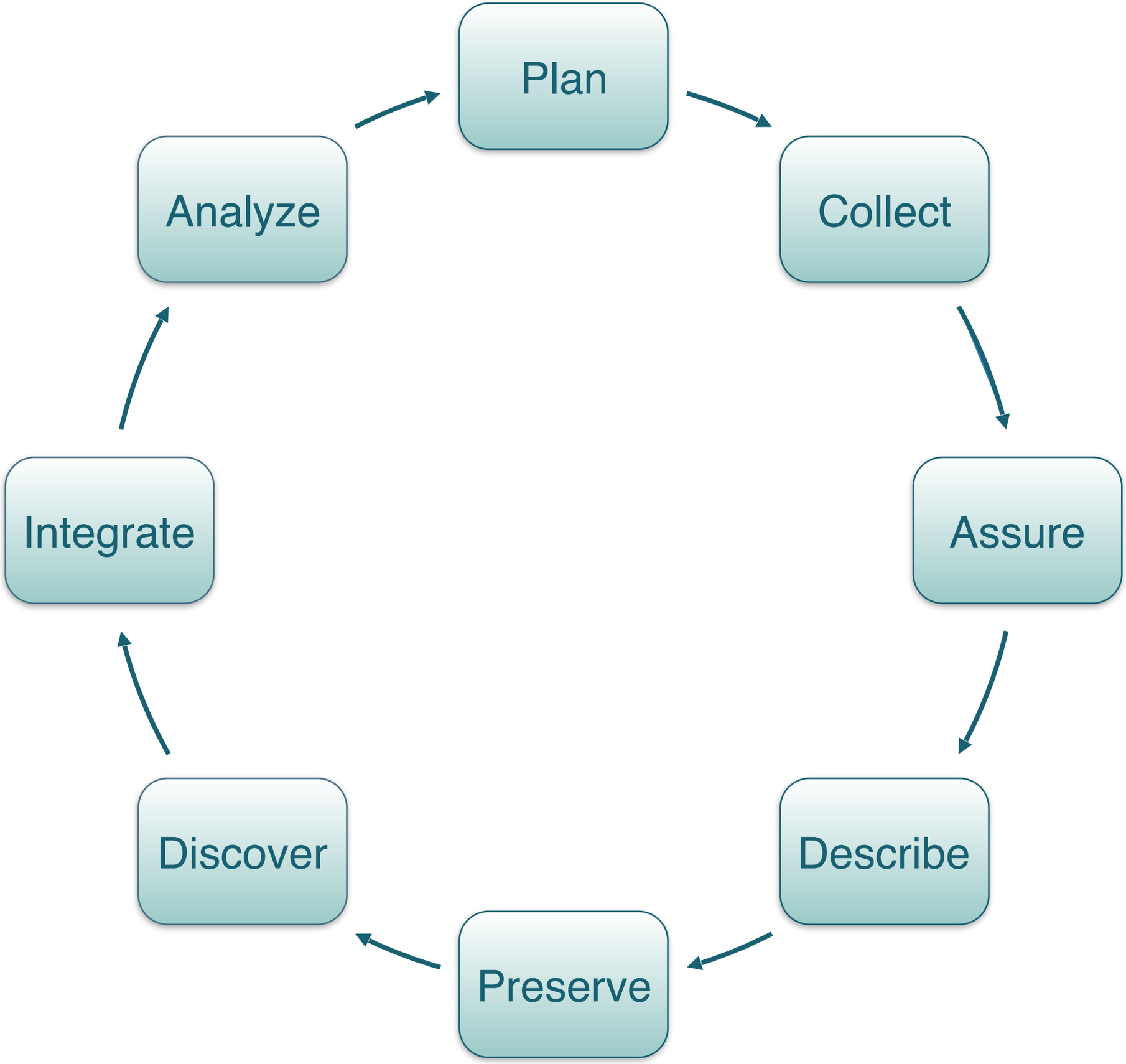
Authoring High Quality Metadata

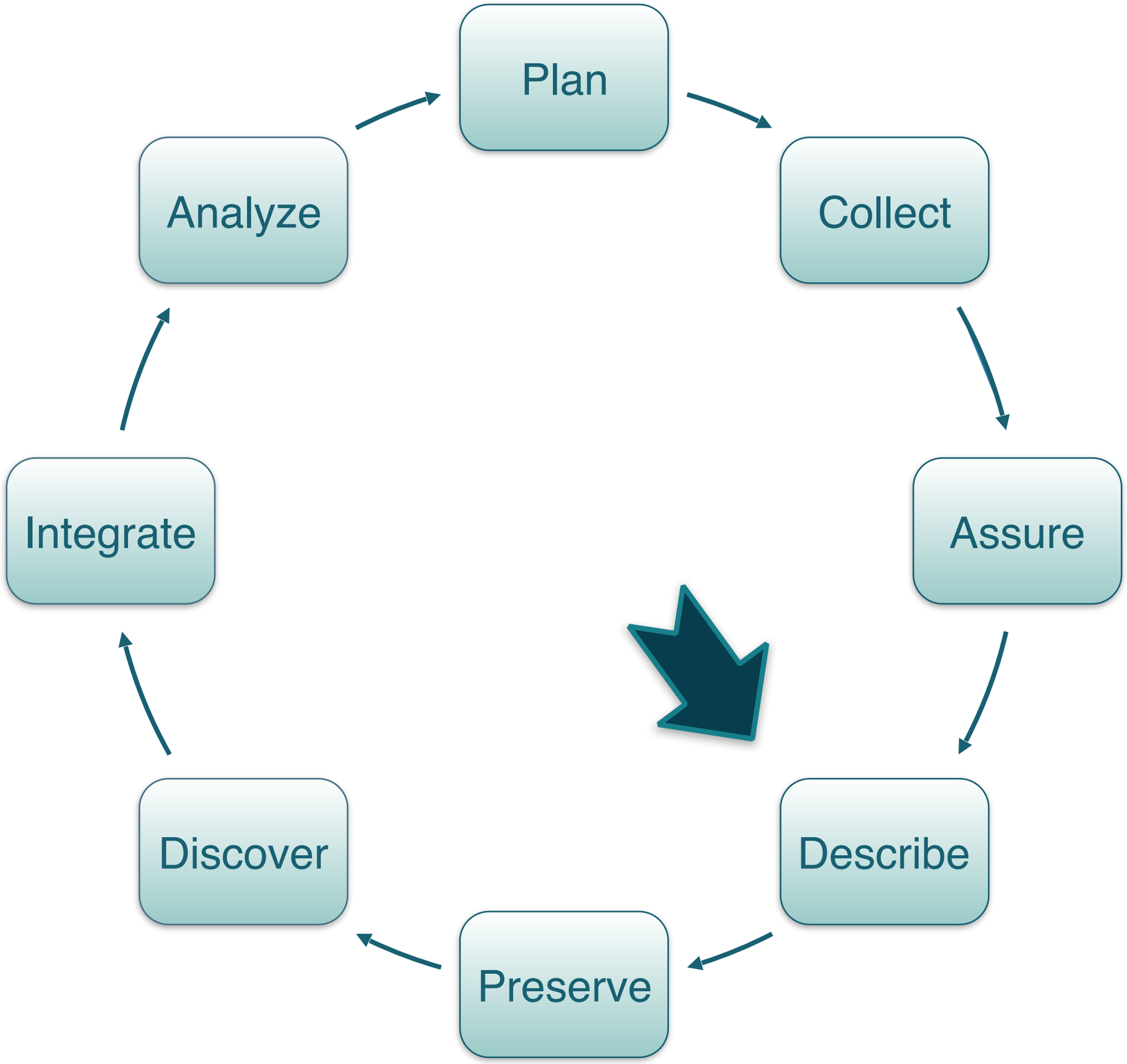
Jesse Goldstein and Jeanette Clark
UC Santa Barbara

JG orcid.org/0000-0002-1006-9496
JC orcid.org/0000-0003-4703-1974

The Data Life Cycle



The Data Life Cycle



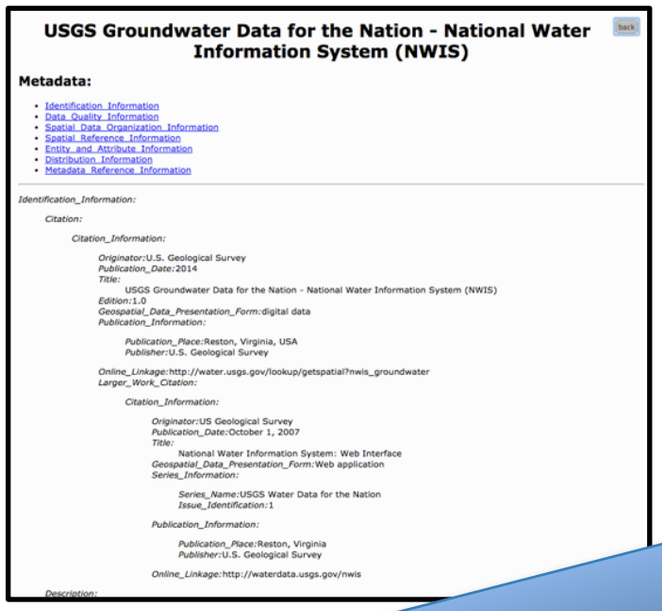
What are metadata?

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

back

Metadata:

- Identification_Information
- Data_Quality_Information
- Spatial_Data_Organization_Information
- Spatial_Reference_Information
- Entity_and_Attribute_Information
- Distribution_Information
- Metadata_Reference_Information

Identification_Information:

Citation:

Citation_Information:

*Originator:*U.S. Geological Survey
*Publication_Date:*2014
Title:
USGS Groundwater Data for the Nation - National Water Information System (NWIS)
*Edition:*1.0
*Geospatial_Data_Presentation_Form:*digital data
Publication_Information:

*Publication_Place:*Reston, Virginia, USA
*Publisher:*U.S. Geological Survey

*Online_Linkage:*http://water.usgs.gov/lookup/getspatial?nwis_groundwater
Larger_Work_Citation:

Citation_Information:

*Originator:*US Geological Survey
*Publication_Date:*October 1, 2007
Title:
National Water Information System: Web Interface
*Geospatial_Data_Presentation_Form:*Web application
Series_Information:

*Series_Name:*USGS Water Data for the Nation
*Issue_Identification:*1

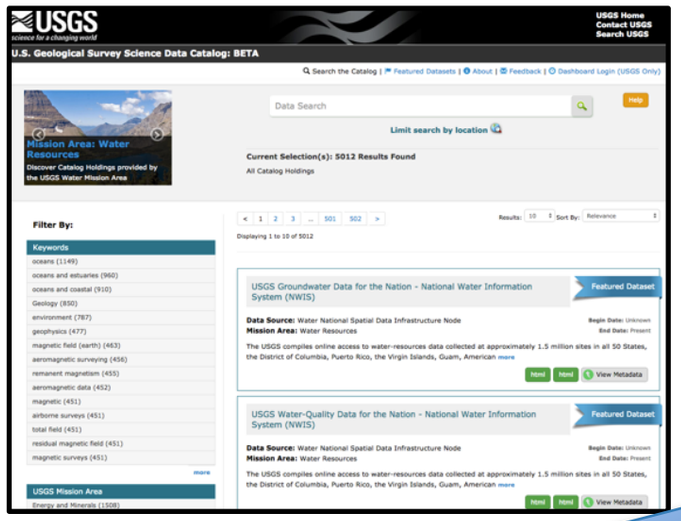
Publication_Information:

*Publication_Place:*Reston, Virginia
*Publisher:*U.S. Geological Survey

*Online_Linkage:*http://waterdata.usgs.gov/nwis

Description:

USGS Science Data Catalog: enables discovery



USGS

USGS Home
Contact USGS
Search USGS

U.S. Geological Survey Science Data Catalog: BETA

Search the Catalog | Featured Datasets | About | Feedback | Dashboard Login (USGS Only)

Mission Area: Water Resources
Discover Catalog Holdings provided by the USGS Water Mission Area

Data Search

Limit search by location

Current Selection(s): 5012 Results Found
All Catalog Holdings

Filter By:

Keywords
oceans (1149)
oceans and estuaries (960)
oceans and coastal (910)
Geology (850)
environment (787)
geophysics (477)
magnetic field (earth) (463)
aeromagnetic surveying (456)
remanent magnetism (455)
aeromagnetic data (452)
magnetic (451)
airborne surveys (451)
total field (451)
residual magnetic field (451)
magnetic surveys (451)
more

USGS Mission Area
Energy and Minerals (1508)

Displaying 1 to 10 of 5012

USGS Groundwater Data for the Nation - National Water Information System (NWIS) — Featured Dataset

Data Source: Water National Spatial Data Infrastructure Node
Mission Area: Water Resources
Begin Date: Unknown
End Date: Present

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 more

html html View Metadata

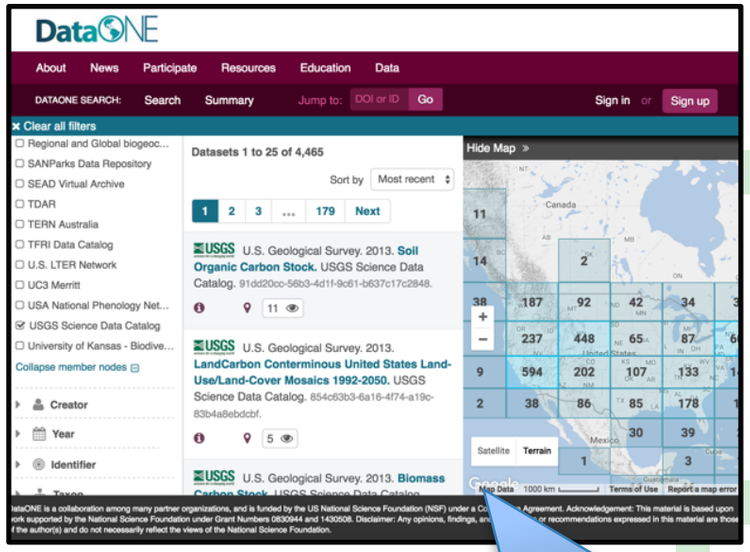
USGS Water-Quality Data for the Nation - National Water Information System (NWIS) — Featured Dataset

Data Source: Water National Spatial Data Infrastructure Node
Mission Area: Water Resources
Begin Date: Unknown
End Date: Present

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 more

html html View Metadata

DataONE: enables exchange



DataONE

About | News | Participate | Resources | Education | Data

DATAONE SEARCH: Search | Summary | Jump to: DOI or ID Go | Sign in or Sign up

Clear all filters

Regional and Global biogeoc...
SANParks Data Repository
SEAD Virtual Archive
TDAR
TERN Australia
TFRI Data Catalog
U.S. LTER Network
UC3 Merritt
USA National Phenology Net...
USGS Science Data Catalog
University of Kansas - Biodive...
Collapse member nodes

Creator
Year
Identifier
Taxon

Datasets 1 to 25 of 4,465

Sort by Most recent

1 2 3 ... 179 Next

USGS U.S. Geological Survey. 2013. **Soil Organic Carbon Stock.** USGS Science Data Catalog. 91dd20cc-56b3-4d1f-9c61-b637c17c2848.

USGS U.S. Geological Survey. 2013. **LandCarbon Conterminous United States Land-Use/Land-Cover Mosaics 1992-2050.** USGS Science Data Catalog. 854c63b3-6a16-4f74-a19c-83b4a8ebdcbf.

USGS U.S. Geological Survey. 2013. **Biomass Carbon Stock.** USGS Science Data Catalog.

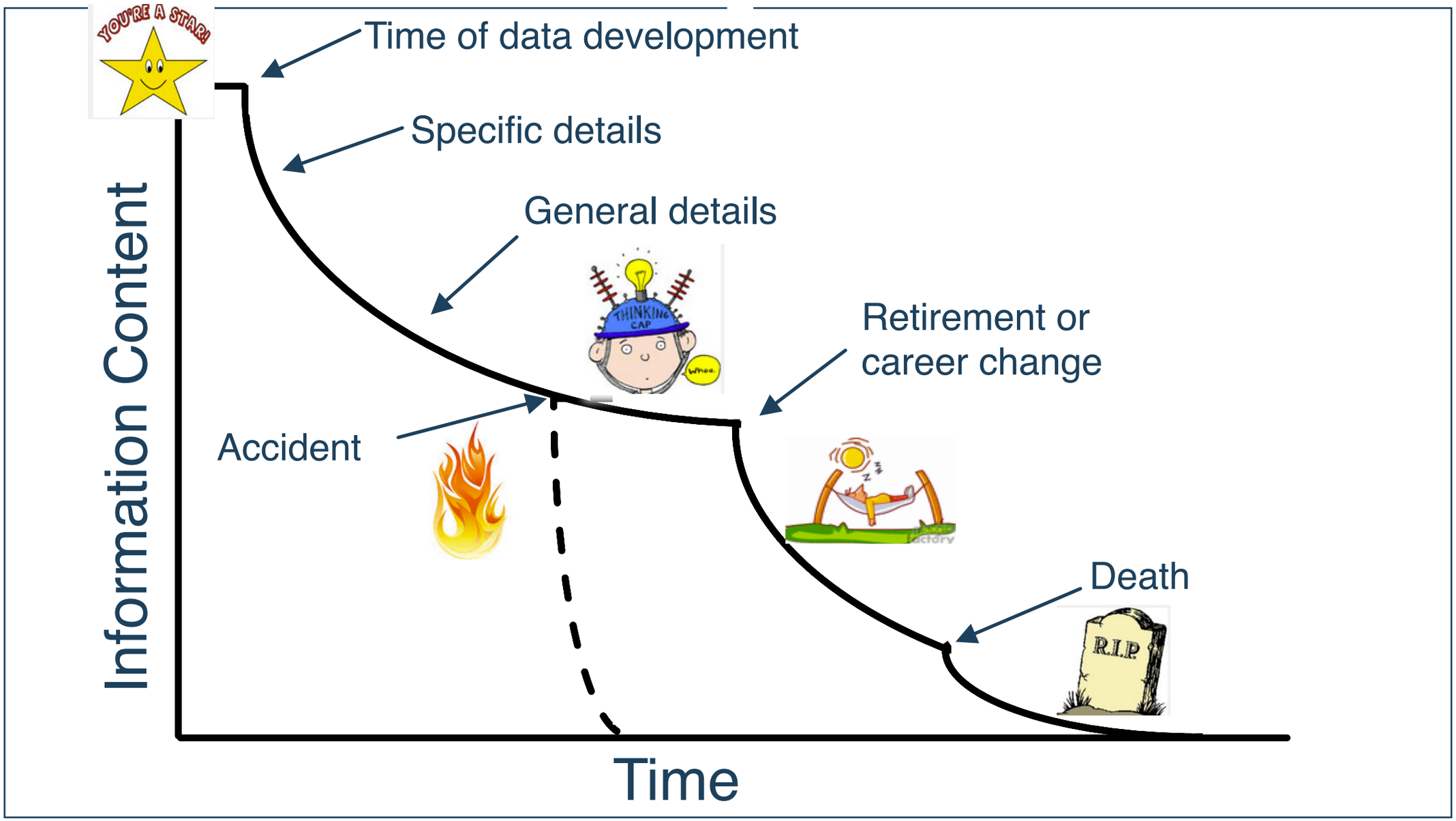
Hide Map

Satellite | Terrain

Map Data 1000 km Terms of Use Report a map error

DataONE is a collaboration among many partner organizations, and is funded by the US National Science Foundation (NSF) under a Cooperative Agreement. Acknowledgement: This material is based upon work supported by the National Science Foundation under Grant Numbers 0830944 and 1430508. Disclaimer: Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Why are metadata important?



(modified from Michener et al. 1997)

Why are metadata important?

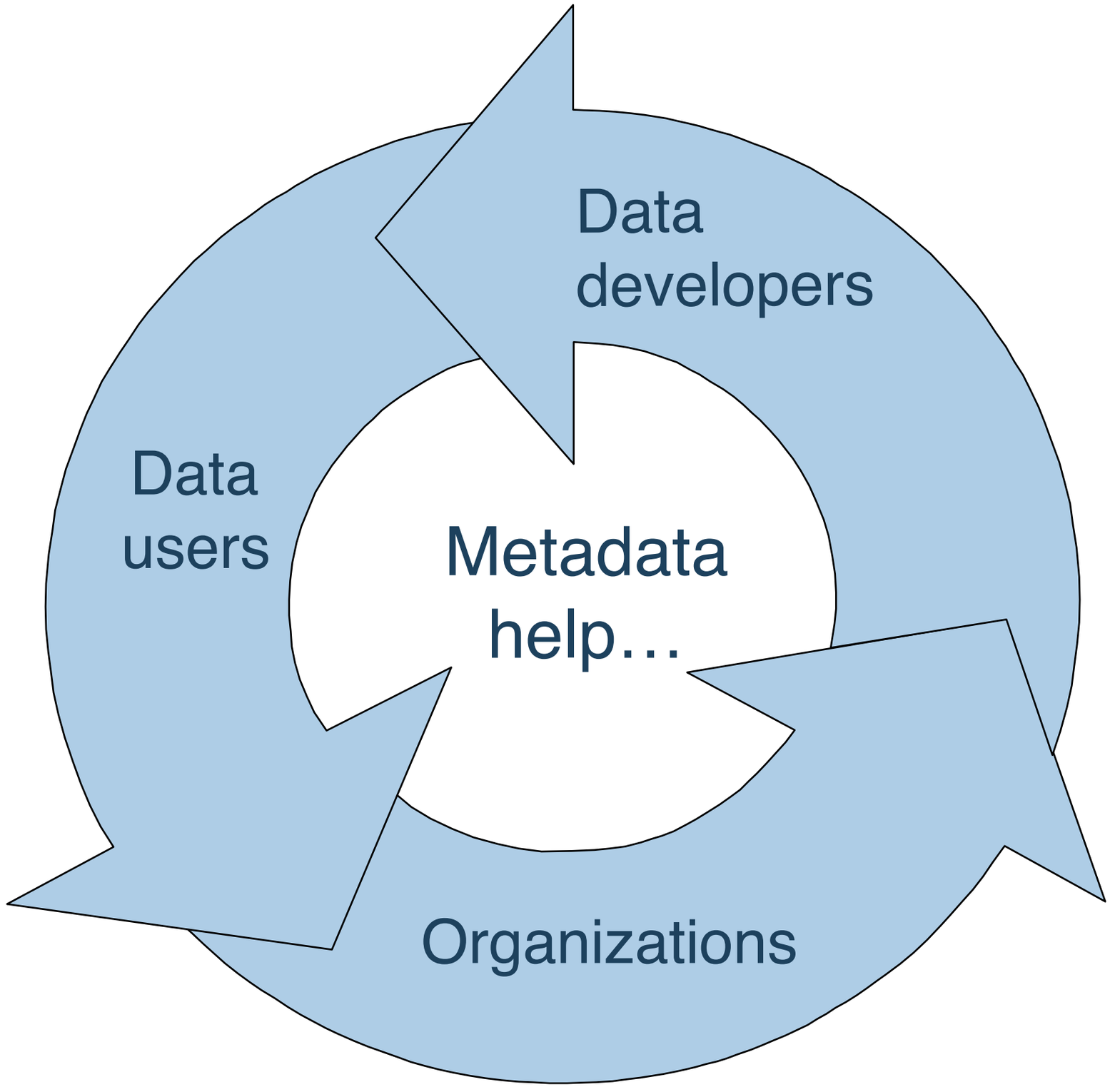
Metadata are important for the short and long-term utility of data

Why are metadata important?

Why are metadata important?

- Metadata support scrutiny of data
  - Motivations
  - Methodologies
  - Conclusions

Who uses metadata?



Metadata for data developers

- Avoid data duplication
  - What data have already been collected?
  - Save time the next time
  - “Hey, I’ve already done this!”
- Share reliable information
  - What methods were used?
  - What methods are in common use in my field?
- Publicize your 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

- Help ensure the organization's investment in the data
  - 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" encoding="UTF-8"?>

<gmi:MI_Metadata xmlns:gmi="http://www.isotc211.org/2005/gmi" xmlns:gco="http://www.isotc211
  <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/schem
  </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>
```

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
  <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/schem
  </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
              </gmd:deliveryPoint>
              <gmd:city>
                <gco:CharacterString
              </gmd:city>
              <gmd:administrativeArea>
```

…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.

Creating Standardized Metadata

- Specialized tools are your friend!

dublincoregenerator.com - a better dublin core generator

Main Page | Simple Generator | Advanced Generator | xZINECOREx Generator | About | Contribute

Directions

- Fill in the fields below and click on "Generate Code!" to convert your input into fully formed Dublin Core metadata code. Additional options for the format of the output code are available below.
- If you need additional copies of a given field, click the plus sign to the upper-right of the tag's name to add an additional copy of it.
- Click the minus sign to delete any unneeded additional copies -- don't worry about removing tags you don't intend to use, the system will ignore any empty tags (and you can't delete the first row anyway).
- If you are unsure how a specific tag works, you can click the question mark next to the tag's name to see the tag's entry in Diane Hilmann's wonderful guide "Using Dublin Core -- The Elements."
- If you would like to use encoding schemes and the more advanced qualified elements of Dublin Core metadata, use the Advanced Generator located here.

Input

Title? [+][-]
My Paper

Creator? [+][-]
Jeanette Clark

Subject? [+][-]
Example

Description? [+][-]

Publisher? [+][-]

Contributor? [+][-]

Date? [+][-]

http://dublincoregenerator.com

Creating Standardized Metadata

- Specialized tools are your friend!

dublincoregenerator.com - a better dublin core generator

Main Page | Simple Generator | Advanced Generator | xZINECOREx Generator | About | Contribute

Directions

- Fill in the fields below and click on "Generate Code!" to convert your input into fully formed Dublin Core metadata code. Additional options for the format of the output code are available below.
- If you need additional copies of a given field, click the plus sign to the upper-right of the tag's name to add an additional copy of it.
- Click the minus sign to delete any unneeded additional copies -- don't worry about removing tags you don't intend to use, the system will ignore any empty tags (and you can't delete the first row anyway).
- If you are unsure how a specific tag works, you can click the question mark next to f… Diane Hilmann's wonderful guide "Using Dublin Core -- The Elements."
- If you would like to use encoding schemes and the more advanced qualified elemer… Advanced Generator located here.

Input

Title? [+][-]
My Paper

Creator? [+][-]
Jeanette Clark

Subject? [+][-]
Example

Description? [+][-]

Publisher? [+][-]

Contributor? [+][-]

Date? [+][-]

Output

```
<dc:title>My Paper</dc:title>
<dc:creator>Jeanette Clark</dc:creator>
<dc:subject>Example</dc:subject>
```

http://dublincoregenerator.com

Creating Standardized Metadata

https://data.gulfresearchinitiative.org/metadata-editor/

Report Issue | Suggest Improvement

GULF OF MEXICO RESEARCH INITIATIVE

Investigating the effect of oil spills on the environment and public health.

GRIIDC

HOME | SEARCH DATA | SUBMIT DATA | TRACKING & STATS | ABOUT US | HELP | RESEARCH

ISO 19115-2 Metadata Editor

Load from File | Load from Submitted Dataset | Save to File | Clear Form | Check and Save to File | ? Help

Dataset Contact | Dataset Information | Keywords | Data Extent | Distribution Info | Distribution Contact | Metadata Contact

NOTE: *Fields with * are required.*

Dataset Information

This section collects identifying and amplifying information about the dataset. Provides future researchers with specific details on the dataset content and additional context regarding the broader purpose of the dataset.

***Title**	Name by which the cited resource is known. It is recommended the title include (where applicable) a description of the data, a date or date range, and geographic area.
Short Title	Short name or other language name by which the cited information is known.
***Date** ...	Reference date for the cited dataset. This date can refer to dataset creation, publication, or revision. Format should be YYYY-MM-DD.
***Date Type** Publication	Creation: Date identifies when the resource was brought into existence. Publication: Date identifies when the resource was issued. Revision: Date identifies when the resource was examined or re-examined and improved or amended.
***Abstract**	Brief narrative summary of the dataset's contents.

Creating Standardized Metadata

https://github.com/ropensci/EML/

```
attributes2 <- data.frame(attributeName = c('Time', 'Wind_Speed'),
                          attributeDefinition = c('Date and time of wind speed reading', 'Measured
                          measurementScale = c('dateTime', 'ratio'),
                          domain = c('dateTimeDomain', 'numericDomain'),
                          formatString = c('YYYY-MM-DD hh:mm:ss', NA),
                          definition = c(NA, NA),
                          unit = c(NA, 'metersPerSecond'),
                          numberType = c(NA, 'real'),
                          missingValueCode = c(NA, NA),
                          codeExplanation = c(NA, NA),
                          stringsAsFactors = FALSE)
attributeList2 <- set_attributes(attributes2)
id2 <- 'PID2'
physical2 <- pid_to_eml_physical(mn, id2)
dataTable2 <- new('dataTable',
                  entityName = 'EagleMtnWindData.csv',
                  entityDescription = 'Wind data from Eagle Mountain',
                  physical = physical2,
                  attributeList = attributeList2)
```

Creating Standardized Metadata

https://github.com/ropensci/EML/

```
attributes2 <- data.frame(attributeName = c('Time', 'Wind_Speed'),
                          attributeDefinition = c('Date and time of wind speed reading', 'Measured
                          measurementScale = c('dateTime', 'ratio')
                          domain = c(
                          formatStrin
                          definition
                          unit = c(NA
                          numberType
                          missingValu
                          codeExplana
                          stringsAsFa
attributeList2 <- set_attributes(attr
id2 <- 'PID2'
physical2 <- pid_to_eml_physical(mn,
dataTable2 <- new('dataTable',
                  entityName = 'Eagle
                  entityDescription =
                  physical = physical
                  attributeList = att
```

```
> attributeList2
<attributeList>
  <attribute>
    <attributeName>Time</attributeName>
    <attributeDefinition>Date and time of wind speed reading</attributeDefinition>
    <measurementScale>
      <dateTime>
        <formatString>YYYY-MM-DD hh:mm:ss</formatString>
      </dateTime>
    </measurementScale>
  </attribute>
  <attribute>
    <attributeName>Wind_Speed</attributeName>
    <attributeDefinition>Measured wind speed</attributeDefinition>
    <measurementScale>
      <ratio>
        <unit>
          <standardUnit>metersPerSecond</standardUnit>
        </unit>
        <numericDomain>
          <numberType>real</numberType>
        </numericDomain>
      </ratio>
    </measurementScale>
  </attribute>
</attributeList>
```

Siddharth Narayan. 2016. SNAPP Coastal Defenses - Effectiveness, Costs and Benefits of Nature-based Defences for Wave Reduction. Knowledge Network for Biocomplexity. doi:10.5063/F1Z31WKX.

Copy Citation | Quality report

Files in this dataset Package: urn:uuid:9fb6de93-c932-4e93-8a1c-b17b462a3d62

Name		File type	Size	Downloads	Download All
Metadata: SNAPP Coastal Defenses - Effectiveness, Costs and Benefits of Nature-based Defences for Wave Reduction		EML v2.1.1	53 KB	120 views	Download
Wave Reduction Data	More info	text/csv	11 KB	12 downloads	Download
Coastal Protection Cost and Benefit	More info	text/csv	9 KB	13 downloads	Download
DatabasePaper_Analyses_Plots.R	More info	plain text (.txt)	36 KB	7 downloads	Download

Show 1 more item in this data set

General

Identifier: doi:10.5063/F1Z31WKX

Abstract:

There is great interest in the restoration and conservation of coastal habitats for protection from flooding and erosion. This is evidenced by the growing number of analyses and reviews of the effectiveness of habitats as natural defenses and increasing funding worldwide for naturebased defences–i.e. restoration projects aimed at coastal protection; yet, there is no synthetic information on what kinds of projects are effective and cost effective for this purpose. This paper addresses two issues critical for designing restoration projects for coastal protection: (i) a synthesis of the costs and benefits of projects designed for coastal protection (naturebased defences) and (ii) analyses of the effectiveness of coastal habitats (natural defences) in reducing wave heights and the biophysical parameters that influence this effectiveness. We (i) analyse data from sixtynine field measurements in coastal habitats globally and examine measures of effectiveness of mangroves, saltmarshes, coral reefs and seagrass/kelp beds for wave height reduction; (ii) synthesise the costs and coastal protection benefits of fiftytwo nature based defence projects and;(iii) estimate the benefits of each restoration project by combining information on restoration costs with data from nearby field measurements. The analyses of field measurements show that coastal habitats have significant potential for reducing wave heights that varies by habitat and site. In general, coral reefs and saltmarshes have the highest overall potential. Habitat effectiveness is influenced by: a) the ratios of wave heighttowater depth and habitat widthtowavelength in coral reefs; and b) the ratio of vegetation heighttowater depth in saltmarshes. The comparison of costs of naturebased defence projects and engineering structures show that saltmarshes and mangroves can be two to five times cheaper than a submerged breakwater for wave heights up to half a metre and, within their limits, become more cost effective at greater depths. Naturebased defence projects also report benefits ranging from reductions in storm damage to reductions in coastal structure costs. http://dx.doi.org/10.1371/journal.pone.0154735

Keywords

Keyword	Type
natural coastal defenses	
Science for Nature and People Partnership (SNAPP)	
salt-marsh	
coral reefs	

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

What makes a good metadata record?

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:

“ITP37”

What makes a good metadata record?

Title:

“ITP37” 😕 ⚠️

What makes a good metadata record?

Title:

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

What makes a good metadata record?

Title:

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

😃 ✅

What makes a good metadata record?

Begin: 2003-04-14
End: 2003-04-13

Sag River

😕⚠️

What makes a good metadata record?

What makes a good metadata record?

“Begin: 2002-04-14
End: 2003-04-13”

“Sagavanirktok River, North Slope, Alaska”

What makes a good metadata record?

“Begin: 2002-04-14
End: 2003-04-13”

“Sagavanirktok River, North Slope, Alaska”

😀 ✅

What makes a good metadata record?

“ ”

What makes a good metadata record?

😕 ⚠️

“ ”

What makes a good metadata record?

Transect

We established three 100-m transects at the Airport Site to quantify differences in micro-topography, soil temperatures, thaw depth, soils, vegetation, permafrost and snow in relationship to distance from the road. Pin flags were placed at 1-m intervals along Transects 3 and 4, and vertical 150-cm PVC posts were placed at 0, 5, 10, 25, 50 and 100 m. The poles have blue stripes at 50, 100 and 150 cm height to help locate the transects in winter. No poles or pin flags were placed along T5, but the plots are permanently marked by wooden corner stakes and an aluminum-capped piece of rebar at the center bearing the plot number.

Vegetation Plots

We established permanent vegetation plots with photo points in polygon centers and troughs at 5, 10, 25, 50 and 100 m from the road along T3 and T4, and at 25, 50 and 100 m from the road on T5. Voucher collections of all vascular plants, mosses and lichens were collected from each plot and are stored at the Alaska Geobotany Center. Species cover was measured using 100 points from a 1 x 1 m2 point-quadrat. Cover of all species was estimated using Braun-Blanquet cover abundance scores. The species at the top of the plant canopy were recorded at 100 grid points within each plot. Leaf Area Index (LAI) was measured using an AccuPAR LP-80 PAR/LAI Ceptometer. Soil temperature loggers were installed at all permanent plots on T3 and T4. Air temperature loggers were installed along the T3 and T4 flag transects.

Topographic Surveys

The location and elevation of all boreholes, transects, vegetation plots and other reference points were surveyed using a combination of a GPS real time kinematic (RTK) system and a robotic imaging system. All measurements where connected to the stable National Geodetic Survey (NGS) benchmark point DF3643 (70˚ 11’ 48.87851” N, 148˚ 25’ 53.20441” W) in order to acquire the exact location and orthoheight of all surveyed points. Since we required two different levels of accuracy, we used two different survey systems for the topographic survey. At the

What makes a good metadata record?


Transect

We established three 100-m transects at the Airport Site to quantify differences in micro-topography, soil temperatures, thaw depth, soils, vegetation, permafrost and snow in relationship to distance from the road. Pin flags were placed at 1-m intervals along Transects 3 and 4, and vertical 150-cm PVC posts were placed at 0, 5, 10, 25, 50 and 100 m. The poles have blue stripes at 50, 100 and 150 cm height to help locate the transects in winter. No poles or pin flags were placed along T5, but the plots are permanently marked by wooden corner stakes and an aluminum-capped piece of rebar at the center bearing the plot number.

Vegetation Plots

We established permanent vegetation plots with photo points in polygon centers and troughs at 5, 10, 25, 50 and 100 m from the road along T3 and T4, and at 25, 50 and 100 m from the road on T5. Voucher collections of all vascular plants, mosses and lichens were collected from each plot and are stored at the Alaska Geobotany Center. Species cover was measured using 100 points from a 1 x 1 m2 point-quadrat. Cover of all species was estimated using Braun-Blanquet cover abundance scores. The species at the top of the plant canopy were recorded at 100 grid points within each plot. Leaf Area Index (LAI) was measured using an AccuPAR LP-80 PAR/LAI Ceptometer. Soil temperature loggers were installed at all permanent plots on T3 and T4. Air temperature loggers were installed along the T3 and T4 flag transects.

Topographic Surveys

The location and elevation of all boreholes, transects, vegetation plots and other reference points were surveyed using a combination of a GPS real time kinematic (RTK) system and a robotic imaging system. All measurements where connected to the stable National Geodetic Survey (NGS) benchmark point DF3643 (70˚ 11’ 48.87851” N, 148˚ 25’ 53.20441” W) in order to acquire the exact location and orthoheight of all surveyed points. Since we required two different levels of accuracy, we used two different survey systems for the topographic survey. At the

What makes a good metadata record?

Abstract

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

What makes a good metadata record?

Abstract

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

These data are ocean water property observations reported from ice-tethered profiler #37.

😕 ⚠️

What makes a good metadata record?

Abstract

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

These data are ocean water property observations reported from ice-tethered profiler #37. Profiler #37 was deployed near Barrow, Alaska from a research vessel. These data are used to characterize upper ocean dynamics to better understand the underlying conditions for sea ice formation. Included in this dataset are …

😀 ✅

What makes a good metadata record?

Documented filesystem artifacts

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

Attribute Information

https://github.com/ropensci/EML/

```
attributes2 <- data.frame(attributeName = c('Time', 'Wind_Speed'),
                          attributeDefinition = c('Date and time of wind speed reading', 'Measured
                          measurementScale = c('dateTime', 'ratio'),
                          domain = c('dateTimeDomain', 'numericDomain'),
                          formatString = c('YYYY-MM-DD hh:mm:ss', NA),
                          definition = c(NA, NA),
                          unit = c(NA, 'metersPerSecond'),
                          numberType = c(NA, 'real'),
                          missingValueCode = c(NA, NA),
                          codeExplanation = c(NA, NA),
                          stringsAsFactors = FALSE)
attributeList2 <- set_attributes(attributes2)
id2 <- 'PID2'
physical2 <- pid_to_eml_physical(mn, id2)
dataTable2 <- new('dataTable',
                  entityName = 'EagleMtnWindData.csv',
                  entityDescription = 'Wind data from Eagle Mountain',
                  physical = physical2,
                  attributeList = attributeList2)
```

Attribute Information

https://github.com/ropensci/EML/

```
attributes2 <- data.frame(attributeName = c('Time', 'Wind_Speed'),
                          attributeDefinition = c('Date and time of wind speed reading', 'Measured
                          measurementScale = c('dateTime', 'ratio')
                          domain = c(
                          formatStri
                          definition
                          unit = c(NA
                          numberType
                          missingValu
                          codeExplana
                          stringsAsFa
attributeList2 <- set_attributes(attr
id2 <- 'PID2'
physical2 <- pid_to_eml_physical(mn,
dataTable2 <- new('dataTable',
                  entityName = 'Eagle
                  entityDescription =
                  physical = physical
                  attributeList = att
```

```
> attributeList2
<attributeList>
  <attribute>
    <attributeName>Time</attributeName>
    <attributeDefinition>Date and time of wind speed reading</attributeDefinition>
    <measurementScale>
      <dateTime>
        <formatString>YYYY-MM-DD hh:mm:ss</formatString>
      </dateTime>
    </measurementScale>
  </attribute>
  <attribute>
    <attributeName>Wind_Speed</attributeName>
    <attributeDefinition>Measured wind speed</attributeDefinition>
    <measurementScale>
      <ratio>
        <unit>
          <standardUnit>metersPerSecond</standardUnit>
        </unit>
        <numericDomain>
          <numberType>real</numberType>
        </numericDomain>
      </ratio>
    </measurementScale>
  </attribute>
</attributeList>
```

Attribute Information

	A	B	C	D	E	F	G	H	I	J
1	DateCollecte	Year	location	sampler	agency	LAT	LONG	SampleType	QCERROR	comment
2	26-Mar-89	1989	ROCKB	KARINEN	NMFS ABL	60.337	-147.124	ENV	GOOD	
3	26-Mar-89	1989	ROCKB	KARINEN	NMFS ABL	60.337	-147.124	ENV	GOOD	
4	26-Mar-89	1989	ROCKB	KARINEN	NMFS ABL	60.337	-147.124	ENV	GOOD	
5	26-Mar-89	1989	ROCKB	KARINEN	NMFS ABL	60.337	-147.124	ENV	GOOD	
6	26-Mar-89	1989	ROCKB	KARINEN	NMFS ABL	60.337	-147.124	ENV	GOOD	
7	26-Mar-89	1989	ROCKB	KARINEN	NMFS ABL	60.337	-147.124	ENV	GOOD	
8	26-Mar-89	1989	CONST	KARINEN	NMFS ABL	60.349	-146.761	ENV	GOOD	
9	26-Mar-89	1989	CONST	KARINEN	NMFS ABL	60.349	-146.761	ENV	GOOD	
10	26-Mar-89	1989	CONST	KARINEN	NMFS ABL	60.349	-146.761	ENV	GOOD	
11	29-Mar-89	1989	SIWAB	KARINEN	NMFS ABL	60.954	-147.681	ENV	GOOD	

	A	B	C	D	E	F
1	attributeName	attributeDefinition	unit	formatString	missingValueCode	missingValueCodeDefinition
2	DateCollected	Date sample was collected		"YYYY-MM-DD"		
3	Year	Year sample was collected		"YYYY"		
4	location	Location of sample				
5	sampler	Person who collected sample				
6	agency	Agency responsible for collection				
7	LAT	Latitude of location where sample was collected	degree		"NA"	no latitude/longitude information was collected
8	LONG	Longitude of location where sample was collected	degree		"NA"	no latitude/longitude information was collected
9	SampleType	Type of sample (F = feather, S = skin, O = other)				
10	QCERROR	Whether there was an error in the quality control process				
11	comment	Sample comments				
12						
13						
14						

Attribute Information

- Column names
- Column name definitions
- Format strings for dates
- Units
- Missing value codes and definitions

What makes a good metadata record?

Involved parties

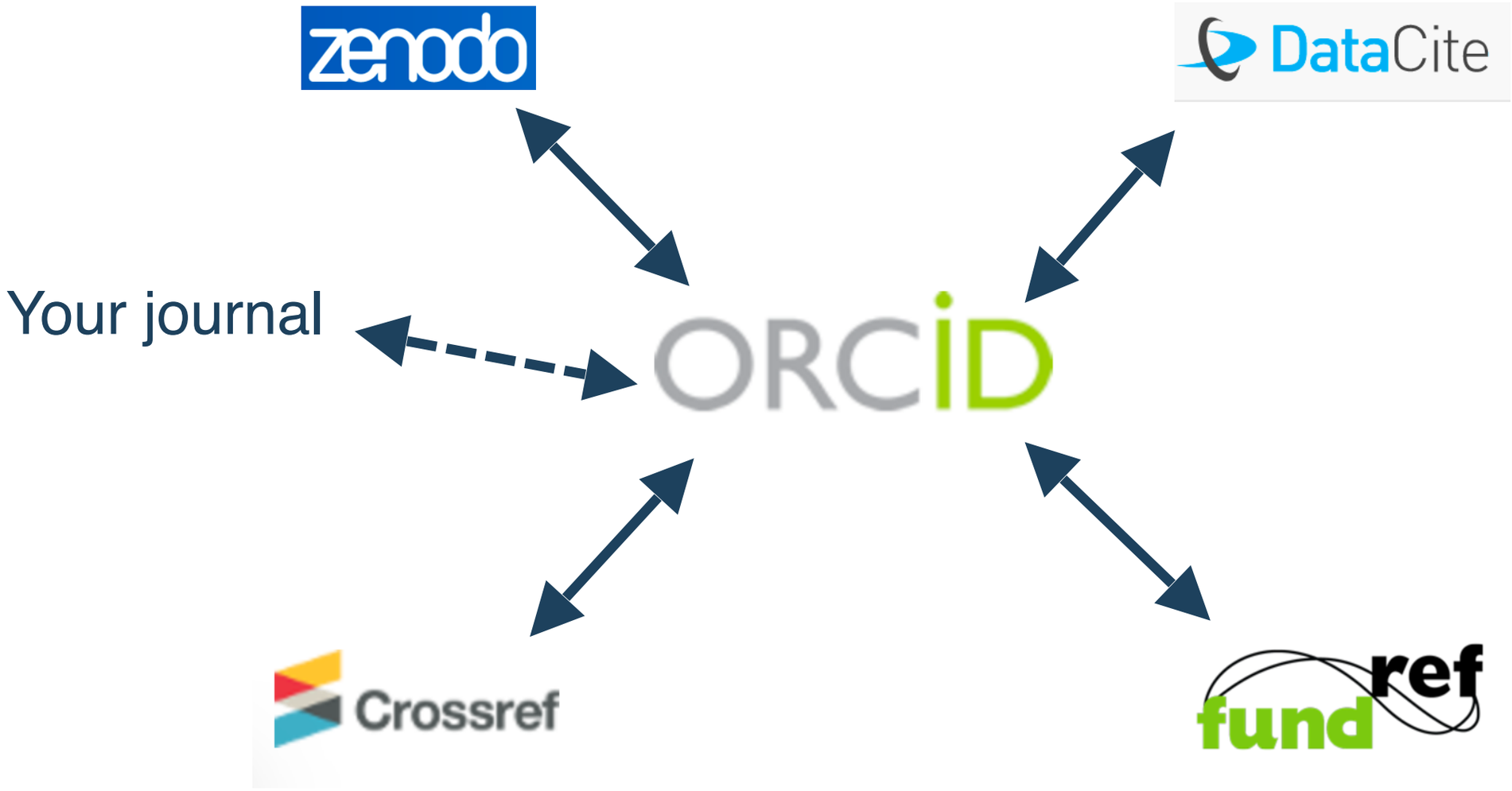
- Name alone is not enough…
  - to assign credit, nor
  - to disambiguate across data sets
- Email addresses help
- Including ORCID iD is best

ORCID iDs: ”Wait, what is an ORCID iD?”

- Like an ISBN for people
  - e.g. mine: orcid.org/0000-0002-1006-9496
- Enables unambiguous reference to humans
- Free
- Becoming a community norm
- Inherently connected…

ORCID iDs

Inherently connected



Activity

Register an ORCID iD:

- orcid.org/register

Sign in to dev.nceas.ucsb.edu/#share