USGS Data Management Training Modules:

Module 1: Metadata for Research Data

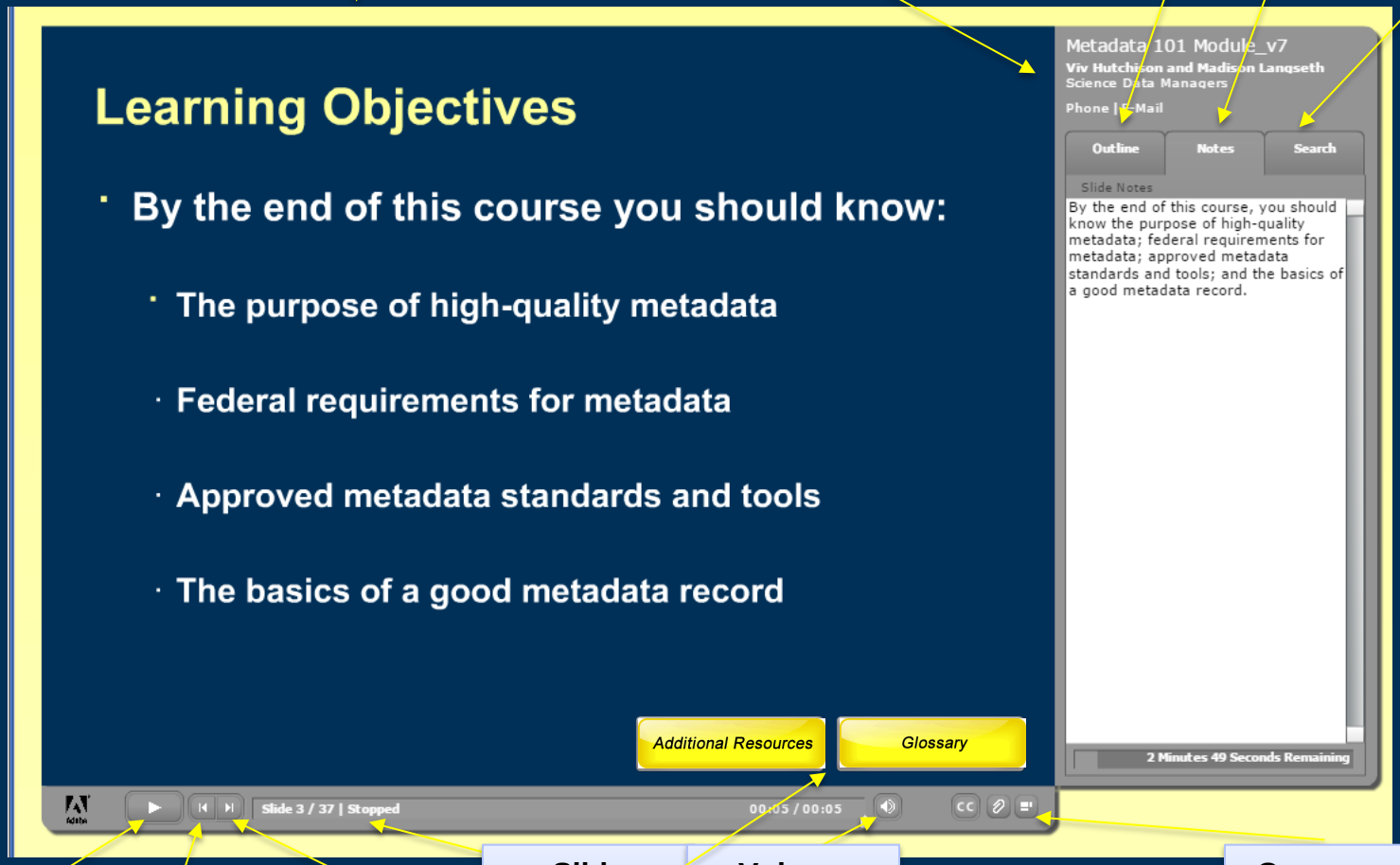
Slides and slide notes are available for download from the Data Management website.

U.S. Department of the Interior
U.S. Geological Survey

Version 1, April 2016

Welcome to the Data Management Training Modules. This module will cover metadata for research data.

Course Navigation 101



Presenter Window

Instructor's Contact Info

Notes Tab

Outline Tab

Search Tab

Navigation Side Bar

Play

Backward

Forward

Slide Number

Volume Control

Compress Nav. Bar

Additional Resources and Glossary buttons

USGS

Course Navigation 101

Each lesson starts with a title screen followed by lecture slides. The screen consists of a presentation window, navigation side-bar, and navigation base bar. The presentation window is used to view the slides that summarize the lecture material. The **gray navigation sidebar** is to the right of the presentation window. **At the top** you will find contact information for the course coordinators. The rest of **the sidebar is divided into three panes:**

- The **Outline** pane lists the slides in the lesson. You can use this pane to go to any slide within the lesson.
- The **Notes** pane shows the **lecture material** for each slide. **The principal content of each lesson is contained in the lecture material and should not be skipped.** However, we recommend reading the lecture material from a downloaded file because the scientific formatting in the downloaded version may be more accurate than in the version shown in the notes pane.
- The **Search** pane allows you to search for any word in the lesson and will show all slides containing that word.

- The **gray navigation** bar below the presentation window contains:

- "Forward" and "Backward" buttons to advance and return to previously viewed slides.
- The "Fast Forward" button.
- The status bar displaying the current slide number.
- Controls for sound and volume.
- The **Paper Clip icon**, used to access attachments. **IMPORTANT:** PDF versions of each lesson are attached here to download for ease of viewing and use in your student workbook. Some lessons contain additional attachments.
- A toggle button (lower right) changes the view of the Navigation Side Bar to full screen or to a compressed icon.

GRAY BUTTONS: Additional Resources and Glossary **Additional Resources** – Provides additional references, suggested training, or other information.
Glossary – Click here to find the definitions of terms used in the course.

Learning Objectives

- By the end of this course you should know:
  - The purpose of high-quality metadata.
  - Federal requirements for metadata.
  - Approved metadata standards and tools.
  - The basics of a good metadata record.

USGS

Additional Resources

Glossary

Learning Objectives

By the end of this course, you should know the purpose of high-quality metadata; federal requirements for metadata; approved metadata standards and tools; and the basics of a good metadata record.

Suggested Citation

- Zolly, L., Henkel, H.S., Hutchison, V.B., Langseth, M.L., Thibodeaux, C.J., 2015, USGS Data management training modules—metadata for research data: U.S. Geological Survey, **http://dx.doi.org/10.5066/F7RJ4GGJ**.

Suggested Citation

Zolly, L., Henkel, H.S., Hutchison, V.B., Langseth, M.L., Thibodeaux, C.J., 2015, USGS Data management training modules—metadata for research data: U.S. Geological Survey, **http://dx.doi.org/10.5066/F7RJ4GGJ**.

Acknowledgments

- These modules were developed with funding from the USGS Office of Organizational and Employee Development's (OED) Distance Learning Program.

USGS

Additional Resources

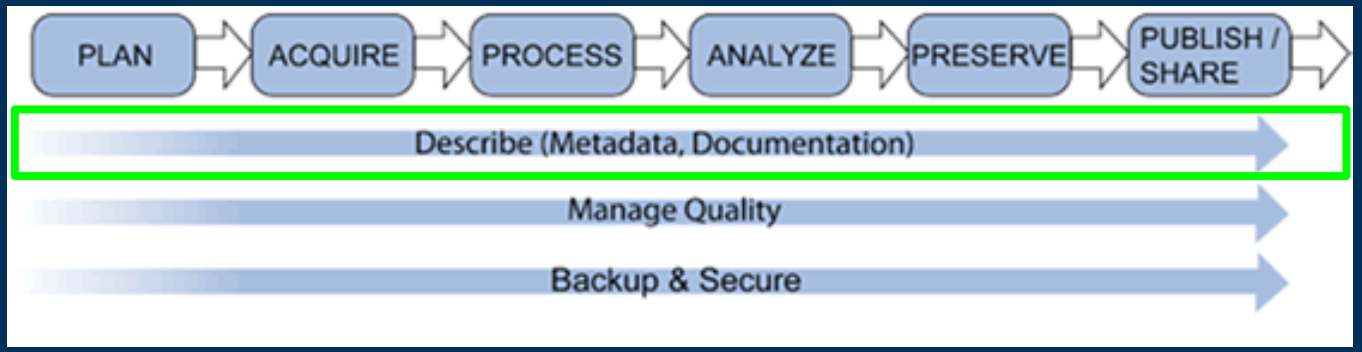
Glossary

Acknowledgments

These modules were developed with funding from the USGS Office of Organizational and Employee Development's (OED) Distance Learning Program.

Data Lifecycle

- **Reference point – This module will cover aspects of the "Describe" stage, which occurs throughout the data lifecycle.**



USGS

Additional Resources

Glossary

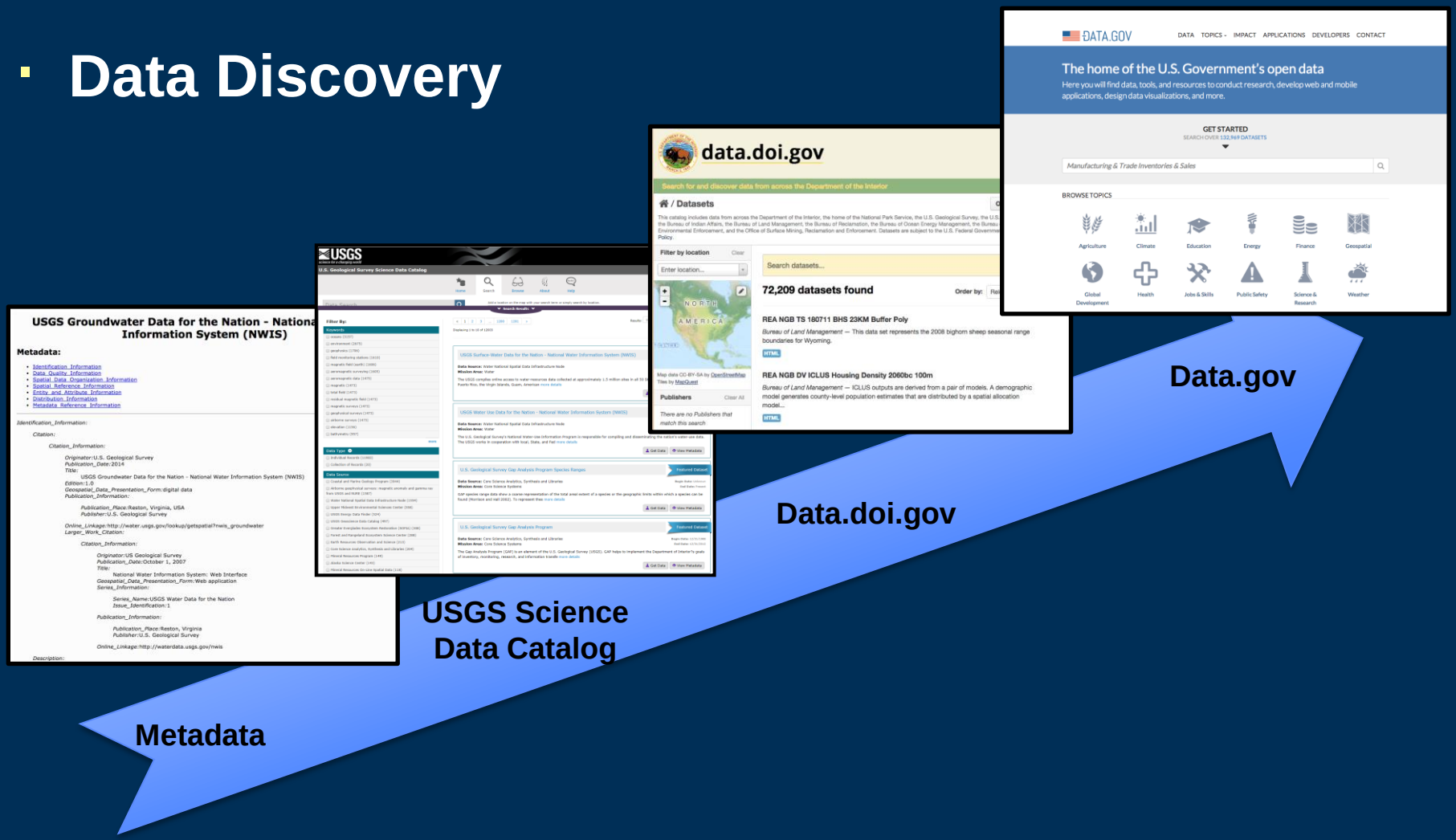
Data Lifecycle

The Data Lifecycle Model shown here was developed by a USGS team to illustrate how data management activities relate to data workflows and to assist with understanding the expectations of proper data management. This module will cover aspects of the multi-stage activity “Describe,” which occurs throughout the data lifecycle. It is important to begin thinking about the metadata tools and standards that you will use as you develop your data management plan and then, utilize them to document and describe the activities and decisions made in each of the other stages.

NOTE:
The USGS Science Data Lifecycle model shows the stages of data management within a project (plan, acquire, process, analyze, preserve, and publish/share) and cross-cutting data management activities that take place throughout the project (describe, manage quality, backup and secure).

Metadata: What are they Good for?

- Data Discovery

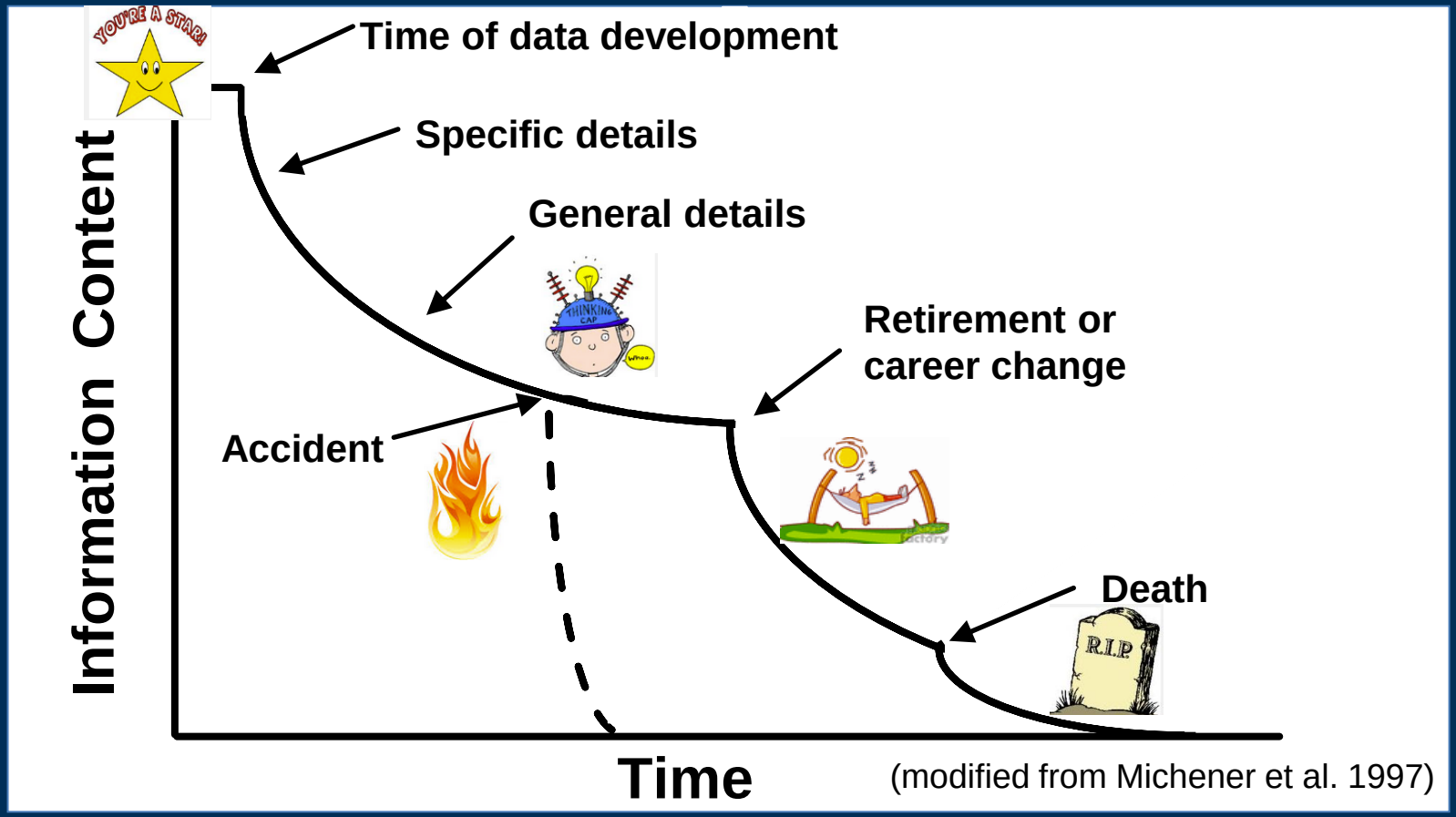


Metadata: What are they Good for?

Metadata are essential for data discovery. Metadata are what underlie the USGS Science Data Catalog and subsequently get rolled up into Data.doi.gov and Data.gov. Elements such as Title, Description, and Keywords enable users to discover data based on topical searches. Spatial and temporal elements allow users to discover data based on the geographical location or the time period of the data.

Metadata: What are they Good for?

- **Scientific Understanding and Reuse**



USGS

Additional Resources

Glossary

Metadata: What are they Good for? (Cont.)

Metadata are also very important for understanding and reusing scientific data. At the time of data development, scientists know the most about their data set and the steps that were taken to create it. Over time, memory of the details begins to fade. Circumstances in life can intervene, and eventually the knowledge about the dataset is gone. Without a metadata record, information about the data set could be lost forever, therefore, making the data unusable.

Metadata: What are they Good for?

- **Defending policy decisions based on data:**
  - Regulatory decisions based on undocumented data are not defensible.
  - Metadata accuracy and details are important.

USGS

Additional Resources

Glossary

Metadata: What are they Good for? (Cont.)

DOI regulatory agencies depend on metadata to defend the science behind regulatory decisions. If regulatory decisions are based on data that have no associated metadata or poor quality metadata, those decisions will not hold up in court. It is important to have accurate and detailed metadata records for supporting claims and policies. For example, the date when data are collected should be as specific as possible. In many situations, simply including the year that data were collected may not be sufficient.

Metadata: What are they Good for?

- **Controversies arise when metadata are incomplete/absent**
  - Bureau of Land Management (BLM) Example

USGS

Additional Resources

Glossary

Metadata: What are they good for? (Cont.)

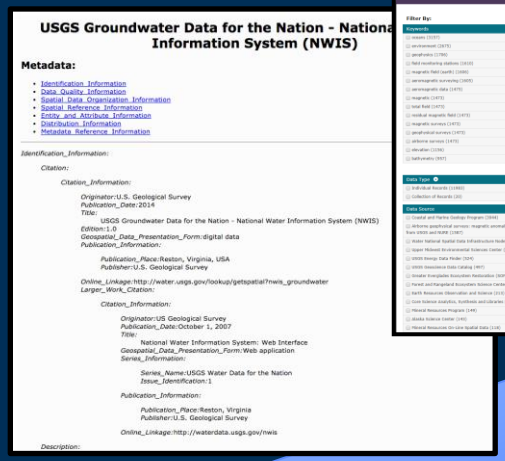
Controversies can arise when metadata are incomplete or absent. The USGS is the research arm of the DOI. We are tasked with providing sound and unbiased data, and data interpretation, for resource management decisions made by our sister agencies in the Department. Those decisions can be controversial and often, they are challenged in court. The data used to make those management decisions must be sufficiently documented so as to be transparent and reproducible to a third party attempting to connect the dots and reach a similar conclusion. If our data lack sufficient documentation, they cannot be used towards decision making.

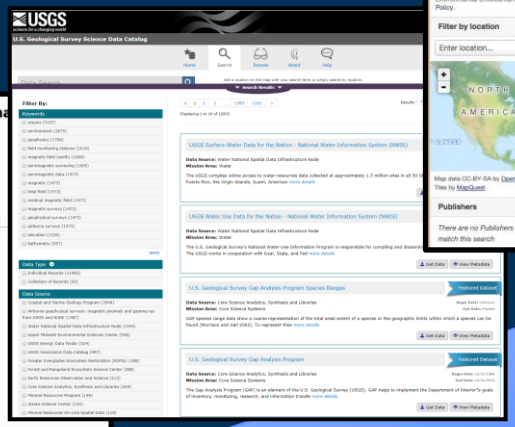
It is difficult to find real-world examples of the costs of poor metadata. Agencies don't like to admit these shortcomings; however, Tom Chatfield, a data management expert with the Bureau of Land Management bravely shared one real example from his own agency. A particular BLM field office developed an overall resource management plan for its region which included a map of oil and gas potential for its area. Naturally, this is an issue that registers strong opinions both for and against oil and gas recovery on public lands. When reviewed in the advent of litigation, BLM discovered that no Data Quality information existed for the map. No data sources could be identified, nor was there any discussion of the analysis methodology by which the oil and gas potential information was developed for the map. No geologic or geographic data were identified.

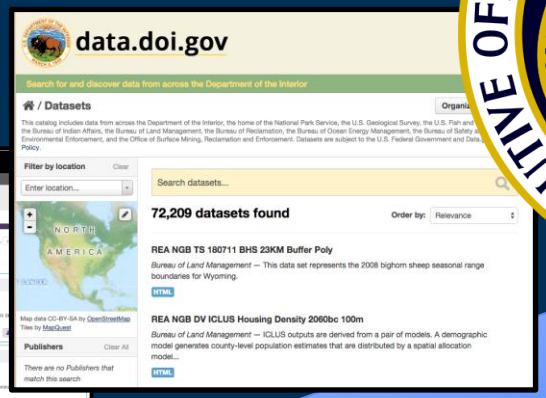
This was potentially disastrous in terms of defending the resource management plan. BLM was forced to develop a brand new map and closely document the provenance of the source data and methodology used to determine the potential ratings on the map. The cost exceeded several thousands of dollars in staff time and production costs, not to mention the intangible costs of loss of trust in the agency by partners and the public.

Metadata: What are they Good for?

- Reporting







Office of Management & Budget

Data.doi.gov

USGS Science Data Catalog

Metadata

[USGS OSQI Instructional Memo / Survey Manual Chapter on Metadata]

USGS

Additional Resources

Glossary

Metadata: What are they Good for? (Cont.)

The method by which the Department of the Interior reports its compliance with the Open Data Policy is through the metadata in its data catalog. The DOI receives metadata from its various bureaus, such as the USGS. The USGS submits its metadata through the Science Data Catalog. The number of complete and compliant metadata records with working links to the actual data within the DOI's data catalog is calculated and reported to the Office of Management and Budget. Please see the OSQI Survey Manual Chapter on Metadata for more information on what constitutes a complete and compliant metadata record for the USGS.

Federal Metadata Requirements

- **Executive Order 12906 (1994)**
  - All agencies should generate FGDC-compliant metadata for all geospatial data that they collect or produce.
- **Department of the Interior (DOI) Metadata Policy (2015)**
  - All DOI bureaus shall create metadata for all data sets for which the DOI is the data owner. Data include: text, numbers, graphics, images, sound, or video.

By Bob McNeely, The White House [1] [Public domain], via Wikimedia Commons

USGS

Additional Resources

Glossary

Federal Metadata Requirements

In addition to the Open Data Policy, there are a number of other federal mandates that require scientists to generate metadata. Some have been around for a number of years, while others are just now coming down the pipeline. In 1994, President Bill Clinton signed an executive order requiring all agencies to generate metadata that comply with standards approved by the Federal Geographic Data Committee for all geospatial data that they collect or produce. The Department of the Interior is requiring that all bureaus create metadata for all data sets that they own including text, numbers, graphics, images, sound, and video.

USGS Metadata Requirements

- **USGS Survey Manual Chapter (2015)**


USGS
science for a changing world
USGS Home
Contact USGS
Search USGS

U.S. Geological Survey Manual

U.S. Geological Survey Instructional Memorandum

No: IM OSQI 2015-02

Issuance Date: February 19, 2015

Expiration Date: Retain Until Suspended

Subject: Metadata for Scientific Data, Software, and Other Information Products

1. Purpose. This Instructional Memorandum (IM) provides interim policy on metadata requirements for U.S. Geological Survey (USGS) scientific data, software, and other information products. It also provides guidance for complying with appropriate USGS and other Federal standards, such as the metadata standards endorsed by the Federal Geographic Data Committee (FGDC), the interagency committee that provides metadata guidance for all Federal Government scientific data (geospatial and non-geospatial). This interim policy is issued to allow the time needed for USGS science activities to fully implement the metadata requirements herein and will be retained until superseded by a permanent Bureau Survey Manual (SM) policy chapter.

2. Background. Clear metadata descriptions answer fundamental questions (who, what, when, where, why, and how) and provide critical information intended to promote scientific collaboration, facilitate discovery, enable effective use, promote accurate interpretation of data, document its nature and quality, and augment the inherent value of USGS data and other products. Metadata also help users make effective use of scientific USGS data and information and aid the USGS in tracking its data and publications. Metadata is one of the required lifecycle elements described in IM-OSQI-2015-01 - Scientific Data Management Foundation.

3. Policy.

A. Metadata must accompany all USGS scientific data, software, and other information products described in this policy that are approved for release. The content and format of metadata depend upon the type of data, software, or information product being described. When data, software, and other information products are digitally released, metadata must include an appropriate persistent identifier for the product, such as a digital object identifier or DOI, unless access to the product is restricted. Metadata records are to be developed in a standardized way that enables users to evaluate the data, software, or information product's fitness for use in a research context. Metadata records must be updated to reflect changes and to ensure that links are functioning and continue to point to the intended files. Machine-readable metadata must be generated to enhance the transfer of information between systems and applications. Refer to http://www.usgs.gov/datamanagement/describe.php for additional guidance related to metadata.

B. A minimum of one metadata review by a qualified reviewer is required for all USGS scientific data, software, and other information products approved for release. A metadata review includes both checking for compliance with metadata standards using a recommended metadata validation tool when available, and performing quality checks. The metadata review can be conducted as part of the peer review, data review, editorial review, or separately as appropriate. A report of all metadata reviews (reviewer comments and how they were reconciled) must be included in the package in the internal USGS Information Product Data System (IPDS) that is submitted for Bureau approval.

4. Creating Metadata: General Guidelines.

A. Data. Metadata must be created for all scientific data, prior to approval and release as a USGS information product (refer to IM-OSQI-2015-03 - Review and Approval of Scientific Data for Release). This includes geospatial and non-geospatial datasets, databases, and Web data services that are created, collected, or compiled by USGS employees, volunteers, contractors, or data from other sources that are subsequently made part of a USGS dataset, database, or service. In addition to the records that describe data, metadata records are required to describe databases and Web data services. The metadata

Metadata must…

- **accompany** all USGS scientific data, software, and other information products prior to approval & release.
- **comply** with an FGDC-approved standard.
- go through **review** for quality and completeness.
- be **deposited** in and **shared** through appropriate catalogs.
- **reflect changes** to the data and **contain working links.**

USGS

Additional Resources

Glossary

USGS Metadata Requirements

The USGS has specified that metadata must accompany all USGS scientific data, software, and other information products prior to approval and release. These metadata records must comply with one of the FGDC-approved standards. Data and metadata must be reviewed for quality and completeness prior to release. A best practice is to have the metadata and data reviewed together by the same person.

Once approved, these metadata must be deposited in and shared through applicable bureau catalogs and larger systems, such as the DOI Catalog, which reports data assets to the Office of Management and Budget and data.gov. Finally, metadata records must be updated to reflect changes and to ensure that links are functioning and continue to point to the intended files.

Approved Metadata Standards

- Federal Geographic Data Committee's (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM)

- ISO 19115 Suite (as of 2015):
  - ISO 19115-2:2009 Geospatial Metadata
  - ISO 19110:2005 Geospatial Feature Cataloging
  - ISO 19119:2005 Geospatial Web Services
  - ISO 19139:2007 Geographic Information – Metadata – XML Schema Implementation

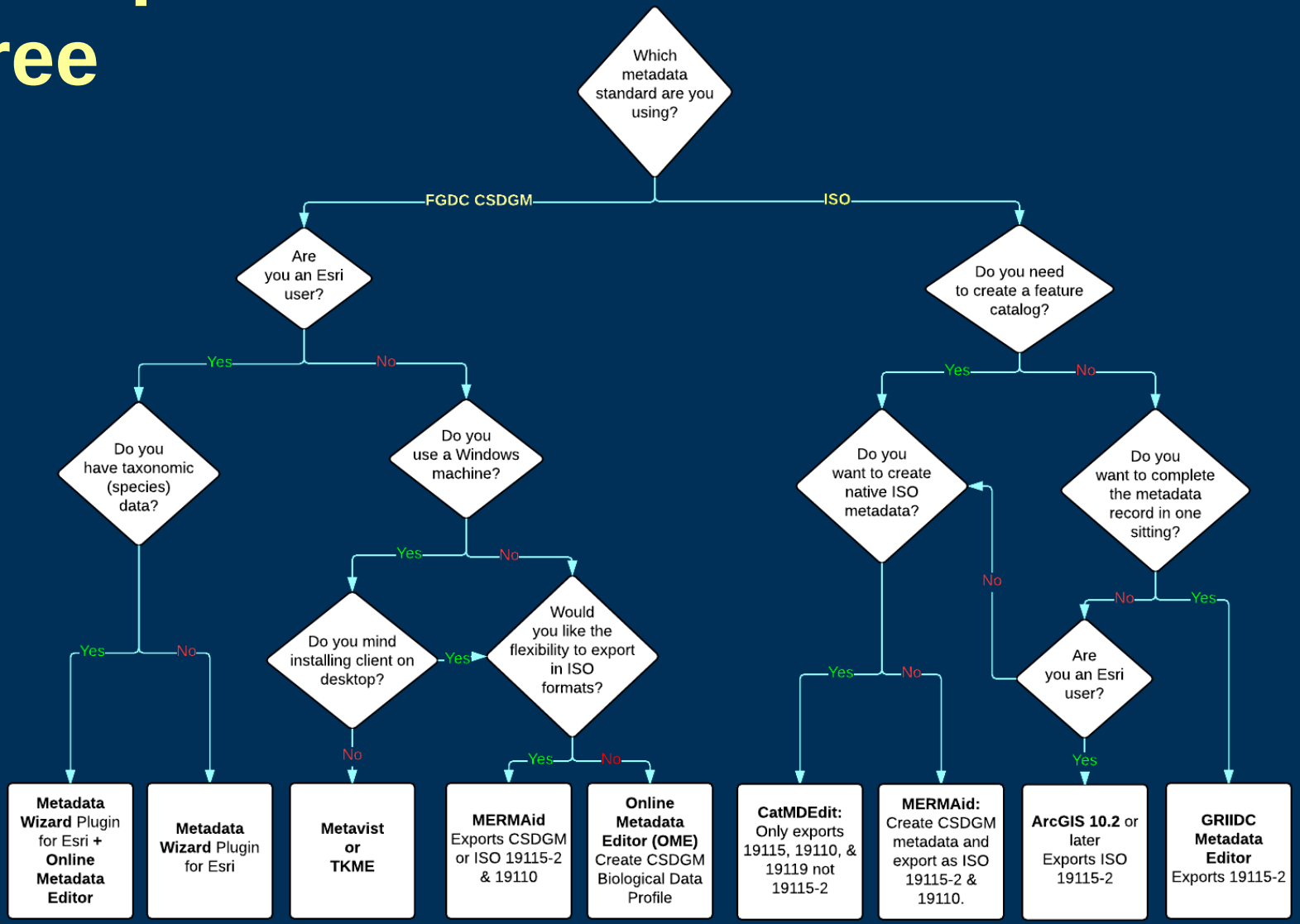
Approved Metadata Standards

Most of the data produced by the USGS are geospatial in nature and, therefore, must have metadata that are compliant with the Federal Geographic Data Committee's (FGDC) approved standards. Geospatial data include not only data that can be imported into GIS applications, but also data such as field photographs and samples for which the location is important. The FGDC adopted the Content Standard for Digital Geospatial Metadata in 1994 and revised it in 1998. In 2010, the FGDC endorsed the ISO 19115 suite of standards. This endorsed suite, as of 2015, includes ISO 19115-2 an extension of 19115 for imagery and gridded data; ISO 19110 the standard for geospatial feature cataloging; ISO 19119 the content standard for geospatial web services; and ISO 19139 the XML implementation schema. Federal agencies are encouraged to transition to ISO 19115; however, this transition will take time and the CSDGM will likely continue to be used for years to come. The USGS does not support one standard over the other, but requires that one of these standards be used to describe geospatial data.

Metadata standards ensure that all the essential metadata fields for discovery, access, and use are included. Standards also enforce organization and structure, which allow people to easily find the information for which they are looking. Additionally, metadata standards enable interoperability and machine readability of the metadata.

Please note that examples provided in this module are from CSDGM records because the USGS is still overwhelmingly using this standard. An eventual transition to ISO is envisioned for USGS metadata.

Geospatial Metadata Tools Decision Tree



USGS

Additional Resources

Glossary

Geospatial Metadata Tools Decision Tree

There are a number of tools that are available for generating FGDC CSDGM and ISO Metadata records. This decision tree includes some examples of commonly used tools. There are a number of elements to consider when choosing a metadata tool. For example, if you are an Esri user, it will be easiest for you to generate your metadata within Esri either through the built-in metadata editor or a plug-in. To build really robust metadata records, you may need to use more than one tool. For example, if you need to include additional information in your metadata record, such as species taxonomic data or other Biological Data Profile elements, you can export your metadata record from Esri and import into a tool like the Online Metadata Editor. Use this decision tree to determine an appropriate metadata tool, or tools, for you. For links to these tools and information on which profiles and extensions they may support, see the Additional Resources page at the end of this module.

Geospatial Metadata Tools Decision Tree

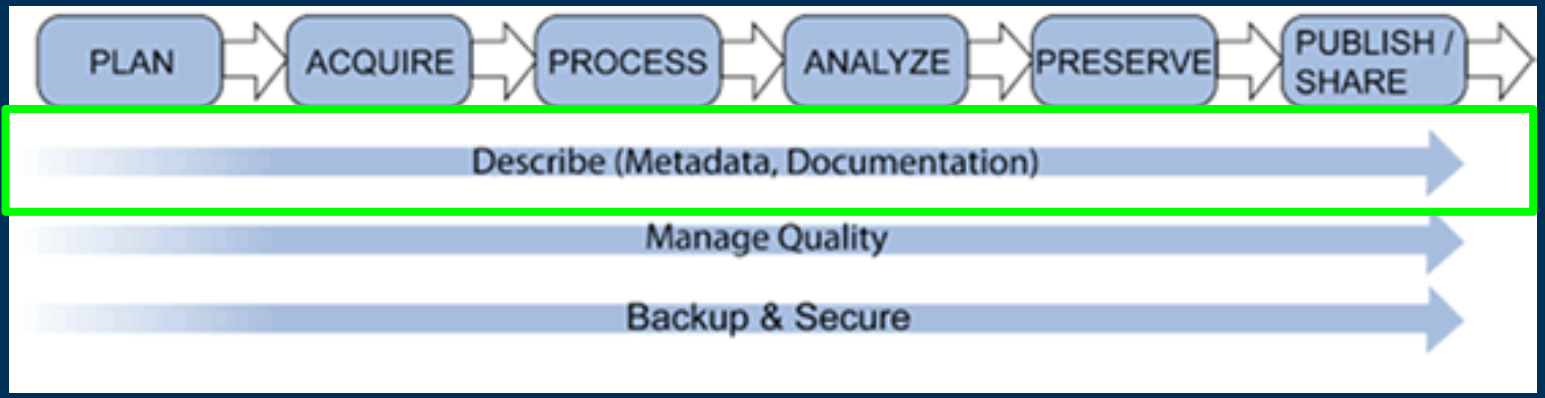
Which metadata standard are you using?

I. FGDC CSDGM
   a. Are you an Esri user?
      i. Yes
         1. Do you have taxonomic (species) data?
            a. Yes
               i. Metadata Wizard Plugin for Esri or Online Metadata Editor
            b. No
               i. Metadata Wizard Plugin for Esri
      ii. No
         1. Do you use a Windows machine?
            a. Yes
               i. Do you mind installing client on desktop?
                  1. Yes
                     a. Go to question 1.a.ii.b.i
                  2. No
                     a. Metavist or TKME
            b. No
               i. Would you like the flexibility to export in ISO formats?
                  1. Yes
                     a. MERMAid Exports CSDGM or ISO 19115-2 & 19110
                  2. No
                     a. Online Metadata Editor (OME) – Create CSDGM Biological Data Profile

II. ISO
   a. Do you need to create a feature catalog?
      i. Yes
         1. Do you want to create native ISO metadata?
            a. Yes
               i. CatMDEdit: Only exports 19115, 19110, & 19119 not 19115-2
            b. No
               i. MERMAid: Create CSDGM metadata and export as ISO 19115-2 & 19110
      ii. No
         1. Do you want to complete the metadata record in one sitting?
            a. Yes
               i. GRIIDC Metadata Editor: Exports 19115-2
            b. No
               i. Are you an Esri user?
                  1. Yes
                     a. ArcGIS 10.2 or later Exports ISO 19115-2
                  2. No
                     a. Go to question II.a.i.1

What Makes a Good Metadata Record?

- **Metadata are developed continuously throughout the entire data lifecycle.**



USGS

Additional Resources

Glossary

What Makes a Good Metadata Record?

Now that you know the appropriate metadata standards to use and some available tools for developing metadata, let’s discuss some aspects of a good metadata record. As mentioned at the beginning of this module, metadata should be developed continuously throughout the entire data lifecycle. This means that metadata should be generated along-side the data. Starting early will ensure that information about your data is fresh in your mind, which in turn saves you time. It will also enable you to provide more details about your data, which in turn leads to a higher-quality metadata record. For example, writing the Entity & Attributes section of your metadata or creating a separate data dictionary while you are developing your database will reduce this burden at the end of the project and can serve as a helpful tool during the project to ensure that everyone understands what each of the entities and attributes means. Writing the processing section of your metadata when you are processing your data ensures that all processing steps are documented. It is most beneficial if the person who is completing the data processing is the same person who writes the processing section for the metadata record.

What Makes a Good Metadata Record?

- **Consistency with commonly used fields**

Examples for a CSDGM record:

Correct ✔	Incorrect ✗
Publisher:	Publisher:
<publish>**U.S. Geological Survey**</publish>	<publish>**USGS**</publish>
Originator:	Originator:
<origin>**John Doe**</origin> <origin>**Jane Doe**</origin>	<origin>**XXXXXXX Science Center**</origin> <origin>**John Doe, Jane Doe**</origin> <origin>**Doe, John**</origin>
Date:	Date:
<pubdate>**YYYYMMDD**</pubdate> <pubdate>**YYYY**</pubdate>	<pubdate>**MM/DD/YYYY**</pubdate> <pubdate>**May 27, 2003**</pubdate>

USGS

Additional Resources

Glossary

What Makes a Good Metadata Record? (Cont.)

As you get started developing your metadata record, it is important that certain fields are used consistently. This means that some metadata fields require values that are written the same way to aid in machine-readability and discovery. For example, the publisher field should contain the standard form of your agency's name. For the USGS, the name should be written as U.S. Geological Survey, not USGS.

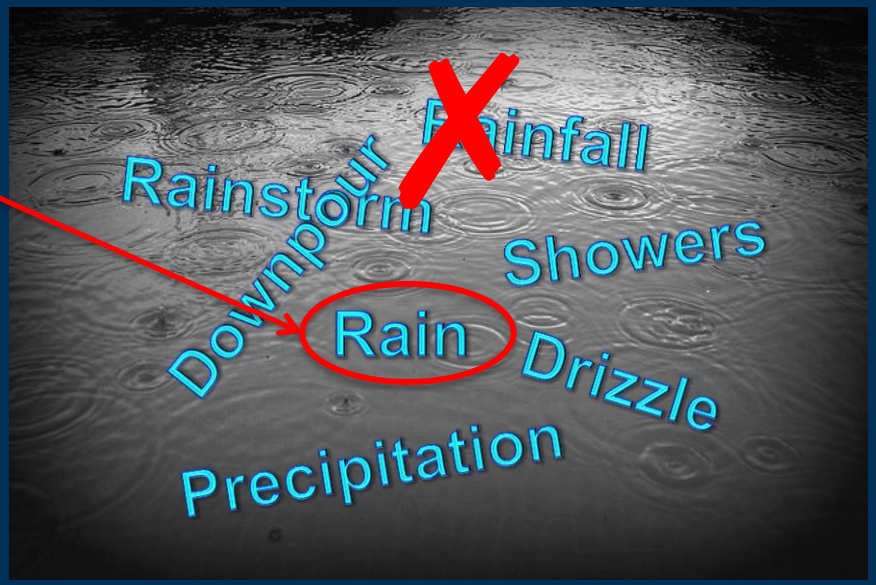
The originator field should contain the name of the researcher using the format first name then last name. If there are multiple authors, the field should be repeated for each author instead of comma separating the authors' names. This field should not be used for the Science Center name.

For CSDGM records, all dates should be written using the format of four digits for year, then two digits for month, followed by two digits for the day (YYYYMMDD). Data.gov is enforcing this format and metadata will fail validation if other formats are used.

What Makes a Good Metadata Record?

- **Use Authority Files and Standard Vocabulary**

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



USGS

Additional Resources

Glossary

What Makes a Good Metadata Record? (Cont.)

When possible, theme keywords should be drawn from thesauri or keyword lists, such as the USGS Thesaurus or the USGS Biocomplexity Thesaurus. The use of authority files and standard vocabulary will help to make your metadata more discoverable, understandable, and interoperable. Make sure you reference the thesaurus used in your metadata record. Here are a few resources to get you started.

What Makes a Good Metadata Record?

- **Consistency with commonly used fields**

Examples for a CSDGM record:

Correct ✔
Theme Keywords:
<themekt>**USGS Thesaurus**</themekt> <themekey>**precipitation (atmospheric)**</themekey>
Place Keywords:
<placekt>**Geographic Names Information System**</placekt> <placekey>**Roosevelt National Forest**</placekey>

Incorrect ✗
Theme Keywords:
<themekey>**rainwater**<themekey>
Place Keywords:
<themekey>**Roosevelt Forest**</themekey>

USGS

Additional Resources

Glossary

What Makes a Good Metadata Record? (Cont.)

Select the keywords that are specific to your dataset. Do not worry about representing broader concepts in your metadata record. The thesaurus used should be documented in the metadata. If a single thesaurus does not provide all of the theme keywords that your metadata require, you may use and reference more than one thesaurus.

Within FGDC CSDGM and ISO 19115 there are specific fields for place keywords. The theme keyword field should not be used for place keywords. A good resource for finding standard place keywords is the Geographic Names Information System.

If you need to include species names in your metadata, use the taxonomic section within the Biological Data Profile instead of the theme keyword fields, to do so.

What Makes a Good Metadata Record?

- **Use Appropriate Profiles and Extensions**
  - CSDGM
    - Biological Data Profile
    - Metadata Profile for Shoreline Data
    - Extensions for Remote Sensing Metadata
  - ISO
    - 19115 North American Profile
    - 19115-2 Extension for imagery and gridded data

What Makes a Good Metadata Record? (Cont.)

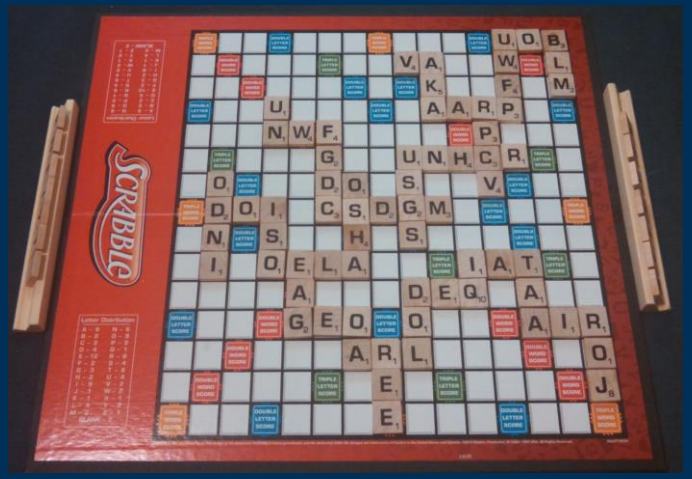
The CSDGM and ISO geospatial metadata content standards have profiles and extensions to allow users to customize their metadata records. Profiles are custom adaptations of the standard that may specify domain values for existing elements or increase conditionality of a specific element. Extensions are a set of added elements that extend the standard to better serve the community or data type.

CSDGM has two profiles, the Biological Data Profile, which broadens the standard to make it more applicable to data that are not explicitly geographic, such as laboratory results, field notes, and specimen collections and the Metadata Profile for Shoreline Data, which provides a standardized set of terms and data elements required to support metadata for shoreline and coastal data sets. CSDGM also has one set of extensions, the Extensions for Remote Sensing Metadata, which adds elements for describing remote sensing platforms and sensors. This extension is to be used for documenting data collected directly from a sensor, not derived data.

ISO 19115 has one profile and one extension. The North American Profile is intended for use with US and Canadian data. It extends some domains and increases conditionality for some elements as well as specifying best practices for populating most elements. The 19115-2 Extension for imagery and gridded data is an extension of 19115 that includes elements for raster data, imagery, and data collected using instruments. It also includes improved descriptions of lineage and processing information. 19115-2 is actually the preferred standard for organizations implementing ISO Metadata because its additional elements tend to be relevant to many geospatial data sets.

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 for file formats (e.g.,TIFF, JPEG, PDF)



USGS

Additional Resources

Glossary

What Makes a Good Metadata Record? (Cont.)

If your metadata include acronyms, you should spell out the acronyms with their first use. Many acronyms have multiple meanings and, therefore, need to be clarified. DOI is an example of an acronym with multiple meanings. It is used for the Department of the Interior & Digital Object Identifiers. The use of widely known acronyms for file formats is acceptable without first spelling them out. For example, TIFF, JPEG, & PDF are all widely known across disciplines as different file types and would be acceptable to use.

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

USGS

Additional Resources

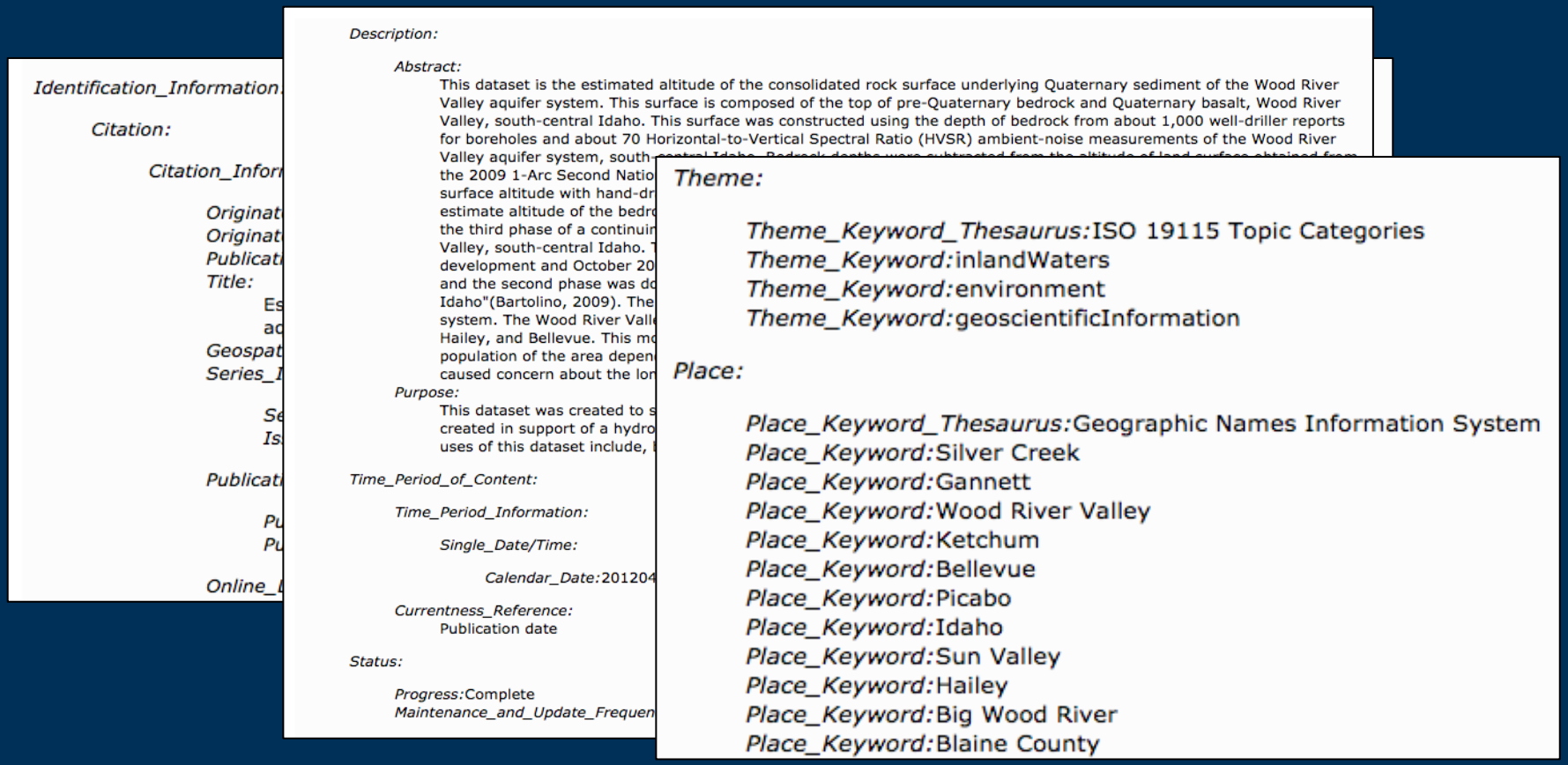
Glossary

What Makes a Good Metadata Record? (Cont.)

While not all metadata fields are mandatory, it is a best practice to provide all of the information that is applicable to your data. The more complete your metadata record is, the better chance your data have of being discovered and understood. This means that there will be less risk of your data being misinterpreted and misused. The critical information for CSDGM records includes: identification information, entities and attributes, data quality, access, use & liability constraints, distribution, and spatial references.

What Makes a Good Metadata Record?

- **Provide all of the Critical Information**
  - Identification Information



USGS

Additional Resources

Glossary

What Makes a Good Metadata Record? (Cont.)

First and foremost, it is essential that your metadata record has complete identification information so that your data can be discovered. Basic citation information the like the author’s name, the dataset title, and publication information allows your dataset to be referenced and attributed appropriately. A good, detailed abstract helps users get a general overview of your data without having to download the entire dataset. Theme keywords and Place keywords enable users to search for and find your data.

What Makes a Good Metadata Record?

- **Provide all of the Critical Information**
  - Entity / Attribute

Entity_and_Attribute_Information:

Detailed_Description:

Entity_Type:

*Entity_Type_Label:*Struct_Point
Entity_Type_Definition:
A feature class representing the location of a building or other structure as a single point.
Entity_Type_Definition_Source:
USGS NGTOC archives: DataDictionary_Structures_v4.xls, draft copy 2013

Attribute:

*Attribute_Label:*Shape
Attribute_Definition:
Feature geometry.
Attribute_Definition_Source:
Esri
Attribute_Domain_Values:

Unrepresentable_Domain:
Coordinates defining the features.

Attribute:

*Attribute_Label:*OBJECTID
Attribute_Definition:
Internal feature number.
Attribute_Definition_Source:
Esri
Attribute_Domain_Values:

Unrepresentable_Domain:
Sequential unique whole numbers that are automatically generated.

USGS

Additional Resources

Glossary

What Makes a Good Metadata Record? (Cont.)

Entity and attribute fields help to describe the often short and cryptic labels that are usually provided in datasets. If you have developed an external data dictionary for your data, that is great and you should link to it from your metadata. If an external data dictionary does not exist, you should describe each entity and attribute individually within your metadata record as is shown on the screen. A detailed attribute definition is essential to making your data usable. The definition should include the units, if applicable, and any special values, such as a no data value, which can be represented by a number such as -999.

What Makes a Good Metadata Record?

- **Provide all of the Critical Information**

- Data Quality

  - Accuracy,
  - Consistency,
  - Completeness

Data_Quality_Information:

Attribute_Accuracy:

Attribute_Accuracy_Report:
EDEN real-time water surfaces are created using real-time (four-day delay) water level data for the EDEN network. Most data relayed by satellite or other telemetry have received little or no review. Inaccuracies in the data may be present because of instrument malfunctions or physical changes at the measurement site. A threshold comparison program eliminates daily values that appear erroneous (i. e. extremely high or low, extremely different from previous days). Subsequent review of the data may result in significant revisions to the data. Users are cautioned to consider carefully the provisional nature of the information when using provisional data. Within approximately 45 days after the end of each quarter (December 31, March 31, June 30, and September 30), finalized and approved water level data are provided by SFWMD and ENP at which time real-time EDEN surfaces will be replaced by provisional surfaces. EDEN surfaces created with final, approved water level data from all agency gages will be available in approximately May of each year for the previous year's water year (October - September). For the provisional water-level surfaces for the period 1990-1999, users are cautioned about the quality of the water-level surfaces for several subareas. 1) WCA2A for the period 1/1/1990 â€“ 12/31/1999, most of the water-level data for gages in WCA2A were hindcasted. Resulting water-level surfaces show inconsistencies compared with surfaces post-2000 and will require further analysis. Water-level surfaces in WCA2A should be used with caution for this period until further analysis is completed and revised surfaces are generated. 2) WCA2B for the period 1/1/1990 â€“ 4/27/1993, no data is available for the northern boundary structures (S144_T, S145_T, S146_T), therefore the water-level surfaces in WCA3B are not considered valid in the northern portion of the subarea. 3) Pennsuco wetlands, The surface for this subarea is modeled only when data for 5 or more gages is available.

Logical_Consistency_Report:
Prior to 5/14/12, the Daily Median Output Files "median" files contain a list of the stations that were used to create water surfaces for that day. The "median reject" files contain a list of the gages that were not used during the creation of that day's water surface. Starting on 5/14/12, a single daily median file combines the information in the previous two daily median files and includes information about the data type at each gage. The Data Type field informs users whether the water-level data is measured, estimated, dry, or missing at each gage. Gages with missing data are not used for that day's surfacing.

Completeness_Report:
Data from all the gages were collected and evaluated for use in calculating the water surface for each day.

Positional_Accuracy:

Horizontal_Positional_Accuracy:

Horizontal_Positional_Accuracy_Report:
Horizontal positions are established by GPS observations and are referenced to the North American Datum of 1983 (NAD83). The desired horizontal accuracy is +/- 15 centimeters. This level of accuracy is consistent with GPS differential techniques which use two stations - a high-quality dual-frequency GPS receiver base station and a roving GPS station. The density and accuracy of a given GPS data observation varies from a few meters to a few centimeters according to the Position Dilution of Precision (PDOP) in the study area. Generally if the PDOP is observed to be excessive, data collection is discontinued or the data are discarded. The PDOP is an indicator of the positional accuracy of the GPS that be can derived from the current GPS satellite geometry, which varies continuously. Generally the smaller the PDOP number, the higher the data quality. The PDOP is a permanent part of the recorded data and is also included in the post processing procedures during reduction of the GPS observations to NAD 83. Where possible, the GPS base station has an ellipsoid height to an accuracy of two centimeters relative to the Continuously Operating Reference Stations (CORS) or the High Accuracy Reference Network (HARN), both operated by the National Geodetic Survey (NGS).

Vertical_Positional_Accuracy:

Vertical_Positional_Accuracy_Report:
Source data (HAED) collected to better than +/- 15cm. Standard errors of cross-validation for the DEM range from ~7cm to 17cm depending on the EDEN subarea. This version is composed of new models created for WCA3N and the BCNP combined with the mosaic of two versions (i.e., eden_v002a for WCA1 and eden_v002c for all other EDEN subregions) that comprised EDEN_EM_JAN07. The previous mosaic was selected based on analysis of model performance in the water depth estimation process. The WCA1 surface was produced by removing all "upland" AHF points as defined by a reclassification of the Florida GAP process. For WCA1, the krigging model was developed based on statistics from the entire EDEN domain (the process for eden_v002a production). For each other EDEN sub-area (i.e., the other WCAs and the National Park) region specific krigging models were developed and applied.

USGS

Additional Resources

Glossary

What Makes a Good Metadata Record? (Cont.)

There are a number of available metadata fields for describing the quality of your data, such as attribute accuracy, logical consistency, completeness, and horizontal and vertical positional accuracy. Provide as much detail in these sections as is possible to enable users to determine if the data are appropriate for their needs.

What Makes a Good Metadata Record?

- **Provide all of the Critical Information**
  - Data Quality Information - Lineage

Lineage:

Source Information:

Source Citation:

Citation Information:

Originator: State and Federal Partners
Publication Date: 20140324
Title: Fire Stations
Geospatial Data Presentation Form: Vector digital data
Other Citation Details: Any location where fire fighters are stationed or based out of, or where equipr
which are Mobile Units and not having a permanent location, are included, in which case their location h
engaged in forest or grasslands fire fighting, including fire lookout towers if the towers are in current us
Locations serving both administrative and operational functions are included.
Online Linkage: http://nationalmap.usgs.gov

Source Scale Denominator: 24000
Type of Source Media: None
Source Time Period of Content:

Time Period Information:

Range of Dates/Times:

Beginning Date: 20060101
Ending Date: 20140310

Source Currentness Reference: ground condition

Source Citation Abbreviation: Structures - Fire Stations
Source Contribution: Image

Source Information:

Source Citation:

Citation Information:

Originator: Office of the Deputy Under Secretary of Defense for Installations and Environment, Busines
Publication Date: 20140324
Title: U.S. Military Installations, Ranges, and Training Areas
Geospatial Data Presentation Form: vector digital data
Other Citation Details: Authoritative locations of Department of Defense sites, commonly referred to
only those in the fifty United States and US Territories were considered for inclusion. Some sites were n
Online Linkage: http://www.acq.osd.mil/ie/index.shtml

Lineage:

Process_Step:

Process_Description:
Surfacing Process for Version 1 of the surface-water model: The steps in preparing a daily surface from Everglades stage gaging stations are to collect the gaging station values as daily medians, create boundary conditions along canals/levees, generate the continuous surface using multiquadric radial basis function, and save the output into 400 meter output grids in netCDF and geoTIFF formats. For each day that the surface is run, values for water stage are obtained from the USGS NWIS server for each gage listed in the EDEN master station list maintained by USGS, St Petersburg, Florida. Median daily values are calculated at each gage and stored in UTM, zone 17, NAD 1983 projection and datum, meter horizontal units. Vertical units are centimeters, NAVD 88. Where gage data is obtained in NGVD 29, it is converted to NAVD 88 using difference values for each gage. The difference values are provided in the EDEN master list. Median stage values along specific canals are linearly interpolated every 200m between gages and extrapolated up to 3 kilometers beyond a gage when canals extend beyond the last gage. This interpolation/extrapolation will be used in the surfacing process to enforce boundaries at the canals between Everglades water conservation areas. The specific canals and the point locations interpolated/extrapolated in the canals are read from files maintained by USGS, St Petersburg, Florida. Median stage for the day at each gage and each interpolated/extrapolated canal point location are combined as input for surface interpolation using the radial basis function routine in ArcGIS geostatistical analyst. The parameters are: Kernel Functions: Multiquadric; Parameter = 16.77 Neighbors to Include = 1 Include at least = 1 shape type = 8 sectors angle = 350 major semiaxis = 31000 minor semiaxis = 30000 A boundary data layer is used to clip the surface interpolation to the EDEN area. The surface interpolation is saved to an ESRI GRID with 400m cell resolution aligned to the EDEN grid data layer. The prediction grid value at each cell is obtained from the interpolation surface by assigning to the grid cell the prediction value that corresponds to the center position of the grid cell. Both the boundary data layer and the EDEN GRID are maintained by USGS, St Petersburg, Florida. Interpolation surface GRID files for each day are converted to geoTIFF and CF compliant netCDF files for achieving. The netCDF files collect 3 months of daily surfaces together in each file.
*Process_Date:*2008
Process_Contact:

Contact_Information:

Contact_Person_Primary:

*Contact_Person:*Leonard Pearlstine
*Contact_Organization:*Everglades and Dry Tortugas Parks

Contact_Address:

*Address_Type:*mailing and physical address
Address: South Florida Natural Resources Center 950 Krome Avenue
*City:*Homestead
*State_or_Province:*FL
*Postal_Code:*33030
*Country:*USA

*Contact_Voice_Telephone:*305 224-4228
*Contact_Facsimile_Telephone:*305 224-4147
*Contact_Electronic_Mail_Address:*Leonard_Pearlstine@nps.gov

USGS

Additional Resources

Glossary

What Makes a Good Metadata Record? (Cont.)

Citations for all source data used in the dataset should be included in the metadata record. These citations will allow the user to understand your dataset better and be able to trace the data back to their origin. The processing steps that are performed on the data should also be documented in detail. Each step should be documented separately, preferably by the person performing that particular processing step. Contact information for that person should be included, so any questions from users can be directed to the most knowledgeable person.

What Makes a Good Metadata Record?

- **Provide all of the Critical Information**
  - 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.

Use_Constraints:

Users are free to use, copy, distribute, transmit, and adapt the work for commercial and non-commercial purposes, without restriction, as long as clear attribution of the source is provided.

Use_Constraints:

Users may use and redistribute these data without explicit written permission from USGS or University of [XYZ], with the exception of roads data associated with countries that are listed in the data documentation as requiring additional credits or holding special restrictions. Users are advised to consult the data documentation for further information and to obtain necessary permissions or adhere to relevant restrictions that apply to each of those data sets.

Distribution Liability: statement of the liability assumed by the distributor with respect to content and accuracy of the data.

Distribution_Liability:

Unless otherwise stated, all data, metadata and related materials are considered to satisfy the quality standards relative to the purpose for which the data were collected. Although these data and associated metadata have been reviewed for accuracy and completeness and approved for release by the U.S. Geological Survey (USGS), no warranty expressed or implied is made regarding the display or utility of the data on any other system or for general or scientific purposes, nor shall the act of distribution constitute any such warranty.

What Makes a Good Metadata Record? (Cont.)

Describe if there are any access or use constraints for your data. Access Constraints refer to any restriction and legal prerequisites for accessing the data. Use Constraints refer to any restrictions and legal prerequisites for using the data after access is granted. Both of these elements include any constraints applied to assure the protection of privacy or intellectual property. Additionally, ensure that you establish what liability you have, if any, with respect to the content and accuracy of the data.

What Makes a Good Metadata Record?

- **Provide all of the Critical Information**
  - Accessing the Data

Standard_Order_Process:

Digital_Form:

Digital_Transfer_Information:

*Format_Name:*WAF
Format_Information_Content:
Web Accessible Folder

Digital_Transfer_Option:

Online_Option:

Computer_Contact_Information:

Network_Address:

*Network_Resource_Name:*http://thor-f5.er.usgs.gov/ngtoc/metadata/waf/structures/nsd/

*Fees:*None.

Standard_Order_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/Struct/FileGDB101/

*Fees:*None.

USGS

Additional Resources

Glossary

What Makes a Good Metadata Record? (Cont.)

It is also important to inform users how and where they can access the data. Are they available for download or are there associated web services for the data? Also, be sure to include information about the format of the dataset, so users can understand what software applications they will need to utilize the data. In a CSDGM record there are a couple locations where you should include online links to your data. First, links can be included in the online linkage element (onlink) in the Citation Information subsection of Identification Information. Links can also be added to the Network Resource element (networkr) in the Distribution section, as shown on this slide. It is all right, and sometimes even encouraged, to have more than one format for your data. The online linkage and network resource elements can be repeated multiple times to accommodate links to these different data access points and formats.

What Makes a Good Metadata Record?

- **Provide all of the Critical Information**
  - Spatial Reference

```
Spatial_Reference_Information:
    Horizontal_Coordinate_System_Definition:
        Planar:
            Grid_Coordinate_System:
                Grid_Coordinate_System_Name:Universal Transverse Mercator
                Universal_Transverse_Mercator:
                    UTM_Zone_Number:17
                    Transverse_Mercator:
                        Scale_Factor_at_Central_Meridian:0.9996
                        Longitude_of_Central_Meridian:-81
                        Latitude_of_Projection_Origin:0
                        False_Easting:500000
                        False_Northing:0
            Planar_Coordinate_Information:
                Planar_Coordinate_Encoding_Method:Coordinate Pair
                Coordinate_Representation:
                    Abscissa_Resolution:400
                    Ordinate_Resolution:400
                Planar_Distance_Units:meters
        Geodetic_Model:
            Horizontal_Datum_Name:North American Datum of 1983
            Ellipsoid_Name:Geodetic Reference System 80
            Semi-major_Axis:6378137
            Denominator_of_Flattening_Ratio:298.257
    Vertical_Coordinate_System_Definition:
        Altitude_System_Definition:
            Altitude_Datum_Name:North American Vertical Datum of 1988
            Altitude_Resolution:0.0001
            Altitude_Distance_Units:centimeters
            Altitude_Encoding_Method:Explicit elevation coordinate included with horizontal coordinates
```

USGS

Additional Resources

Glossary

What Makes a Good Metadata Record? (Cont.)

Finally, spatial reference information is essential for integrating your data with other geospatial data. Spatial reference information includes information about both horizontal and vertical coordinate systems.

My Metadata is Complete, Now What?

- Review your metadata
- Package metadata with data
- Submit or expose metadata to relevant bureau and federal catalogs
- Ensure that persistent identifiers in metadata resolve to landing pages for your published data

USGS

Additional Resources

Glossary

My Metadata is Complete, Now What?

Once you have finished adding content to your metadata record, there are a few more steps that you will need to complete. First of all, be sure to review your metadata for quality and completeness. Additionally, ask a colleague who is familiar with your data to review it as well. Once you are certain that your metadata record is accurate and complete, package it with your data. Zipping the metadata with the data will ensure that when the data are shared, the metadata will be shared along with them.

You should also expose your metadata to relevant systems and catalogs. USGS personnel should expose their data by registering metadata assets in the USGS Science Data Catalog. Exposing your metadata records allows them to be discoverable and allows the USGS to comply with Open Data Policies.

Over time, it is important to ensure that online links in your metadata continue to resolve to the landing page for the data. If your data have been moved, you should ensure that any persistent identifiers are kept up-to-date.

Student Practice/Exercises

- To try an exercise in metadata, visit this web page https://www.dataone.org/education-modules

- Find Lesson 7: Metadata
- https://www.dataone.org/sites/all/documents/L07_L08_Exercise.pdf

- Supporting data are also included on the DataONE webpage for this exercise.

USGS

Additional Resources

Glossary

Student Practice/Exercises

Writing accurate and complete metadata is a skill that is improved with practice. Metadata practice exercises are available through DataONE at the links shown in the slide.

Summary/Recap

- Metadata are essential for discovering, understanding, and reusing scientific data.
- Metadata are mandated through an Executive Order and USGS policy.
- Metadata submitted to the USGS Science Data Catalog are used for required reporting to Office of Management and Budget.
- Use approved metadata standards and tools to create metadata.
- Aspects of a good metadata record include consistency, the use of standard keywords, and completeness of all critical information.

Summary/Recap

Here is a summary of what we covered in this module:

-Metadata are essential for discovering, understanding, and reusing scientific data.

-Metadata are mandated through an Executive Order and USGS policy.

-Metadata that are submitted to the USGS Science Data Catalog are used for required reporting to the Office of Management and Budget.

-You should use approved metadata standards and tools to create your metadata.

-And finally, the aspects of a good metadata record include consistency, the use of standard keywords, and completeness of all critical information.

References

- 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, v. 7, no. 1, p. 330-342.

References

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, v. *7*, no. 1, p. 330-342.