data should have already gone through rigorous quality checks before submission, or the metadata must clearly note why.
 - Things to double-check (common issues we see)
 - Does your data contain outliers? If there are outliers and there is a reason, please note this in the Attribute Accuracy section of the metadata. If the outlier(s) has no discernible reason, it should be reviewed and/or changed and noted in the metadata.
 - Ensure that all categorical variables have valid values, and the descriptions included in the entity and attributes section of the metadata.
 - Verify missing data are in fact missing, and note why in the metadata Completeness Report.
 - Ensure that zeros, blank cells, and/or missing data are clearly defined in the metadata and consistently used.
- **Use transparent filenames**
- **Convert files to stable and useable format**
 - Can be submitted in virtually any format and archivists can help with conversion
 - Common formats: CSV, TXT, JPG, PDF or PDF/a

2. DEVELOP METADATA

A metadata document is the documentation for a data set(s). It provides the user with the information needed to completely understand the data, why it was collected, how it was collected, and the quality of the data.

- **Use proper metadata standard**
 - Standard we currently use: [Federal Geographic Data Committee's \(FGDC\) Biological Data Profile of the Content Standard for Digital Geospatial Metadata \(BDP\), version 2](#) (see the [BDP workbook](#) for complete details)
 - Other standards can be used if data necessitates
 - We can accept metadata in almost any format and the Archive Team can help with conversion

- **Use tools to help create metadata**
 - Metavist software (<https://www.fs.usda.gov/rds/archive/Metadata>)
 - ArcGIS software (if archiving GIS data)
 - Microsoft Word form (<https://www.fs.usda.gov/rds/archive/Metadata>)
- **Use help**
 - Appendix 1 contains a short description of the contents expected in the most common metadata sections.
 - Training slides: [Creating Metadata using Metavist Software](#)
 - Training slides: [Introduction to Metadata](#)
- **Review metadata to ensure it is correct and complete**

3. COMPILE DATA PACKAGE = data files + supplemental files + metadata document + file index

- **Gather all files**
 - Data set(s)
 - Metadata
 - Additional files that should be archived with the data if applicable (e.g. maps, research notes, study plans)
- **Use directory structure** to help organize files
- **Create a file index**, which is a list of all files to be archived and a short description of their format and content (see Appendix 2)

4. SUBMIT DATA PACKAGE

- **Contact member of the Archive Team** (<https://www.fs.usda.gov/rds/archive/contactus>) to determine how to submit files
- **Include the following files in your submission:**
 - Data package as defined in Step 3 above
 - Associated publication(s) or a link to them if available online
 - Signed submission form - <https://www.fs.usda.gov/rds/archive/SubmittingData> (see bullet 5)

Appendix 1: Description of important metadata fields

IDENTIFICATION

- **ABSTRACT** - Published on the landing page for the data, as well as on the metadata page, and needs to clearly detail what is included in the data, including a quick description of the data with dates, locations, and any other info that will help the user understand what is in the data.
- **PURPOSE** - Why were these data collected? What can be deduced from these data?
- **SUPPLEMENTAL INFORMATION** - Is there anything else the user should know. (OPTIONAL)
- **SPATIAL DOMAIN** - Where did the data collection take place?
- **BOUNDING COORDINATES** - Bounding coordinates for the data collection location(s).
- **KEYWORDS** - Under theme, attach keywords not relating to locations of data collection. Under Place, put location keywords including state.
- **TAXONOMY** - Include a list of any species associated with the data (OPTIONAL)

DATA QUALITY

- **ATTRIBUTE ACCURACY REPORT** - Information detailing methods used for assigning values, as well as why there might be any outlying values in the data
- **LOGICAL CONSISTENCY REPORT** - Information about the relationships between the data and the tests used to get the data.
- **COMPLETENESS REPORT** - Information about how the data set was created/how the data is defined.
- **METHODOLOGY** - VERY important. Should clearly explain how data was collected concisely so that ANYONE regardless of background, education level, etc., should they have the proper equipment, could recreate the data. It needs to be thorough and understandable.

DATA DESCRIPTION

- **ENTITY AND ATTRIBUTES** - List the data files included in the data package along with a description of each variable which should include the unit of measurement, the format the value is in (i.e. MMDDYYYY, MM-DD-YY, etc.), and a full description of any categorical data values (i.e. T=True and F=False; 1=low, 2=medium, 3=high; etc.). Be very specific. Unless dealing with GIS data, we typically just use the Entity and Attribute Overview section to describe the data.

MISCELLANEOUS

- **CONTACT INFO** - Providing contact info for someone who is willing and able to answer questions about the data is optional, but recommended. Consider using a phone number for the building/lab instead of an individual phone number that may change.
- **REFERENCES** - If there are any cross references or methodology citations, provide a thorough citation. The cross-reference section is for citations that utilize the data being published, or are associated with these data. The methodology citation section is for publications that provide information about the methods for the data being archived.

Appendix 2: File Index example

(File can be submitted as DOCX, TXT, XLXS, HTML, etc.)

Data Publication title goes here

Research Data Publication File Index

File	Folder	Description
_metadata_RDS-YYYY-XXXX.html		Metadata file in HTML format containing a description of the content, quality, and other characteristics of the data.
_metadata_RDS-YYYY-XXXX.xml		Metadata file in Extensible Markup Language (XML) format containing a description of the content, quality, and other characteristics of the data.
Region1.csv	\Data	Comma-delimited ASCII text file containing meteorological data from region 1.
Region2.csv	\Data	Comma-delimited ASCII text file containing meteorological data from region 2.
Region_Units.shp	\Data	ESRI shapefile and associated files containing all units associated with data.
Region_Units.shp.xml	\Data	Metadata file in XML format containing a description of the content, quality, and other characteristics of the Region-Units shapefile.
Image1.jpg	\Supplements	JPEG image file containing a map of region 1.
Image2.jpg	\Supplements	JPEG image file containing a map of region 2.
Publication.pdf	\Supplements	Adobe Acrobat PDF file containing the publication "Title of publication".