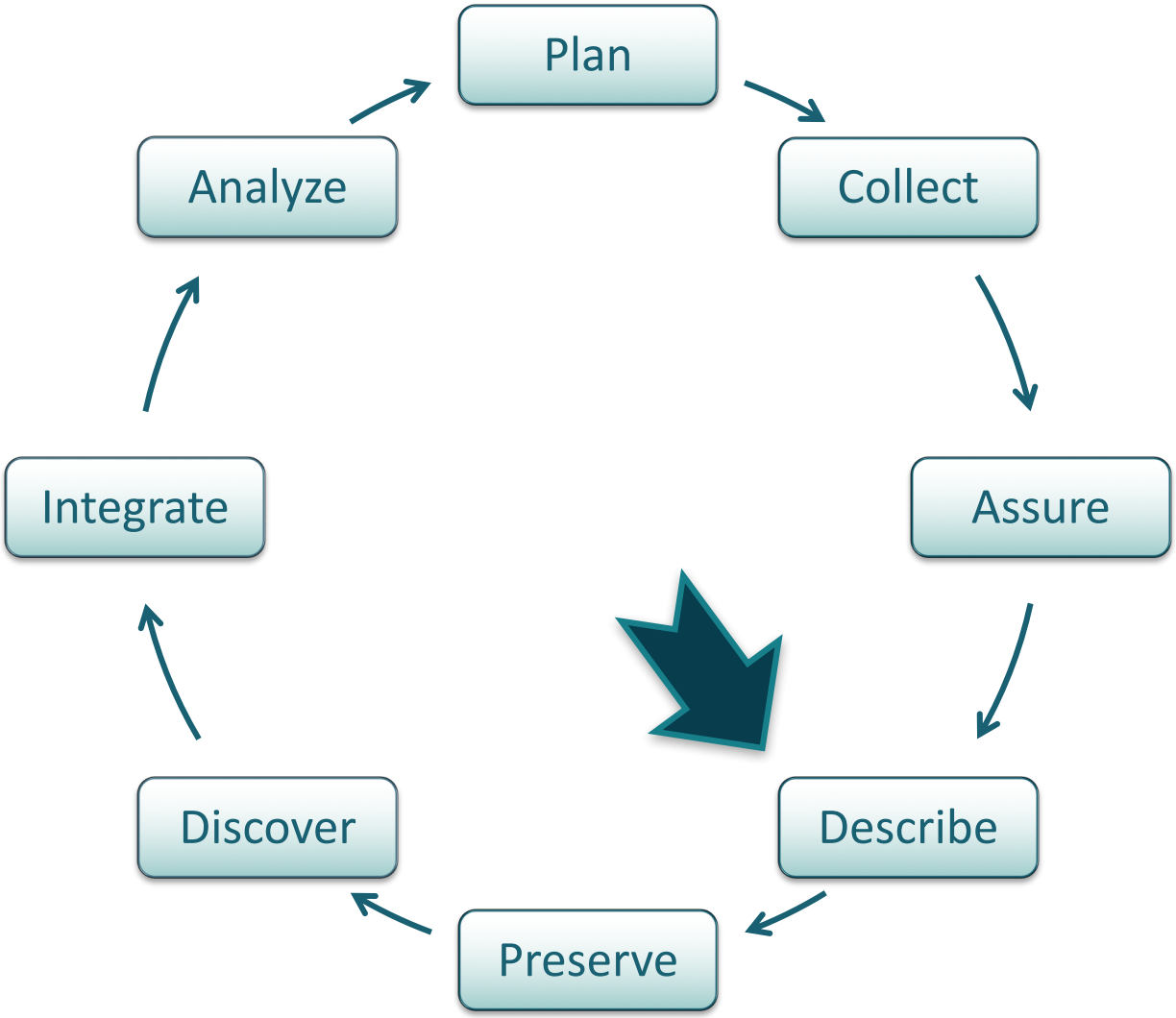
Authoring High Quality Metadata

Jeanette Clark
UC Santa Barbara

orcid.org/0000-0003-4703-1974

DataONE NSF Arctic Data Center NCEAS National Center for Ecological Analysis and Synthesis

The Data Life Cycle



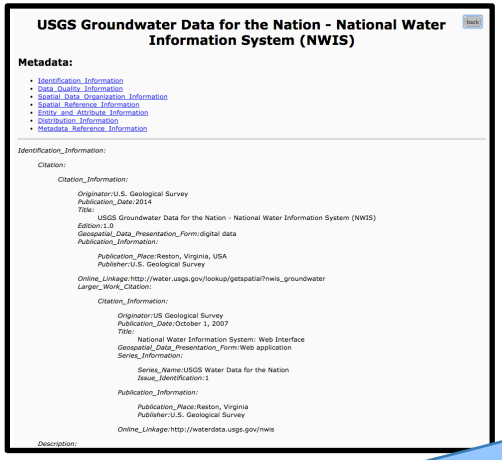
What are metadata?

Think of metadata as "data reporting"

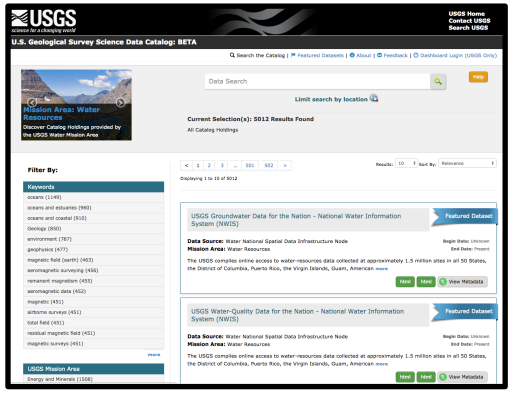
- **WHO** created the data?
- **WHAT** is the content of the data?
- **WHEN** were the data collected?
- **WHERE** are the data from?
- **HOW** were the data developed?
- **WHY** were the data developed?

Why are metadata important?

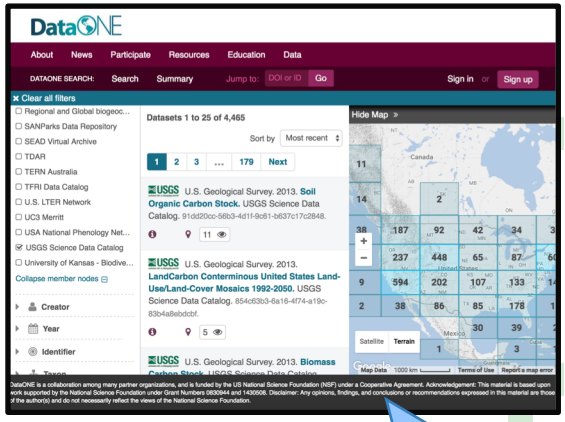
Metadata capture information



USGS Science Data Catalog: enables discovery



DataONE: enables exchange

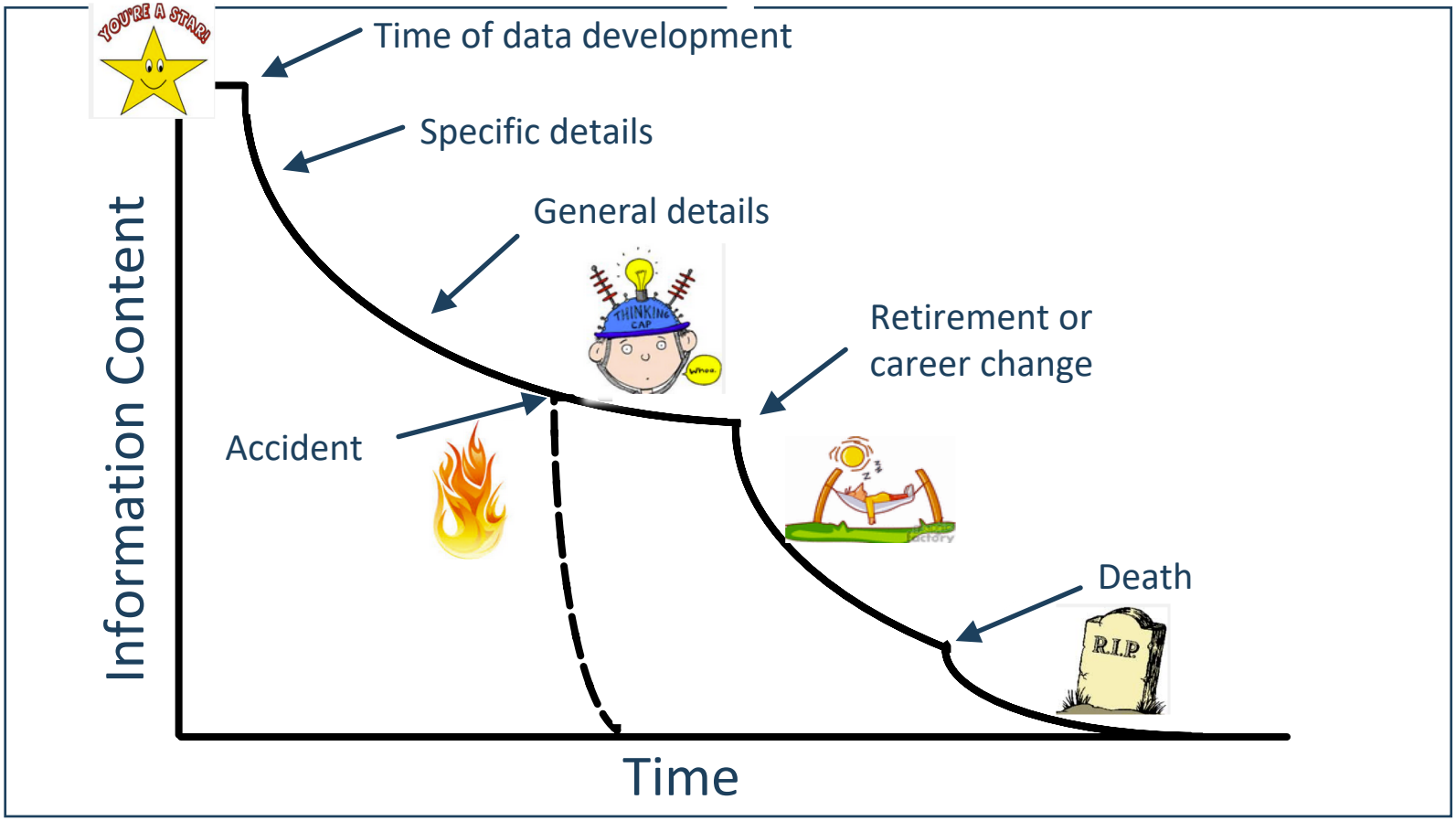


DataONE

NSF Arctic Data Center

NCEAS
National Center for Ecological Analysis and Synthesis

Why are metadata important?



(modified from Michener et al. 1997)

Why are metadata important?

Metadata are important for the short and long-term utility of data

Why are metadata important?

DataONE

NSF Arctic Data Center

NCEAS
National Center for Ecological Analysis and Synthesis

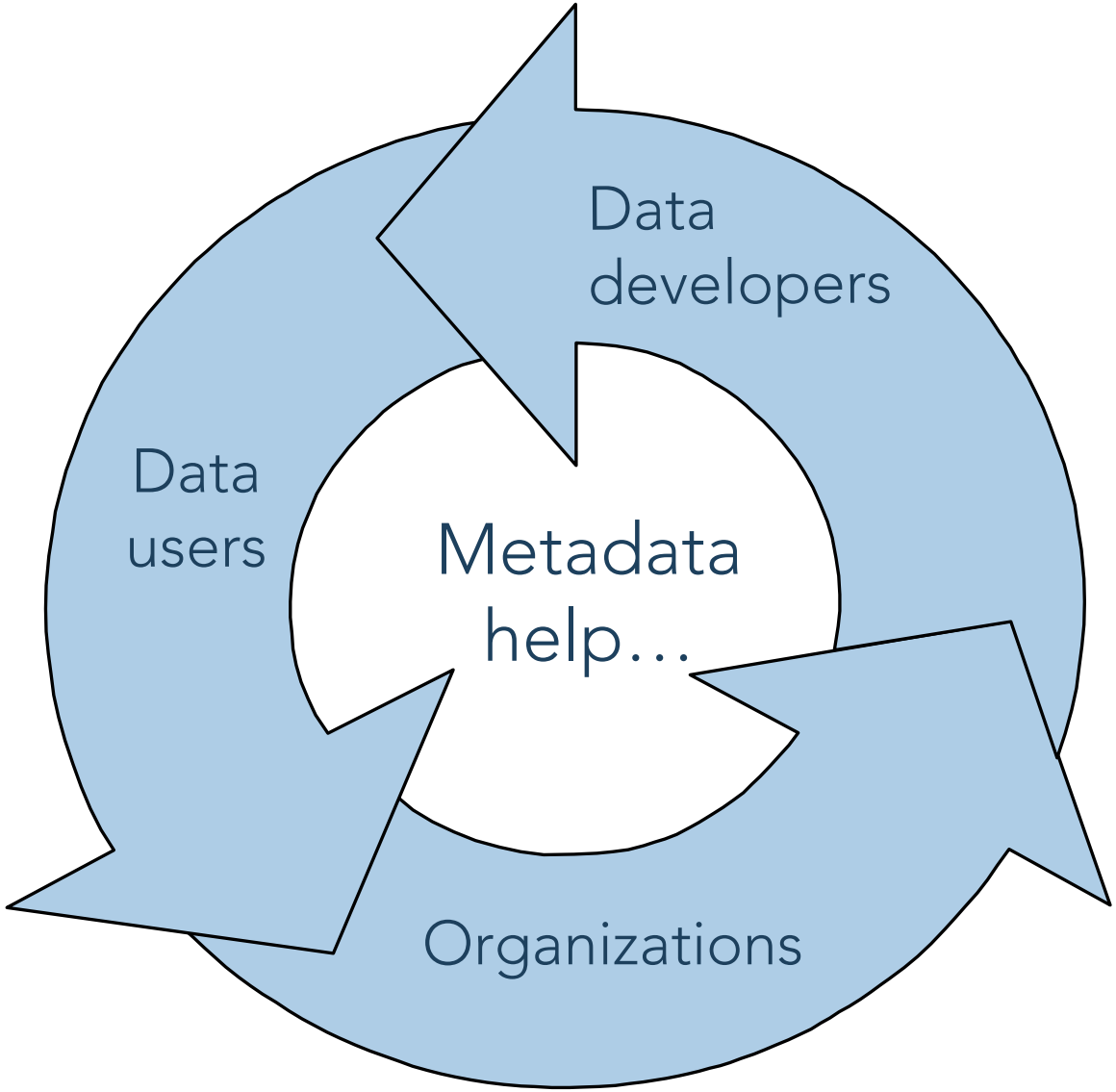
Why are metadata important?

Metadata support scrutiny of data

- Motivations
- Methodologies
- Conflicts of interest

DataONE NSF Arctic Data Center NCEAS National Center for Ecological Analysis and Synthesis

Who uses metadata?



DataONE NSF Arctic Data Center NCEAS National Center for Ecological Analysis and Synthesis

Metadata for data developers

Avoid data duplication

- What data have already been collected?

Share reliable information

- What methods were used?
- What methods are in common use in my field?

Publicize and share work

- "Hey, I made this!"

Metadata for data users

- Find relevant data
- Evaluate what is suitable for use in your work
- Retrieve the data you've found
- Understand if and how to actually use the data

Metadata for organizations

Insure the organization's investment in the data

- Documentation for sampling and data processing methods are recorded
- Ability to use data after initial intended purpose
- Track data re-use and citation

Transcend people and time

- Data are not lost when researchers or labs leave
- Avoid duplication in new work

Advertise organization's research

- What data has our organization produced?

Concerns about creating metadata

Even if the value of data documentation is recognized, researchers are often concerned about the effort required to create metadata that effectively describe their data.

Concerns about creating metadata

Concern	Solution
Workload required to capture accurate robust metadata	Incorporate metadata creation into data development process – distribute the effort
Time and resources to create, manage, and maintain metadata	Include in grant budget and schedule
Readability / usability of metadata	Use a standardized metadata format
Discipline specific information and ontologies	Use a standard 'profile' that supports discipline specific information

Metadata standards

A metadata standard provides a uniform structure to describe data:

- Machine readable (usually XML)
- Common terminology
- Common structure

Metadata standards

Example standards:

- Dublin Core (emphasis on publications)
- Darwin Core (emphasis on collections)
- FGDC (emphasis on spatial data)
- ISO19115 (emphasis on spatial data and services)
- Ecological Metadata Language (general, but emphasis on filesystem artifacts, attributes, taxonomy)

Metadata standards

```
<?xml version="1.0"?>
<eml:eml xmlns:eml="eml://ecoinformatics.org/eml-2.1.1" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlr
xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.1 eml.xsd" packageId="urn:uuid:71bfb8d4-a692-4a08-aaad-d9e1f4
<dataset>
    <title>Example dataset showing metadata submission tools, Ecological Society of America, 2019</title>
    <creator id="1407671562528544">
        <individualName>
            <givenName>Jeanette</givenName>
            <surName>Clark</surName>
        </individualName>
        <organizationName>National Center for Ecological Analysis and Synthesis</organizationName>
        <positionName>Projects Data Coordinator</positionName>
        <electronicMailAddress>jclark@nceas.ucsb.edu</electronicMailAddress>
        <userId directory="https://orci
    </creator>
    <abstract>
        <para>This is an example of how                                                    nter,
        Society of America conference i                                                    for c
        workflow.</para>
    </abstract>
    <intellectualRights>
        <para>This work is dedicated to                                                    main D
        https://creativecommons.org/pub
    </intellectualRights>
    <coverage>
        <geographicCoverage>
            <geographicDescription>Lou
            <boundingCoordinates>
                <westBoundingCoordinate>85.7585</westBoundingCoordinate>
                <eastBoundingCoordinate>85.7585</eastBoundingCoordinate>
                <northBoundingCoordinate>38.2527</northBoundingCoordinate>
                <southBoundingCoordinate>38.2527</southBoundingCoordinate>
            </boundingCoordinates>
        </geographicCoverage>
        <temporalCoverage>
            <rangeOfDates>
```

The 'creator' element provides the full name of the person, organization, or position who created the resource. The list of creators for a resource represent the people and organizations who should be cited for the resource.

What makes a good metadata record?

Overall goal: Could a reasonable scientist make sense of your data in 10, 20, 30+ years without contacting you?

When in doubt, be more specific:

- Spell out acronyms
- Use full names, emails, addresses, etc.

Include as much information as possible directly in the metadata record

What makes a good metadata record?

Target multiple user groups:

- Someone looking directly for your data
- Someone who does not know about your work but should
- Someone looking to scrutinize your work
- Someone trying to reproduce your work
- Someone looking to give you credit for your work

Dataset Title

Good titles include:

- What
- When
- Where

The title is often the first way a user will evaluate your data set

What makes a good metadata record?

Title:

"Ocean water property observations reported from ice-tethered profiler #37, Transpolar Drift, 2009"

DataONE NSF Arctic Data Center NCEAS National Center for Ecological Analysis and Synthesis

Abstract

- Distinct from publication abstract
- Should provide more context for the title
- Should give a high-level summary of methodologies, data formats, coverages, etc.

Data Files

- File formats
- File sizes
- Checksums ("Do I have the same file?")
- Where to download (web address)
- Attributes used (variables)

Involved Parties

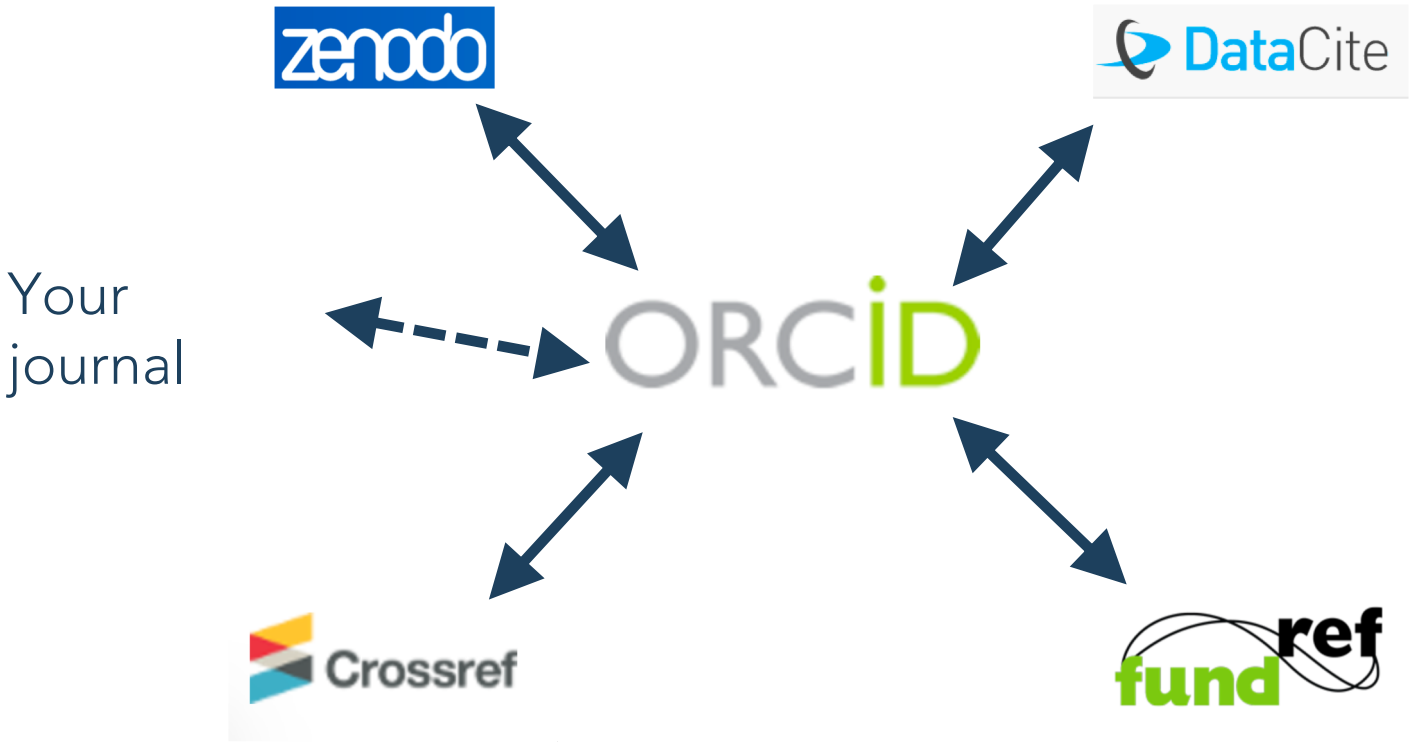
- Name alone is not enough…
  - to assign credit, nor
  - to disambiguate across data sets
- Email addresses help
- Including ORCID is best

ORCID: "Wait, what is an ORCID?"

- Like an ISBN for people
  - e.g. mine: orcid.org/0000-0003-4703-1974
- Enables unambiguous reference to humans
- Free
- Becoming a community norm
- Inherently connected…

ORCID

Inherently connected



Activity

Register an ORCID iD:
https://orcid.org/register

Sign in to
https://test.arcticdata.io/share

Creating a metadata record

Web forms

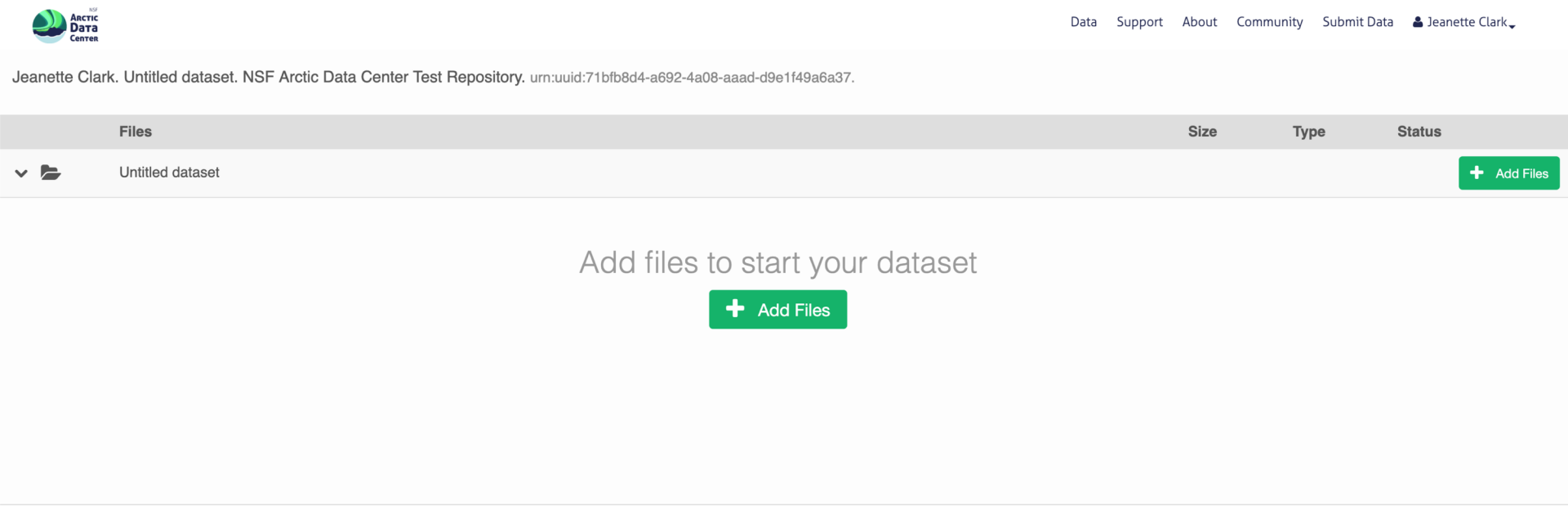
- Frequently linked to a data archive
- eg: https://arcticdata.io/submit

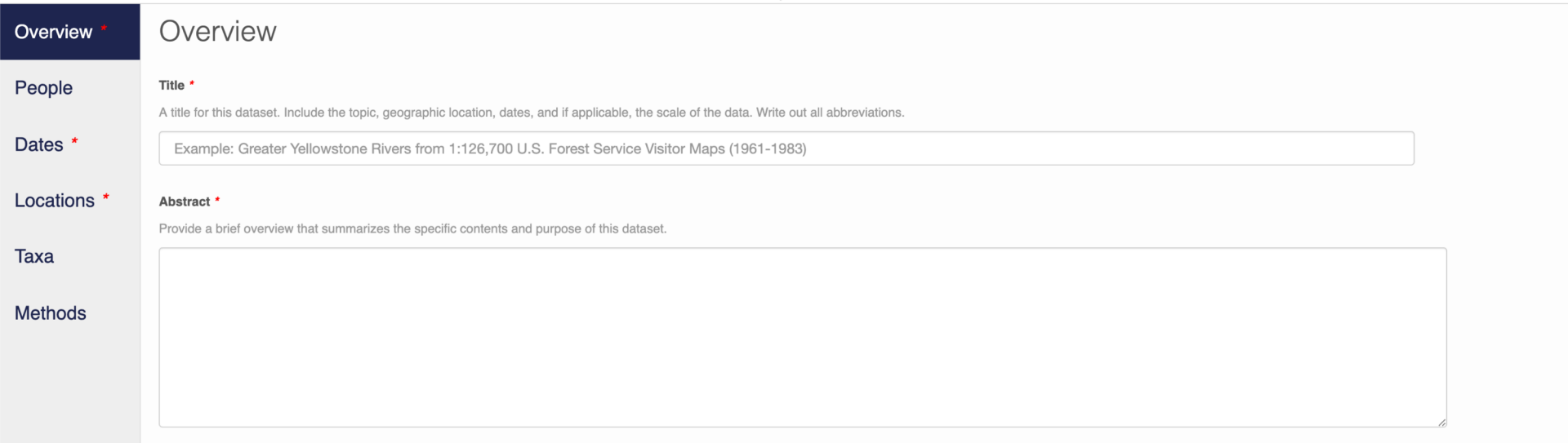
Software packages

- Run locally
- eg: https://github.com/ropensci/EML

Submitting to the Arctic Data Center

- https://test.arcticdata.io/submit





Jeanette Clark. Example dataset showing metadata submission tools, Ecological Society of America, 2019. NSF Arctic Data Center Test Repository. urn:uuid:71bfb8d4-a692-4a08-aaad-d9e1f49a6a37.

Files	Size	Type	Status	
Example dataset showing metadata submission tools, Ecological Society of America, 2019				Add Files
figure1.jpg	159 KB	Data		Describe
example_script.Rmd	21 KB	Data		Describe
dummy_data.csv	62 KB	Data		Describe

Overview *
People
Dates *
Locations *
Taxa
Methods

Overview

Title *

A title for this dataset. Include the topic, geographic location, dates, and if applicable, the scale of the data. Write out all abbreviations.

Example dataset showing metadata submission tools, Ecological Society of America, 2019

Abstract *

Provide a brief overview that summarizes the specific contents and purpose of this dataset.

This is an example of how to write a metadata record and archive data using the Arctic Data Center, for use in a presentation for the Ecological Society of America conference in Louisville, Kentucky, 2019. This dataset contains three dummy files for demonstration purposes, that represent a scientific workflow.

People

Dates *

Locations *

Taxa

Methods

Keywords

Add keywords that accurately categorize these data. Add one keyword per line.

Keyword suggestions for the Earth Sciences can be found on the NASA Global Change Master Directory (GCMD), particularly under Earth Science > Biosphere.

Keyword	Keyword Thesaurus
Add one new keyword	None

Funding *

To link this dataset to your research project, identify the appropriate funding identification. To search for an NSF award, start typing and choose your award from the search result list.

12345

Search for NSF awards by keyword or enter custom funding information

Publication Date

Set a citation date for this data set. This can be a year (YYYY) or an exact date (YYYY-MM-DD).

YYYY-MM-DD or YYYY

Usage Rights *

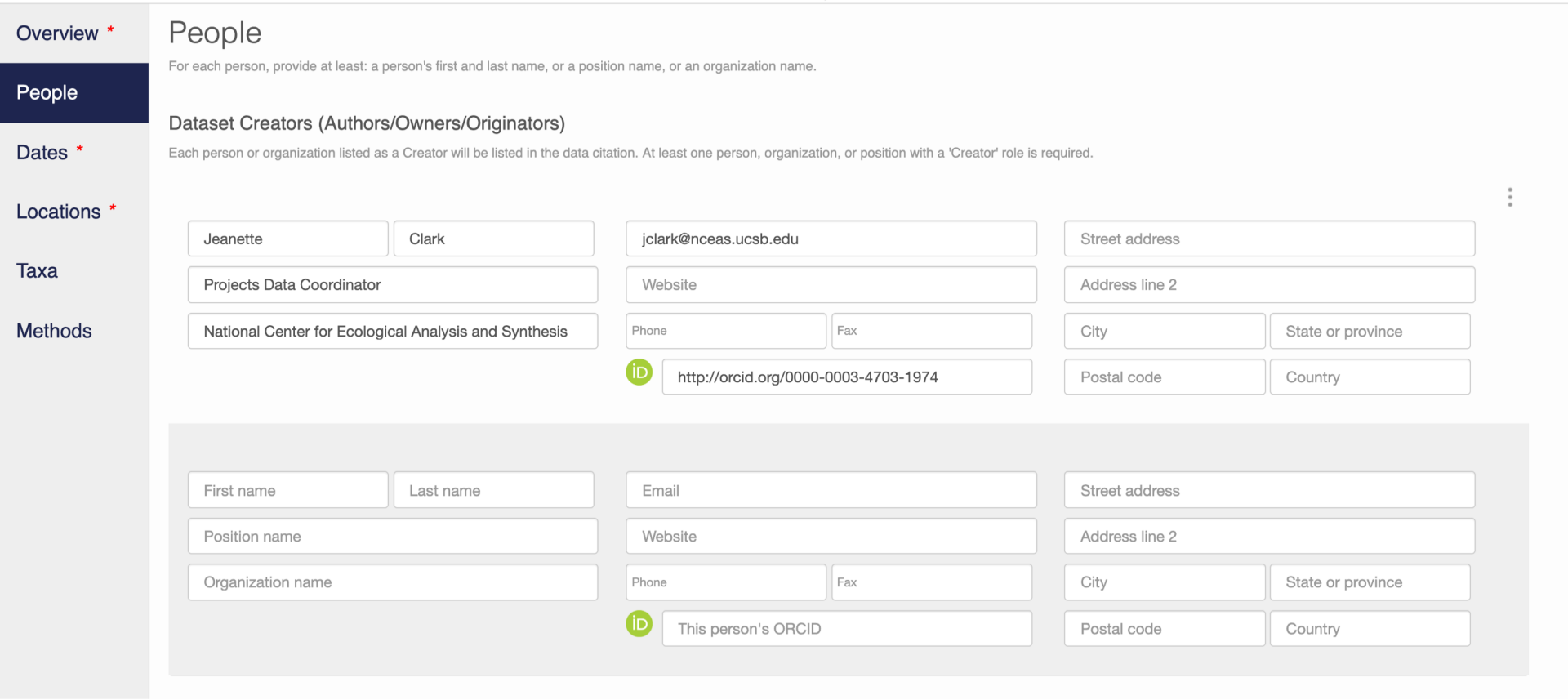
Choose how you wish your data to be shared and reused.

- (•) Creative Commons Public Domain 
- ( ) Creative Commons Attribution 

Alternate Identifiers

If this dataset is archived in another location, such as a different data management system, list any additional identifiers that can be used to locate or label the dataset here.

Add a new alternate identifier


Overview *
People
Dates *
Locations *
Taxa
Methods
People
For each person, provide at least: a person's first and last name, or a position name, or an organization name.
Dataset Creators (Authors/Owners/Originators)
Each person or organization listed as a Creator will be listed in the data citation. At least one person, organization, or position with a 'Creator' role is required.
Jeanette
Clark
jclark@nceas.ucsb.edu
Street address
Projects Data Coordinator
Website
Address line 2
National Center for Ecological Analysis and Synthesis
Phone
Fax
City
State or province
http://orcid.org/0000-0003-4703-1974
Postal code
Country
First name
Last name
Email
Street address
Position name
Website
Address line 2
Organization name
Phone
Fax
City
State or province
This person's ORCID
Postal code
Country

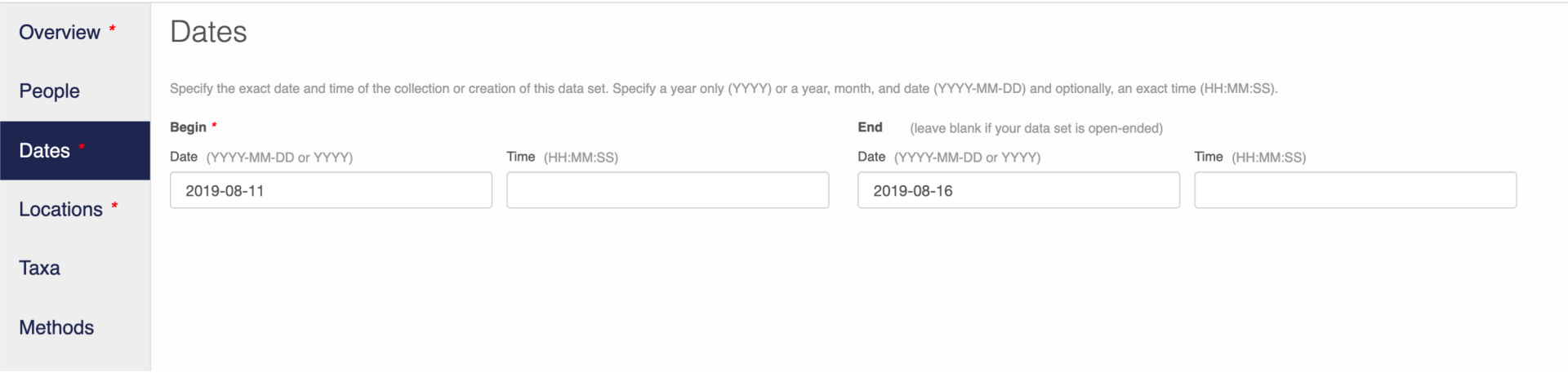
Contacts

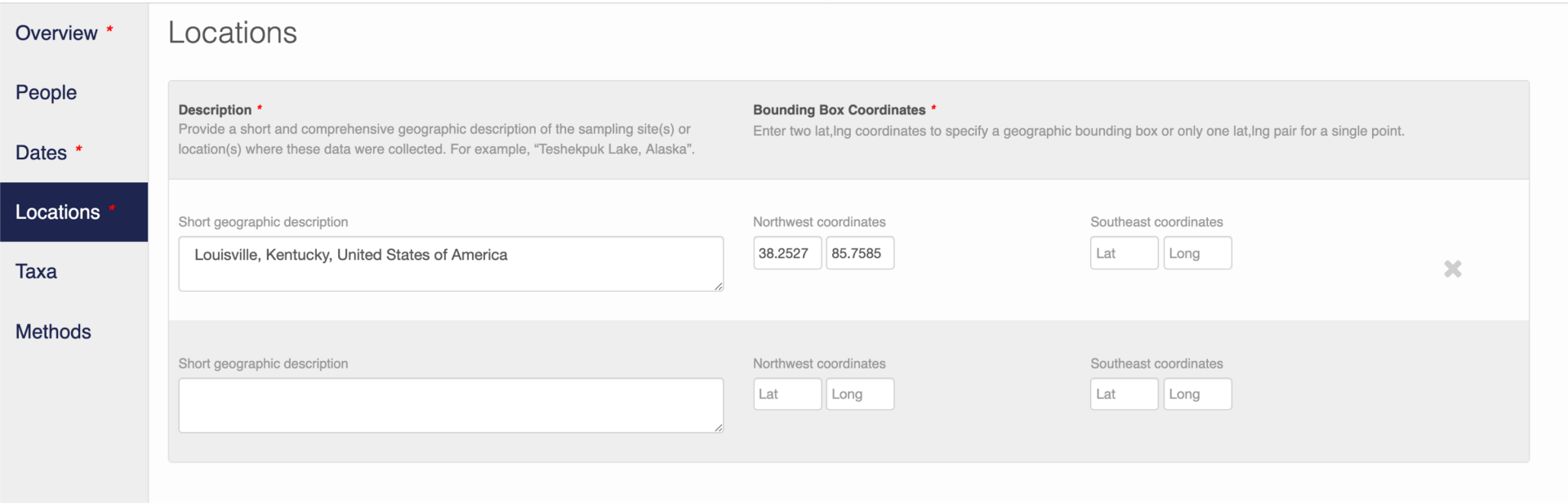
Jeanette | Clark | jclark@nceas.ucsb.edu | Street address
Projects Data Coordinator | Website | Address line 2
National Center for Ecological Analysis and Synthesis | Phone | Fax | City | State or province
http://orcid.org/0000-0003-4703-1974 | Postal code | Country

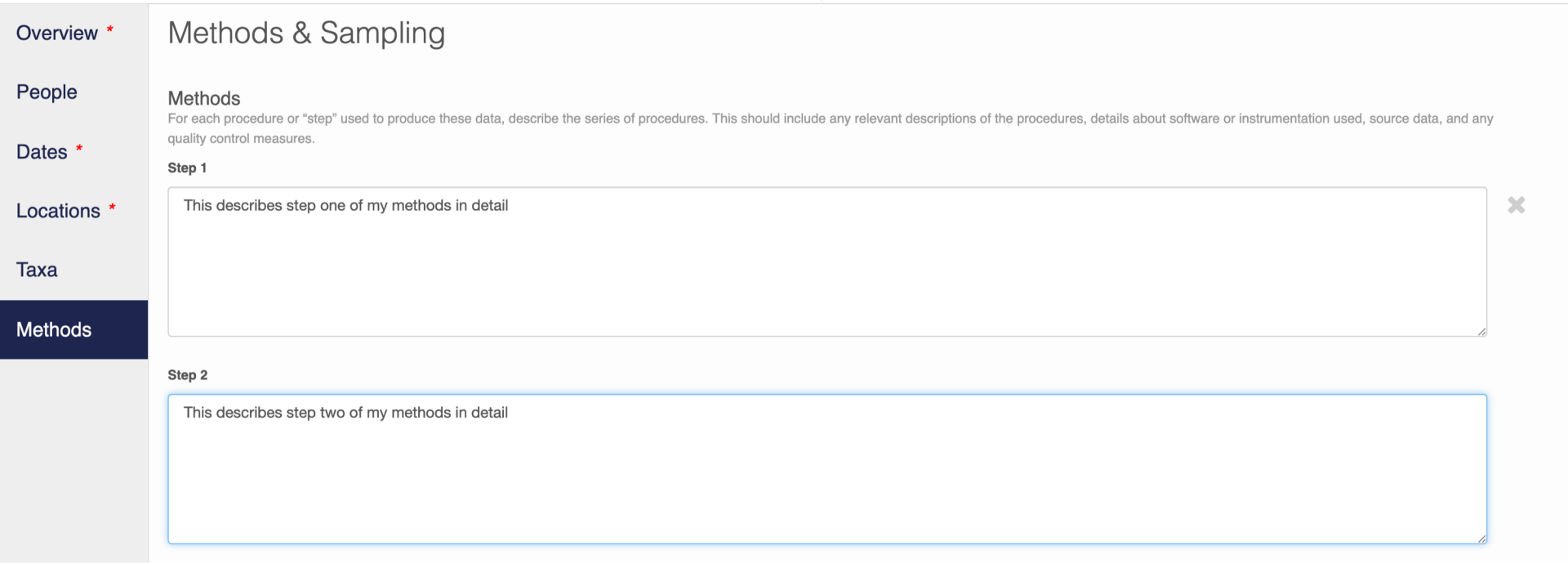
First name | Last name | Email | Street address
Position name | Website | Address line 2
Organization name | Phone | Fax | City | State or province
This person's ORCID | Postal code | Country

Choose new person or organization role ...

First name | Last name | Email | Street address
Position name | Website | Address line 2
Organization name | Phone | Fax | City | State or province
This person's ORCID | Postal code | Country

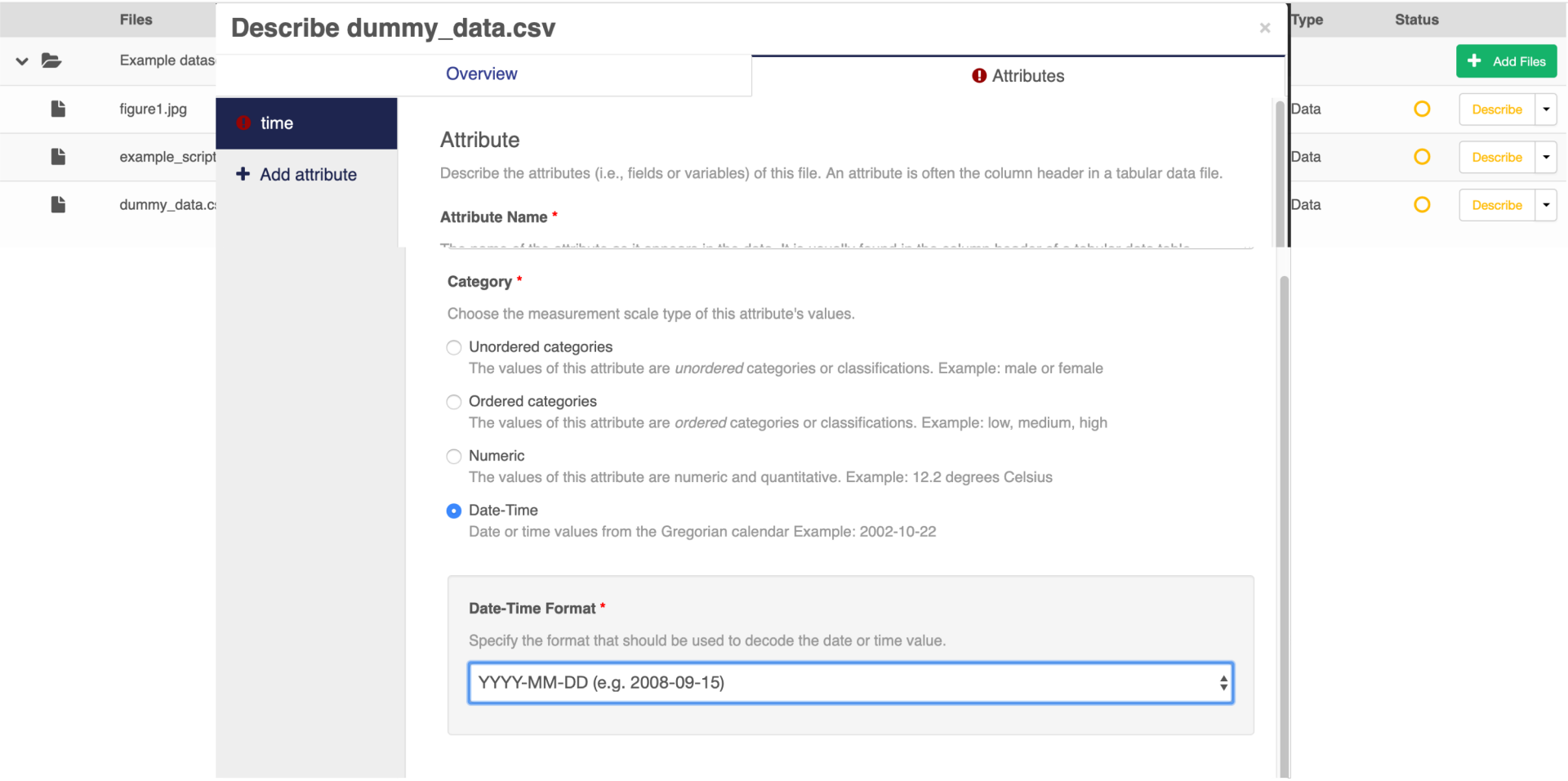

Overview *
People
Dates *
Locations *
Taxa
Methods
Dates
Specify the exact date and time of the collection or creation of this data set. Specify a year only (YYYY) or a year, month, and date (YYYY-MM-DD) and optionally, an exact time (HH:MM:SS).
Begin *
Date (YYYY-MM-DD or YYYY)
2019-08-11
Time (HH:MM:SS)
End (leave blank if your data set is open-ended)
Date (YYYY-MM-DD or YYYY)
2019-08-16
Time (HH:MM:SS)

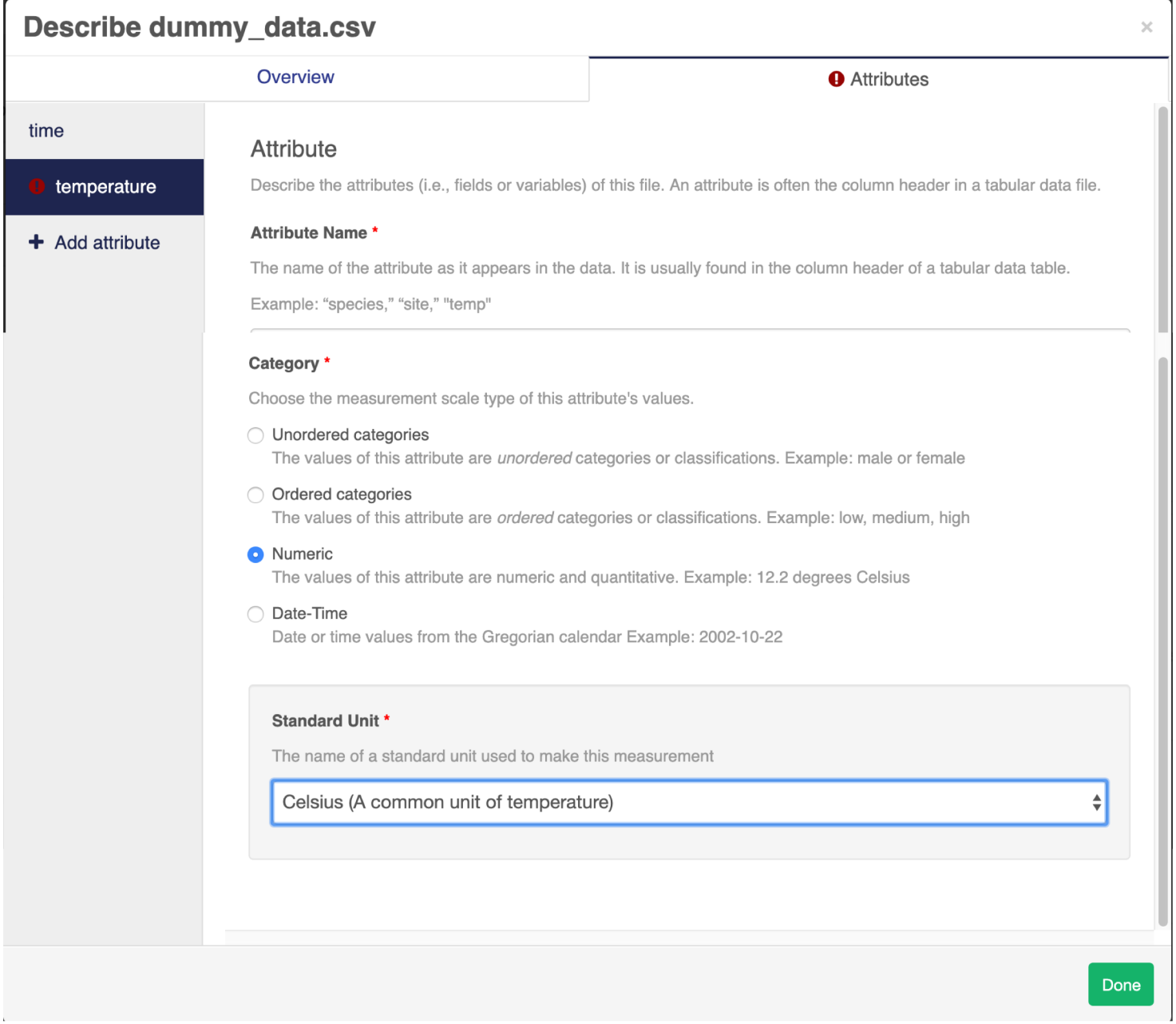

Overview *
People
Dates *
Locations *
Taxa
Methods
Locations
Description *
Provide a short and comprehensive geographic description of the sampling site(s) or location(s) where these data were collected. For example, "Teshekpuk Lake, Alaska".
Bounding Box Coordinates *
Enter two lat,lng coordinates to specify a geographic bounding box or only one lat,lng pair for a single point.
Short geographic description
Louisville, Kentucky, United States of America
Northwest coordinates
38.2527
85.7585
Southeast coordinates
Lat
Long
Short geographic description
Northwest coordinates
Lat
Long
Southeast coordinates
Lat
Long


Overview *
People
Dates *
Locations *
Taxa
Methods
Methods & Sampling
Methods
For each procedure or “step” used to produce these data, describe the series of procedures. This should include any relevant descriptions of the procedures, details about software or instrumentation used, source data, and any quality control measures.
Step 1
This describes step one of my methods in detail
Step 2
This describes step two of my methods in detail

Entity and Attribute Information

Jeanette Clark. Example dataset showing metadata submission tools, Ecological Society of America, 2019. NSF Arctic Data Center Test Repository. urn:uuid:71bfb8d4-a692-4a08-aaad-d9e1f49a6a37.




Describe dummy_data.csv
Overview
Attributes
time
temperature
Add attribute
Attribute
Describe the attributes (i.e., fields or variables) of this file. An attribute is often the column header in a tabular data file.
Attribute Name *
The name of the attribute as it appears in the data. It is usually found in the column header of a tabular data table.
Example: “species,” “site,” "temp"
Category *
Choose the measurement scale type of this attribute's values.
Unordered categories
The values of this attribute are unordered categories or classifications. Example: male or female
Ordered categories
The values of this attribute are ordered categories or classifications. Example: low, medium, high
Numeric
The values of this attribute are numeric and quantitative. Example: 12.2 degrees Celsius
Date-Time
Date or time values from the Gregorian calendar Example: 2002-10-22
Standard Unit *
The name of a standard unit used to make this measurement
Celsius (A common unit of temperature)
Done

Jeanette Clark. Example dataset showing metadata submission tools, Ecological Society of America, 2019. NSF Arctic Data Center Test Repository. urn:uuid:71bfb8d4-a692-4a08-aaad-d9e1f49a6a37.

	Files	Size	Type	Status	
	Example dataset showing metadata submission tools, Ecological Society of America, 2019				Add Files
	figure1.jpg	159 KB	Data		Describe
	example_script.Rmd	21 KB	Data		Describe
	dummy_data.csv	62 KB	Data		Describe

Overview *

People

Dates *

Locations *

Taxa

Methods

Overview

Title *

A title for this dataset. Include the topic, geographic location, dates, and if applicable, the scale of the data. Write out all abbreviations.

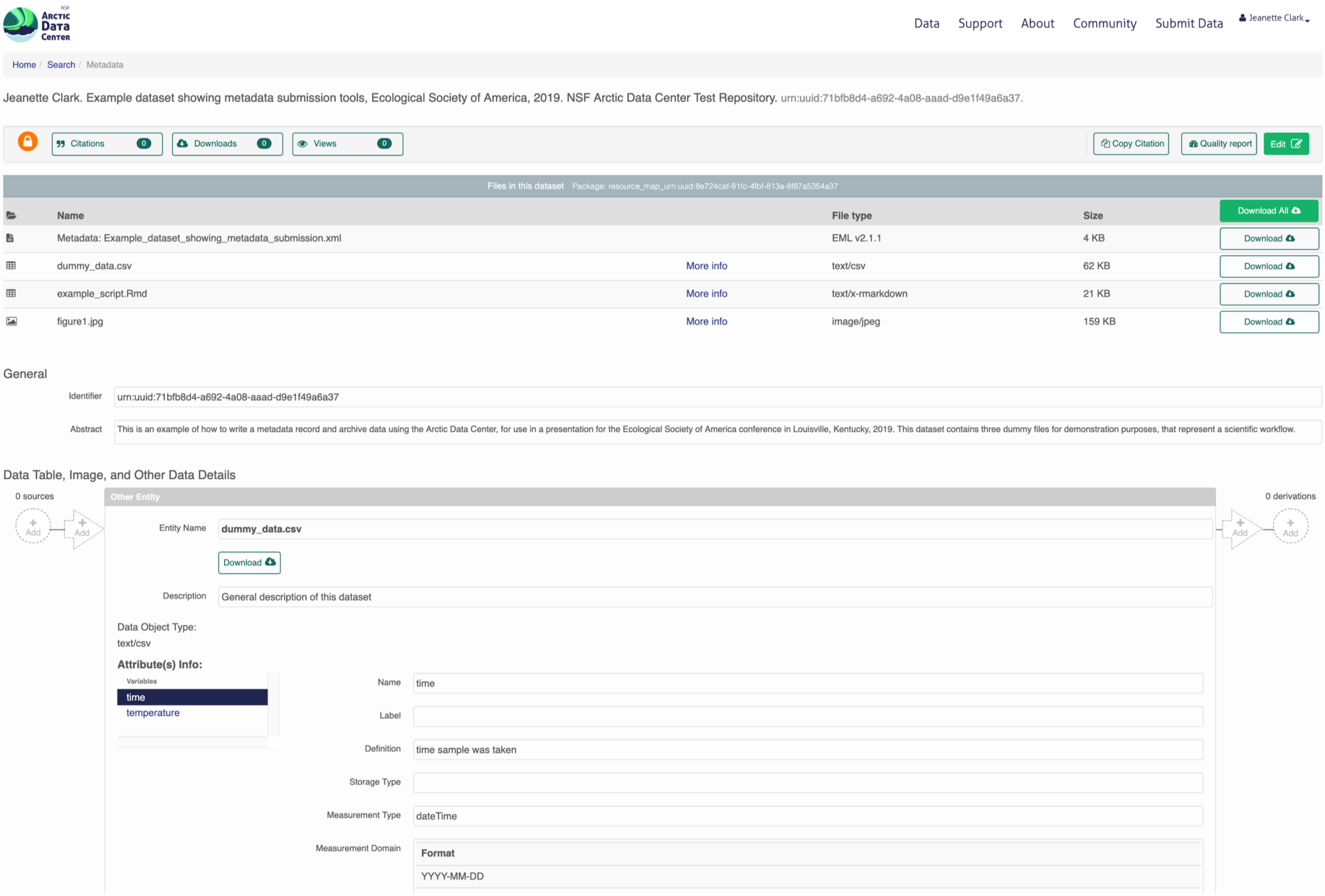
Example dataset showing metadata submission tools, Ecological Society of America, 2019

Abstract *

Provide a brief overview that summarizes the specific contents and purpose of this dataset.

This is an example of how to write a metadata record and archive data using the Arctic Data Center, for use in a presentation for the Ecological Society of America conference in Louisville, Kentucky, 2019. This dataset contains three dummy files for demonstration purposes, that represent a scientific workflow.

Submit Dataset


Arctic Data Center
Data
Support
About
Community
Submit Data
Jeanette Clark
Home / Search / Metadata
Jeanette Clark. Example dataset showing metadata submission tools, Ecological Society of America, 2019. NSF Arctic Data Center Test Repository. urn:uuid:71bfb8d4-a692-4a08-aaad-d9e1f49a6a37.
Citations 0
Downloads 0
Views 0
Copy Citation
Quality report
Edit
Files in this dataset Package: resource_map_urn:uuid:8e724caf-81fc-4fbf-813a-8f87a5354a37
Name
File type
Size
Download All
Metadata: Example_dataset_showing_metadata_submission.xml
EML v2.1.1
4 KB
Download
dummy_data.csv
More info
text/csv
62 KB
Download
example_script.Rmd
More info
text/x-rmarkdown
21 KB
Download
figure1.jpg
More info
image/jpeg
159 KB
Download
General
Identifier
urn:uuid:71bfb8d4-a692-4a08-aaad-d9e1f49a6a37
Abstract
This is an example of how to write a metadata record and archive data using the Arctic Data Center, for use in a presentation for the Ecological Society of America conference in Louisville, Kentucky, 2019. This dataset contains three dummy files for demonstration purposes, that represent a scientific workflow.
Data Table, Image, and Other Data Details
0 sources
Add
Add
Other Entity
0 derivations
Add
Add
Entity Name
dummy_data.csv
Download
Description
General description of this dataset
Data Object Type:
text/csv
Attribute(s) Info:
Variables
time
temperature
Name
time
Label
Definition
time sample was taken
Storage Type
Measurement Type
dateTime
Measurement Domain
Format
YYYY-MM-DD

0 sources

Add Add

Other Entity

1 derivations

</> Add

Entity Name: **dummy_data.csv**

Download

Description: General description of this dataset

Data Object Type:
text/csv

Attribute(s) Info:

Variables
- time
- temperature

Name: time

Label:

Definition: time sample was taken

Storage Type:

Measurement Type: dateTime

Measurement Domain:

Format

YYYY-MM-DD

Precision

Missing Value Code:

Accuracy Report:

Accuracy Assessment:

Coverage:

Method:

1 inputs

Add

Other Entity

1 outputs

Add

Entity Name: **example_script.Rmd**

Download

Data Object Type:
application/octet-stream

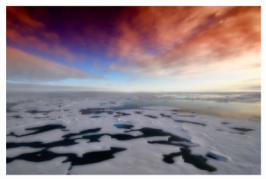
1 sources

Add </>

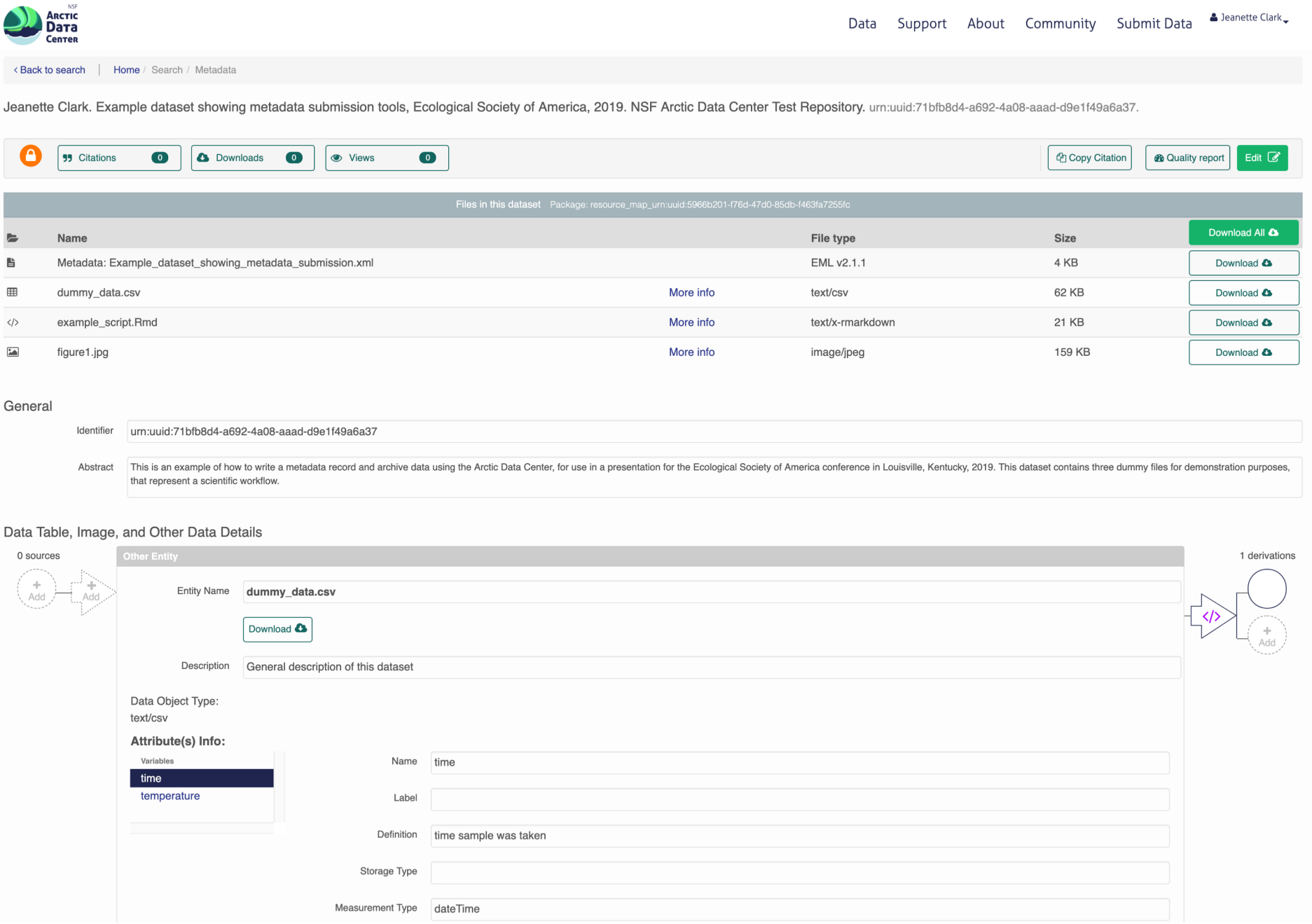
Other Entity

0 derivations

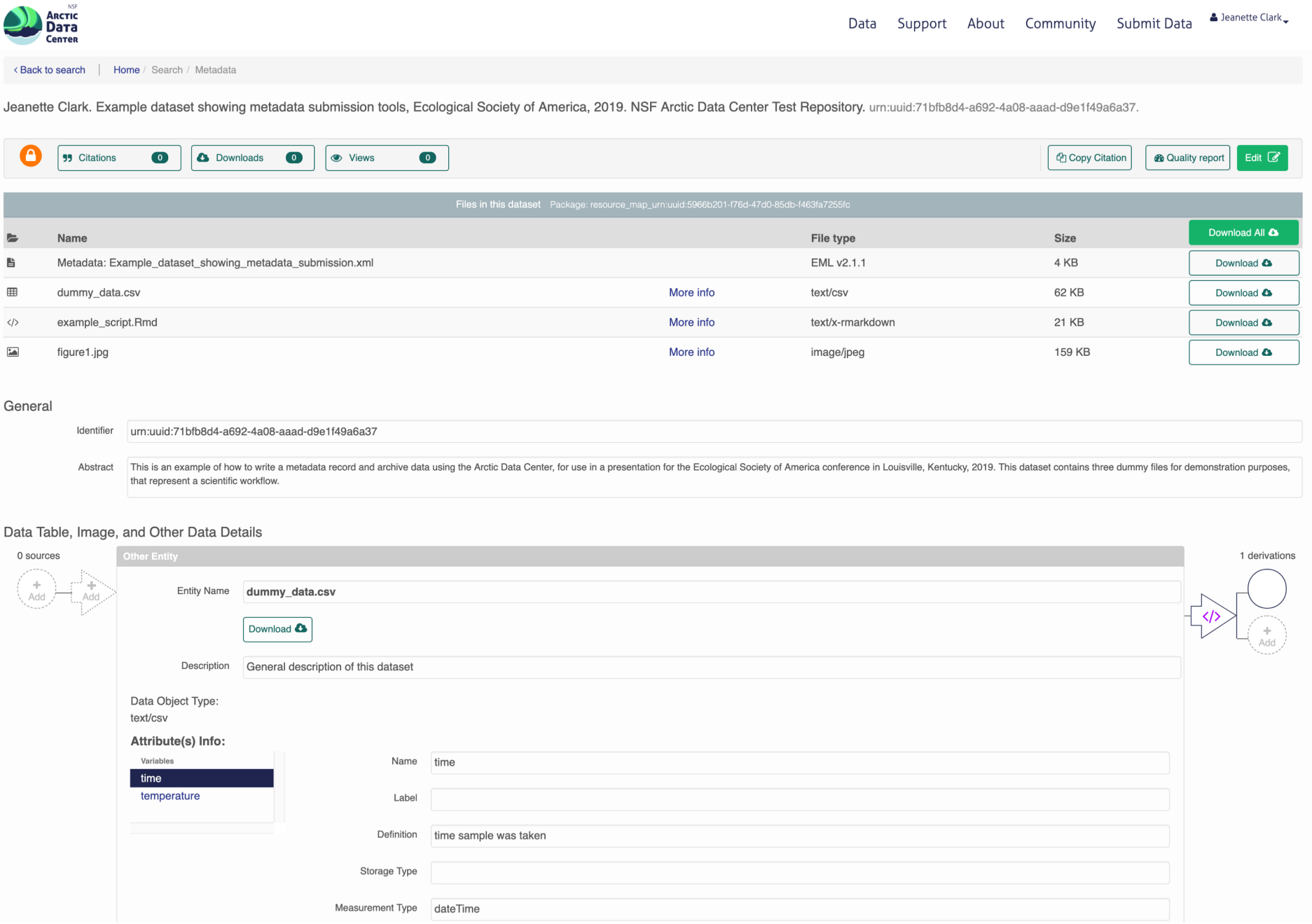
Add Add



Entity Name: **figure1.jpg**

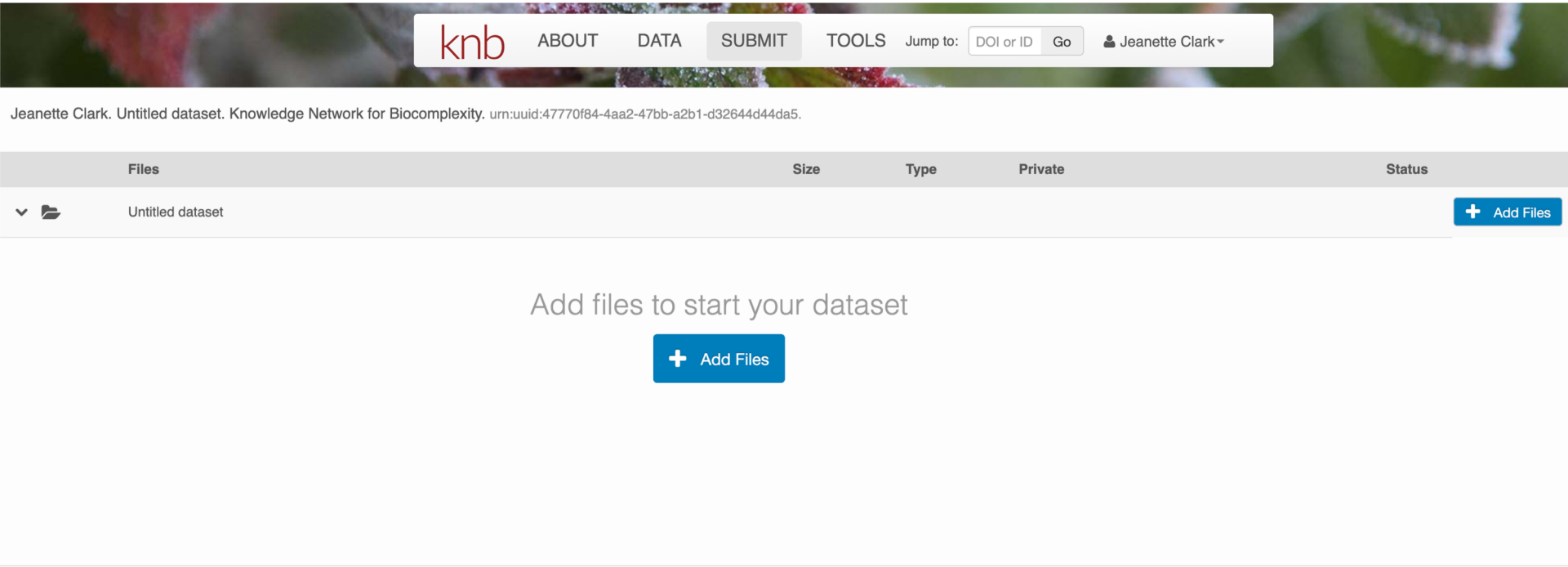

Arctic Data Center
Data
Support
About
Community
Submit Data
Jeanette Clark
Back to search
Home / Search / Metadata
Jeanette Clark. Example dataset showing metadata submission tools, Ecological Society of America, 2019. NSF Arctic Data Center Test Repository. urn:uuid:71bfb8d4-a692-4a08-aaad-d9e1f49a6a37.
Citations 0
Downloads 0
Views 0
Copy Citation
Quality report
Edit
Files in this dataset Package: resource_map_urn:uuid:5966b201-f76d-47d0-85db-f463fa7255fc
Name
File type
Size
Download All
Metadata: Example_dataset_showing_metadata_submission.xml
EML v2.1.1
4 KB
Download
dummy_data.csv
More info
text/csv
62 KB
Download
example_script.Rmd
More info
text/x-rmarkdown
21 KB
Download
figure1.jpg
More info
image/jpeg
159 KB
Download
General
Identifier
urn:uuid:71bfb8d4-a692-4a08-aaad-d9e1f49a6a37
Abstract
This is an example of how to write a metadata record and archive data using the Arctic Data Center, for use in a presentation for the Ecological Society of America conference in Louisville, Kentucky, 2019. This dataset contains three dummy files for demonstration purposes, that represent a scientific workflow.
Data Table, Image, and Other Data Details
0 sources
Add
Add
Other Entity
Entity Name
dummy_data.csv
Download
Description
General description of this dataset
Data Object Type:
text/csv
Attribute(s) Info:
Variables
time
temperature
Name
time
Label
Definition
time sample was taken
Storage Type
Measurement Type
dateTime
1 derivations
Add

```
<?xml version="1.0"?>
<eml:eml xmlns:eml="eml://ecoinformatics.org/eml-2.1.1" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlr
xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.1 eml.xsd" packageId="urn:uuid:71bfb8d4-a692-4a08-aaad-d9e1f4
<dataset>
    <title>Example dataset showing metadata submission tools, Ecological Society of America, 2019</title>
    <creator id="1407671562528544">
        <individualName>
            <givenName>Jeanette</givenName>
            <surName>Clark</surName>
        </individualName>
        <organizationName>National Center for Ecological Analysis and Synthesis</organizationName>
        <positionName>Projects Data Coordinator</positionName>
        <electronicMailAddress>jclark@nceas.ucsb.edu</electronicMailAddress>
        <userId directory="https://orcid.org">https://orcid.org/0000-0003-4703-1974</userId>
    </creator>
    <abstract>
        <para>This is an example of how to write a metadata record and archive data using the Arctic Data Center,
        Society of America conference in Louisville, Kentucky, 2019. This dataset contains three dummy files for d
        workflow.</para>
    </abstract>
    <intellectualRights>
        <para>This work is dedicated to the public domain under the Creative Commons Universal 1.0 Public Domain D
        https://creativecommons.org/publicdomain/zero/1.0/.</para>
    </intellectualRights>
    <coverage>
        <geographicCoverage>
            <geographicDescription>Louisville, Kentucky, United States of America</geographicDescription>
            <boundingCoordinates>
                <westBoundingCoordinate>85.7585</westBoundingCoordinate>
                <eastBoundingCoordinate>85.7585</eastBoundingCoordinate>
                <northBoundingCoordinate>38.2527</northBoundingCoordinate>
                <southBoundingCoordinate>38.2527</southBoundingCoordinate>
            </boundingCoordinates>
        </geographicCoverage>
        <temporalCoverage>
            <rangeOfDates>
```


Arctic Data Center
Data
Support
About
Community
Submit Data
Jeanette Clark
Back to search
Home / Search / Metadata
Jeanette Clark. Example dataset showing metadata submission tools, Ecological Society of America, 2019. NSF Arctic Data Center Test Repository. urn:uuid:71bfb8d4-a692-4a08-aaad-d9e1f49a6a37.
Citations 0
Downloads 0
Views 0
Copy Citation
Quality report
Edit
Files in this dataset Package: resource_map_urn:uuid:5966b201-f76d-47d0-85db-f463fa7255fc
Name
File type
Size
Download All
Metadata: Example_dataset_showing_metadata_submission.xml
EML v2.1.1
4 KB
Download
dummy_data.csv
More info
text/csv
62 KB
Download
example_script.Rmd
More info
text/x-rmarkdown
21 KB
Download
figure1.jpg
More info
image/jpeg
159 KB
Download
General
Identifier
urn:uuid:71bfb8d4-a692-4a08-aaad-d9e1f49a6a37
Abstract
This is an example of how to write a metadata record and archive data using the Arctic Data Center, for use in a presentation for the Ecological Society of America conference in Louisville, Kentucky, 2019. This dataset contains three dummy files for demonstration purposes, that represent a scientific workflow.
Data Table, Image, and Other Data Details
0 sources
Add
Add
Other Entity
Entity Name
dummy_data.csv
Download
Description
General description of this dataset
Data Object Type:
text/csv
Attribute(s) Info:
Variables
time
temperature
Name
time
Label
Definition
time sample was taken
Storage Type
Measurement Type
dateTime
1 derivations
Add

Knowledge Network For Biocomplexity

- https://knb.ecoinformatics.org/submit

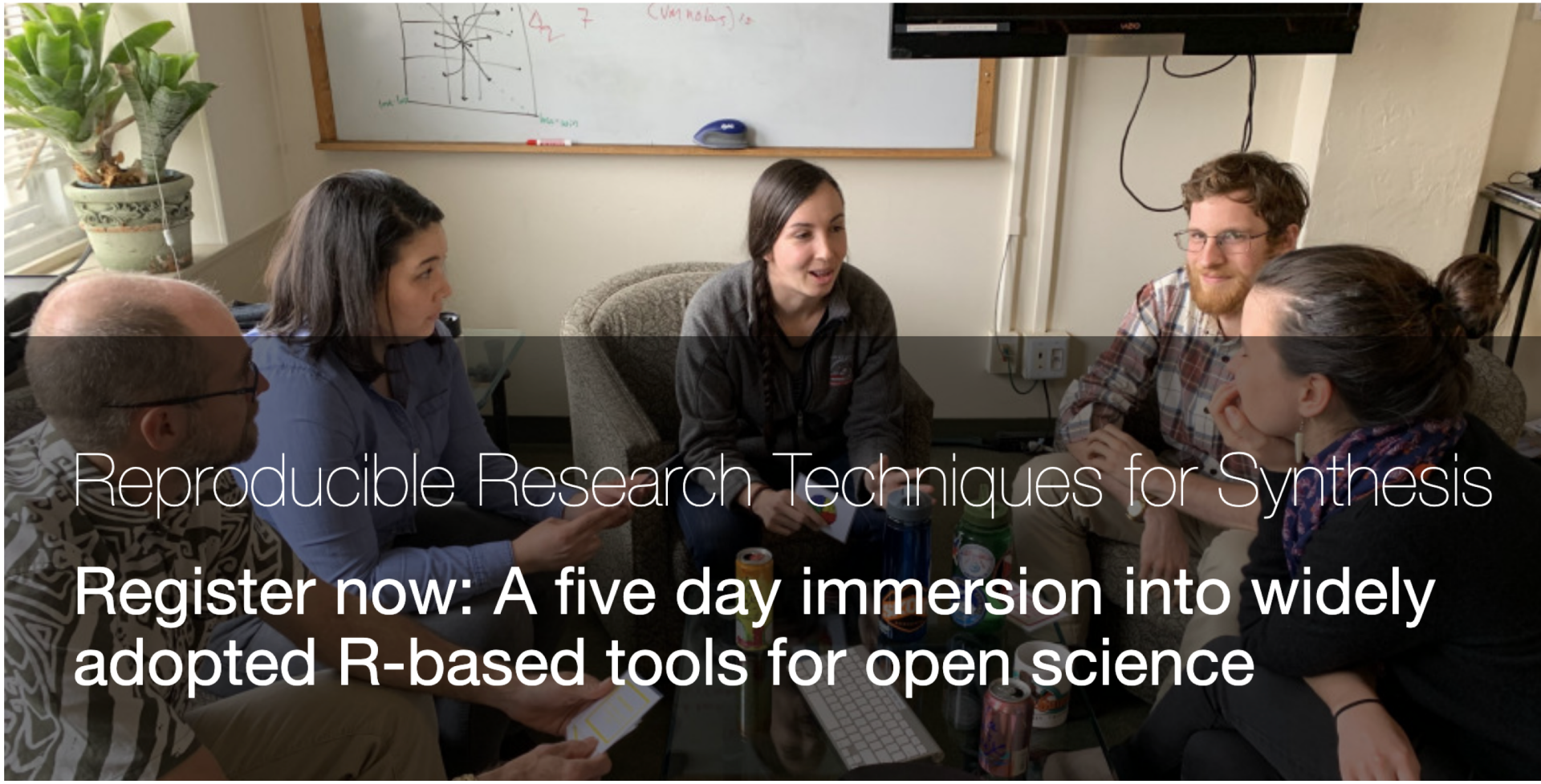


For more information…

Data Science Help Desk

Exhibitors Hall

Booth 219



https://nceas.ucsb.edu/learning-hub/short-course