Tutorials on Data Management: Metadata

DataONE Community Engagement & Outreach Working Group

Metadata

Image Attribution: CC image by bonus on Flickr

Lesson Topics

- Explanation of metadata
- Illustrate the value and utility of metadata to data users, data providers, and organizations
- Examine information included in a metadata record
- Examples of metadata standards and how to choose
- Preparing to write metadata
- Tips for writing a quality metadata record

Photo Attribution: CC image by Alec Couros on Flickr

Learning Objectives

After completing this lesson, the participant will be able to:

- Identify & list the types of **information typically included in metadata records** for environmental datasets
- Identify 3 **reasons metadata is of value** to data users, data developers, and organizations
- List 3 **uses for metadata**, beyond discovery of data
- Identify and describe factors that may determine which **metadata standards** are most appropriate for a given dataset
- List **steps to prepare** to write metadata
- Explain how to **write good metadata**

DataONE

The Data Life Cycle
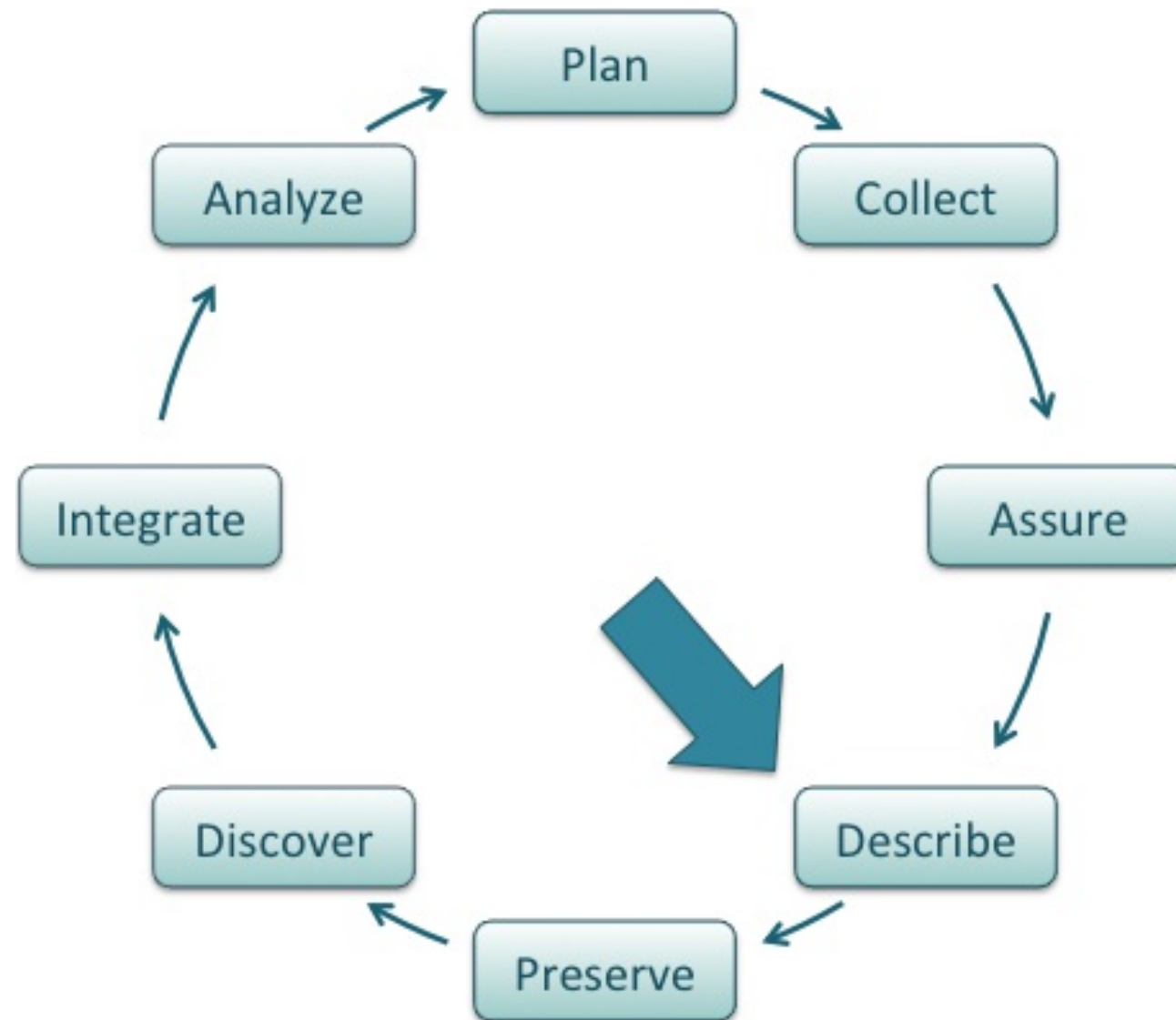
What is metadata?

Metadata is: Data 'reporting'

- **WHO** created the data?
- **WHAT** is the content of the data?
- **WHEN** were the data created?
- **WHERE** is it geographically?
- **HOW** were the data developed?
- **WHY** were the data developed?

Photo Attribution: Photo by Michelle Chang. All Rights Reserved
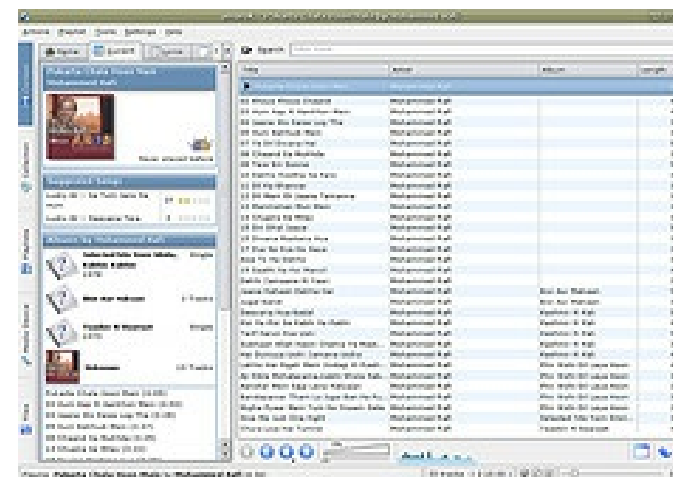
DataONE

Metadata in Real Life

Metadata is all around...

Moby-Dick, or, the Whale / by Herman Melville.

Main Author:	Melville, Herman, 1819-1891.
Language(s):	English
Published:	New York : Harper & Brothers ; 1851.
Subjects:	Whales > Fiction. Whaling > Fiction. Ship captains > Fiction. Whaling ships > Fiction. Ahab, Captain (Fictitious character) > Fiction.
Note:	Publisher's advertisements on final 6 p. First American edition.
Physical Description:	xxiii, 634, [8] p. ; 20 cm.
Locate a Print Version:	Find in a library

Hathi Trust

Nutrition Facts	Amount/serving	%DV*	Amount/serving	%DV*
Serv. Size 1 oz (28g)	**Total Fat** 13g	**19**%	**Total Carb.** 8g	3%
Serv. Per Cont. 4	Sat. Fat 1.5g	**7**%	Fiber 6g	**23**%
Calories 150	*Trans* Fat 0g		Sugars 0g	
Fat Cal. 110	**Cholest.** 0mg	**0**%	**Protein** 5g	
	Sodium 180mg	**7**%		

*Percent Daily Values (DV) are based on a 2,000 calorie diet.

Vitamin A 0% • Vitamin C 0% • Calcium 15% • Iron 15%

CC image by Mskadu on Flickr

DataONE

Metadata: What are they good for?

Metadata: captures information

USGS Science Data Catalog: enabling discovery

DataONE: enables exchange

Metadata: Why are they important?
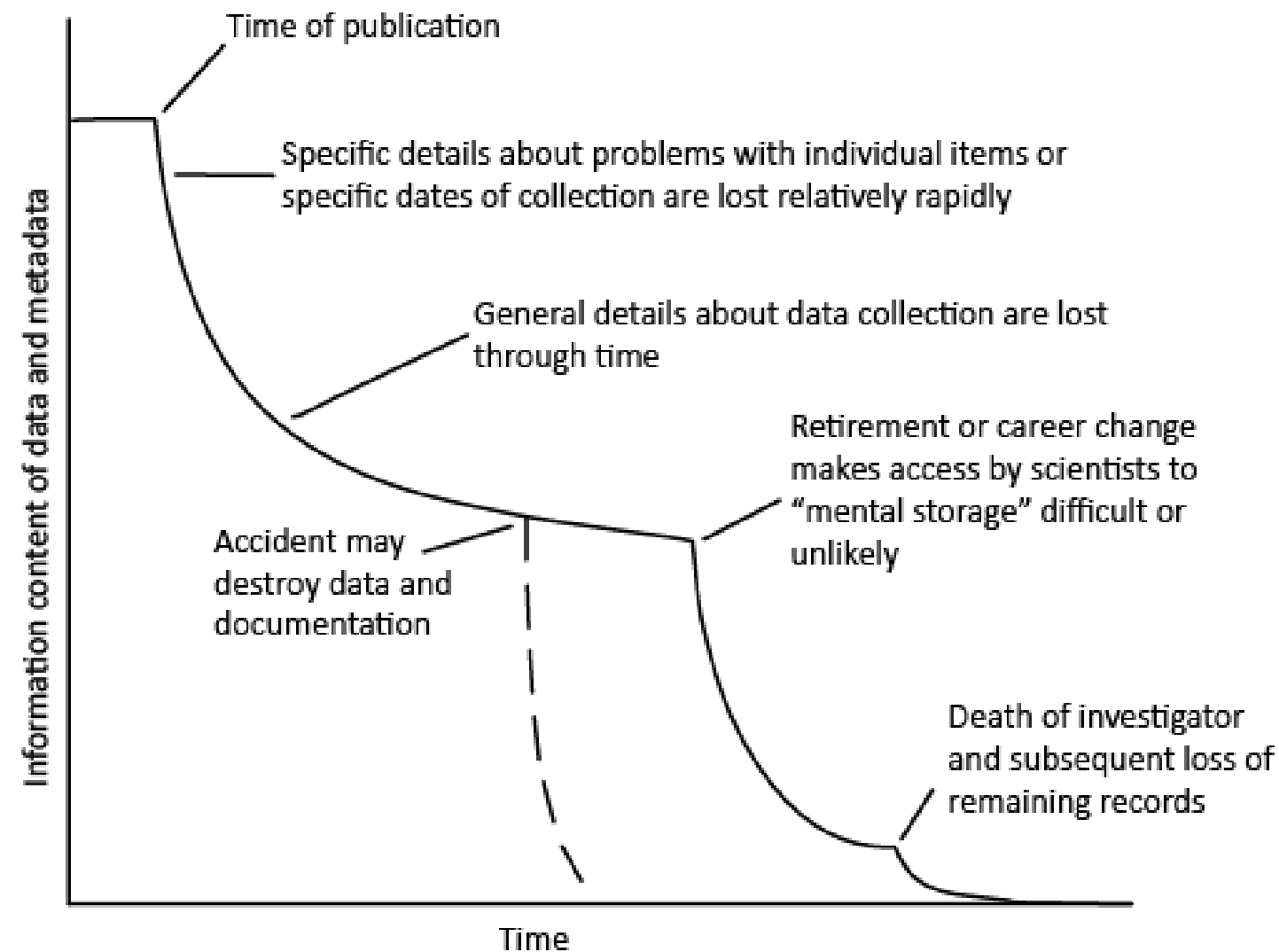
Michener, W. K., Brunt, J. W., Helly, J. J., Kirchner, T. B. and Stafford, S. G. (1997), NONGEOSPATIAL METADATA FOR THE ECOLOGICAL SCIENCES. Ecological Applications, 7: 330–342. doi:10.1890/1051-0761(1997)007[0330:NMFTES]2.0.CO;2]

DataONE

Metadata: Why are they important?

Policy decisions based on data can only be defended if the metadata are good quality

- Regulatory decisions based on undocumented data are not defensible
- Metadata accuracy and details are important as supporting evidence for the science and policy

Controversies arise when metadata are incomplete and/or absent

Public domain image from Unsplash.com

The Value of Metadata
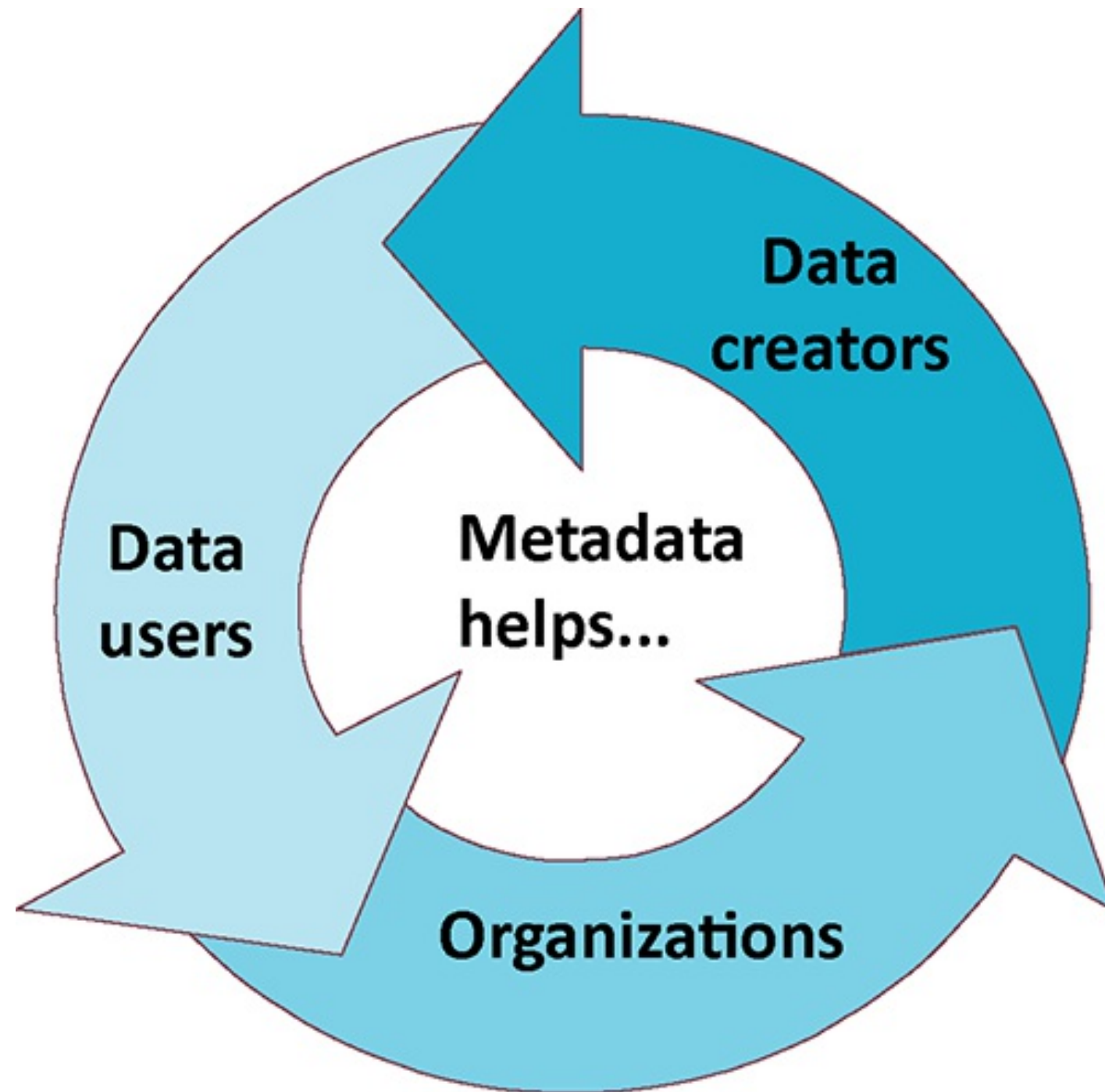
DataONE Community Education & Outreach Working Group

What is the Value to Scientists, Researchers, and other Data Creators?

Metadata allows data developers to:

- Avoid data duplication
- Share reliable information
- Publicize efforts – promote the work of a scientist and his/her contributions to a field of study
- Metadata reuse saves time and resources in the long-run

CC image by US Embassy Guyana on Flickr

What is the Value to Data Users?

Metadata gives a user the ability to:

- **Search, retrieve, and evaluate** dataset information from both inside and outside an organization
- **Find data**: Determine what data exists for a geographic location and/or topic
- **Determine applicability**: Decide if a dataset meets a particular need
- **Discover** how to acquire the dataset identified; process and use the dataset
- **Understand** the dataset, including definitions of column names, or expected numerical ranges found in the data

CC image by ASEE on Flickr

What is the Value to Organizations?

Metadata helps ensure an organization's investment in data:

- Documentation of data **processing steps**, quality control, definitions, data uses, and restrictions
- Ability to **use** data after initial intended purpose
- Allows organization to **track data use** and facilitates publication

Transcends people and time:

- Offers data **permanence**
- Creates institutional memory

Advertises an organization's research:

- Creates possible new **partnerships** and collaborations through data sharing

The Utility of Metadata

Metadata can support:

- data distribution (discovery, catalogs)
- data management (provenance, accountability, liability)

Public domain image from Unsplash.com

Data○**NE**

Data Distribution: Discovery

The descriptive content of the metadata file can be used to identify, assess, and access available data resources.
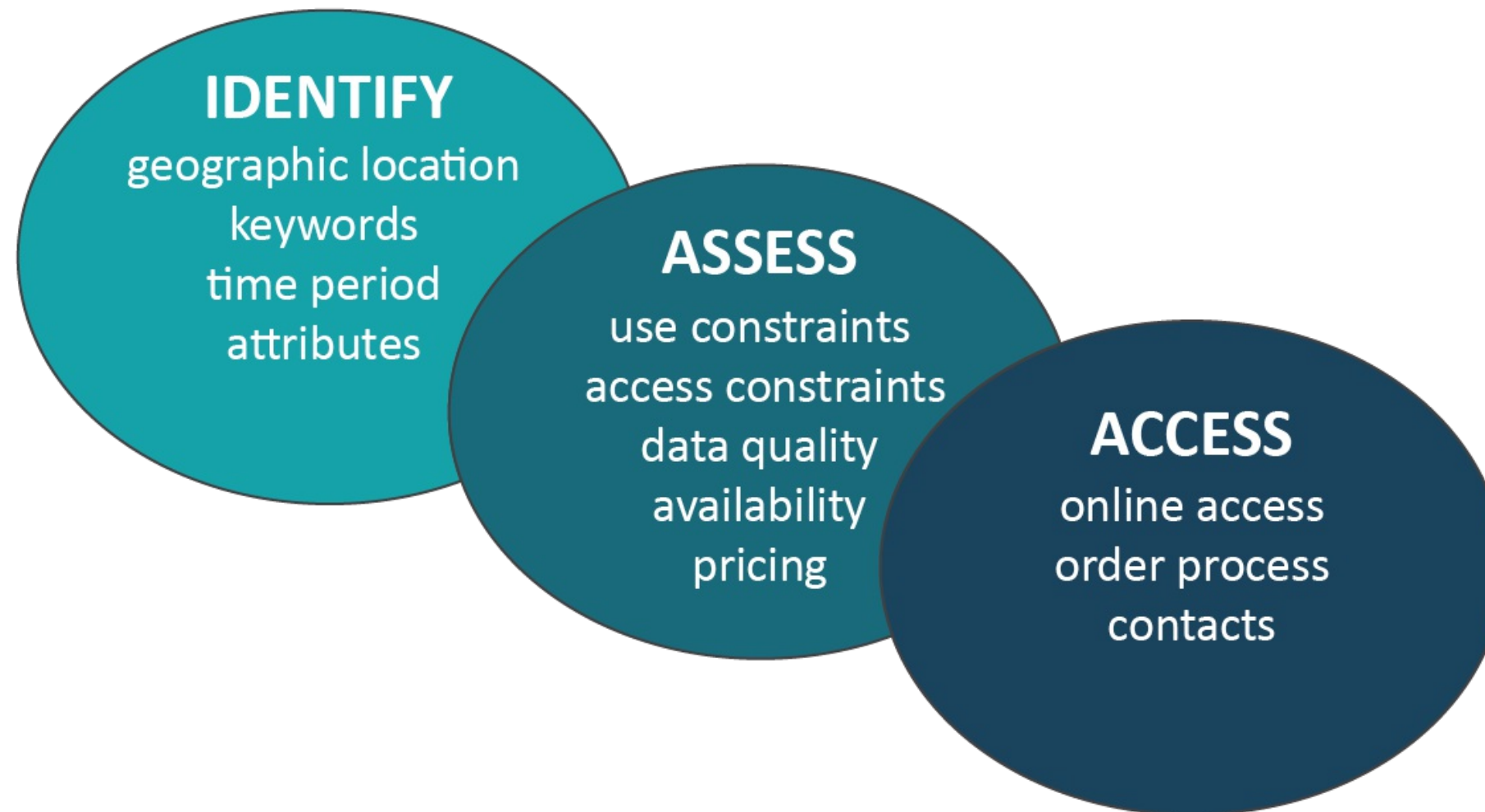
DataONE Community Education & Outreach Working Group

Data Distribution: Catalogs

- **DataONE**
  - Data discovery, knowledge, community…for a sustainable future
  - https://search.dataone.org
- **Data.gov**
  - Federal e-gov geospatial data portal
  - http://www.geo.data.gov
- **Metacat**
  - Repository for data and metadata
  - http://knb.ecoinformatics.org/index.jsp

- **US Geological Survey**
  - USGS Science Data Catalog
  - http://data.usgs.gov/datacatalog
- **ArcGIS Online**
  - ESRI sponsored national geospatial data portal
  - http://www.geographynetwork.com

Data Distribution Example: DataONE

Data Management: Provenance

Metadata records can be used to track data provenance accurately

- Are the data **current**?
- Are the data in a **reliable format**?
- Where are the data **stored**?
- **Contact** information
- **Distribution** policies, availability, pricing, URLs
- New **derivations** of the dataset

Public domain image from Unsplash.com

Data Management: Accountability

Metadata allows you to repeat a scientific process if:

- methodologies are defined
- variables are defined
- analytical parameters are defined

Metadata allows you to defend your scientific process:

- process is documented
- increasingly data savvy public requires metadata for consumer information

DataONE

Data Management: Liability

Metadata is a declaration of:

- **Purpose** – the originator's intended application of the data
- **Use Constraints** - inappropriate applications of the data
- **Completeness** - features or geographies excluded from the data
- **Distribution Liability** - explicit liability of the data producer and assumed liability of the consumer

Public domain image from Unsplash

Concerns About Creating Metadata (1)

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.

Public domain image from Unsplash

Concerns About Creating Metadata (2)

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

DataONE

What is a Metadata Standard?

A Standard provides a structure to describe data with:

- Common terms to allow consistency between records
- Common definitions for easier interpretation
- Common language for ease of communication
- Common structure to quickly locate information

In search and retrieval, standards provide:

- Documentation structure in a reliable and predictable format for computer interpretation
- A uniform summary description of the dataset

What does a metadata standard include?

Components of metadata:

- A metadata standard is made up of defined **elements**, including the type of information the user should enter (e.g. text, numbers, date).
- Examples of elements include:
  - title, abstract, keywords
  - persistent link, usage rights
  - variables, units
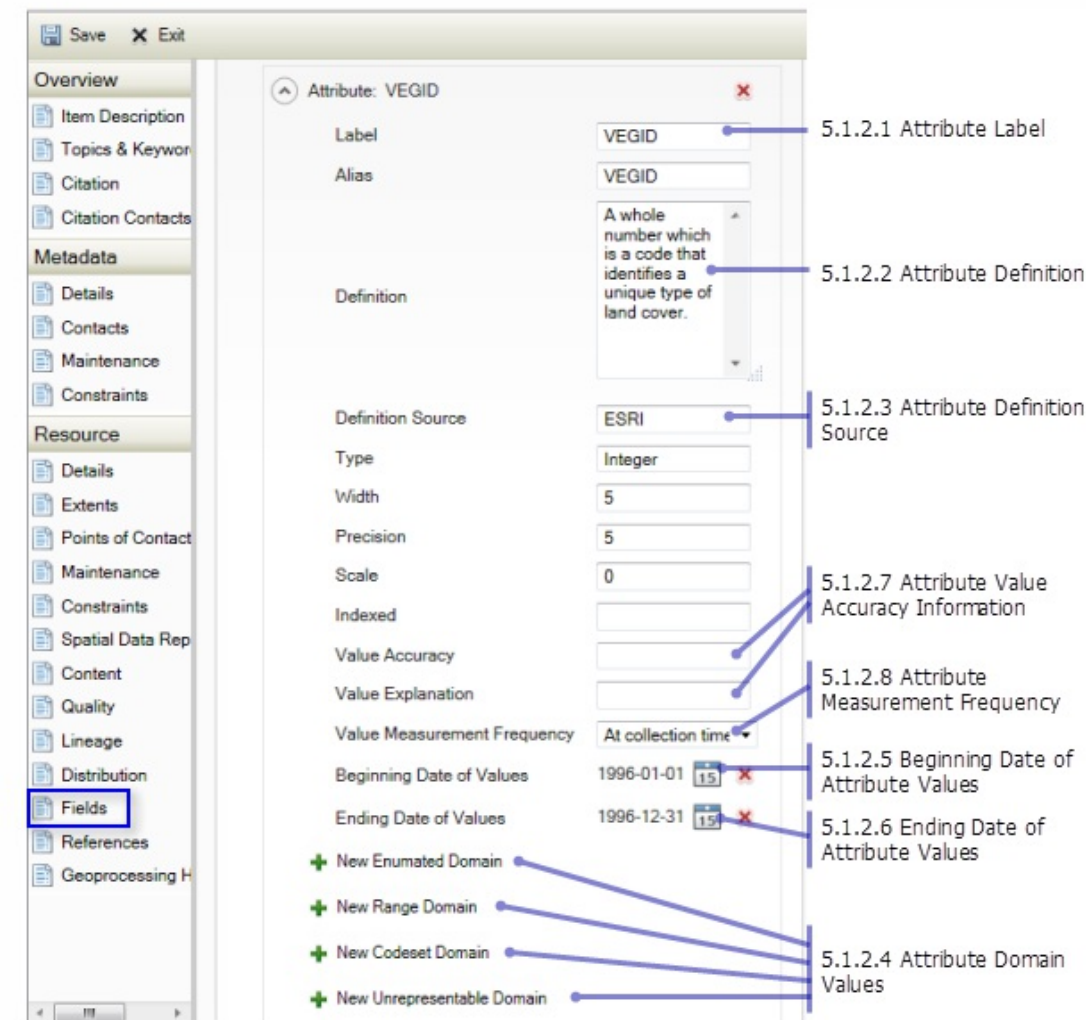
ArcGIS/Esri

What does a metadata record look like?

Metadata from the Santa Barbara Channel LTER program in kelp forest ecology

This is the descriptive metadata. Notice tabs for other, more granular use metadata

Data Set (knb-lter-sbc.19.22)

SBC LTER: Reef: Kelp Forest Community Dynamics: Invertebrate and algal density

Summary and Data Links	People and Organizations	Temporal, Geographic and Taxonomic Coverage	Methods and Protocols

Data Set General Information:

Identifier:	knb-lter-sbc.19.22 (in the *knb* Catalog System)
Alternate Identifier:	DOI: 10.6073/pasta/62803c95783c4e771695d1c6cc3d23ac
Alternate Identifier:	knb-lter-sbc.19
Abstract:	These data describe the abundance and size of a select group of about 50 species of benthic invertebrates and understory algae in fixed plots (either 1mx1m or 20mx1m) along permanent transects. These data are part of SBCLTER?s kelp forest monitoring program to track long-term patterns in species abundance and diversity of reef-associated organisms in the Santa Barbara Channel, California, USA. The sampling locations in this dataset are at nine reef sites along the mainland coast of the Santa Barbara Channel and at two sites on the north side of Santa Cruz Island. These sites reflect several oceanographic regimes in the channel and vary in distance from sources of terrestrial runoff. Sampling began in 2000, and these data are updated annually.
Short Name:	KFCD Reef quad-swath counts
Publication Date:	2016-09-06
Language:	english

Time Period:

Begin:	2000-08-01
End:	2016-07-26

Detailed Data Description and Download:

Data Table — Benthic community survey, inverts and understory algae, all years
abundance and size of selected species of benthic invertebrates and understory algae in fixed plots along permanent transects.

People and Organizations:

View complete information for all parties

Contact:	Information Manager, Santa Barbara Coastal LTER [email]
Organization	Santa Barbara Coastal LTER
Owner:	Reed, Daniel C

Data Set Citation

How to cite this data set:	Reed, D. C. . 2016. SBC LTER: Reef: Kelp Forest Community Dynamics: Invertebrate and algal density. Santa Barbara Coastal LTER. doi:10.6073/pasta/62803c95783c4e771695d1c6cc3d23ac

Key Words and Terms

By Thesaurus:

Kelp_forest_monitoring	
Santa Barbara Coastal LTER Places	Arroyo Burro, Arroyo Hondo, Arroyo Quemado, Bullito, Carpinteria, Goleta Bay, Isla Vista, Mohawk, Naples, Santa Cruz Island, Diablo, Santa Cruz Island, Twin Harbor West
Knowledge Network for Biocomplexity	invertebrate
Global Change Master Directory (GCMD) v6.0.0.0.0	Macroalgae, Marine Invertebrates, Marine Plants
LTER Controlled Vocabulary v1: Core Research Areas	populations
LTER, Santa Barbara Coastal	

Data Set Usage Rights

- The user of SBC LTER data agrees to contact the data owner (i.e., the SBC investigator responsible for data) prior to publishing. Where appropriate, users whose projects are integrally dependent on SBC LTER data are encouraged to consider collaboration and/or co-authorship with the data owner.
- The user agrees to cite SBC LTER in all publications that use SBC LTER data by including the following statement in the Acknowledgments: "Data were provided by the Santa Barbara Coastal LTER, funded by the US National Science Foundation (OCE-1232779)".
- The user agrees to send the full citation of any publication using SBC LTER data to sbclter@msi.ucsb.edu
- Users are prohibited from selling or redistributing any data provided by SBC LTER.
- Extensive efforts are made to ensure that online data are accurate and up to date, but SBC LTER will not take responsibility for any errors that may exist.
- The user agrees also to adhere to the Data Use Agreement of the Long Term Ecological Research Network.
- Any violation of the terms of this agreement will result in immediate forfeiture of the data and loss of access privileges to other SBC LTER data sets.
- SBC LTER is committed to protecting the privacy and accuracy of your confidential information. See our Privacy Policy for more information.

DataONE

Choosing Metadata Standards

Public domain image from Unsplash

DataONE

Multiple Metadata Standards Exist: Examples (1)

- Dublin Core Element Set
  - Emphasis on web resources, publications
- FGDC Content Standard for Digital Geospatial Metadata (CSDGM)
  - Emphasis on geospatial data
  - The Biological Data Profile (BDP) of the CSDGM is a profile to the CSDGM with an emphasis on biological data (and geospatial)
- ISO 19115/19139 Geographic information – metadata
  - Emphasis on geospatial data and services

Public domain image from Unsplash

Multiple Metadata Standards Exist: Examples (2)

- Ecological Metadata Language (EML)
  - Focus on ecological data
- Darwin Core (DwC)
  - Emphasis on museum specimens
- Geography Markup Language (GML)
  - Emphasis on geographic features (roads, highways, bridges)

Public domain image from Unsplash

Comparing Metadata Standards

Terminology for the same concepts may vary across standards

Ecological Metadata Language (EML)	FGDC Content Standard for Digital Geospatial Metadata
Title	Title
Abstract	Abstract
Entity Description	Entity Type Definition
Intellectual Rights	Use Constraints

DataONE

Choosing a Metadata Standard: Factors (1)

Your data type

- Are you working mainly with GIS data? Raster/vector or point data? Do you have biological or shoreline information in your dataset?
  - Consider the FGDC Content Standard for Digital Geospatial Metadata with one of its profiles: the Biological Data Profile or the Shoreline Data Profile.
- Are you working with data retrieved from instruments such as monitoring stations or satellites? Are you using geospatial data services such as applications for web-mapping applications or data modeling?
  - If so, then consider using the ISO 19115-2 standard
- Are you mainly working with ecological data?
  - Consider Ecological Metadata Language (EML)

Choosing a Metadata Standard: Factors (2)

Your organization's policies

- Do they state which standard to use?
- What resources are available to create metadata?

Examples of Tools:

- FGDC CSDGM tools from FGDC.gov
- EML: Morpho
- ISO metadata editor from FGDC.gov
- XML: Spy or Oxygen
- CatMD

Other factors: Availability of human support; instructional materials; use of controlled vocabularies; output formats

DataONE

What Makes a Good Metadata Record?

Metadata are developed continuously throughout the entire data lifecycle
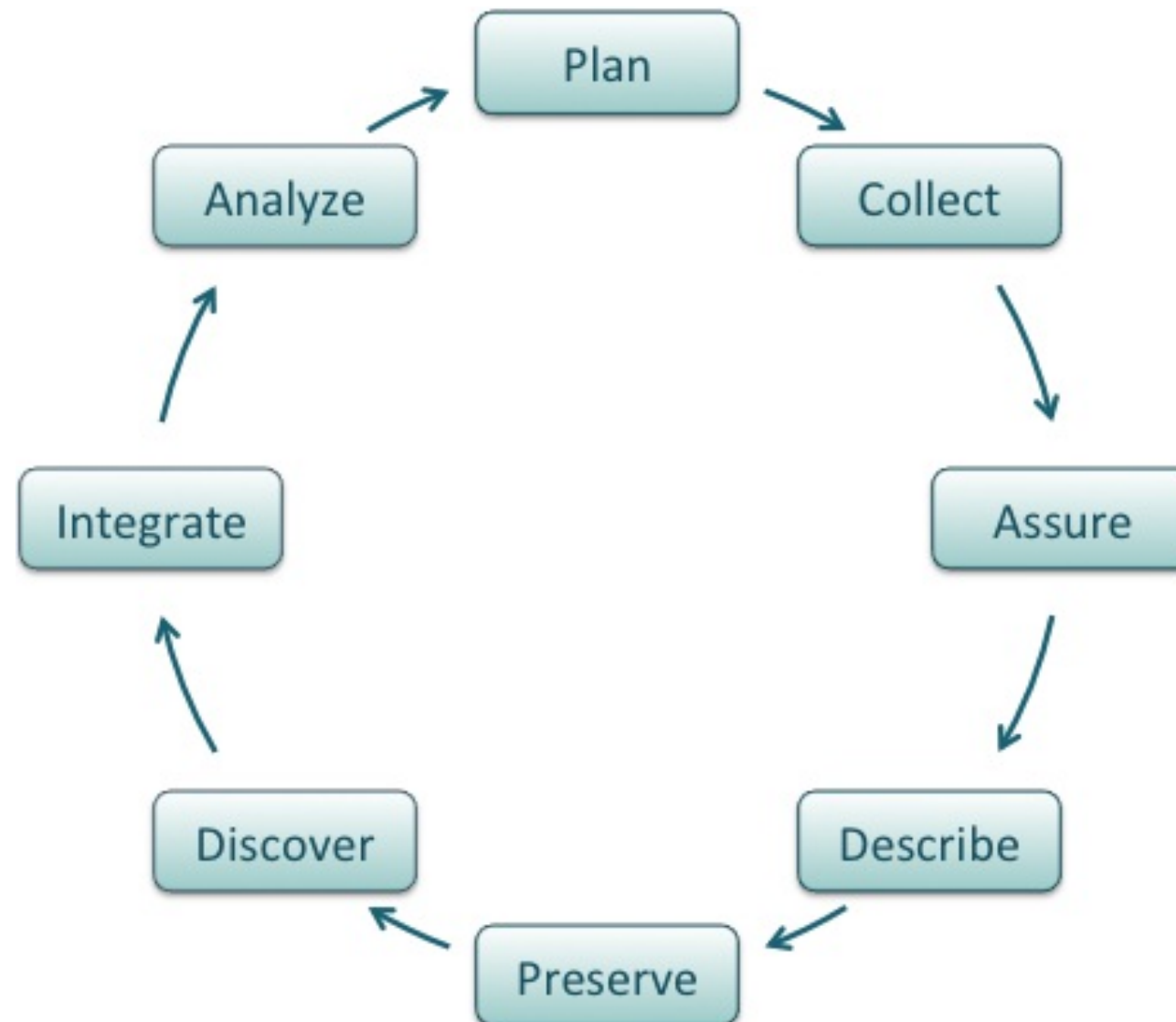
What Makes a Good Metadata Record?

Consistency with commonly used fields Here are some examples for a FGDC CSDGM record:

Field	DO THIS	NOT THIS
Publisher	`<publish>U.S. Geological Survey</publish>`	`<publish>USGS</publish>`
Date	`<pubdate>YYYYMMDD</pubdate>` `<pubdate>YYYY</pubdate>`	`<pubdate>MM/DD/YYYY</pubdate>` `<pubdate>May 27, 2003</pubdate>`
Keywords	`<placekt>Geographic Names Information System</placekt>` `<placekey>Roosevelt National Forest</placekey>`	`<themekey>Roosevelt Forest</themekey>`

DataONE

What Makes a Good Metadata Record?

Use Authority Files and Standard Vocabulary

- Global Change Master Directory
- Geographic Names Information System
- Getty Thesaurus of Geographic Names
- ISO 19115 Topic Category Thesaurus

Public domain image from Unsplash

DataONE

What Makes a Good Metadata Record?

Acronyms

- Spell out acronyms with first use. Many acronyms have multiple meanings (e.g., DOI)
- Use widely known acronyms only when it corresponds to specific metadata fields such as file formats (e.g., TIFF, JPEG, PDF)

Photo Attribution: Photo by Michelle Chang. All Rights Reserved

What Makes a Good Metadata Record?

Provide all of the critical information for **discovery**, **understanding**, and **reuse**:

- Identification Information
- Entities & Attributes
- Data Quality
- Access, Use & Liability Constraints
- Distribution
- Spatial References

Public domain image from Unsplash

DataONE

What Makes a Good Metadata Record?

Provide all of the critical information for: Identification

Data Set (knb-lter-sbc.77.1)

SBC LTER: Reef: Abundance, size and fishing effort for California Spiny Lobster (Panulirus interruptus), ongoing since 2012

Summary and Data Links	People and Organizations	Temporal, Geographic and Taxonomic Coverage	Methods and Protocols

Data Set General Information:

Identifier:	knb-lter-sbc.77.1 (in the *knb* Catalog System)
Alternate Identifier:	DOI: 10.6073/pasta/41b7ec14a5ea42ace9e81fa8de9d4568
Alternate Identifier:	knb-lter-sbc.77
Abstract:	Data on abundance, size and fishing pressure of California spiny lobster (Panulirus interruptus) are collected along the mainland coast of the Santa Barbara Channel. Spiny lobsters are an important predator in giant kelp forests off southern California. Two SBC LTER study reefs are located in or near the California Fish and Game Network of Marine Protected Areas (MPA), Naples and Isla Vista, both established as MPAs on 2012-01-01. MPAs provide a unique opportunity to investigate the effects of fishing on kelp forest community dynamics. Sampling began in 2012 and is ongoing.

This dataset contains two tables. 1) Abundance and size data collected annually by divers in late summer before the start of the fishing season at five SBC LTER long term kelp forest study sites: two within MPAs (Naples and Isla Vista) and three outside (Arroyo Quemado, Mohawk and Carpinteria). 2) Fishing pressure, as determined by counting the number of commercial trap floats. Data are collected every two to four weeks during the lobster fishing season (October to March) at nine sites along the mainland, eight of which are also SBC LTER long-term kelp forest study reefs. |
Short Name:	lobster abundance and fishing effort
Publication Date:	2014-01-01
Language:	english

Detailed Data Description and Download:

Data Table	Time-series of lobster abundance and size
Time-series of abundance and size of California spiny lobster (Panulirus interruptus) at SBC LTER reefs
Data Table	Time-series of lobster trap buoy counts
Time-series of counted trap marker buoys in spatially-discrete subsections inshore of the 15-m isobath

People and Organizations:

View complete information for all parties

Contact:	Information Manager, Santa Barbara Coastal LTER [email]
Organization	Santa Barbara Coastal LTER
Owner:	Reed, Daniel C

What Makes a Good Metadata Record?

Provide all of the critical information for: Entity / Attribute

Table Column Descriptions

	Year	Month	Date	Site	SBC LTER Transect name	Lobster Transect name	Size	Count	Observer code	Sample Area
Column Name	YEAR	MONTH	DATE	SITE	SBC_LTER_TRANSECT	LOBSTER_TRANSECT	SIZE	COUNT	OBS_CODE	AREA
Definition	Calendar year	Calendar month	Date of survey, (d)d-mon-yy	Code for site sampled	40m transects defined by six permanent markers (stainless steel eyebolts or rebar stakes) at 0, 8, 16, 24, 32, and 40m	Code for the lobster transect. 4 lobster transects (60m x 2m) are associated with an SBC LTER permanent transect. Transects A-D are arranged south to north (A is southern-most and D is the northern-most)	Visual estimate of lobster carapace length in mm	Total number of lobsters of this size in the lobster transect	Numeric code indicating the SBCLTER data collector	Lobster transect sampling area meters squared
Storage Type	integer	integer	date	string	string	string	integer	integer	string	integer
Measurement Type	dateTime	nominal	nominal	nominal	nominal	nominal	ratio	ratio	nominal	ratio
Measurement Values Domain	Format YYYY / Precision 1	Definition any text	Definition any text	Allowed values and definitions	Definition any text	Definition any text	Unit millimeter / Precision 1 / Type integer	Unit number / Precision 1 / Type integer	Definition any text	Unit meterS / Precision 1 / Type integer

Santa Barbara LTER Program

What Makes a Good Metadata Record?

Provide all of the critical information for: Data quality

- Accuracy
- Consistency
- Completeness

	Cruise	station_name	Type	mon/day/yr	hh:mm	Lon (degE)	Lat (degN)
Column Name	odv_cruise	odv_station	odv_type	odv_mm_dd_yy	odv_hh_mm	odv_lon	odv_lat
Definition	column required by ODV.	2-letter code for this station name	cast type at this station, column required by ODV	format is mm/dd/yyyy, column required by ODV	time, UTC, in 24 hr format, column required by ODV	Longitude recorded by the first ctd sample at the station, if available. If not available, then nominal longitude for this location. East is positive. Column required by ODV: all samples which are to be considered at same station must have the same value in this column.	Latitude recorded by the first ctd sample at the station, if available. If not available, then nominal latitude for this location. North is positive. Column required by ODV: all samples which are to be considered at same station must have the same value in this column.
Storage Type	string	string	string	string	string	float	float
Measurement Type	nominal	nominal	nominal	dateTime	dateTime	ratio	ratio
Measurement Values Domain	**Definition** any text	**Allowed values and definitions**	**Allowed values and definitions**	**Format** mm/dd/yyyy **Precision** 1	**Format** hh:mm **Precision** 1	**Unit** degree **Precision** 1e-05 **Type** real **Min** -180 **Max** 180	**Unit** degree **Precision** 1e-05 **Type** real **Min** -90 **Max** 90
Missing Value Code	**Code** 99999 **Expl** no measurement collected at this point, or not available	**Code** 99999 **Expl** no measurement collected at this point, or not available	**Code** 99999 **Expl** no measurement collected at this point, or not available	**Code** 99999 **Expl** no measurement collected at this point, or not available	**Code** 99999 **Expl** no measurement collected at this point, or not available	**Code** 99999.00000 **Expl** no measurement collected at this point, or not available	**Code** 99999.00000 **Expl** no measurement collected at this point, or not available

Santa Barbara LTER Program

DataONE

What Makes a Good Metadata Record?

Provide all of the critical information for: Data lineage

Data Set Citation

How to cite this data set:	Washburn, L. , MA. Brzezinski, CA. Carlson and DA. Siegel. 2014. SBC LTER: Ocean: Ocean Currents and Biogeochemistry: Nearshore water profiles (monthly CTD and chemistry). Santa Barbara Coastal LTER. doi:10.6073/pasta/4e9255de0662434ded1f50b6a310510c

These methods, instrumentation and/or protocols apply to all data in this dataset:

Protocols and/or Procedures

Description:	**CTD Instrumentation, collection and analysis** CTD data are collected and analyzed according to the following protocols and instruments
Protocol:	Author: Washburn Title: SBC Monthly Water Sampling - CTD and bottles Available Online: View: http://sbc.lternet.edu/external/Ocean/Protocols/Monthly_Water_Sampling_CTD_bottles_textonly.pdf
Protocol:	Author: Washburn Title: SBC LTER Monthly CTD processing -- All data Available Online: View: http://sbc.lternet.edu/external/Ocean/Protocols/Washburn_2007-10-23_SBCLTER_monthly_CTD_processing.pdf
Instrument(s):	SBE19-Plus with rosette sampler (Go-Flo) and bottles drained under pressure (june 2008 - present)
Instrument(s):	SBE19-Plus with rosette sampler (Go-Flo) (Feb 2003 - Jun 2008)
Instrument(s):	SBE19 Seacat Profiler, with pump (2001 - 2002)

What Makes a Good Metadata Record?

Provide all of the critical information for: Access, Use & Liability Constraints

- *Access Constraints*: restrictions and legal prerequisites for access the data.
- *Use Constraints*: restrictions and legal prerequisites for using the data after access is granted.
- *Distribution Liability*: statement of the liability assumed by the distributor with respect to content and accuracy of the data.

Data Set Usage Rights

- The user of SBC LTER data agrees to contact the data owner (i.e., the SBC investigator responsible for data) prior to publishing. Where appropriate, users whose projects are integrally dependent on SBC LTER data are encouraged to consider collaboration and/or co-authorship with the data owner.
- The user agrees to cite SBC LTER in all publications that use SBC LTER data by including the following statement in the Acknowledgments: "Data were provided by the Santa Barbara Coastal LTER, funded by the US National Science Foundation (OCE-1232779)".
- The user agrees to send the full citation of any publication using SBC LTER data to sbclter@msi.ucsb.edu
- Users are prohibited from selling or redistributing any data provided by SBC LTER.
- Extensive efforts are made to ensure that online data are accurate and up to date, but SBC LTER will not take responsibility for any errors that may exist.
- The user agrees also to adhere to the Data Use Agreement of the Long Term Ecological Research Network.
- Any violation of the terms of this agreement will result in immediate forfeiture of the data and loss of access privileges to other SBC LTER data sets.
- SBC LTER is committed to protecting the privacy and accuracy of your confidential information. See our Privacy Policy for more information.

DataONE

What Makes a Good Metadata Record?

Provide all of the critical information for: Accessing the data

Standard_<mark>Order</mark>_Process:

   Digital_Form:

      Digital_Transfer_Information:

         Format_Name: FileGDB
         Format_Version_Number: 10.1
         Format_Information_Content:
            ESRI File Geodatabase

      Digital_Transfer_Option:

      Online_Option:

         Computer_Contact_Information:

         Network_Address:

            *Network_Resource_Name:*ftp://rockyftp.cr.usgs.gov/vdelivery/Datasets/Staged/LndCvr/FileGDB101/

   Fees: None.

DataONE

What Makes a Good Metadata Record?

Provide all of the critical information for: Spatial reference

Steps to Create Quality Metadata

1. **Organize** your information
   - Did you write a project abstract to obtain funding for your proposal? Re-use it in your metadata!
   - Did you use a lab notebook or other notes during the data development process that define measurements and other parameters?
   - Do you have the contact information for colleagues you worked with?
   - What about citations for other data sources you used in your project?
2. Write your metadata using a **metadata tool**
3. **Review** for accuracy and completeness
4. Have **someone else** read your record
5. **Revise** the record, based on comments from your reviewer
6. **Review** once more before you publish

DataONE

Tips for Writing Quality Metadata

Titles, Titles, Titles…

- **Titles are critical** in helping readers find your data

    - While individuals are searching for the most appropriate datasets, they are most likely going to use the title as the first criteria to determine if a dataset meets their needs.
    - *Treat the title as the opportunity to sell your dataset.*

- A complete title includes: *What, Where, When, Who,* and *Scale*

- An **informative title** includes: topic, timeliness of the data, specific information about place and geography

DataONE

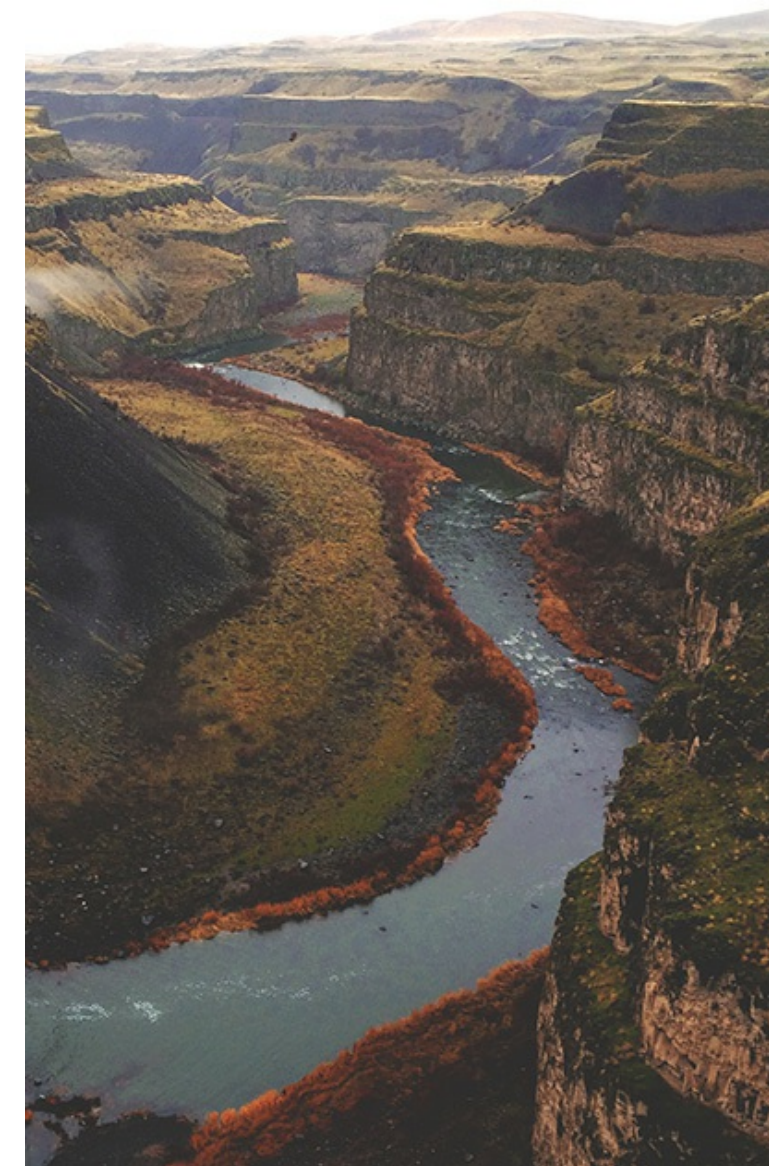# Tips for Writing Quality Metadata

A clear choice: Which title is better?

Rivers

OR

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

Greater Yellowstone **(where)** Rivers **(what)** from 1:126,700 **(scale)** U.S. Forest Service **(who)** Visitor Maps (1961-1983) **(when)**

Public domain image from Unsplash

DataONE

Tips for Writing Quality Metadata

Be specific and quantify when you can! The goal of a metadata record is to give the user enough information to know if they can use the data without contacting the dataset owner.

Vague: *We checked our work and it looks complete.*

Specific: *We checked our work using a random sample of 5 monitoring sites reviewed by 2 different people. We determined our work to be 95% complete based on these visual inspections.*

Public domain image from Unsplash

Tips for Writing Quality Metadata

- Use **descriptive & clear** writing
- Fully document **geographic** locations
- Select **keywords** wisely
- Use **thesauri** for keywords whenever possible
- **Be detailed**: there's no such thing as too much metadata!

Public domain image from Unsplash

DataONE

Tips for Writing Quality Metadata

- **Remember**: a computer will read your metadata.
- Do not use **symbols** that could be misinterpreted by software.
  - ! @ # % { } | / \ < > ~
- Don't use **tabs**, **indents**, or **line feeds/carriage returns**
- When copying and pasting from other sources, use a **text editor** (e.g., Notepad) to eliminate hidden characters

Public domain image from Unsplash

DataONE

Summary

Metadata is documentation of data

- A metadata record **captures critical information** about the content of a dataset
- Metadata allows data to be **discovered, accessed, and re-used**
- A metadata **standard** provides structure and consistency to data documentation
- **Standards and tools vary** – select according to defined criteria such as data type, organizational guidance, and available resources
- Metadata is of **critical** importance to data developers, data users, and organizations Metadata completes a dataset.

DataONE

Additional Resources

Federal Policies

- Executive Order 12906 (PDF)
- M-13-13 Open Data Policy (PDF)

Data Catalogs

- DataONE
- USGS Science Data Catalog
- Data.gov

More about CSDGM & ISO 19115

- FGDC Geospatial Metadata Website

Metadata Tools

- Metadata Wizard
- TKME
- CatMDEdit
- GRIIDC Metadata Editor
- ArcGIS 10.2

Standard Vocabularies

- USGS Thesaurus
- Global Change Master Directory
- Geographic Names Information System
- Getty Thesaurus of Geographic Names

DataONE

About

Participate in our GitHub repo: https://dataoneorg.github.io/dataone_lessons/

The full slide deck (in PowerPoint) may be downloaded from:
http://www.dataone.org/education-modules

Suggested citation: DataONE Education Module: Metadata. DataONE. Retrieved March 12, 2017. From
http://www.dataone.org/sites/all/documents/L07_Metadata.pptx

DataONE