



# NSF Arctic Data Center

## Authoring High Quality Metadata

**Bryce Mecum**  
Scientific Software Engineer  
**NCEAS**  
[orcid.org/0000-0002-0381-3766](https://orcid.org/0000-0002-0381-3766)

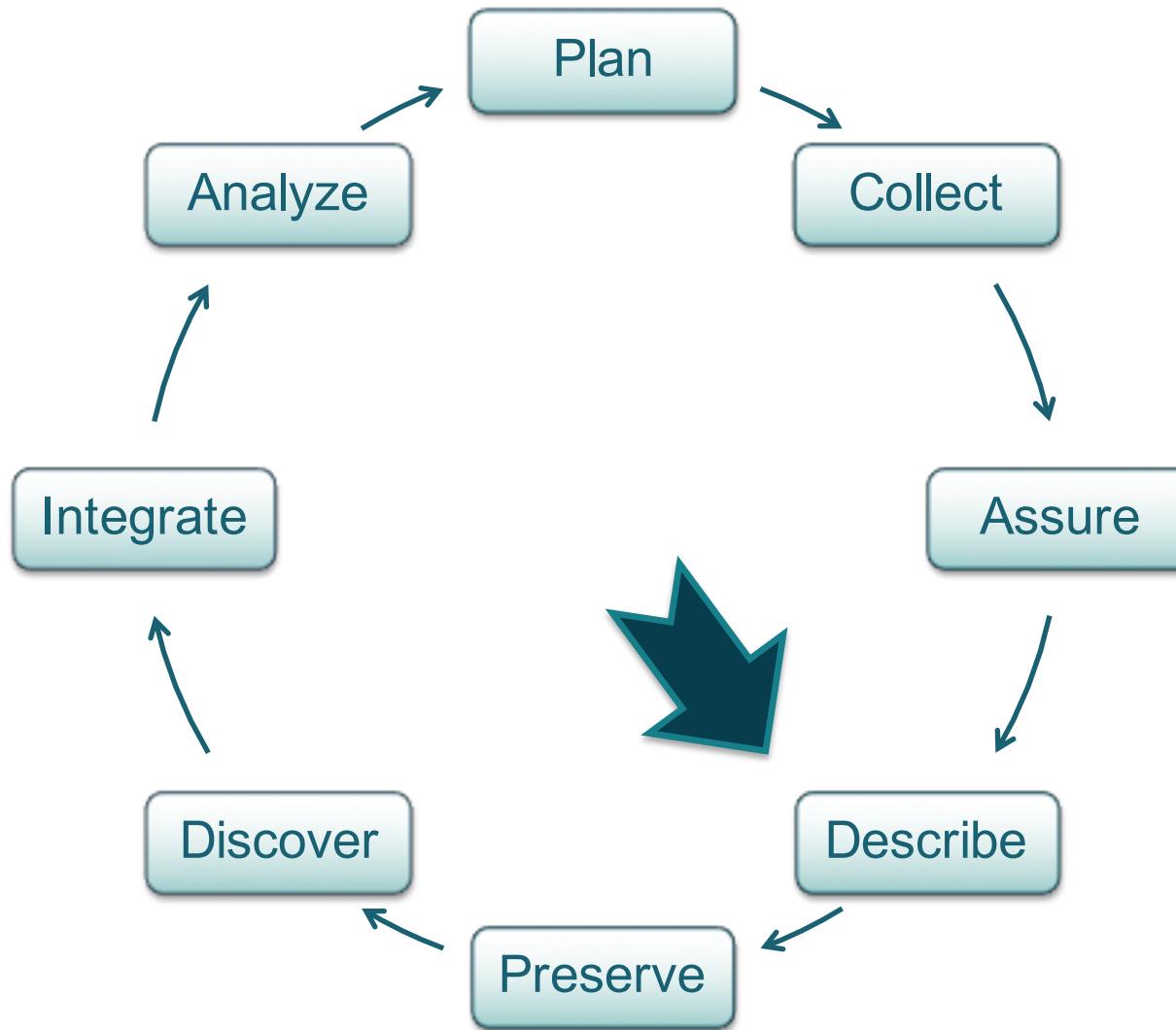


<https://arcticdata.io>  
NSF Award #: 1546024





# The Data Life Cycle





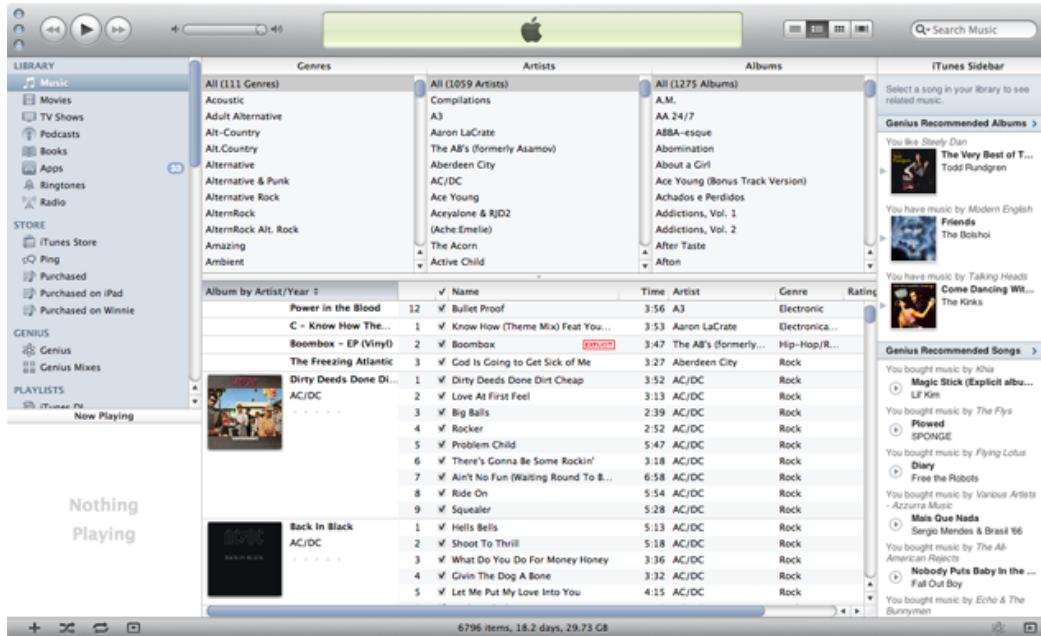
# What is metadata?

Think of metadata as “data reporting”

- **Who** created the data?
- **What** is the content of the data?
- **When** were the data created?
- **Where** are the data from?
- **How** were the data developed?
- **Why** were the data developed?



# Metadata powers our world



<https://arstechnica.com/gadgets/2011/07/mac-os-x-lion-a-visual-introduction/>

<b>Author(s)</b>	Boullosa, Carmen.
<b>Title(s)</b>	They're cows, we're pigs / by Carmen Boullosa
<b>Place</b>	New York : Grove Press, 1997.
<b>Physical Descr</b>	viii, 180 p ; 22 cm.
<b>Subject(s)</b>	Pirates Caribbean Area Fiction.
<b>Format</b>	Fiction



CC image by USDAOgov on Flickr



# What are metadata good for?

An example...

**USGS Groundwater Data for the Nation - National Water Information System (NWIS)**

**Metadata:**

- Identification\_Information
- Data\_Quality\_Information
- Spatial\_Bounding\_Operation\_Information
- Temporal\_Bounding\_Information
- Entity\_and\_Attribute\_Information
- Relationship\_Information
- Usage\_Information
- Metadata\_Reference\_Information

**Identification\_Information:**

Citation:  
Originator: U.S. Geological Survey  
Publication\_Date: 2014  
Title: USGS Groundwater Data for the Nation - National Water Information System (NWIS)  
Edition: 1.0  
Geographic\_Data\_Presentation\_Form: digital data  
Publisher\_Information:  
Publication\_Place: Reston, Virginia, USA  
Publisher: U.S. Geological Survey  
Online\_Linkage: [http://waterdata.usgs.gov/lookup/gspatial/nwis\\_groundwater](http://waterdata.usgs.gov/lookup/gspatial/nwis_groundwater)  
Language: English  
Citation:  
Originator: US Geological Survey  
Publication\_Date: October 1, 2007  
Title: National Water Information System: Web Interface  
Geographic\_Data\_Presentation\_Form: Web application  
Series\_Identifier:  
Series\_Name: USGS Water Data for the Nation  
Issue\_Identifier: 1  
Publication\_Place: Reston, Virginia  
Publisher: U.S. Geological Survey  
Online\_Linkage: <http://waterdata.usgs.gov/nwis>

**Description:**

**USGS Groundwater Data for the Nation - National Water Information System (NWIS)**

**Filter By:**

**Keywords:**  
oceans (193)  
oceans and coastal (940)  
oceans and coastal (910)  
deposits (93)  
minerals (179)  
geophysics (477)  
magnetic field (343)  
magnetic field (343)  
geomagnetic field (45)  
geomagnetic field (45)  
aeromagnetic data (452)  
magnetic (41)  
altimetry (41)  
total field (41)  
residual magnetic field (41)  
magnetic surveys (41)

**USGS Mission Area:**  
Energy and Minerals (1508)

**USGS Groundwater Data for the Nation - National Water Information System (NWIS)**  
**Mission Area:** Water Resources  
The USGS compiles online access to water-resources data collected at approximately 1.5 million sites in all 50 States, the District of Columbia, Puerto Rico, the Virgin Islands, Guam, American Samoa, and the Commonwealth of the Northern Mariana Islands.

**USGS Water-Quality Data for the Nation - National Water Information System (NWIS)**  
**Mission Area:** Water Resources  
The USGS compiles online access to water-resources data collected at approximately 1.5 million sites in all 50 States, the District of Columbia, Puerto Rico, the Virgin Islands, Guam, American Samoa, and the Commonwealth of the Northern Mariana Islands.

**DataONE**

**About News Participate Resources Education Data**

**DATACINE SEARCH:** Search Summary Jump to: DOI or ID Go

**Sign in or Sign up**

**Clear all filters**

Regional and Global biogeoc...  
 SANPath Data Repository  
 SEAD Virtual Archive  
 TDAR  
 TERI Australis  
 TERRI Data Catalog  
 U.S. LTER Network  
 UC3 Merritt  
 US National Phenology Net...  
 USGS Science Data Catalog  
 University of Kansas - Biodive...  
 Collapse member nodes ▾

**1 2 3 ... 179 Next**

**USGS** U.S. Geological Survey. 2013. **Soil Organic Carbon Stock**. USGS Science Data Catalog. 916d22c0-5603-4d11-9eb1-bc307c17c2848. [View Metadata](#)

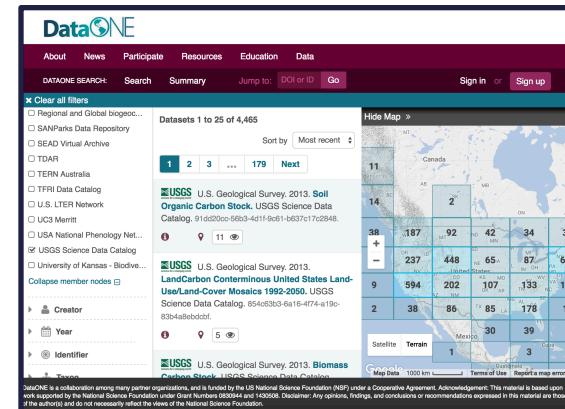
**USGS** U.S. Geological Survey. 2013. **LandCarbon Contiguous United States Land-Use/Land-Cover Mosaics 1990-2005**. USGS Science Data Catalog. 854d83b3-6a16-4f74-e19c-83b4e8ebddc. [View Metadata](#)

**USGS** U.S. Geological Survey. 2013. **Biomass Carbon Stock**. USGS Science Data Catalog. [View Metadata](#)

DataONE is a collaboration among many partner organizations, and is funded by the US National Science Foundation (NSP) under a Cooperative Agreement. Acknowledgment: This material is based upon work supported by the National Science Foundation under Grants 0830806 and 1405008. Disclaimer: Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the view of the National Science Foundation.

**Hide Map**

**Map** Satellite Terrain **Zoom** **Map Data** 100 km **Terms of Use** Report a map error



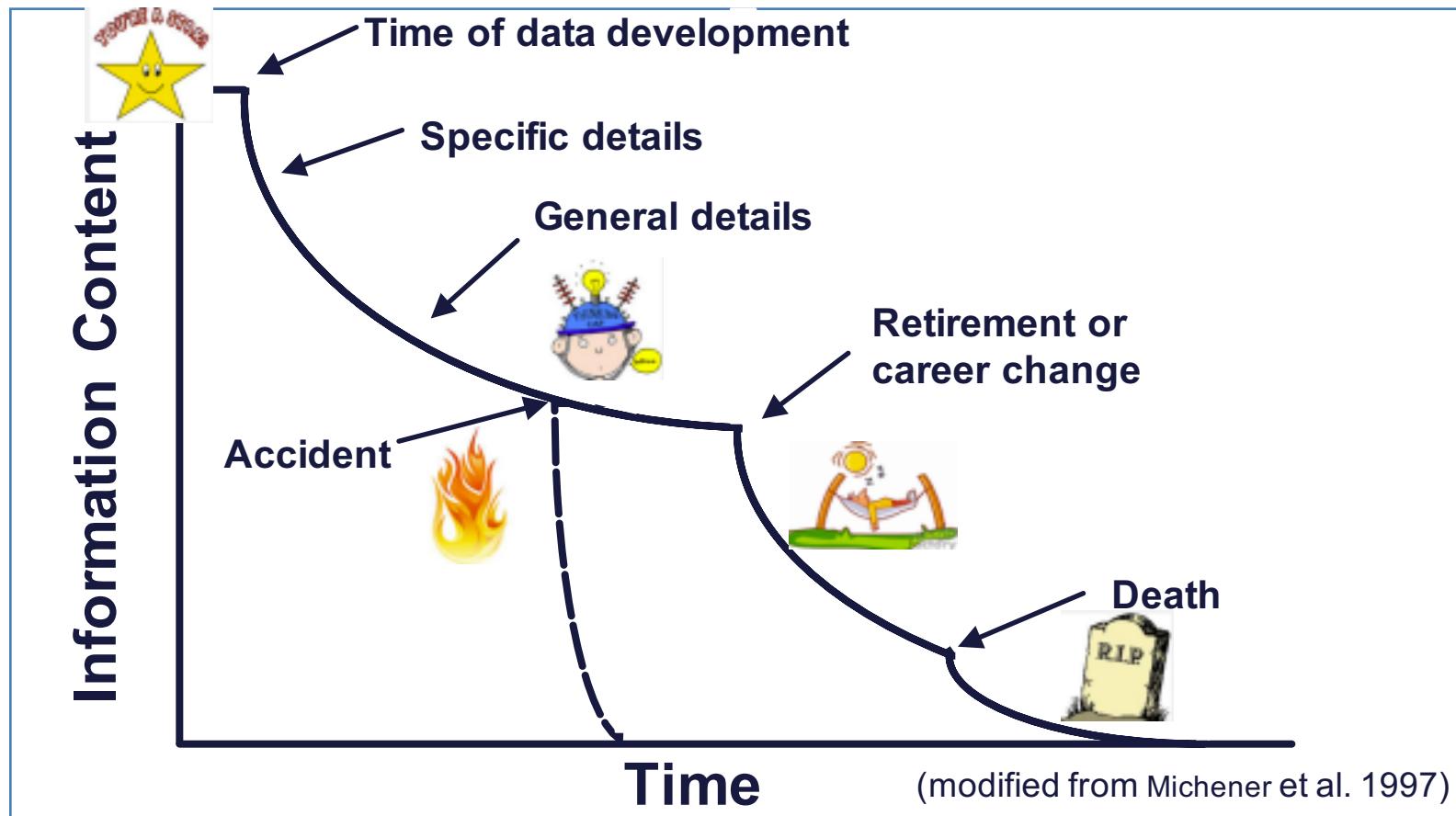
**USGS Science Data Catalog: enabling discovery**

**Metadata: captures information**

**DataONE: enables exchange**



# The importance of metadata





# The importance of metadata

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



# The importance of metadata



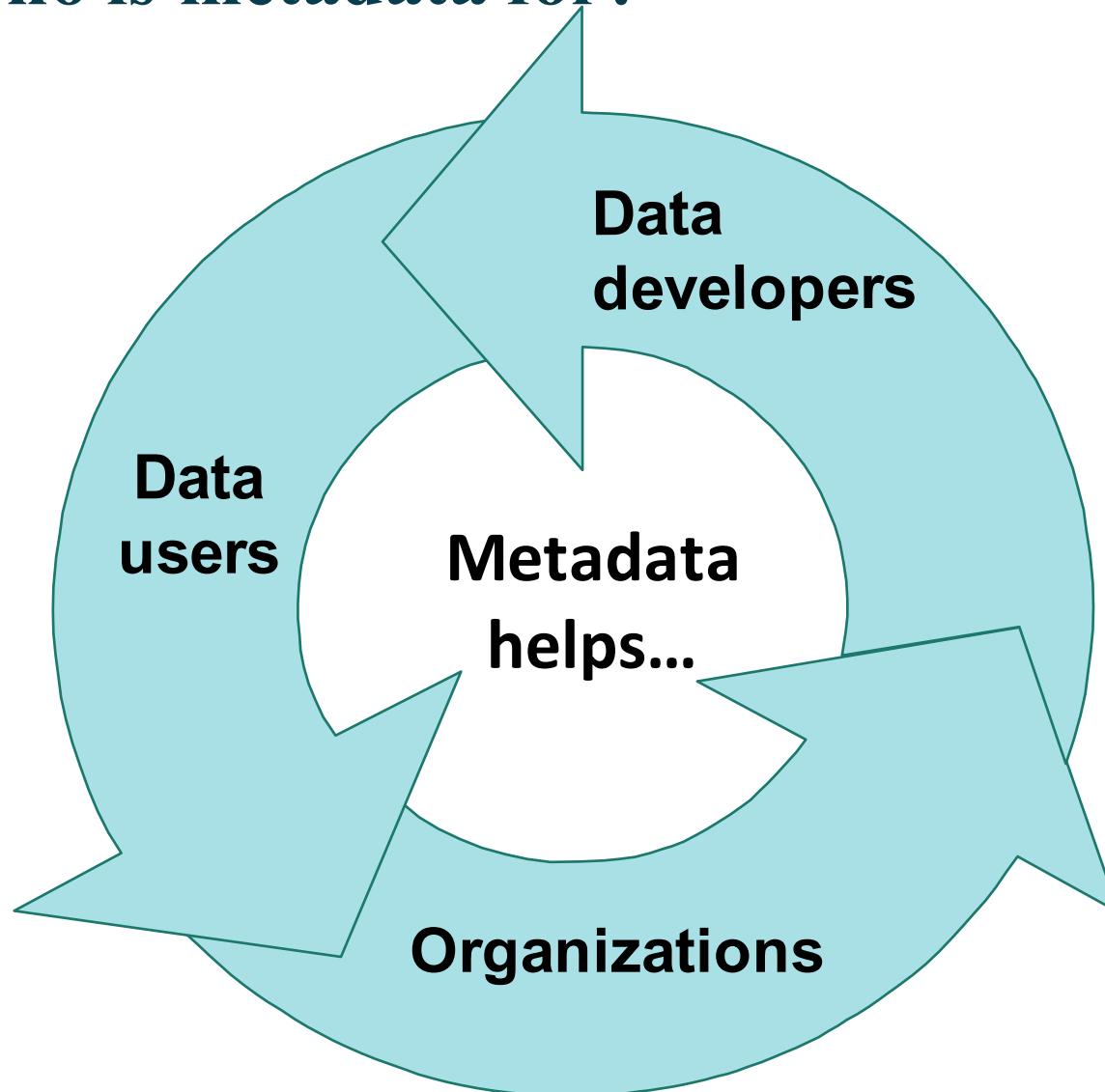


# The importance of metadata

- Metadata are essential for policy work
  - Discovering data relevant to policy questions in the first place
  - Metadata support scrutiny of our data
    - Motivations
    - Methodologies
    - Conflicts of interest



# Who is metadata for?





# Metadata for data developers

- Avoid data duplication
  - What has been collected already?
- Share reliable information
  - What method was used?
  - What methods are in common use in my field?
- Publicize our work
  - Hey, I made this!
- Save time the next time
  - Hey, I've already done this!



# Metadata for data users

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



# Metadata for organizations

- Help ensure the organization's investment in the data
  - Documentation for sampling & data processing methods get recorded
  - Ability to use data after initial intended purpose
  - Track data re-use and citation



# Metadata for organizations

- Transcend people and time
  - Don't lose our data when researchers/labs leave
  - Avoid duplication in new work



# Metadata for organizations

- Advertise organization's research
  - What data has our organization produced?



# Metadata for organizations

Metadata is all about scale



# 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 (emph. publications)
- Darwin Core (emph. collections)
- FGDC (emph. spatial)
- ISO19115 (emph. spatial & services)
- Ecological Metadata Language (general, but emphasis on filesystem artifacts, attributes, taxonomy)



# Metadata standards

```
<?xml version="1.0" encoding="UTF-8"?>

<gmi:MI_Metadata xmlns:gmi="http://www.isotc211.org/2005/gmi" xmlns:gco="http://www.isotc211.org/2005/gco"
  <gmd:fileIdentifier gco:nilReason="missing"/>
  <gmd:language>
    <gco:CharacterString>eng;USA</gco:CharacterString>
  </gmd:language>
  <gmd:characterSet>
    <gmd:MD_CharacterSetCode codeList="http://www.ngdc.noaa.gov/metadata/published/xsd/schemacodeList.xml" codeListValue="UTF-8"/>
  </gmd:characterSet>
  <gmd:contact>
    <gmd:CI_ResponsibleParty>
      <gmd:organisationName>
        <gco:CharacterString>Axiom Data Science</gco:CharacterString>
      </gmd:organisationName>
      <gmd:positionName>
        <gco:CharacterString>Metadata Specialist</gco:CharacterString>
      </gmd:positionName>
      <gmd:contactInfo>
        <gmd:CI_Contact>
          <gmd:address>
            <gmd:CI_Address>
              <gmd:deliveryPoint>
                <gco:CharacterString>1016 W 6th Ave, Ste 105</gco:CharacterString>
              </gmd:deliveryPoint>
              <gmd:city>
                <gco:CharacterString>Anchorage</gco:CharacterString>
              </gmd:city>
              <gmd:administrativeArea>
                <gco:CharacterString>AK</gco:CharacterString>
              </gmd:administrativeArea>
              <gmd:postalCode>
                <gco:CharacterString>99501</gco:CharacterString>
              </gmd:postalCode>
            </gmd:CI_Address>
          </gmd:address>
        </gmd:CI_Contact>
      </gmd:contactInfo>
    </gmd:CI_ResponsibleParty>
  </gmd:contact>
</gmi:MI_Metadata>
```



# Metadata standards

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<gmi:MI_Metadata xmlns:gmi="http://www.isotc211.org/2005/gmi" xmlns:gco="http://www.isotc211.org/2005/gco">
  <gmd:fileIdentifier gco:nilReason="missing"/>
  <gmd:language>
    <gco:CharacterString>eng;USA</gco:CharacterString>
  </gmd:language>
  <gmd:characterSet>
    <gmd:MD_CharacterSetCode codeList="http://www.ngdc.noaa.gov/metadata/published/xsd/schemacodeList.xml" codeListValue="UTF-8"/>
  </gmd:characterSet>
  <gmd:contact>
    <gmd:CI_ResponsibilityParty>
      <gmd:organisationName>
        <gco:CharacterString>Axiom Data Science</gco:CharacterString>
      </gmd:organisationName>
      <gmd:positionName>
        <gco:CharacterString>Metadata Specialist</gco:CharacterString>
      </gmd:positionName>
      <gmd:contactInfo>
        <gmd:CI_Contact>
          <gmd:address>
            <gmd:CI_Address>
              <gmd:deliveryPoint>
                <gco:CharacterString>123 Main Street</gco:CharacterString>
              </gmd:deliveryPoint>
              <gmd:city>
                <gco:CharacterString>Anchorage</gco:CharacterString>
              </gmd:city>
              <gmd:administrativeArea>
                <gco:CharacterString>AK</gco:CharacterString>
              </gmd:administrativeArea>
              <gmd:postalCode>
                <gco:CharacterString>99501</gco:CharacterString>
              </gmd:postalCode>
            </gmd:CI_Address>
          </gmd:address>
        </gmd:CI_Contact>
      </gmd:contactInfo>
    </gmd:CI_ResponsibilityParty>
  </gmd:contact>
</gmi:MI_Metadata>
```

...is a person that creates and manages metadata for resources and services. This person generally has expertise in documentation standards and has enough experience and understanding of the resource to document it in partnership with the originator or resource contact.



# What makes a good metadata record?

- Overall goal: Could a reasonable scientist make sense of our data in 10, 20, 20+ years *without contacting you?*
- When in doubt, be more specific:
  - Spell out acronyms
  - Use full names, emails, addresses, etc.
- Include as much info 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 doesn't 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



# What makes a good metadata record?

- Good titles include:
  - Who
  - What
  - When
  - Where
  - Why

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



# What makes a good metadata record?

- Titles: Which is preferable?

River Data

or...

Greater Yellowstone Rivers from 1:126,700  
U.S. Forest Service Visitor Maps (1961-1983)



# What makes a good metadata record?

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



# What makes a good metadata record?

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



# What makes a good metadata record?

- Involved parties
  - Name alone is not enough
    - To assign credit
    - To disambiguate across datasets
  - Email helps
  - ORCiD (w/ above) is best



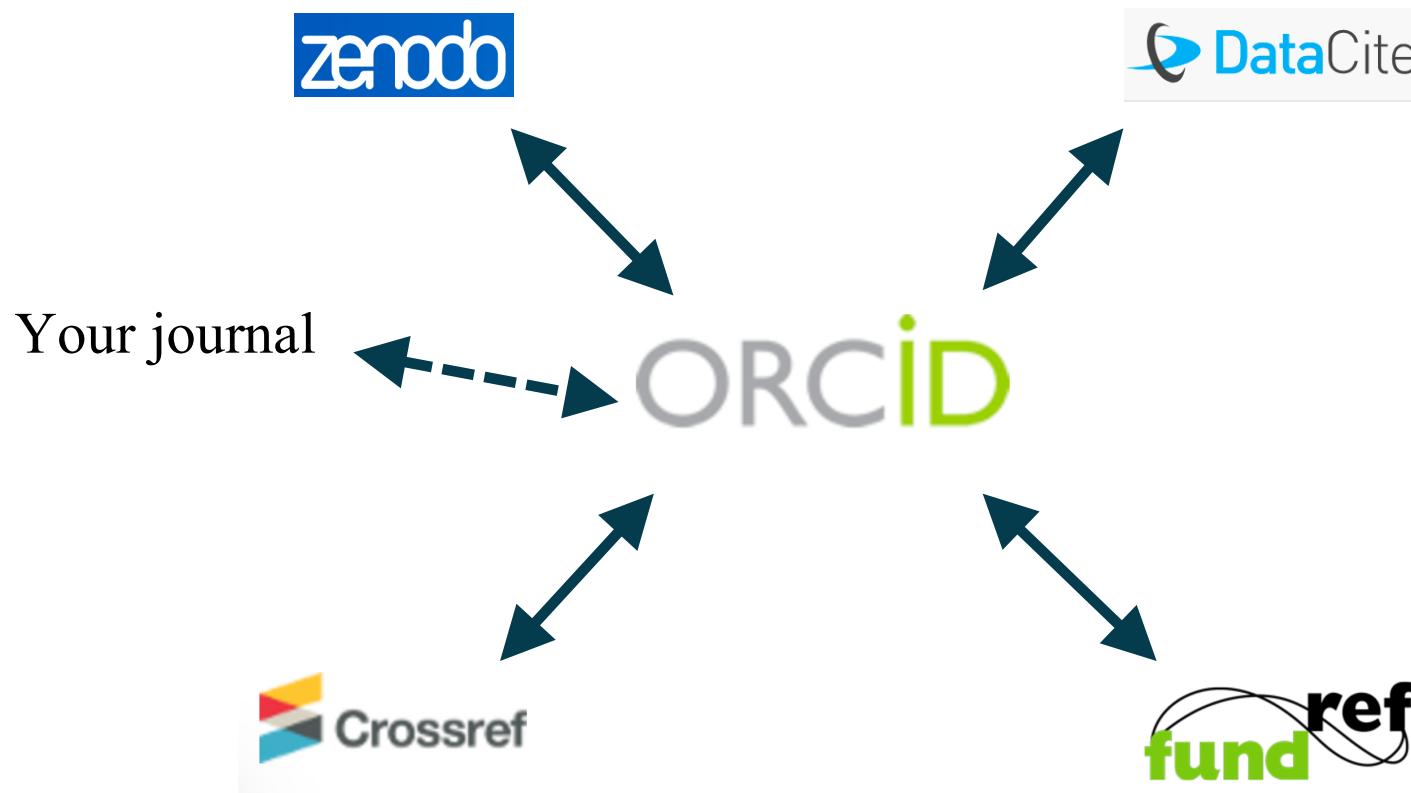
# ORCiDs: Wait, what is an ORCiD?

- Like an ISBN for people
  - e.g. mine: 0000-0002-0381-3766
- Enables unambiguous reference to humans
- Free
- Becoming a community norm
- Inherently connected...



# ORCiDs

- Inherently connected





# Activity

- Get an ORCiD:
  - <https://orcid.org/>
- Sign in to <https://dev.nceas.ucsb.edu/>