

FS R&D, Research Data Services

# FS RESEARCH DATA ARCHIVE

*How to prepare data and metadata for submission*

## Research Data Archive

Roots of our Research



Search catalog



Data Catalog

Using Our Formats

Terms of Use

Submitting Data

Metadata & Tools

About Us



### Spatial wildfire occurrence data for the United States, 1992-2015

This spatial database of wildfires that occurred in the United States from 1992 to 2015. It includes 1.88 million geo-referenced wildfire records, representing a total of 140 million acres burned during the 24-year period. Wildfire records were acquired from federal, state, and local fire reporting systems. Core data elements are: discovery date, final fire size, and a point location at least as precise as Public Land Survey System (PLSS) section (1-square mile grid). The data were transformed to conform, when possible, to the data standards of the National Wildfire Coordinating Group (NWCWG). Basic error-checking was performed and redundant records were identified and removed, to the degree possible.

Ever wonder how science knows what it knows? Mostly by thinking about observations made on the natural world, also known as data. The roots of our research are in the data we collect. Short-term sharing of research data is important for the transparency and reproducibility of the research results our scientists publish in research papers. Preserving these data for long-term use is important to being able to re-use the data for multiple purposes. The FS Research Data Archive offers a catalog of hundreds of research datasets funded by Forest Service Research and Development ([FS R&D](#)) or by the Joint Fire Science Program ([JFSP](#)). Of special interest, our collection includes long-term datasets from a number of Forest Service Experimental Forests, Ranges, and Watersheds.

Data publications listed in the catalog are publicly available either from us or another scientific data repository (the publication description page will include a link to the dataset's location if in a different repository). Each published dataset comes with a complete set of metadata (data documentation) and other information the scientist considered important to understanding and successfully using the data. Most data publications that we have released are downloadable; others have a database query interface to help you get that part of the large dataset which is of interest to you. Some data are also available through a map service. Check out the catalog and discover if there are data of interest to you!

# Forest Service Research Data Archive (FS-RDA)

[www.fs.usda.gov/rds/archive](http://www.fs.usda.gov/rds/archive)

- Purpose

- Publish digital research data funded by

- FS Research & Development
- Joint Fire Science Program
- APHIS National Wildlife Research Center

- What we publish

- Long-term research data
- Data from specific studies
- Supplements to FS publications
- Image libraries

The screenshot displays the Forest Service Research Data Archive (FS-RDA) website. The header features the title "Research Data Archive" with the tagline "Roots of our Research" and a search bar. A navigation menu includes links for "Data Catalog", "Using Our Formats", "Terms of Use", "Submitting Data", "Metadata & Tools", and "About Us". The main content area shows a search interface with a sidebar for filtering by "Publication Year", "Authors", "Locations", "Funder", "Keywords", and "GIS". The search results display a list of items, with the first three visible: 1. "Glacier Lakes Ecosystem Experiments Site (GLEES) hourly meteorology tower data (3rd Edition)", 2. "United States hurricane accumulated cyclone energy and its potential impacts to forest basal area and urban tree canopy", and 3. "Granivory and germination data for: 'Tree encroachment impacts on seed predator selection and seedling establishment in degraded pine woodlands'". Each item includes the author(s) and a link to the data.

**Research Data Archive**  
Roots of our Research

Search catalog

Data Catalog Using Our Formats Terms of Use Submitting Data Metadata & Tools About Us

Limit your search

Publication Year +

Authors +

Locations +

Funder +

Keywords +

GIS +

Search

(To search for a phrase use "", e.g. "experimental forest")

Displaying items 1 - 10 of 595

Sort by [Date\(desc\)](#) | [Relevance](#) 10 | 20 | 50 | per page

1. [Glacier Lakes Ecosystem Experiments Site \(GLEES\) hourly meteorology tower data \(3rd Edition\)](#)  
**Author(s):** Korfmacher, John L.; Hultstrand, Douglas M.; Doebley, Valerie T.; Claybrook, Tanner P.; Ellsworth, Alan C.
2. [United States hurricane accumulated cyclone energy and its potential impacts to forest basal area and urban tree canopy](#)  
**Author(s):** Cole, Jason A.; Nowak, David J.; Greenfield, Eric J.
3. [Granivory and germination data for: "Tree encroachment impacts on seed predator selection and seedling establishment in degraded pine woodlands"](#)  
**Author(s):** Willis, John L.; Schnake, David K.; DePerno, Christopher S.; Lashley, Marcus A.; Wetzstein, Branson; Yow, Justin
4. [Subalpine seedling establishment and climate data provide broad scale drivers of subalpine tree establishment patterns](#)  
**Author(s):** Schapira, Zoe H.; Stevens-Rumann, Camille S.

# How to prepare files for the FS-RDA

## Step 1: Determine what files to include

- Raw research data files
  - Data directly associated with a particular publication
  - Observational or historical data related to a short or long-term study
  - Secondary data that has been substantially modified (with documented concurrence of originating sources)
- Supplemental files, examples include...
  - Maps
  - Photos
  - Unpublished reports
  - Lab notes
  - Study plan
  - Data analysis documentation (helps meet journal requirements)

# How to prepare files for the FS-RDA

## Step 2: Prepare data and supplemental files

- Ensure data 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 so, please note why if known.
    - Ensure all categorical variables have valid values and descriptions for each value.
    - Verify missing data are in fact missing and note why if known.
    - Ensure zeros, blank cells, and/or missing data are clearly defined and consistently used.
- Use transparent filenames
- Convert files to stable and usable format
  - Can be submitted in virtually any format and archivists can help with conversion
  - Common formats: CSV, XLSX, TXT, JPG, PDF or PDF/a

# How to prepare files for the FS-RDA

## Step 3: Develop metadata

*Data documentation (metadata) is required with all submissions. This document provides the user with the information needed to completely understand the data, why it was collected, how it was collected, the quality of the data, and who to contact if they have questions.*

- Understand metadata standards
  - CSDGM
    - Federal Geographic Data Committee (FGDC): Content Standard for Digital Geospatial Metadata
    - Designed for geospatial data
  - 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

# How to prepare files for the FS-RDA

- Understand metadata contents

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 all variables within each file

6. Distribution

How to get data, data formats

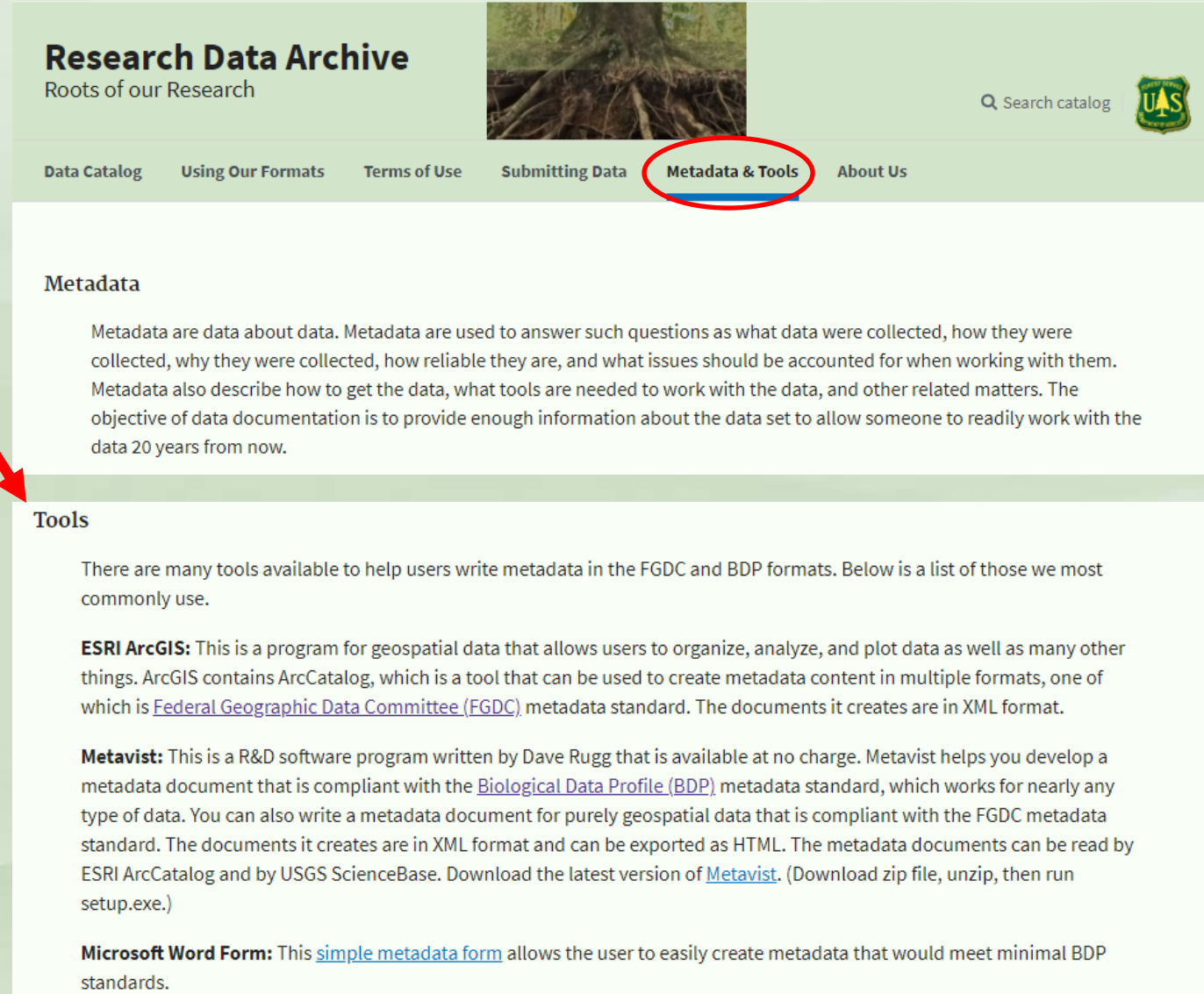
7. Metadata Reference

Metadata currentness, responsible party




# How to prepare files for the FS-RDA

- Use tools to generate metadata
  - ESRI ArcGIS
    - For geospatial data only – doesn't understand BDP
    - May not be complete form of metadata
    - Default standard is ISO 19115, but knows CSDGM
  - Metavist software (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
  - Microsoft Word form
    - Easy to fill out
    - Requires no prior knowledge of FGDC standards
    - Works for all data (geospatial data will require additional info)
    - Works for CSDGM (except spatial sections) or BDP metadata



**Research Data Archive**  
Roots of our Research

Search catalog 

Data Catalog Using Our Formats Terms of Use Submitting Data **Metadata & Tools** About Us

### Metadata

Metadata are data about data. Metadata are used to answer such questions as what data were collected, how they were collected, why they were collected, how reliable they are, and what issues should be accounted for when working with them. Metadata also describe how to get the data, what tools are needed to work with the data, and other related matters. The objective of data documentation is to provide enough information about the data set to allow someone to readily work with the data 20 years from now.

### Tools

There are many tools available to help users write metadata in the FGDC and BDP formats. Below is a list of those we most commonly use.

**ESRI ArcGIS:** This is a program for geospatial data that allows users to organize, analyze, and plot data as well as many other things. ArcGIS contains ArcCatalog, which is a tool that can be used to create metadata content in multiple formats, one of which is [Federal Geographic Data Committee \(FGDC\)](#) metadata standard. The documents it creates are in XML format.

**Metavist:** This is a R&D software program written by Dave Rugg that is available at no charge. Metavist helps you develop a metadata document that is compliant with the [Biological Data Profile \(BDP\)](#) metadata standard, which works for nearly any type of data. You can also write a metadata document for purely geospatial data that is compliant with the FGDC metadata standard. The documents it creates are in XML format and can be exported as HTML. The metadata documents can be read by ESRI ArcCatalog and by USGS ScienceBase. Download the latest version of [Metavist](#). (Download zip file, unzip, then run setup.exe.)

**Microsoft Word Form:** This [simple metadata form](#) allows the user to easily create metadata that would meet minimal BDP standards.

# Metavist software

Metavist 2017 - \_metadata\_RDS-2012-0008-2.xml

File Options Help

Help / Examples

Identification

Data Quality

Spatial Data Org

Spatial Reference

Entity & Attribute

Distribution

Metadata Ref

Basic Information

Spatial Domain

Keywords

Taxonomy

Access

Analytical Tools

Miscellaneous

Citation for the data set

Author(s)

Kenefic, Laura S.

Publication Date

2015

Title

Overstory tree and regeneration data from the "Silvicultural Effects on Composition, Structure, and Growth" study at Penobscot Experimental Forest

Edit

Abstract

This data publication contains overstory tree measurements, regeneration data, and permanent sample plot location information collected between 1952 and 2014 under the study plan: FS-NRS-07-08-01 "Study Plan: Silvicultural effects on composition, structure and growth of northern conifers in the Acadian Forest Region: Revision of the Compartment Management Study on the Penobscot Experimental Forest" (see Methodology citation section). Data are available in six data sets. 1) Overstory tree measurement data include tree species, condition code (e.g., merchantability status and cause of mortality, if applicable), and diameter at breast height (dbh), 1952 to 2014. 2) Regeneration data include tree seedling species, presence, and count by height class, 1964 to 2014. 3) Spatial location data include location of a subsample of trees, 2000 to 2014. 4) Height and crown measurement data include tree height, height to crown base, and crown radii for a subsample of trees, 2000 to 2014. 5) Understory vegetation data include percent cover by substrate and non-tree vegetation categories, 2000 to 2014. 6) Permanent plot location data include the geospatial coordinates for permanent sample plots.

Purpose

The primary objective of the long-term silvicultural study, called the Compartment Management Study, conducted by the USDA Forest Service at the Penobscot Experimental Forest (PEF) is to quantify tree and stand response to silvicultural treatment. Response variables include regeneration, species composition, and tree and stand growth, productivity, and quality. Data provide information about the interaction of natural and human disturbances and their effects on forest stand dynamics. A secondary objective of the study is to provide a variety of forest structures at one location to be used as the framework for short-term experiments in ecology and silviculture.

Supplemental Information (Optional)

A number of revisions have been made to this long-term study over time, including but not limited to changes in silvicultural prescriptions (e.g., target residual basal areas, tree size thresholds for removals, and species composition goals), sampling protocols (e.g., numbers and sizes of sample plots, tree species and condition codes, and frequency of inventory), and response variables (e.g., overstory tree attributes, regeneration, and deadwood).

The first edition of this data publication was published in 2012 (see Cross-Reference section). The second edition (published on 08/27/215) includes three years of additional data and corrections to data errors. The metadata have been revised to reflect these changes and a complete list of differences between the first and the second edition is located in the supplemental files (available with the full data publication download). On 12/12/2016 the metadata were updated to include more details to the methods section and other minor metadata updates.

Time Period of Content

Data is from:

☐ A single date

☐ Multiple dates

☒ A range of dates

☐ Unknown

The calendar is:

☒ Gregorian

☐ Geologic

Time Period(s)

begin: 1952 end: 2014

Add

Edit

Delete

Currentness Reference

Ground condition

Status

Progress

Planned

In progress

Complete

Updating

As needed



# Microsoft Word form

## METADATA DOCUMENT

### 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/submitdata/Keywords\\_for\\_FS-RDA\\_archive.pdf](https://www.fs.usda.gov/rds/archive/submitdata/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)

# How to prepare files for the FS-RDA

## Step 4: Sign submission forms (required to publish in FS-RDA)

- First author **must** sign form
  - There is a form for Forest Service (FS) authors and another for non-FS authors.
  - FS submission forms also require the signature of a supervisor.
- If first author is not a FS author, then we **also** need the first FS author to sign a submission form.

### Important Notes:

- Electronic signatures are acceptable, please include the email showing the form being “signed” and forwarded.

Forest Service Research Data Archive Submission Form for Forest Service Authors	
Author(s):	
Submitter (if not author):	
Research Station(s):	
Group/Program/Unit(s):	
<i>Data Package Information</i>	
Working title of data package:	
Full citation and status of manuscript(s) if applicable:	
Data package being delivered via:	
Approx. size of data package (GB):	
<i>Handling</i>	
Author – I request that this data set be archived in the: <input type="checkbox"/> Open Public Repository (metadata: public; data set access: public) <input type="checkbox"/> Private Repository (metadata: private; data set access: user completes custom U	
Repository delivers archive data set via: (if applicable)	<input type="checkbox"/> Download <input type="checkbox"/> Link to alternate archive or website (p
Special Instructions: (URL for alternate archive or website, justify private repository request, etc.)	
<i>Approval</i>	
1. Author – I believe the data are correct, and the documentation is correct, complete, comprehensible. This data package is ready for publication by the Forest Service R	
Signature: <input type="text"/>	
2. Science Manager (or Supervisor) – I approve submission of these data and their documentation to the Forest Service Research Data Archive as requested by the author.	
Check One: <input type="checkbox"/> N/A <input type="checkbox"/> Data and/or associated publication contain <b>Influential Scientific Information</b> Scientific information FS R&D has determined will or does have a clear and substantial impact on important public policies or private sector decisions. <input type="checkbox"/> Data and/or associated publication are considered to be a <b>Highly Influential Scientific Assessment</b> Potential impact >\$500 million in any one year on either the public or private sector; dissemination is novel, controversial, or precedent-setting; or dissemination has significant interagency interest.	
Signature: <input type="text"/>	

Forest Service Research Data Archive Submission Form for non-Forest Service Authors	
Author(s):	
Submitter (if not author):	
Organization(s):	
FS Affiliation (if any):	
<i>Data Package Information</i>	
Working title of data package:	
Full citation and status of manuscript(s) if applicable:	
Data package being delivered via:	
Approx. size of data package (GB):	
<i>Handling</i>	
Author – I request that this data set be archived in the: <input type="checkbox"/> Open Public Repository (metadata: public; data set access: public) <input type="checkbox"/> Private Repository (metadata: private; data set access: user completes custom Use Agreement)	
Repository delivers archive data set via: (if applicable)	<input type="checkbox"/> Download <input type="checkbox"/> Link to alternate archive or website (provide description below)
Special Instructions: (URL for alternate archive or website, justify private repository request, etc.)	
<i>Approval</i>	
Author – I believe the data are correct, and the documentation is correct, complete, comprehensive, and comprehensible. This data package is ready for publication by the Forest Service Research Data Archive.	
Signature: <input type="text"/>	Date: <input type="text"/>

# How to prepare files for the FS-RDA

## Step 5: Compile submission package

- Gather all files
  - Data set(s)
  - Metadata
  - Supplemental files
- Use directory structure to help organize files, this is structure we will use
  - “Data” – all data files will go in this folder, subfolders allowed
  - “Supplements” – all supplemental files will go in this folder, subfolders allowed
- Create a file index
  - List files in submission package and short description of their format and content
  - This can be submitted in any format (Excel, Word, text file, email, etc.)

*You are now ready to submit package to the Archive Team!*

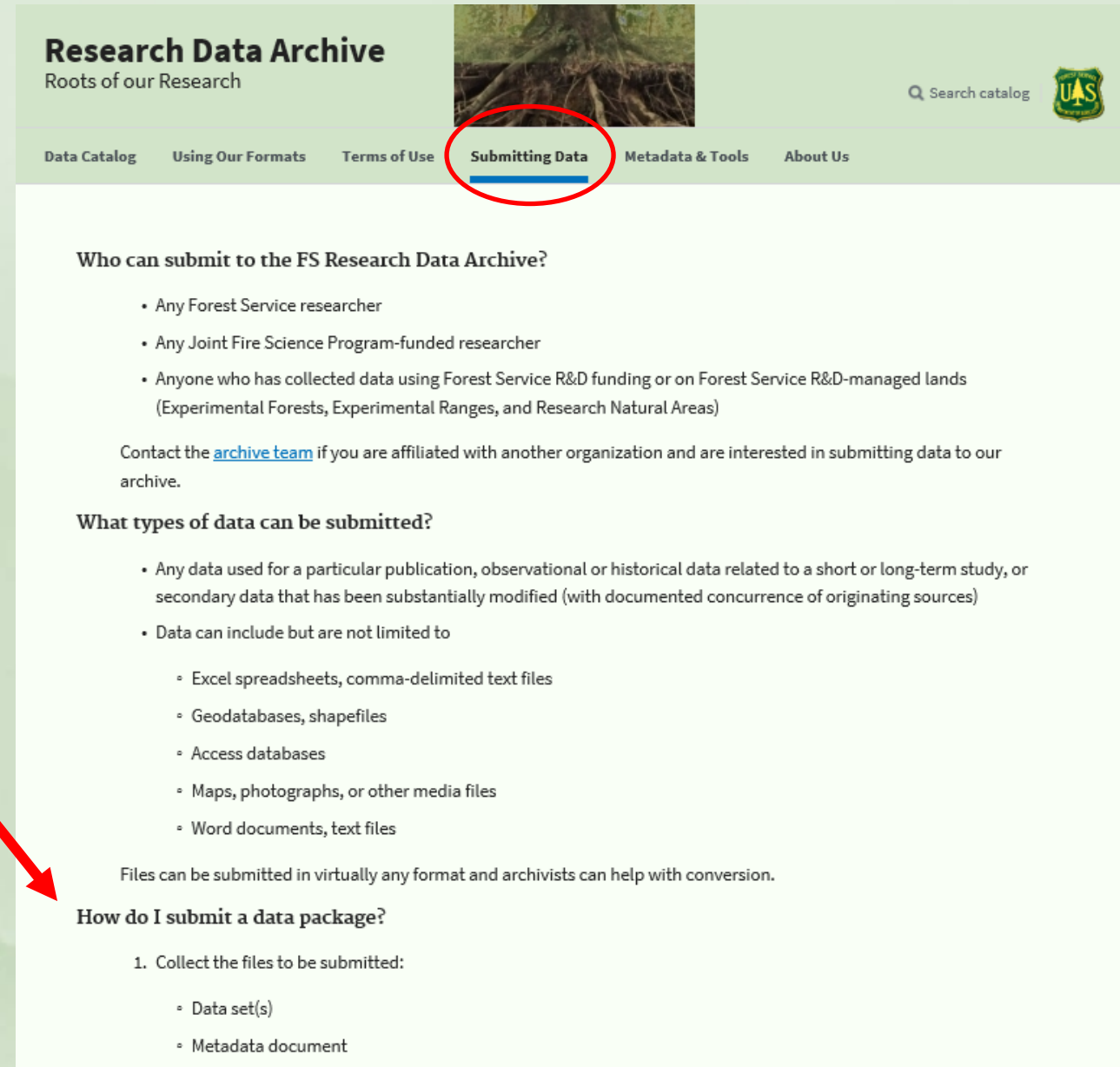
# How to submit and publish data in the FS-RDA

Step 1: Contact a member of the Archive Team and submit package for **thorough** review


Step 2: Archive Team reviews submission and sends authors questions and comments

Step 3: Archive Team finalizes package, assigns DOI, and data are published!

*Reviews help ensure data are usable...*



**Research Data Archive**  
Roots of our Research

Search catalog 

Data Catalog Using Our Formats Terms of Use **Submitting Data** Metadata & Tools About Us

### Who can submit to the FS Research Data Archive?

- Any Forest Service researcher
- Any Joint Fire Science Program-funded researcher
- Anyone who has collected data using Forest Service R&D funding or on Forest Service R&D-managed lands (Experimental Forests, Experimental Ranges, and Research Natural Areas)

Contact the [archive team](#) if you are affiliated with another organization and are interested in submitting data to our archive.

### What types of data can be submitted?

- Any 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)
- Data can include but are not limited to
  - Excel spreadsheets, comma-delimited text files
  - Geodatabases, shapefiles
  - Access databases
  - Maps, photographs, or other media files
  - Word documents, text files

Files can be submitted in virtually any format and archivists can help with conversion.

### How do I submit a data package?

1. Collect the files to be submitted:
  - Data set(s)
  - Metadata document

The background of the slide features a series of rolling, misty mountains in shades of light green and white, creating a serene and atmospheric landscape. The text is centered over this background.

# Additional Info

# Need a DOI early for journal article?

- What is required
  - Signed submission form
  - Draft (or final) version of data
  - Draft (or final) version of metadata
- How it works
  - Submit submission form, data, metadata to Archive Team and request early DOI
  - DOI assigned early, will point to temporary out-of-stock page
  - Get final files to Archive Team, so they can begin review
  - Work with Archive Team to ensure data are published before journal article goes live



# Benefits of publishing in FS-RDA

- FS RGEG direction says FS-RDA publications are refereed!
- Broad audience
  - ≈35,000 global customer visits in FY2020
  - 7,925+ downloads in FY2020 (>25% growth)
  - Domestic customer base composed of government agencies, academic institutions, K-12, variety of businesses
- Other repositories are providing access to FS-RDA
  - Data.gov: catalog for U.S. Government data
  - Ag Data Commons (USDA Nat. Ag. Library): catalog for all USDA scientific data
  - Science.gov: <https://www.science.gov/> (+ Spanish: <https://ciencia.science.gov/>)
  - Data Citation Index:  
<https://clarivate.com/webofsciencegroup/solutions/webofscience-data-citation-index/>