Meetings of the Metadata Reviewers Community

(Like sedimentary layers: the most recent meeting is on top, then reverse chronological to oldest meeting at the bottom. No folds or faults so far.)

December 6, 2021

Topics: Lessons learned for creating metadata templates, maintaining point of contact fields in the metadata.

November 1, 2021

Topics: How to update metadata when a related primary publication is found after the data is published. Accessibility, what it means for a data release.

Mon. October 4, 2021, 2 pm - 3 pm EDT

Topic: Metadata for data releases that include self-describing data files such as netCDF or hdf5 files. What should be included in FGDC metadata accompanying these files?

Mon. September 6, 2021: Labor Day, no meeting.

Mon. August 2, 2021, 2 pm - 3 pm EDT

Topic 3: Data and Metadata Review Checklists

- Reminder about the updated Data Review Checklist on the Data Management Website
- Marc Hunter noted that his center updated their metadata review checklist based off of the format for the updated data review checklist: https://doimspp.sharepoint.com/:w:/s/GS-AstroCenter/EQ9f26ERMfFCidMxO1pQktQBKeyZmuvY2gAps4voln29wg?e=BosHm8 and GGGSC said they have done something similar.
- Question: Should we update the metadata review checklist to be in the same format as the data review checklist and provide a content update to include core metadata requirements that are applicable across disciplines?
    - There was interest in doing this. Madison would be happy to help coordinate next FY like she did for the data review checklist.
- Question: Should we try to compile discipline-specific data and metadata review checklists in one location (or have pointers to their original location) so that they are more broadly available across the Bureau?
    - There was also interest in doing this. This may be a point for future discussion. Should this type of resource be on the Data Management Website (publicly accessible) or in a SharePoint space that is only available to DOI folks?

Topic 2: Science Publishing Network Data Reports

- What metadata is necessary for the new Data Report series?
- Where is the line between what can be included in a Data Release versus what needs to go through a Data Report?
- We agreed that it would be beneficial to have someone from SPN and a BAO speak about the new Data Report:
    - how it differs from other USGS series publications?
    - where the line is between additional data documentation in a data release and a data report?
    - how will SPN keep up with the demand for fast turn around times on data releases that also require a Data Report?
    - what is the current funding model for SPN? It seems to differ from science center to science center.

Topic 1: Department of the Interior Draft Records Schedule.

- Spent 15 minutes reading through the Draft Records Schedule
- Agreed it still seems very drafty.
- Individuals added their own comments to the document. As a group, we added a comment that whenever data are mentioned, they should also reference the accompanying metadata, which should follow the same records schedule.
- There was a discussion about what metadata should be included for large databases such as NWIS.

Mon. June 7, 2021, 2 pm - 3 pm EDT

2 Topics:

1. **"Disclaimers" in the metadata – where do they go?**

USGS has a list of approved "disclaimers" at Fundamental Science Practices (FSP) Guidance on Disclaimer Statements Allowed in USGS Science Information Products.

It would be useful if these were consistently found in the same place in our metadata records. Can we recommend where to put them?

2. **Advise the ScienceBase team on intervention to improve proper use of USGS Thesaurus keywords.**

The ScienceBase Data Release Team has the ability to check metadata records for proper use of USGS Thesaurus keywords using the MP secondary validation tools. We have been noticing that many records do not properly use USGS Thesaurus keywords. We are wondering if we should intervene in any way and if so, how. Here are some options:

1. Inform the author that they have not used any USGS Thesaurus keywords in their metadata record and that it is a recommended practice to do so. (We would ignore cases where they properly use some USGS Thesaurus keywords but add in extraneous non-USGS Thesaurus keywords to the USGS Thesaurus keyword section.)
2. If no USGS Thesaurus keywords are listed, automatically update the metadata record, when possible, to include USGS Thesaurus keywords based on the following criteria:
   a. MP suggested keywords could be added as a new section in the metadata. Essentially, the keywords would show up twice, once in a "None" section and once in the "USGS Thesaurus" section.
   b. Add science topic keywords that are selected when a user completes the ScienceBase Data Release form.
3. Inform the author and ask if they would like us to update their metadata, if possible, using the criteria in number 2.
4. Do nothing...this is not our responsibility.

Notes from discussion

Topic 2: We suggest that ScienceBase use option 3.

Tamar was present and explained that USGS Thesaurus keywords in metadata records will be used by the Science Data Catalog for categorizing data, and it's also just a good practice.

Ideally, metadata reviewers would discover these keyword failures and get them fixed before they got to ScienceBase.

Automatically correcting metadata could cause problems in situations when there are other copies of the metadata record. Changes to the record should start with the upstream copy that others propagate from. So ask whether it is okay before making the change in the ScienceBase copy. This also educates the author to the need for using USGS Thesaurus keywords correctly in future metadata.

In addition, it would be useful if Metadata Wizard encouraged the best practice of including USGS Thesaurus keywords. We understand that OME already does this.

In addition, it would be good to communicate the value of USGS Thesaurus keywords to the data managers in the groups ScienceBase works with, so there will be fewer metadata records that need changing in the future.

Topic 1:

Fran says this question came from Fundamental Science Practices, but she doesn't remember the reason for asking. After the discussion, she agreed to report out consensus and ask how it will be used.

We discussed the list of FSP disclaimers, which are provided below along with our preferred placement in metadata records.

We would put the following in the distribution - liability section. ....

1. Approved data released to the public:

- "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 for other purposes, nor on all computer systems, nor shall the act of distribution constitute any such warranty."

5. Databases and software:

- Approved database:
  - "This database, identified as [database name], has been approved for release by the U.S. Geological Survey (USGS). Although this database has been subjected to rigorous review and is substantially complete, the USGS reserves the right to revise the data pursuant to further analysis and review. Furthermore, the database is released on condition that neither the USGS nor the U.S. Government shall be held liable for any damages resulting from its authorized or unauthorized use."

We prefer to put the following in the supplemental information section, because it is not about distribution. We also sometimes combine it with no. 1, the approved data statement, and put them together in distribution - liability. ....

3. Nonendorsement of commercial products and services (refer to SM 1100.3, Appendix A):

- "Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government."

We would put the following in access and use constraints. We prefer the disclaimers that do not imply that the data cannot or legally must not be used for certain purposes, and instead only state that the data are not intended for certain purposes. ....

5. Databases and software:

- Processed data:

- "Although these data have been processed successfully on a computer system at the U.S. Geological Survey (USGS), no warranty expressed or implied is made regarding the display or utility of the data for other purposes, nor on all computer systems, nor shall the act of distribution constitute any such warranty. The USGS or the U.S. Government shall not be held liable for improper or incorrect use of the data described and/or contained herein."

6. Report and map products:

- Map interpretation to prevent misuse:
  - "Not for navigational use."
- Map or report prepared under a contract or grant:
  - "This report [map] was prepared under [contract to] [a grant from] the U.S. Geological Survey (USGS). Opinions and conclusions expressed herein do not necessarily represent those of the USGS."
- Products that provide flood-inundation information:
  - "Inundated areas shown should not be used for navigation, regulatory, permitting, or other legal purposes. The U.S. Geological Survey provides these maps "as-is" for a quick reference, emergency planning tool but assumes no legal liability or responsibility resulting from the use of this information."
- Products that provide natural hazards information:
  Note that following disclaimer can be modified for use with any natural hazard (such as earthquakes, slides, debris flows, avalanches, and volcanic hazards) by inserting the applicable wording.
  - "The suggestions and illustrations included in this [information or map product] are intended to improve [hazard name] awareness and preparedness; however, they do not guarantee the safety of an individual or structure. The contributors and sponsors of this product do not assume liability for any injury, death, property damage, or other effects of the [hazard name]."

The following provisional data statements belong in the distribution liability section of the metadata, but provisional or preliminary data should also be clearly labeled on the landing page and in the data title. Further explanations are appropriate in the metadata describing a database that includes provisional or preliminary data. ....

5. Databases and software:

- Provisional database:
  - "The data you have secured from the U.S. Geological Survey (USGS) database identified as [database name] have not received USGS approval and as such are provisional and subject to revision. The data are released on the condition that neither the USGS nor the U.S. Government shall be held liable for any damages resulting from its authorized or unauthorized use."

11. Preliminary (provisional) data, information, or software:

- "These data are preliminary or provisional and are subject to revision. They are being provided to meet the need for timely best science. The data have not received final approval by the U.S. Geological Survey (USGS) and are provided on the condition that neither the USGS nor the U.S. Government shall be held liable for any damages resulting from the authorized or unauthorized use of the data."

Mon. May 3, 2021, 2 pm - 3 pm EDT

Topic: **What will the FAIR Principles mean for our metadata?**

FAIR isn't just a snazzy acronym, it's a set of principles that describe the characteristics of data and metadata that meet scientific and Federal Government expectations for public access and scientific transparency.

On May 3, the Metadata Reviewers Community of Practice will delve into the FAIR principles. To help us prioritize our conversation, Leslie has set up a Confluence forum at

https://my.usgs.gov/confluence/display/cdi/Delving+into+the+FAIR+principles+as+they+apply+to+metadata

Please use this Confluence forum to enter your questions and comments about the individual principles, and also "like" the principles that you are most interested in discussing or learning more about. On May 3, we'll use these comments and "likes" to choose what to talk about first.

Mon. Apr. 5, 2021, 2 pm - 3 pm EDT

Topic: **Which link do you provide in the Network Resources section?**

This question is alive for Susie. Here's her summary:

Historically, PCMSC has always offered the direct download links for data in the <networkr> tag of the metadata file. In the past few years, we've given the direct link to the download, the direct link to the repository page, and then the doi link for the entire data release. All of the different <networkr> links are then explained in the Access Instructions.

We are considering changing up our methods and templates and only offering the doi link. I'm interested in hearing thoughts from others, both pros and cons. Do you think this topic would be a good one for the Metadata community of practice (or would that just lead to too many options)?

The main pros/cons I can think of:

Pro: user ease of access to the data

Con: static links may change; better to use the permanent DOI link

Mon. Mar. 1, 2021, 2 pm - 3 pm EST

Topic: Peeking into the future of metadata

Fran has been reading two interesting reports about the future of metadata:

- Smith-Yoshimura, Karen. 2020. *Transitioning to the Next Generation of Metadata.* Dublin, OH: OCLC Research. https://doi.org/10.25333/rqgd-b343.
- Leah Riungu-Kalliosaari (CSC) and others 2020. D2.4 2nd Report on FAIR requirements for persistence and interoperability (Version 1.0 DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION). FAIRsFAIR. https://doi.org/10.5281/zenodo.4001631 1631

Both reports emphasize the importance of linked data and persistent identifiers.

On Mar. 1, Fran will summarize some findings in the reports and we will discuss how they might be relevant to USGS metadata.

Mon. Feb. 1, 2021, 2 pm - 3 pm EST

This is a special session of the Metadata Reviewers Community of Practice monthly meeting. Madison will be introducing the updated Data Review Checklist and requesting feedback. During this session, we will be using breakout groups to review each data check and documenting if it should be kept in the final checklist. Breakout groups will document decisions in the Metadata Reviewers Data Review Checklist Feedback spreadsheet.

Agenda:

- Data Review Checklist Introduction (10 minutes)
- Breakout Groups (40 minutes)
- Breakout Group Report Outs (10 minutes)
    - Section you worked on (e.g., Geospatial)
    - Did your group get through all the checks?
    - Did you primarily elect to keep the checks, reword them, or remove them?

Mon. Dec. 7, 2020, 2 pm - 3 pm EST

This month we will have a Microsoft Teams meeting. Please use the link on the calendar invitation.

The topic of discussion this month is how much metadata reviewers need to investigate the data which the metadata describe, in order to review the metadata.

Erika Sanchez-Chopitea will introduce this discussion with a demo of a Python Notebook that she, Stephanie Galvan, Stephanie, and Ed Olexa developed a while back. You can find in via the Teams site. The notebook will ingest huge csv files, e.g., 24+ million records, iterate through the columns, and provide a summary. Look for "A Tool for Mining CSV Files 20201008.ipynb"

Summary: The group decided we would like to put this Notebook on GitLab where we can collaborate on improvements. Erika will check with Ed to make sure this is okay. The new, improved, data review checklist might be a useful source of ideas. Sometimes instead of having separate qualifier fields, the qualifiers are encoded into data values (negative numbers or decimals). It would be good if these tools helped reviewers discover these embedded codes in case the data owners forgot to mention them in the metadata.

Mon. Nov. 2, 2020, 2 pm - 3 pm EST

This month we will have a Microsoft Teams meeting. Please use the link on the calendar invitation.

Topics:

1. Presented by Ray Obuch.
   I can share a data dictionary template we are beginning to use in Energy and referenced as a pdf within the FDGC MD xml.
   We are also using this with our Uzbekistan group at GOSCOMGEOLGY.
   Seems to keep things simple and easy to translate into other MD standards like ISO. referencing a pdf is more complete and easier to translate then the embedded dictionary within FDGC.
2. Requested by Stu Giles:

   I would like to discuss the topic of successful methodologies and procedures for tracking data releases and reviews through the publication process, specifically, using apps and connectors currently or potentially available in Microsoft Teams to their fullest extent, and the best ways to get scientists and systems to communicate the start and completion of critical review steps.

Does anyone have other topics we should talk about?

Mon. Oct. 5, 2020, 2 pm - 3 pm EDT

This month we will have a Microsoft Teams meeting. Please use the link on the calendar invitation.

Topics:

1. The message in the Microsoft Teams space from Andrea S. Medenblik, on 8/19, about templates for metadata, and additionally how to get advice. The specific challenge is metadata for data collected by Autonomous Underwater Vehicles (AUVs).
2. The message and document that Lisa Zolly sent to the Data Management Working Group on 9/23, about the new requirement for persistent identifiers in metadata records. Specific concerns include:
   a. The need for an actual Metadata Contact (individual or group name), and <u>not</u> ASK USGS, in the Metadata Contact field. ASK USGS has nothing to do with metadata production and cannot address metadata issues. As well, the Federal Catalogs are requiring an actual <u>email address</u> in this field for Metadata Contact, and ASK USGS does not have an email address.
   b. Normalizing the responses for 'no data' in Enumerated Domains. Currently we see values including NULL, NA, n/a, na, N/A, no data, none, -9999, etc. This makes our data less interoperable, because we don't have a standard convention for conveying this information across datasets; as well, machine readability can become a problem if the value is interpreted as '0.' It would be great if we could collectively determine a USGS convention for this in our data and metadata, and ensure in data and metadata review that the convention is implemented.

Mon. Aug 3, 2020, 2 pm - 3 pm EDT

This month we will have a Microsoft Teams meeting. The link is on the calendar invitation.

Topic: Metadata for public release of legacy data, for which full documentation is not available.

Discussion Leaders: Tara Bell, Matt Arsenault, Sofia Dabrowski

Mon. July 6, 2020, 2 pm - 3 pm EDT

This month we will have a Microsoft Teams meeting. The link is on the calendar invitation.

Topic: Continue discussion of metadata for software and code. Special guest: Eric Martinez.

Here are the URLs that Eric shared on chat:

https://www.usgs.gov/products/software/software-management

https://code.chs.usgs.gov/software/software-management/-/issues/new?issue%5Bassignee_id%5D=&issue%5Bmilestone_id%5D=

https://github.com/GSA/code-gov-data/blob/master/schemas/schema-2.0.0.json

https://code.usgs.gov/emartinez/test-inventory-validation

Some notes from the discussion (please correct or add to these):

USGS is responding to OMB 16-12, which sets policy for custom developed source code that is publicly available. In particular, it requires an inventory and up-to-date metadata.

The software management website is a good source for information.

Code can be an approved release or a preliminary release. For approval, there must be three reviews: subject matter, security, and technical (code). For code to be actually made public, there must also be a license, disclaimer, and metadata.

Currently there are two recommended GitLab instances for tracking versions during code development and also for serving as repositories for code release. code.chs.usgs.gov is internal to the USGS network, so if you are developing code that will ever be public, it would be good to use code.usgs.gov. The admin for this site checks the metadata, disclaimer, and license before making a project public. But once it has been made public, the whole project is public, so the new code versions become preliminary releases. The recommendation is that the version that is approved for release should be made a "tag" which is immutable as of that point in time. Then the main branch can continue to change through bug fixes and also through incorporating new science or other improvements that change the code output. You can also "fork" the code and work on it privately.

OMB provided a JSON schema called code.json that documents releases. It does not have a defined place to put the DOI (digital object identifier) of the release, but a tags array can be used for the DOI. Some projects will have a homepage that is separate from the software landing page, so putting the DOI in the homepage element will not always work.

To find the controlled vocabularies for code.json fields, go to the schema file and search for "enum".

Mon. June 1, 2020, 2 pm - 3 pm EDT

This month we will have a Microsoft Teams meeting. The link is on the calendar invitation.

Topic: Why does each of us review metadata? And how does that affect the way we review metadata?

This is a different kind of topic. Not a search for the best answer, but an opportunity for each of us to give our own answer.

Benefits:

- We will get to know each other better, and better understand why we sometimes have different approaches to questions that require a best answer from our group.
- We will get to know ourselves better, from the experience of articulating what we care about and how that influences the way we do the job of reviewing metadata.
- If someone else expresses something, and we recognize that we also care about it, we will get to know ourselves better.
- With greater ability to remind ourselves of our good reasons, we will have more stability, self-assurance, and resistance to taking personally those unhappy things that data producers say.

Notes:

Some of the things we care about in reviewing metadata are:

- Is the data reviewed properly? Is it consistent with the companion publication? Is it complete? Is it accurate?
- Important details: references to permits, product disclaimers, cross references, accuracy and completeness fields.
- Entity and attribute section.
- Keeping it simple, making it easier for scientists to do well.
- Helping authors with entity and attribute section, and with keywords.
- Keywords, standard vocabularies
- Provide necessary information for future use in improved catalog queries.
- For legacy data, saying what information we don't have.
- Consistency of the metadata.
- Abstracts "that cut to the chase".

We spent some time on discussing where, in the metadata, to put the reference to a journal article associated with the data. Most often this could be one of the cross references, especially if there is a different larger work.

Some future discussion topics:

- How to make metadata easier for scientists to do well.
- What metadata is needed for future use in improved catalog queries.
- What is good enough metadata for legacy data when we don't have all the information? When should we just give up on releasing legacy data with metadata?
- The possibility of assuming that the DOI record will provide some of the metadata so it doesn't need to be provided in the XML metadata record.

Mon. May 4, 2020, 2 pm - 3 pm EDT

This month we will have a simple phone bridge conversation. 703-648-4848 (or 1-855-547-8255) code 64914.

Topics:

A question offered by John McCoy:

> The Question: **What type information (in the metadata) is necessary for a data publication vs research publication?**
>
> Considering **anyone** can use these data. Giving too much information can lead to problems of interpretation by the reader if it is not presented properly and even then may cause the reader to doubt since there is too much explanation with too many caveats. "If you can't explain it to a six year old, you don't understand it yourself." Einstein.
>
> Data release only. These are I am a minimalist. Keep it simple but with enough explanation to replicate the data only. The data can be presented as is. Measurements of any kind and qualified only by the data integrity statement… that there were standard calibrations and good QA/QC protocols were followed.
>
> Important details that could include: depth of sampling, any treatments of the sample, how the sample is transported. In general anything that will modify the outcome when measuring the sample. Measurement details such as the type of instrument may not matter since there are many instruments that do the same thing but get there differently. The number of significant figures matter.
>
> Research publications and data releases point to each other. The publication should have all the necessary information to replicate the study. However, abstracts for publications are in not meant to support a data release. Additionally the data release may be best if it were treated as a stand alone data release.

Notes from discussion of John's question:

- Data abstracts should be different from the abstracts of associated scientific publications, especially if the data are given without analysis or interpretation.
- To make search results in catalog interfaces easier to read, some argue that data abstract should begin with a tweet size paragraph telling the "what, where, when" of the data set and providing slightly more information than the title.
- The "why" of the data belongs in the "purpose" section.
- Data should be a stand-alone product, not dependent on the interpretive publication because the data, if collected well and documented well, will outlast interpretations of them.
- It's good to cite a publication that documents standards like the data dictionary or QA/QC activities. (Is there a persistent identifier for those data dictionaries?)
- When providing a citation for published standards, also summarize this information in the metadata record, as be sure to describe variances from the standards.

Ongoing topic: What about metadata for software or code? What have you learned? What issues should we address? The following resources were contributed by Community members:

- may be helpful: https://code.chs.usgs.gov/software

- Explanation of code.json elements: https://code.gov/about/compliance/inventory-code
- also, here is the top-level page that contains the link to the template: https://www.usgs.gov/products/software/software-management
- Here's a journal article pre-print that has some information about what software metadata is needed for findability, interoperability, and re-usability. See p 7-8 and p. 11-14. Data Science -1 (2019) 1–23, DOI 10.3233/DS-190026, IOS Press

Notes from discussion of metadata for software or code:

- Question to consider: What do you need to know about software, and how much of that needs to be added outside the Code.json record that is required by the repository system?
- Probably Code.json could be expanded. What scheme is Code.json based on?
- Is Code.gov driving Code.json, so that it can have a broadly searchable catalog?
- Eric Martinez might be able to tell us more.
- USGS Science Data Catalog is working on a code catalog using very minimal metadata from the digital object identifier tool.
- We need objective standards for code releases that can be used by reviewers and approvers.
- Useful next steps for metadata reviewers community: explain discovery aspects of code metadata; collect ideas in a Microsoft Team document; share workflows (for review and approval of code? for creating metadata for code?); share review documents we are using, like checklists; somebody investigate what is possible in the Datacite DOI record.

Mon. April 6, 2020, 2 pm - 3 pm EDT

This month we will have a simple phone bridge conversation. 703-648-4848 (or 1-855-547-8255) code 64914.

Topic: A question offered by the FSPAC Subcommittee on Scientific Data Guidance:

Would it work to include, in metadata titles, the date of the revision or release of the data? This would be helpful to people using systems like the Science Data Catalog, who are usually looking for the latest version of the data but might be looking for the one they used last year.

Our conclusion:

Two metadata records should not have the same title in their citation elements.

Discussion leading to the conclusion:

- Catalog users are complaining because they can't tell from the short catalog entry in its search results list, which is the version of the data they want. They have to go look at the metadata records, sometimes several of them. (Are these really data users, if they don't want to read the metadata?)
- The edition number in the metadata should be updated with the metadata revision date.
- We advise authors to have the "when" of the data in the title, and that would cause confusion if there were another date in the title (the revision or release date).
- For the changes in title to work as needed, we need a systematic routine process that everybody follows. But what about all the old metadata records?
- Can we only put metadata into the catalogs for the most recent version of the data? We need use cases of the situations when this would not be good.
- Who is responsible for archiving and making available old versions of data for which the landing page and DOI are used by a new version?
- No matter how we answer this particular question, USGS will need to clean up the records in Science Data Catalog. Maybe we should wait until the persistent identifiers for metadata records are implemented, and then focus on the problem records.
- Why can't the catalog just display dates, versions, etc. from other elements of the metadata records?

Mon. March 2, 2020, 2 pm - 3 pm EST

This month we will have a simple phone bridge conversation. 703-648-4848 (or 1-855-547-8255) code 64914.

Topics:

At your office, how much do you create a single metadata record for? Individual data files, items in a database, collections of data, whole data releases, or what?

- Discussion: it depends on how the data will be used. Should the data files be stand-alone, so each with their own metadata record? Do several data files need to be used together, and a single metadata record might be written for them all? Often a collection of data files provides important context for single files that might be used individually, so parent and child metadata records that link to each other in the Cross Ref section and Larger Work section of the metadata.
- SBDR Team put together presentation on this a few years ago. It's posted on the Data Management Website's Training page and titled "Structuring and Documenting a USGS Public Data Release."
- Examples on instructions page: https://www.sciencebase.gov/about/content/data-release#Decide%20how%20to%20organize%20and%20display%20data
- ScienceBase Data Release Network - a group for SB points of contact in science centers. Learn more at ScienceBase Data Release Network.
- For general questions about ScienceBase data release: sciencebase_datarelease@usgs.gov

What about metadata for software or code? How can we prepare to think together about that, maybe on our next phone call? Should we invite a speaker? Bring in reference materials? Bring in good examples?

- may be helpful: https://code.chs.usgs.gov/software
- also, here is the top-level page that contains the link to the template: https://www.usgs.gov/products/software/software-management
- Here's a journal article pre-print that has some information about what software metadata is needed for findability, interoperability, and re-usability. See p 7-8 and p. 11-14. Data Science -1 (2019) 1–23, DOI 10.3233/DS-190026, IOS Press

Mon. February 3, 2020, 2 pm - 3 pm EST

This month we will use Zoom. Here's the link. https://zoom.us/j/3496622825

Topic: Continue our November conversation about the need for persistent unique identifiers in metadata records that can be used to identify the data in the Drupal CMS as well as downstream data catalogs such as Data.gov. This solution would be inclusive of legacy data.

Lisa Zolly presented some Powerpoint slides to frame the conversation.

Here are some things we learned. (Add your own items to the list!)

- For many inescapable reasons, our USGS metadata will need to have unique persistent identifiers (PIDs) within the metadata record. This will be required when the new version of the Science Data Catalog (SDC) goes live at the end of fiscal 2020.
- There is a tool being developed to allow metadata authors to get a PID, either one at a time or in bulk using an API. The tool should be ready in about 2 months, and we can start using it immediately for all metadata records being submitted to SDC.
- Metadata reviewers should add the PID to our list of things to check.
- The SAS team will run programs to add PIDs to old metadata records. We will need to update our authoritative copies to be the ones with PIDs, and use them for any future modifications or versions.
- ISO metadata has a field for the PID, MI_fileIdentifier.
- CSDGM metadata does not have a field for the PID, so it will be put in a Theme Keyword field, using something like "USGS Metadata Identifier" as the Keyword Thesaurus value.
- Metadata tool developers and maintainers are going to try to make this as easy as possible.
- More details will be available.
- The PowerPoint deck includes a cool slide about where our metadata goes after we submit it to SDC, which is related to why we need to start using PID's, although it's not the whole story.

Why USGS Metad...IDs_rev_sm.pdf

Mon. January 6, 2020, 2pm - 3 pm EST

This month we will use Zoom. Here's the link. https://zoom.us/j/3496622825

Topic: Digital Object Identifier (DOI) Tool: new features and use with dataset revisions.

Lisa Zolly presented some Powerpoint slides to frame the conversation.

Whats New With ...ol 20200106.pdf

Here are some things we learned. (Add your own items to the list!)

- We can now get DOI's for data that was released before 2016 and don't have IPDS numbers.
- We can also provide CrossRef DOI's for related publications, as part of the DOI Tool record.
- The DOI tool will be keeping track of data citations, which will be displayed in a new version of the Science Data Catalog. These will also be provided in the DataCite record: don't delete them!
- DataCite and the RDA are working on a system of identifiers for particular scientific instruments, which can be used in our metadata files and provide a link to many details that wouldn't need to be included in the metadata files.
- USGS has an organizational identifier as part of the Research Organization Registry. When we use the DOI Tool, that identifier is automatically added to our DataCite record.
- ORCid's cannot be automatically harvested from the ORCid site by the USGS Active Directory, but we can ask our local Active Directory managers to add them to Active Directory records. That will allow the DOI Tool to automatically insert them into our DOI records.

Mon. December 2, 2019, 2pm - 3 pm EST

Topic: review questions from our forum.

Mon. November 4, 2019, 2pm - 3 pm EST

Topic: identifying different types of metadata.

Participants in the FAIR Roadmap workshop once again reminded ourselves that "metadata" is a word that refers to a variety of things. When each of us says "metadata" we know exactly what we're talking about, but listeners might think we're talking about something else. That miscommunication can make it hard to collaborate.

It would be good for USGS to develop accepted terms for different kinds of metadata. If the Metadata Reviewers Community agreed on those terms and their meanings, we could lead USGS just by the way we talk and write. Let's see what we can agree on!

And, no, the FAIR Roadmap workshop report isn't ready to be shared yet.

Links suggested at the meeting

ESIP work:

https://github.com/NCEAS/metadig-checks/wiki/Clarify-Nomenclature-and-Revise-Check-Names

https://blog.datacite.org/metadig-recommendations-for-fair-datacite-metadata/

https://github.com/NCEAS/metadig-checks/issues

http://jennriley.com/metadatamap/ (metadata visualization)
Also from the Digital Curation Centre: http://www.dcc.ac.uk/resources/metadata-standards/list%20?page=1

Example record from GenBank: https://www.ncbi.nlm.nih.gov/genbank/samplerecord/

Example of the ISO 19110 (collection level) - 19115 (item level) relationship that may help bridge the separate-but-related persistent identifier issue: https://geo-ide.noaa.gov/wiki/index.php?title=ISO_19110_(Feature_Catalog)

Highlights of discussion

Yes, there are many kinds of metadata, and many opportunities for miscommunication. We need to simply be clear about what we are talking about everytime we talk about metadata. One handy way of being clear is to say "standard compliant metadata."

Examples: metadata in the DOI (digital object identifier) record; XML format records for FGDC or ISO metadata; ScienceBase metadata that appears on a landing page; publications metadata that goes into a Pubs Warehouse database; the version of metadata used by google data search or data.gov; encapsulated metadata inside data records, like in netCDF or GenBank (not really self-documenting because you have to put the metadata into the data record); use metadata VS discovery metadata VS administrative metadata; metadata in SDC that is used to give us credit for having released data; data dictionaries!

We had a long conversation about the desirability of identifiers for metadata, since a single DataCite DOI might lead to a landing page with multiple metadata records. The use case is keeping track of whether revised (maintained, updated, improved) metadata records refer to a new data set or the same one that a previously harvested metadata record was for. This also can help with the need for an authoritative source, when down-stream metadata users are creating their own versions of our metadata records. We're not sure if using the ISO format will solve this problem. It might be something that we could do in-house, building on the expertise of the DOI tool and ScienceBase.

Persistent identifiers would be very useful for other things as well as metadata.

Mon. October 7, 2019, 2pm - 3 pm EST

We'll talk about the new FSP guidance, a revision of Guidance on Documenting Revisions to USGS Scientific Digital Data Releases.

Note: there was a request for examples of revised data releases in ScienceBase. Here are links to a few examples:

https://doi.org/10.5066/F7Q23XDH
https://doi.org/10.5066/P9RRBEYK
https://doi.org/10.5066/F77M076K
https://doi.org/10.5066/F79C6VJ0
https://doi.org/10.5066/P9Q8GCLM

Mon. August 5, 2019, 2pm - 3 pm EST

Madison will lead a discussion about the proposed page on the Data Management Website about reviewing metadata.

Reviewed user stories for Reviewing Metadata page on DM Website:

- As a *technical reviewer of someone else's metadata*, I need to know where to start, how to approach the review, what to look for, so that I can know that I've done a good enough job in my review.
- As a *dataset creator who is writing metadata* for my dataset, I need to know what kinds of things the reviewer will be looking at, what they will want to see in the metadata, and how they will judge what I've written, so that I can do a good job and so that the review process doesn't take too long.
- As a *manager* who needs to make sure the data releases in my center are reviewed and approved well and in a timely fashion, I need to understand what "reviewing metadata" entails so that I can assign the right people to the job and so that I can have realistic expectations about how much time and effort that job will take.
- As a *member of a group developing policy* for my organization, I need to know what aspects of metadata are consistent enough that they can be reviewed with rigor and which are harder to review, so that our policy can support a realistic work process for both producers and reviewers of metadata.

Current resources on Peter's site:

- Checking format & structure
- Substantive review of metadata

Discussion:

- Group has experience with people needing reviewers and guidance on reviewing
- Paul: Considered developing more specific checklist for sections, with boxes to tick
    - Ie. Is the date in the right format?
    - Dennis: Has to stay general - identify common things to address and include
    - Tom: Concerns that the checklist would get too long
- Would it be worth it to put examples/checklists more specific to individual centers/mission areas on the website or should we keep it general?
    - Public facing website keeping it general, thematic topic lists "behind closed doors" on the confluence site
    - If some mission areas want to develop additional standards, they could be supplemental documents to be included on the wiki
- Checklists for review on Confluence have not been updated since 2016 (at least for Woods Hole)
- How to find reviewers/time taken is too specific for public-facing website
    - Agreed, and data review takes a substantial amount of time - try to emphasize that since the process can be lengthy, the scientist /reviewer should start early
    - Include this information in the Plan and Metadata Review section of the DM Website
- Concern that content under data release is currently buried (result of Druple/wret migration)
    - Creating a new Reviewing Metadata page should help with making some of this content more accessible.
- Different processes for metadata authors that are new and those that are experienced. Depends on the type of data release as well - different scientific subjects will contain different elements.

- The level of detail of the review depends on the experience of the author
- Providing tools/example workflows for metadata review?
    - Peter - converts XML to TXT and then pastes in a word processor and adds comments there
    - Sofia - looks in XML at text editor and metadata wizard - use screenshots to show errors and where edits need to be made
    - Colin - in the review tab of metadata wizard, a button will generate a "pretty print" (word document that is exported, opens in word directly from metadata wizard) of the XML, has a list of the schema errors. This is usually the document that goes into XML view. Combines structural review with content review.
    - Mikki - Review checklist (word document) has a place to make comments that looks like review memo, goes into IPDS along with word doc of metadata with comments
    - Andy: Preview of metadata wizard is copied and pasted into a word doc, have places to put comments and replies (reconciliation).
    - Kitty: Uses OME - when you download it comes out as a couple different formats, likes outline format that it creates. Writes comments and scans it back to the creator. Outline format is HTML.
    - Dennis: Copy/past one of the formats out of metadata parser and copy/paste into a doc, includes validation as part of it. Nice to provide mp as an option that you don't have to download
- Other tips/tricks for new reviewers
    - Something that could be noted on the dm website
    - Notes for finding a reviewer - note that the reviewer should have written a metadata record and be familiar with the format
    - (difference between content and structural review)
    - Someone new might want to write a metadata record as a training method
    - People who had never reviewed metadata were asked at speaker's center - can have teams on a release that has people with different skill sets - some people review the GIS component, someone else with water quality experience, etc.
    - Emphasize that the metadata review can be split up by expertise (someone familiar with metadata, someone familiar with subject matter, someone familiar with GIS)

Mon. July 1, 2019, 2pm - 3 pm EST

What did we learn from our breakout session at the CDI Workshop? The notes page is here: https://tinyurl.com/CDI0605-Lightsom

We discussed the answers to the first question from the breakout session, and decided that (1) some clean-up is needed before this is a FAQ, and (2) we have at least two FAQ's, one for beginners at writing metadata, one for experienced metadata writers who are starting to review metadata. The information for beginning metadata authors should be on the USGS Data Management Website, but we're not ready to provide it yet. We will begin by collaboratively developing the FAQ for metadata reviewers in the forum section of our Confluence place. Leslie agreed to put in a first topic as an example, and to invite others to work on it.

Another topic was the frequent need to coax people into writing good metadata, or metadata at all. Fran was reminded that the requirement for metadata comes not from Reston but from OMB, the White House Office of Science and Technology Policy, and probably also the National Archives and Records Administration. Fran wants to look into those policies to see if they are useful for coaxing metadata authors, perhaps because they spell out the purposes of metadata.

Other resources we use: "Ten Common Mistakes" was useful but probably needs updating. Tom Burley has materials from the NBII metadata training that he can share. Several of us like the graphical representations at the FGDC website.

Mon. May 6, 2019, 2pm - 3 pm EST

This month we will test the technology for virtual participation in our breakout session at the June CDI Workshop.

Join Zoom Meeting
https://zoom.us/j/472209309

One tap mobile
+16699006833,,472209309# US (San Jose)
+14086380968,,472209309# US (San Jose)

Dial by your location
        +1 669 900 6833 US (San Jose)
        +1 408 638 0968 US (San Jose)
        +1 646 876 9923 US (New York)
Meeting ID: 472 209 309
Find your local number: https://zoom.us/u/acLWIyw37O

We will also have a presentation by VeeAnn Cross and Peter Schweitzer about how the USGS Science Data Catalog could use the keywords in metadata records to improve data discovery, and what that means for those who are authoring, reviewing, and revising USGS metadata records.

Mon. Apr. 1, 2019, 2pm - 3 pm EST

Sheryn demonstrated the metadata collecting system used by MonitoringResources.org to encourage discussion of how it might be simpler and easier to use, as well as good ideas that the rest of us can copy. Sheryn's slides are available.

MonitoringResources.org is part of the Pacific Northwest Aquatic Monitoring Partnership (PNAMP) and uses the metadata to provide an index of monitoring activities, especially the ecology of streams of the U.S. Pacific Northwest, and the procedures, protocols, and monitoring designs that are in use. Currently Sheryn reviews the metadata that are submitted through the site and used in the index. The site could be used for other types of monitoring and other regions, but there are not currently enough metadata reviewers to handle a larger volume of submissions.

Community discussion included questions about connections with the USGS Quality Management System (QMS) and for using the MonitoringResources. org metadata elements to build ISO standard metadata records. Community members are welcome to email Sheryn with additional ideas about ways the site could be made simpler and easier to use.

Mon. Mar. 4, 2019, 2pm - 3 pm EST

We are satisfied with the answers we received from FSPAC and glad they are posted on the FSPAC FAQ pages. We might ask some more questions later.

Related to our long-term goal of providing more complete guidance for data and metadata review, as well as tips and tricks for data and metadata authors, we agreed to host a breakout session at the 2019 CDI Workshop. We hope participants will bring questions that we can answer or at least discuss, which will be useful in the future for developing responsive online guidance. A fall-back agenda would be to step through the review checklists and talk about how we address each item on the list. Many of our members will be unable to travel to the workshop, so a virtual participation option is important. Fran agreed to put the session proposal on the Wiki, immediately, since it was due last week.

We discussed what location parameters need to be in a metadata system as opposed to being in the data itself, and came to no answer that fits every case. One guideline is that a metadata system needs to provide the parameters users need to locate the data in the associated database.

Ed mentioned a Jupyter Notebook that he, Erika, and Stephanie have developed for quick evaluation of large data files. The tool is available for others to use, and will be demonstrated at a future meeting, and at the CDI workshop. If you would like to try it sooner, contact Ed Olexa.

The ISO Content Specs project will be hosting workshop sessions on Friday of the CDI Workshop. The sessions will focus on collecting requirements for metadata specification modules, most likely modules for experimental data, computational data, and observational data. We are encouraged to plan to stay through Friday, if we can travel to the workshop.

Mon. Feb. 4, 2019, 2pm - 3 pm EST

Two major questions came up at today's meeting that we would like to pass along to the FSPAC subcommittee and/or the BAOs for guidance.

Question 1: Is there updated guidance on the volume of data necessary to trigger a separate data release?

Discussion Notes:

- Some authors have been caught off guard when they are told they need to do a data release during the publication phase of their project. They were under the impression that having the data in the paper would be sufficient.
- A couple of years ago, the idea was floated that if data could fit in a table within a manuscript that it wouldn't need a separate data release. Someone mentioned a 3 page limit, however, specific guidance never came out.
- There is likely a reason that this was left vague. Sometimes just because data could meet that threshold doesn't mean that it should. It is often necessary to have more knowledge about the specific data when determining if a separate data release is necessary. Even if the volume of data is small, they still might benefit from additional documentation that comes with metadata. We certainly don't want to bypass our due diligence with data by just stuffing it into a manuscript.
- Do we also need to ask SPN if they have a size threshold for what can be included in the main body of a publication?
- Kevin Breen would always say to publish it and making it a data release. Anytime an author would ask if they needed to do a data release, Kevin Breen would say yes. Always wanted to see it as a data release. Never had anything really short (e.g. just a few lines).
- If an author just has a few samples, they could just put it in the paper. Often it is still a judgement call – you know if it is overkill to have a data release.
- Best practice is to address this question during the proposal review process. Authors should just plan to do a data release from the start of the project.
- Sometimes these things don't come out until the publication phase and then it's a hard argument to make when funds are all used to do a data release. Put a DR in the proposal bc most projects it would make sense to do one.
- Tom Burley: Our cooperators really like DRs because they are citable and they don't have to go into NWIS to get the information.

UPDATE

Answer from FSPAC:

The original guidance about tables/pages has been removed, and more flexibility is now available to authors. There was, in the past, a conversation with OCAP that involved page numbers in relation to data. Now, there is an FAQ that addresses it - refer to the "with or without a data release" FAQ. Having the data in the paper is ok - however, if data is big enough to be moved into a supplemental section of the paper, it has to be a data release.

New FAQ (from FAQ):

Is there a size cutoff for data tables within the body of a publication or in associated appendixes and supplemental files?

- The size of a data table presented solely within the body of a publication depends on publisher requirements. Journals and other outside publishers generally have a size cutoff for within-article tables. For USGS series publications, authors should contact the Bureau Approving Official (BAO) or local Publishing Service Center Chief during the early stages of product development for guidance on the maximum sizes for tables and associated appendix and supplemental files. Although an appendix or supplemental file may contain a summary of the data that support the publication, the complete dataset may not be contained solely within an appendix or supplemental file, regardless of size.

_____

Question 2: How should authors reference data that is not publicly available when writing a manuscript?

Discussion Notes:

- Collaborations between USGS scientists and private entities has caused some confusion about how to publish papers when the private entities are responsible for the data. If the private entity is responsible for publishing the data but they haven't been published at the time that the USGS scientist is set to publish a manuscript of the analysis of the data, how should they reference the data? Likewise, if the data will remain proprietary, what is the best way to reference them in the paper?
- For sensitive data, we can reference it in manuscripts but can't release it. Same should be true to proprietary data and data that is the responsibility of other entities. Then, all of the data sources are cited and it becomes the responsibility of the user to negotiate with the external entities to get the data.
- People on the call discussed the importance of establishing roles in the DMP. We may need better guidance for what questions to ask during the data management planning phase to address data sharing and private data.
- Not only an issue with collaboration with private entities, also an issue with BLM collaborations since BLM doesn't have to put their data through peer review.
- Does FSP site have information on how to handle private data? People are having some trouble finding information as the site is being migrated to the new Drupal environment. There used to be an FSP page that had tangible scenarios related to this topic (e.g. if this is the case, do this...) Is this decision tree still available?

UPDATE

Answer from FSPAC:

Refer to https://www2.usgs.gov/fsp/guide_to_datareleases.asp for updated guidance.

For example, 'data statements' can be included in the manuscript. (In the FAQ, look for: "What statement(s) must be used to indicate the availability and, if applicable, the location of data that support the conclusions in a publication, and where should the statement(s) be placed?" for further information)

New FAQ (from FAQ):

What statement(s) must be used to indicate the availability and, if applicable, the location of data that support the conclusions in a publication, and where should the statement(s) be placed?

- Below are examples of statements to be used in various cases to describe where the data reside or to clarify disposition of any data or reasons for partial release or lack of release. Add the applicable data statement(s) to the internal USGS Information Product Data System (IPDS) Notes Tab and the publication manuscript before peer review.
  Insert appropriate text for bracketed information and retain parentheses where indicated. See "Data Citation" and USGS Publishing Standards Memorandum 2014.03 for additional citation guidance.

  Case 1. Data are available from an acceptable repository (includes USGS data release products).

  - **IPDS:** Data generated during this study are available from the [*acceptable repository*], [*DOI URL*].
  - **Manuscript:** Data generated during this study are available as a USGS data release ([*author*], [*date*]).
            Data generated during this study are available from the [*acceptable repository*] ([*author*], [*date*]).

  Case 2. Data are partially available from an acceptable repository.

  - **IPDS:** Data generated during this study are partially available from the [*acceptable repository*], [*DOI URL*]. Funding for this study was provided by [*responsible agency*]. [*Describe funding and responsibility for data release*].
  - **Manuscript:** Data generated during this study are partially available from the [*acceptable repository*] ([*author*],[*date*]). Funding for this study was provided by [*responsible agency*]. [*Describe funding and responsibility for data release*].

  Case 3. Data are not available at time of publication.

  - **IPDS and Manuscript:** At the time of publication, data are not available from the [*responsible non-USGS agency*].

  Case 4. Data either are not available or have limited availability owing to restrictions (proprietary or sensitivity).

  - **IPDS and Manuscript:** Data either are not available or have limited availability owing to restrictions ([*state reason for restrictions, such as proprietary interest or sensitivity concern*]). Contact [*third party name*] for more information.

  Case 5. Data generated or analyzed are included in the main text of the publication.

  - **IPDS and Manuscript:** All data generated or analyzed during this study are included in the main text of this publication.

  Case 6. Data were not generated or analyzed for this publication.

  - **IPDS and Manuscript:** No datasets were generated or analyzed for this publication.

- Refer also to "Data Release" in the "Definitions/Explanations/Requirements" section of this guidance for additional requirements.

_____

_____

A few months ago, this group talked about ways to improve the metadata/data review guidance documents. What are the next steps to get things updated? Can we address this at a future meeting?

Mon. Dec. 3, 2018, 2pm - 3pm EST

We will GSTalk again this month. 703-648-4848 (or 1-855-547-8255) code 64914. If we need to share screens, use Internet Explorer to go online at https://gstalk.usgs.gov/64914.

Proposed agenda:

1. Introductions: Welcome new members and address any questions they bring with them.
2. Follow up on the Nov. 8 email thread about using links to publications in the Process Step.
3. Are there items from last month's discussion of checklists for data and metadata review that need follow-up?

Mon. Nov. 5, 2018, 2pm - 3pm EST

We will GSTalk again this month. 703-648-4848 (or 1-855-547-8255) code 64914. If we need to share screens, use Internet Explorer to go online at https://gstalk.usgs.gov/64914.

Proposed agenda:

1. Introductions: Welcome new members and address any questions they bring with them.
2. Review checklists for data review and metadata review and discuss how we review data and metadata in our science centers. The checklists are online at https://www.usgs.gov/products/data-and-tools/data-management/data-release#checklists.

Report from meeting:

Tamar started two new Google Docs here:

Comments about Guidelines for Metadata Review:

https://docs.google.com/document/d/1g14C5fusPeGHP3mxERtBLUgxoz-sxqkbnehiJzD98cc/edit

Metadata review tips:

https://docs.google.com/document/d/1IqAl70nKGTK71KL1gLvihMrrmfErcBZKVefWozJx8E8/edit

Please feel free to add comments and content.

General Comments about the document "[Guidelines for Metadata Review](#)"

- Could be helpful for new reviewers: recommendations on how to start, what are the steps in the process. For example, copy/paste into Word doc for comments
    - Benefit of Word doc – can add comments, include doc in reconciliation materials in IPDS
    - Can be helpful to view with stylesheet as opposed to XML tags
- A supervisor checking a data release wondered if there was a more specific checklist for supervisors
    - In addition: the supervisor wasn't sure how to navigate multiple data and metadata files bundled together on landing page of ScienceBase data release.
    - Another meeting participant noted that supervisory review isn't technically required, just data and metadata peer review.
- A couple comments addressed the fact the doc is very general and high-level:
    - Meeting participant noted: splitting it into sections would make it FGDC-specific and document is meant to be broad and applicable to all metadata
    - MP validation is often the first step for metadata review – information about MP isn't separated out from intro paragraphs or marked by a bullet – should it be more emphasized here?
    - Suggestion: maybe split bullets into categories by CSDGM section?
- A participant's example of their process: they don't reference checklist for most reviews – instead, read metadata as a whole, check for coherence, completeness etc.
- Comment: it would be nice to be able to actively use this doc as a checklist, with spaces for checks next to the bullet points
    - Maybe have a box at the end that says "see attached sheet for additional checks"? Or a note that says this is the start of a checklist?
    - Maybe an editable PDF?
    - Participant response: this might suggest to users that the bullet list is a comprehensive check. The language in the doc is important to notice: "For example, verify that:"

Specific Comments about bullet points in the document

- Questions about recommended elements in the title and whether they are required
    - Note: important to keep in mind that there can be outliers
    - Example: can be helpful to include when there is cooperative work for various agencies

- A few authors didn't want to put the "when" in the title because they didn't think it was necessary there (and might be unclear)
- Why is "who" included in the title?
- Possible solution: add "if applicable" to all the elements of the title (instead of just "scale")
- Question about coordinate system and datum – what to do if you have an aggregation over many years – do you generalize coordinate system information
  - Example from Woods Hole – data compilations of datasets going back 50 or 60 years. There's a variety, can sometimes make guesses based on dates collected
- Note about entity and attribute section review: you can use the Metadata Parser to output the entity/attribute section as a CSV table
- Idea: create a 3rd document that contains tips and tricks that could be helpful for reviewers
  - Can contain: options of secondary validation in MP, different ways people can work on a review (e.g., Word doc methods)
- Future discussion topic: keywords (wait until Peter Schweitzer can join us)
- Should contact information check be included in the checklist?
  - Content in these fields is variable, so may not fit in the list a bullet point check.
  - Some centers have this standardized.
  - Future topic of discussion?
- Discussion topic: network resource name field – what do people use?
  - This can be helpful for large data releases with complex structures, multiple child items
  - Note: this is the default in the Metadata Wizard, so it's what folks often use
  - Woods Hole Coastal and Marine Science Center enters a direct download link to the data and the URL of the child item.
  - Many others use data release DOI (that resolves to landing page) for the networkr field
- Question: how many child items do most data releases have?
  - key consideration: does adding child items/folders improve accessibility? Varies by case
  - Majority have none or only two/three
  - A few have subfolders, but this is rare

Mon. Oct.1, 2018, 2pm - 3pm EDT

We will GSTalk again this month. 703-648-4848 (or 1-855-547-8255) code 64914. If we need to share screens, use Internet Explorer to go online at https://gstalk.usgs.gov/64914.

Proposed agenda:

1. Introductions: Welcome new members and address any questions they bring with them.
2. Discussion: What is the current state of metadata for data releases in USGS? Is there anything we could do as a community to improve the situation?
3. News: Status of ISO Content Specs project, new effort to enable FAIR principles in USGS, what else do we know?

Report from meeting:

Items from the discussion: Generally, the quality of USGS metadata is much improved in the past two years. Several science centers are trying to enlarge the pool of qualified metadata reviewers. Data reviewers are generally required to be familiar with the data type (geospatial, for example) but might be experienced data users or data producers. Metadata reviewers need specific expertise, and the time required to develop this skill depends on multiple factors, such as the background of the new reviewer, their workload, and the degree of variety in the data they will need to review. Similarly, the time an experience metadata reviewer needs for a single job can vary from days to months, depending on the complexity of the data and metadata as well as their condition (number of errors). It is difficult to teach our scientists to write good metadata, even something as simple as consistently providing the necessary information in the data set title. Abstract and purpose are also hard to teach.

Community actions that would help:

- Share a list of qualified reviewers for specific data types so that reviews can be shared among science centers when nobody qualified is available at the home science center.
- Post a list of people who can answer questions, to help new reviewers get started.
- Collect and share training materials. Several science centers have some to share. This is being done on a confluence page at this site.
- Revisit our checklists for data review and metadata review and discuss how we do each element. (Planned for Nov. 5)
- Sharyn will demonstrate the metadata collecting system used by MonitoringResources.org to encourage discussion of how it might be simpler and easier to use, as well as good ideas that the rest of us can copy. (Planned for Jan. 7)

Mon. July 2, 2018, 2pm - 3pm EDT

We're trying GSTalk again this month. 703-648-4848 (or 1-855-547-8255) code 64914. To view the slides, go online at https://gstalk.usgs.gov/64914.

Proposed agenda:

This week we need to advise the USGS Web Re-engineering Team ("WRET") on the proposed metadata requirements for the old "legacy" data sets that have been traditionally released on USGS web sites. Lisa Zolly will introduce the topic.

Mon. June 4, 2018, 2pm - 3pm EDT

We're trying GSTalk again this month. 703-648-4848 (or 1-855-547-8255) code 55793. Also online at https://gstalk.usgs.gov/55793.

Proposed agenda:

1. Questions, news or announcements?
2. Ray Obuch will provide an overview of the new Department of the Interior Metadata Implementation Guide. It uses a lot of the same words we use, but perhaps for different meanings. Do we want to help USGS get the implementation right?
3. I suggest that we look at the proposed FAIR metrics, which have a lot to do with metadata. These links from Leslie:

FAIR metrics: https://github.com/FAIRMetrics/Metrics

Leslie and I prefer this view: http://htmlpreview.github.io/?https://github.com/FAIRMetrics/Metrics/blob/master/ALL.html
A preprint: https://www.biorxiv.org/content/early/2017/12/01/225490

Report from meeting:
The meeting started with unfortunate delays caused by a typo in the calendar item. Fran apologizes.

Ray's presentation was very interesting, although the connection to the metadata review process was not clear.

After Leslie's overview of the FAIR metrics, Peter shared this link about a similar but different way of thinking about the problem, "5 Star Open Data".

Mon. May 7, 2018, 2pm - 3pm EDT

We're trying GSTalk again this month. 703-648-4848 (or 1-855-547-8255) code 64914. To view the slides, go online at https://gstalk.usgs.gov/64914.

Proposed agenda:

Burning questions? Metadata nightmares? Brilliance to brag about?

Barbara Pierson will join us to continue our discussion of the USGS Genetics Metadata Working Group (wiki page, **Genetics Guide to Data Release and Associated Data Dictionary**).

The project to create metadata content specifications for easing USGS transition to the ISO metadata standard has started planning their workshop. We hope for an informal progress report.

Report from meeting:

GSTalk did not work for sharing desktops. We suspect that we need to be more conscientious about installing updates frequently.

We had a good discussion about the Genetics Guide to Data Release, and agreed to provide some comments to enable the working group to take this document to the next stage. Everybody, please try to do this before our June 4 meeting!

The content specifications project is having trouble finding a workable date for their workshop. They are thinking of a modular approach to the specifications, starting with a basic module that includes identification and discovery information, a biological module, a process steps module, and at least one geospatial module. Any given metadata record would use the set of modules that were appropriate. It seems that quality descriptions might be part of multiple modules.

Mon. April 2, 2018, 2pm - 3pm EDT

We're trying GSTalk again this month. 703-648-4848 (or 1-855-547-8255) code 64914. Also online at https://gstalk.usgs.gov/64914

Proposed agenda:

Burning questions? Metadata nightmares? Brilliance to brag about?

Let's look together at this wiki page created by the USGS Genetics Metadata Working Group, especially the **Genetics Guide to Data Release and Associated Data Dictionary.**

Report from meeting:

Burning questions centered on review and release of "legacy data sets" that no longer are supported by the project that created them, yet still have scientific value. Some centers are using the IPDS process, while being careful in metadata to identify limitations of the data. Others are updating metadata that was published with old data sets, when possible with the participation of the originating scientists. Advice: be sure to add a process step to the metadata record when you modify it, and update the metadata date. Unresolved question: can legacy data go into the WRET Drupal environment?

About the Genetics Metadata Working Group materials: much of this seems more general, so that a similar document might be useful beyond the genetics community. Dennis will ask Barbara to meet with us next month.

Topic for a future meeting: How can will develop a collection of quality statements to suggest, similar to those in the Genetics guide? Do we want to provide examples, or the sample questions that the statements should answer? Madison is interested in this.

Mon. March 5, 2018, 2pm - 3pm EST

We're trying GSTalk again this month. 703-648-4848 (or 1-855-547-8255) code 64914. Also online at https://gstalk.usgs.gov/64914

Proposed agenda:

Burning questions? Metadata nightmares? Brilliance to brag about?

News: The ISO metadata project going for a full proposal to CDI, the CMGP meeting with USGS Thesaurus team, what else?

Report from meeting:

Leslie will be adding some email discussions about metadata topics to our community forum.

The project team will be submitting a full CDI proposal to create specifications for USGS data products so that ISO standard metadata records can be created in tools like the ADIwg metadata toolkit. The proposal is due at the end of March, so next week community members will have a chance to look at a draft of the proposal and suggest improvements.

Coastal and Marine Geology metadata specialists had a meeting last week with the USGS Thesaurus team to improve the usefulness of the Thesaurus as a source of metadata keywords that will improve data discovery.

The USGS data management website is improving its page about data dictionaries and would like our comments. You can comment on a draft of the page at https://docs.google.com/document/d/1140npvNsCb-ixQ-e-dDws2AOU7q2yHtk4pJ1_Ce5HHI/edit

Mon. February 5, 2018, 2pm - 3pm EST

Proposed agenda: Demo of the new ADIwg metadata editor by Dennis Walworth and Josh Bradley.

Report from meeting:

Thanks to the FWS WebEx, we had a great presentation and demonstration of the ADIwg metadata toolkit, which is finally fully functional and ready for widespread use. I had thought of it as a way to make ISO19115 metadata, but it can also be used to make the old CSDGM. Our metadata authors can keep using the same tool through the USGS eventual transition to the ISO standard. It seems like it would work well for metadata review, as well, because it can produce an html output that is easy to read. The slides are attached, but the demonstration – you had to be there. Thank you, Josh and Dennis!

Tue. January 9, 2018, 2pm - 3pm EST (note temporary change in schedule)

Proposed agenda: Demo of the new ADIwg metadata editor by Dennis Walworth and Josh Bradley.

Report from meeting:

We were unable to make GSTalk work for the demonstration. We will try again next month.

Eventually, we discussed working together to propose a CDI project to lay the groundwork for use of ISO metadata in USGS. Dennis, Lisa, Fran and Tara volunteered to work on this proposal.

Mon. December 4, 2017, 2pm - 3pm EST

Proposed agenda: That pesky data quality information.
- Can you share an example of a correct and helpful data quality assessment?
- Can you share an example of a data set for which you would like advice on how data quality should be stated?
- How should data quality items be handled in a metadata tool like Metadata Wizard? (Is it counter-productive to suggest boilerplate answers?)

- See the work on revising the Data Management Website, via Madison's forum post: Data Quality Documentation Examples

Report from meeting:

At the meeting we talked about items in Madison's collection of Data Quality Documentation Examples. We didn't finish talking about the collection.

Some things that were said include:

Are these all supposed to be good examples? In any case, they were good discussion starters.

Unanswered question: Should information be given only once in a metadata record, or is redundancy useful? Specifically, should quality control measures be described only as a data processing step, with the data quality elements providing only the quality standards used or the resulting accuracy/precision of the data?

It's important to give a definition of how the project identified "outliers" and what was done with them – flagged? deleted? replaced with interpolations? This could go in completeness report, attribute definitions, or logical consistency.

Completeness report should say what is known to be missing from the data, or what is missing intentionally.

Users like a lot of information to evaluate data before they use it.

One approach to a completeness report is a table that provides the number of missing values for each attribute, but it is a large table and some metadata tools might not allow the table formatting.

Logical consistency is a good place to include mismatches between data from different sources that were merged or compiled or tiled into a data set. Do the data always mean the same thing regardless of which record you look at? In some cases it might be more useful to include measures of data quality as data values associated with observations or measurements.

When metadata was mostly used for digital geospatial data, logical consistency was mostly used for topological correctness.

Some possible conclusions:

- A set of "good examples" only makes sense if it includes information about what kind of data each example would be good for. There is no universal right answer to data quality descriptions.
- There being no universal right answer to data quality descriptions, metadata tools probably shouldn't suggest any boilerplate.
- Perhaps, instead of good examples, it would be more useful to provide additional, easier to understand, definitions of each field. What questions should be answered by the information in each field? (The Metadata Workbook did this, but is out of date.)

Mon. November 6, 2017, 2pm - 3pm EST

Proposed agenda: Discussion of the Biological Data Profile, led by Pai, and Erika, and Robin.

Report from meeting:

Pai and Erika shared examples of using the Biological Data Profile for data from Sea Otter Surveys. (See https://www.sciencebase.gov/catalog/item/55b7a980e4b09a3b01b5fa6f.)

Robin raised the question of how to format taxonomy data when the data involve hundreds of species. Lisa said that the metadata would be okay if you just do the taxonomy on a more generalized level and provide a complete listing of taxa that can be downloaded. Validation of metadata doesn't require that taxonomy are complete to the species level.

Robin said that her group is using the CAS registry for identifying chemical substances, which led to a discussion of the usefullness of similar authority files and codesets. We agreed to add the authority files and codelists that we find useful to a list that will be on the Data Management website.

Mon. October 2, 2017, 2pm - 3pm EDT

Proposed agenda:

1. Burning questions? Metadata nightmares? Brilliance to brag about?
2. Status report on new checklists
3. Metadata training
   a. Summary of Google form survey results
   b. Do we know enough, or should we collect more data?
   c. What can we do for online training for Metadata Wizard and OME?
   d. What can we do to provide shadowing/mentoring for metadata review?
   e. What else do we want to do?

Report from meeting:

The new checklists were changed slightly by the FSPAC Scientific Data Guidance Subcommittee and sent to the webteam to be posted on the USGS Data Management Website.

Slides summarizing the survey results are attached.

Points made during the discussion:

NOAA training is available for ISO metadata, but it assumes willingness and capability to edit an XML file.

USGS people probably need two levels of training:

- Beginners who never wrote metadata before.
- Intermediate learners who need best practices.

Metadata Wizard is in the software release process. The plan is to provide live training at FORT then publish that as a tutorial. We offered to help, when the time comes.

OME was not represented at our meeting.

What do we notice that people need to learn?

- What links go where. (Lisa's presentation)
- What information goes in the different metadata elements.

Typical metadata shortcomings

- Title doesn't describe data.
- Abstract is from paper, not about data.
- Dates used inconsistently or wrong.
- Keywords missing or misused
- Data quality elements not robust. "It's all described in the paper" or obtuse FGDC definitions.
- How to communicate in plain language a description of their scientific work.
- … 25% of elements have misunderstandings, no gaping holes

What helps

- Having written a first record, beginning to see how to use templates.
- SOP's for different data types.

What we would like as metadata reviewers

Our general idea is to meet in small groups that share experience with particular data or formats. What things in the metadata do we look for that lead to conversations with authors and better metadata. Or ways to go beyond the boilerplate offered by tools – recognize boilerplate responses and ask the authors if something more customized to the data might be possible.

- How to create proper FGDC metadata for netCDF files, in addition to the CF information that is already there. But need to provide information to people who don't know how to use netCDF, so they can evaluate whether to download or learn to use it. Votes:3
- How to create metadata for SegY formatted seismic data.
- What to do with records that include reference to or attributes from the lidar base specification (current version is 1.2 from November 2014).
- How to use the biological data profile.

Dennis, Lisa, Pai, and Erika volunteered to start off this new kind of "learning together" with a session about the biological data profile on Nov. 6.

Tue. September 5, 2017, 2pm - 3pm EDT

Meeting on a Tuesday because Labor Day is our regular meeting day, and because there is some urgency for us to recommend new checklists for data review and metadata review!

Proposed agenda:

1. Burning questions? Metadata nightmares? Brilliance to brag about?
2. Discuss, possibly revise, and hopefully approve new USGS checklists for data review and metadata review.

    - Revision for the introductory text on the Data Management Website (this should replace the text in section 5 of https://www2.usgs.gov /datamanagement/share/datarelease.php). We need to find a good utility for converting metadata records to UTF-8. Does anybody have one?
    - Recommended revision of the Data Review Checklist.
    - Recommended revision of the Metadata Review Checklist. Note the comment for discussion, attached to the second paragraph.

Report from meeting:

We discussed the need to check xml metadata records to make sure the text is encoded in a way that will not cause trouble. We don't have an adequate tool for checking or converting to UTF-8 encoding, so we will engage in a use case process to clarify what the tool needs to do, and then likely one of us can develop it. Those interested in participating in this use case process should contact Fran Lightsom, flightsom@usgs.gov.

We reviewed the draft checklists and made some improvements. Community members have until September 12 to double-check the following documents and speak up (email Fran, or the whole group) about any remaining problems or omissions. The new versions are here:

- Revision for the introductory text on the Data Management Website (this should replace the text in section 5 of https://www2.usgs.gov /datamanagement/share/datarelease.php).
- Recommended revision of the Data Review Checklist.
- Recommended revision of the Metadata Review Checklist.

Mon. August 7, 2017, 2pm - 3pm EDT

Proposed agenda:

1. Burning questions? Metadata nightmares? Brilliance to brag about?
2. Discussion with Lisa Zolly about metadata requirements for good functioning in the USGS Science Data Catalog.

Report from meeting:

One burning question: Will the new Metadata Wizard release be open source? Answer: Yes, it will be stand-alone (independent of ArcGIS).

Presentation by Lisa Zolly (CSASL): "Metadata Tips for Better Discoverability of Data in the USGS Science Data Catalog" (see attached PowerPoint)

The focus of Lisa's presentation was twofold:

1. Optimal use of theme and place keywords in Science Data Catalog (SDC)
    a. SDC browse function utilizes coarsely granular keywords from USGS Thesaurus
    b. SDC search function utilizes finely granular keywords from various disciplinary controlled vocabularies (CV)
    c. Controlled vocabulary resources: Controlled Vocabulary Server maintained by Peter Schweitzer and the USGS Data Management website

2. Optimal placement of links related to data releases. Specifically, the preferred use of: <onlink> in <citeinfo>; <onlink> in <lworkcit>; <onlink> in <crossref>; and <networkr> in <distinfo>. See PowerPoint for more detail on the problems the SDC team has had in deciphering links (data link? publication link?), and how metadata authors and reviewers can help alleviate those problems.

The ensuing discussion continued for a second hour and addressed topics such as: the distinction between the USGS Science Data Catalog and ScienceBase; different strategies for constructing ScienceBase landing pages with child items; the ways that metadata records are added to the SDC; recommendations from Force11, DataCite, and other organizations that DOIs point to landing pages (not to XML metadata files); etc.

NEXT MEETING: Tuesday September 5 (to avoid Labor Day, September 4). See proposed agenda, above.

Mon. July 3, 2017, 2pm - 3pm EDT

Proposed agenda:

We anticipate a small group that will work on cleaning up comments that have been made on the google doc versions of the data review checklist and metadata review checklist.

Report from meeting:

It was a good, productive working meeting, with two additional working meetings to finish the job. Results of our work are attached.

- Revision for the introductory text on the Data Management Website (this should replace the text in section 5 of https://www2.usgs.gov /datamanagement/share/datarelease.php). We need to find a good utility for converting metadata records to UTF-8. Does anybody have one?
- Recommended revision of the Data Review Checklist.
- Recommended revision of the Metadata Review Checklist. Note the comment for discussion, attached to the second paragraph.

Mon. June 5, 2017, 1pm - 2pm EDT (Note time change to accommodate Werkheiser Q&A WebEx.)

Proposed agenda:

1. Burning questions? Metadata nightmares? Brilliance to brag about?
2. Discussion of ideas arising from CDI Workshop, see notes below.
3. Form a subcommittee to propose improvements to data and metadata review checklists
    a. Data review checklist and metadata review checklist on google docs for collecting our comments and suggestions.
    b. Metadata Review Checklist on USGS Data mgmt website (CDI, 2014)

Notes from meeting:

1. Curtis Price had some announcements from EGIS:
    a. EsriUC Metadata SIG 7/12 will be on WebEx - Esri will give an update on ISO metadata at this mtg
    b. Updated metadata cookbook on draft EGIS website
    c. Metadata Wizard (the last version which works inside ArcGIS) was released in USGS with ArcGIS 10.5.
2. The workshop idea of most interest is regular program of training and mentoring for both writing and reviewing metadata. (Cian Dawson suggestion at the CDI Workshop.) The purpose would be to provide assistance to new metadata writers and reviewers, although Peter reminded us that it would also be a learning experience for the teachers or mentors. Fran agreed to follow up on this idea, and will get back to the community for guidance on scheduling and syllabus.
3. We did not form a subcommittee to work on the data and metadata review checklists, but instead will make progress through this alternate path:
    a. Community members will leave suggestions and comments on the google doc versions of the lists during June. Be careful to switch to suggesting mode. (Look in the upper right corner of the page, a pencil icon means you're still in editing mode!)
    b. Our next community meeting is scheduled for July 3, and looks like a small group. Those of us working that day will constitute a committee to deal with the suggestions and comments on the google doc lists and create clean documents that the community can fine-tune at our August meeting.
    c. We need to get Lisa Zolly involved in identifying metadata requirements because USGS metadata must be functional in the Science Data Catalog.
    d. Discussion lingered over the possibility of having multiple lists that were customized for different types of data, or lists of "submission guidelines" that would be provided to metadata or data authors so that submissions would be higher quality, there would be fewer surprises during review, and the review checklists could be much shorter. Water was mentioned as a source of good lists.
    e. We were in agreement that it would be a very good thing if the general quality of USGS metadata were uniformly excellent, but did not see a path forward to achieve that, with the current policy and management situations.
    f. We need to share information sources about specific disclaimers or similar statements that should be put in specific places in metadata records, and not in others. The ScienceBase data release checklist is one source.
    g. Andy LaMotte and Alan Allwardt are volunteers to help get this done.

Fri. May 19, 2017, 9am - 12pm MDT

In the "Open Lab" at the CDI Workshop

After a spirited discussion of the larger work citation and the best place in the metadata record for a citation of an associated publication, we had three demonstrations: the new stand-alone version of Metadata Wizard (Colin Talbert), secondary validation of metadata records at https://mrdata.usgs.gov/validation/, and the Alaska Data Integration work group (ADIwg) metadata editor.

Notes from the session:

Rose has code to pull entities and attributes from an Access database, which could be inserted into a Metadata Wizard record.

ADIwg will soon release mdEditor which works on ISO type metadata expressed as mdJSON rather than XML. (A mdJSON schema validator has already been created.)

mdEditor works for ISO metadata, which might be more compatible with data that doesn't fit cleanly into the FGDC CSGDM.

Ideas:

1. Make mdJSON the internal USGS standard for writing ISO metadata.
2. Develop profiles for guidance about which ISO fields should be provided for different kinds of data.
3. A controlled vocabulary service is needed for GNIS.
4. Start investigating how we would review ISO metadata using mdJSON.
5. Can contact items be available as a service for inclusion in mdJSON?
6. Can data dictionaries be available as a service for inclusion in mdJSON?
7. Metadata reviewers CoP have a session to experiment with mdEditor when it is ready.

Suggestion from Cian Dawson: interactive training using WebEx training center. Hands-on in small groups with instructor checking in. Separate tracks for metadata creation and metadata review.

Project idea: A database and interface for a collection of data dictionaries or data dictionary items, for use in designing data collection, and then in metadata, and then in data integration.

Mon. May 1, 2017, 2pm - 3pm EDT

Proposed agenda:

1. Burning questions? Metadata nightmares? Brilliance to brag about?
2. New developments with the Metadata Wizard (Colin Talbert)
3. How much is enough for Data Quality Information? Are there good examples for different situations?

Notes from meeting:

1. A community member expressed the opinion that the keyword list of the Global Change Master Directory (GCMD) "is a pain." Painful aspects are the fact that the list is not monohierarchical and the question of whether one uses the whole string of terms or just the last term. The group seemed to be in general agreement with the pain, with one member saying that there were not useful terms in the GCMD list anyway, and suggesting that we need to bring back the USGS Biocomplexity Thesaurus.
2. Colin Talbert presented a few slides (MetadataWizard2.0.pdf) and a demonstration of the new version of the Metadata Wizard. The Wizard will have many new features:

- Users will no longer need ArcGIS installed. Instead, an installer will be provided to use Wizard as a stand-alone application
- The entity and attribute builder works on CSV or Excel files.
- The error report is friendlier.
- Users can copy and paste whole sections from one record to another.
- A map is provided for the spatial domain, which can be used to modify the domain in the record.
- Users can easily switch between the biological profile and the basic FGDC CSDGM standard.
- The components of the application can be used as a Python library.

Colin hopes to have an "early adopter" version of the new Wizard available at the CDI Workshop, with actual release in late summer.

3. The real question about Data Quality Information turned out to be about the "Logical Consistency" item. Peter clarified that when the metadata standard was only used in the GIS community, this item was used to state how much topology was enforced. In general, the idea is to state any inconsistencies between parts of the data that might arise from compiling data from different sources. Dennis further suggested that "Logical Consistency" is a good place to specify exceptions to the values that are expected in data fields. Drew shared the explanations of these fields that are offered by Metadata Wizard:

- Do all values fall within expected ranges?
- Has data been checked for omission or commission? Has topology been verified for geographic data to ensure data integrity?
- What checks have been performed to ensure that the data set matches up with the description provided in the 'abstract' and 'purpose'?
- Have you verified that features and data entries are not duplicated?
- Do all values fall in a valid range (for example, a data set of precipitation values should not have negative values?) Provide as much information as possible.

Announcement: the AdiWG Metadata Editor will be displayed at the CDI Workshop.

Mon. April 3, 2017, 2pm - 3pm EDT

Proposed agenda:

1. Burning questions? Metadata nightmares? Brilliance to brag about?
2. Dealing with the suggestion that some data isn't worth the time and trouble to write complete metadata records. What is our response as individual reviewers, and as a community?
3. Data review checklist and metadata review checklist: review suggestions from the group.

Notes from meeting:

1. Question: I'm reviewing data that is associated with a publication, and the author says "all process steps are in the report."
Responses:
The bureau has said that the data release is a separate thing from the publication, so it must be able to stand alone.
In particular, the metadata should allow future data users to know if their data download was successful.
Metadata process steps might be either a summary of the method section of the report, or they might be more detailed than the method section of the report. The metadata process steps should be a succinct statement of how the data were developed. Use plain language but don't allow the plain language to reduce clarity.

   Question: A PI, after completing the data review step of the FSP process, decided to add additional fields and cells to the data. The metadata needs to be changed to reflect this, but the PI doesn't want to start the review process all over again.
Responses:
The PI can check with the reviewer to make sure that the metadata is still good after the changes, which the reviewer could document in the notes section of IPDS. Then the changes, to both the data and the metadata, can be considered responses to review, rather than a new data product. (Alaska has a formal data acceptance step before the product goes to the approving official.)

   Question: What do we do with data from projects that have ended – the scientist might even be gone – and the metadata record is incomplete. The problem is that the organization doesn't have support for anyone to complete the metadata, and existing people give the problem a low priority.
Responses:
If a new publication uses the data, the authors will be forced to improve the metadata.
If the data were made available on a website, the visibility would encourage improvement of the metadata in order to look good.
A case could be made of the value of data re-use, of taking pride in the work the organization has done in the past, and sustaining the value of that work into the future.

2. Discussion of the author who suggests that their data is too insignificant to bother with a complete metadata record.
Example given was a USGS coauthor who contributed 6 numbers to a much larger data table that was published by another organization.
Responses:
Could the data be released as part of a larger collection of similar data, which would obviously need metadata?
Could a minimal metadata record provide only the information that is known?
The Bureau says that we must make metadata. In IPDS, all released data is considered to be worth metadata.
In data archives or scientific case files, there are likely to be data sets, such as versions produced during data processing, which do not need a formal metadata record. A question list such as the headers in the "plain English" metadata format would be a good way of collecting the information that will be needed about that data set in the future.
If the data are not worth metadata, then why were they worth collecting?
The metadata, at a minimum, need to tell people what the data are.
The metadata, at a minimum, need to provide the necessary information for future scientists who will re-purpose the data. The goal is to help them do their jobs and achieve their goals.

3. We ran out of time, again, and didn't deal with the revisions to the checklists.

Mon. March 6, 2017, 2pm - 3pm EST

Proposed agenda:

1. Burning questions? Metadata nightmares? Brilliance to brag about?
2. Metadata community face-to-face will be held during the CDI Annual Workshop. Current plan is a social gathering. Do we want to do something more substantial? If so, what and when?
3. Ray Obuch's proposed Energy Program standards for metadata quality.
4. Can we help science fields that don't mesh with FGDC: genomics, those who contribute to big integrated databases, others?
5. Data review checklist and metadata review checklist: review suggestions from the group.

Notes from meeting:

1. We had a question about position and projection and data information in a metadata record for data formatted as an ascii grid. Consensus that the information is important and should go into the same location as in a metadata record for an ArcGIS presentation of the data.
2. At the Metadata community, we will try to have our social gathering on Tuesday evening so that we can use one of the Wednesday afternoon breakout times for discussing issues that we identify at the social gathering.

3. A summary of Ray's proposal is attached. Discussion touched on the value of standard data dictionaries, a need to clarify the minimum required set of metadata elements, and what to do with data sets, for example laboratory experiments, for which no geographic coordinates are really appropriate. Is there a null value for spatial location? Some offices use the global domain.
4. The groups that "don't mesh with FGDC" were not represented. We seem to have our hands full with our own metadata challenges.
5. Checklist revision isn't moving forward, at least not fast.

Mon. February 6, 2017, 2pm - 3pm EST

Proposed agenda:

1. Burning questions? Metadata nightmares? Brilliance to brag about?
2. Geospatial Metadata Validation Service (online version of mp): new default setting for Upgrade function; see <https://mrdata.usgs.gov/validation/about-upgrade.php>.
3. USGS Science Data Catalog to accommodate ISO metadata in near future: implications for USGS metadata reviewers?
4. Data review checklist and metadata review checklist: review suggestions from the group.

Notes from meeting:

Alan Allwardt and Peter Schweitzer leading, in Fran's absence
Notes by Alan Allwardt

1. Burning questions and comments: Janelda asked about the practice of creating an "umbrella" metadata record for a collection of datasets (each with its own metadata record). Madison presented an example from ScienceBase (http://dx.doi.org/10.5066/F7M043G7) in which the parent landing page links to child pages that provide the data (in this case, for different subregions); the parent page has a metadata record and so do the child pages. Generalizing this case: the parent page metadata might have an entity and attribute overview, whereas the child page metadata might have more specific entity and attribute information (especially useful if the child pages present different data types).

   This led to a discussion of other models for relating individual datasets to one another: the "Associated Items" feature in ScienceBase is one option (example: http://dx.doi.org/10.5066/F7GQ6VXX), but this kinship will not be reflected in the metadata records for those associated datasets.

   Dennis: his group uses Larger_Work_Citation to point to parent landing pages and Cross_Reference to point to related publications. (NOTE: Metadata Wizard currently does not accommodate Cross_Reference; Madison will bring this issue to the attention of the developers.)

   Dennis again: his group puts ORCID in the Originator value as follows: <origin>Dennis Walworth (ORCID:0000-0003-1256-5458)</origin>. Seems like a simple and effective solution, but the follow-up discussion centered on possible downstream impacts of this practice: in the USGS Science Data Catalog, for instance, "Dennis Walworth" with and without the ORCID would be listed as separate authors in the browsable sidebar. No deal-killers emerged from this discussion, however.

   Peter demonstrated his internal website for maintaining authority control of authors in Science Topics and Mineral Resources On-Line Spatial Data.
2. Peter reviewed recent changes in the upgrade function for his Geospatial Metadata Validation Service (see https://mrdata.usgs.gov/validation/about-upgrade.php). The change in the default setting makes the online validator work in the same way as the command-line version of mp -- and, by changing the upgrade function from an opt-out procedure to an opt-in procedure, makes users aware of certain types of errors in the input file that used to be fixed without their knowledge.
3. Alan reported this news from Lisa Zolly: the USGS Science Data Catalog will begin harvesting and indexing ISO metadata at the end of February (or so). The group discussed potential impacts on USGS metadata reviewers (primarily: lack of experience and relevant tools).

   Peter likes the practice of Dennis and his group: do as much writing and reviewing as possible before converting to XML (that is, use JSON as an intermediate step).
4. Data review checklist and metadata review checklist: a few members of the group have begun reviewing and suggesting changes; the others were encouraged to take up the task.

   The importance of having the data reviewer also look at the metadata (and the metadata reviewer also looking at the data) was stressed -- we need to make sure that the checklists get this message across loud and clear. At Alan's request, VeeAnn described how this works in Woods Hole: for instance, when a metadata reviewer is not well-versed in a particular data type.

   For the data review, Janelda wondered if there might be a way for authors to indicate the expected range of values for any given parameter, so that the reviewers could easily identify outliers. Peter suggested using Range_Domain (while acknowledging that there is some difference of opinion about what this element should represent: the range of all conceivable values for the parameter, or the range of actual values within the dataset).

   Peter pointed to some of his handy tools for evaluating datasets: <https://geology.usgs.gov/tools/metadata/>. If you select the "Web services" tab on this page you'll see tools for analyzing DBF and CSV files.

   Finally, there Peter and VeeAnn have developed guidance that is too detailed to include in the metadata checklist, but should be referenced in the checklist:

   Primary validation using mp: <https://mrdata.usgs.gov/validation/how-to-review/>
   Substantive review of metadata elements: <https://mrdata.usgs.gov/validation/how-to-review/elements.html>

POSTSCRIPT: If you are using customized data and metadata review checklists or templates in your science center or program, please share your experiences here: <How have you adapted the data review and metadata review checklists for use in your science center?>

Mon. January 9, 2016, Report to CDI Data Management Working Group

This is not really a meeting of the Community, but the Data Management Working Group asked for a progress report. Attached are the slides prepared for that report.

Mon. December 5, 2016, 2pm - 3pm EST

Proposed agenda:

1. Burning questions? Metadata nightmares? Brilliance to brag about?
2. Look at suggested revision to the USGS Data Management website > Publish/Share > Data Release <https://www2.usgs.gov/datamanagement/share/datarelease.php> > Section 5.
3. Are we ready to start reviewing the data and metadata review checklists? (Or wait until January?)
4. Do we want to sponsor training at the CDI Workshop?
5. Next meeting (not Jan. 2).

Notes from meeting:

1. Colin reports seeing good metadata at FORT. Bill is working with data that will be contributed to the California environmental data repository, and reports that our standards for metadata and review are more stringent than theirs.
   There has been some recent email discussion of where data producers' ORCIDs should go in the metadata record. There seems to be no place to put ORCIDs where they would be immediately useful in systems like the Science Data Catalog or data.gov, but there are several places where they might reasonably be found and would not cause a record to fail standard validation checks. Peter Schweitzer will start a discussion at our community confluence site so that we can decide on a consistent approach.
2. Alan introduced the revision to section 5 of the website, explaining that much of the information in the present section is off subject, for example, about publications that are not data. Fran added that the new USGS policy is that no data is interpretive, so we decided to drop the sentence about interpretive data from the revised section. We would like to add specifics of how IPDS is used for data and metadata reviews, and Fran made a suggestion for that in the Google document. We would also like the webpage to provide more easily found links to guidance and policy. Susie showed the IP record in progress for a Santa Cruz data release in ScienceBase. The record shows original metadata files and reviewed metadata files, as well as reviewed ScienceBase pages. This case does not have the metadata harvested from ScienceBase to Science Data Catalog. Conversation continued on the question of whether the short metadata records for ScienceBase project pages, which do not include data but provide a description of a collection of data, need to be compliant with metadata validation, for example, by mp. Tamar shared information that in the future such metadata will be harvested and thus will need to be validated. Peter said that a basic metadata record that only has sections 1 & 7 could be validated. ISO metadata more intrinsically accounts for relationships between collections and the items they contain.
   Decisions: We will leave the revision on Google docs and encourage community members to suggest improvements, using "suggesting" mode instead of "editing" (the mode choice is available in the upper right corner, under the Comments button). Also suggest guidance and policy links that should be provided on the webpage. Fran will negotiate the webpage changes with Viv Hutchison.
3. Peter will put the data and metadata review checklists on Google docs so that community members can start suggesting modifications (see links below). Our goal is to have fairly generic checklists, helpfully grouped and chunked, with links to more detailed lists for particular kinds of data.
4. We did not have time for discussion of the CDI workshop.
5. We decided to skip the January phone call, since Jan. 2 is a holiday and the Data Management Working Group is likely to be meeting on Jan. 9.

Other discussion topics:

Briefly raised, what about data that is included in administrative reports and proprietary data? POSTSCRIPT – January 4, 2017: Alan spoke with Keith Kirk (FSP committee) and he says this issue is currently under consideration by FSP. He also said that the USGS report series called "Administrative Report" will be renamed/redefined in the near future. Stay tuned.

Briefly raised, how can we deal with the issue of links to files changing in ScienceBase, when the data is modified, and the challenge of keeping links correct in metadata?

Google docs for community review before our Feb. meeting:

Data Review Checklist is a copy of the existing checklist formatted as a Google Docs and shared for edit and comment.

Guidelines for Metadata Review of Data is a copy of the existing checklist formatted as a Google Docs and shared for edit and comment.

Mon. November 7, 2016, 2pm - 3pm EST

Proposed agenda:

1. Burning questions? Metadata nightmares? Brilliance to brag about?
2. Start reviewing the data and metadata review checklists on the Data Management Website.
3. Any ideas about how we might get together at the CDI Workshop?
4. Next steps?

Notes from meeting:

Metadata Reviewers Community
Meeting: 20161107

Peter Schweitzer leading, in Fran's absence, with input from Alan Allwardt, VeeAnn Cross, and the group
Notes by Alan Allwardt

Agenda Item 1. Burning questions

Peter Schweitzer: told story of someone asking him what to do about a non-geospatial dataset for which the metadata failed mp because there was no spatial domain information. In the past Peter would have recommended ignoring the mp error, but now he recommends entering a global extent to avoid validation errors in downstream catalogs like data.gov.

Lisa Zolly: confirmed that data.gov will flag and quarantine CSDGM records lacking a spatial domain (USGS Science Data Catalog will not).

Members of the group shared their strategies in dealing with metadata for non-spatial data: some create global spatial extents; others will use the bounding box of the parent project for non-geospatial, supplementary or lab data. It was generally agreed that using the coordinates for the science center where non-geospatial lab results were obtained is a BAD idea.

ACTION ITEM: Peter will add a paragraph to his "Substantive review of metadata" training page <http://geo-nsdi.er.usgs.gov/validation/how-to-review/elements.html> to deal with spatial domain conundrums.

Agenda Item 2. Revising the data review and metadata review checklists

Peter suggested stepping back from the checklists and look at the context in which they are presented: USGS Data Management website > Publish/Share > Data Release <https://www2.usgs.gov/datamanagement/share/datarelease.php> > Section 5.

ACTION ITEM: After extensive discussion, the group decided that the text of Section 5 -- which provides context for the checklists -- should be revisited and revised as necessary FIRST, and only then should we consider how to revise the checklists themselves. (Revising the text of Section 5 will inform the process of revising the checklists.) This plan met with general approval. Alan will begin revising Section 5 and get input from Peter, VeeAnn, and Fran before it is posted on Google Docs for the group to consider.

Highlights of the discussion leading to the action item above:

Peter: data review and metadata review not clearly separated (lots of agreement on that point from the group).

VeeAnn: noted that the revision dates of the checklists (March/April 2014) predate the OSQI IM on data management, data release and metadata (IM 2015-01 through 2015-04): <https://www2.usgs.gov/usgs-manual/95imlist.html>. We need to examine the checklists and, at the very least, bring them in alignment with these IM. NOTE: IM OSQI 2015-03, Section 5A <https://www2.usgs.gov/usgs-manual/im/IM-OSQI-2015-03.html> links directly to the checklists, so we are constrained to revising the checklists individually (we can't combine them, for instance).

Several members of the group shared how they've used the data review and metadata review checklists in their science centers: they've used the checklists as a starting point for creating more specific guidance documents for their particular science centers. Alan created a thread in the Metadata Reviewers Forum where members can share their experiences in adapting the checklists (with encouragement to upload examples of specialized checklists, review templates, etc.): <https://my.usgs.gov/confluence/pages/viewpage.action?pageId=558860218>.

Peter created another thread in the Forum for members to share their thoughts on how the data/metadata review process might be documented for IPDS: <https://my.usgs.gov/confluence/pages/viewpage.action?pageId=558860180>.

Peter: What about revising "Metadata in Plain Language" <http://geology.usgs.gov/tools/metadata/tools/doc/ctc/> so that it is less CSDGM-specific?

VeeAnn: noted that two reviews are necessary -- of data and metadata -- although they can be performed by the same person. She proposed another strategy: use two people. The first would emphasize the data review (but also look at the metadata), the second would emphasize the metadata review (but also look at the data).

Agenda Item 3. 2017 CDI Workshop

Brief discussion at the top of the hour, will continue next time.

Peter suggested considering hands-on training, in one of the following areas:

- Helping metadata reviewers who are new to the USGS
- Strategies for documenting the review process
- Keywords (utilizing controlled vocabularies)
- Strategies for integrating data and metadata reviews
- Sharing useful tricks of the trade

Mon. October 3, 2016, 2pm - 3pm EDT

Proposed agenda:

1. Burning questions? Metadata nightmares? Brilliance to brag about?
2. Keywords in Metadata, a presentation from the USGS Thesaurus Team
3. Next steps?

Notes from meeting:

Question: How can we deal with metadata records that use the EML standard?

- Lisa Zolly: EML metadata will need to be converted to the CSDGM standard, or in the future to ISO. There is a XSL transform to do that; email Lisa directly if you need it.
- Peter Schweitzer: If you're worried about losing information in the format conversion, you can link to the original EML record.

Question: Metadata records are being written by project members who are not USGS employees but are students at a university, so they are unable to authenticate with OME which only uses USGS active directory credentials. Can we get them guest user permission?

- Lisa Zolly: The OME team intends to do that some day, but does not have the resources to do it any time soon. OME relies on Active Directory for authentication; a separate module could be leveraged for external accounts, but the database would need to be built for it, and CSASL would have to dedicate staff to supporting management of non-AD accounts.
- Tom Burley: Could the students sign up as USGS volunteers? (Well, no, that is expensive.) Or use Metavist.
- Aaron Freeman: Could the students use Metadata Wizard in the Esri context? The record could be exported and more information could be added in Metavist.
- Follow-up: Isn't Metadata Wizard being de-coupled from the Esri environment? (That's the hope, but no resources to do it yet. You could import a CSV into a geodatabase, though.)
- NOAA has shut down Mermaid, and EPA has also shut down its CSDGM metadata editor, because both are going to the ISO metadata standard.

Presentation, see Peter Schweitzer's outline linked in the agenda above.

- Peter's concern is how well metadata works in situations where people are using it to find information. Because people often don't know what to ask for, what to call it, or who to ask, Peter was led to the use of controlled vocabularies.
- He prefers keywords that say what the data are, instead of those that say what purposes the data could be used for.
- He cautions that names can be interpreted in different ways, so more keywords are necessary to clarify what the data are.

Lisa Zolly shared the list of USGS Thesaurus terms that is being used in the USGS Science Data Catalog to provide a browse interface. SDC also allows full-text searches of metadata records, with some fields being weighted more heavily than others. As more metadata records provide one of the keywords on the browse list, the interface will be quicker and better.

General tips for metadata reviewers:

- Some accurately spelled keywords from accurately identified thesauri are an important part of good metadata.
- Metadata keywords should include some general terms that allow people to narrow down their search, and also specific terms that allow people to rule out the data sets that are not what they need, and rule in data sets that might be what they need.
- USGS has some tools to help reviewers compare keywords to thesauri, and more thesauri could be added to them.
- If you want your data to show up well in the USGS Science Data Catalog, make sure there is a keyword from Lisa's list.

First meeting, Mon, September 12, 3pm – 4pm EDT (after the CDI Data Management Working Group meeting).

Proposed agenda:

1. Burning questions? Metadata nightmares? Brilliance to brag about?
2. Our Community
    a. Focus: Review of USGS metadata
    b. Community: Share knowledge, questions, and puzzles
    c. Knowledge: Develop, share, and maintain know-how for review of USGS metadata
3. Community Resources
    a. Confluence Site: Member list, link to training, examples, discussion "forum"
    b. Data Management Website
4. Next steps?
    a. A session on keywords?
    b. Distinguish clear USGS requirements from matters for criteria and considerations?
    c. Help desk? Monthly meetings? Review & revise online checklist?

Notes from meeting:

Question: Will there be a similar group for data review, or does this group include data review?

- We agreed to expand our scope to include data review, especially technical aspects of data release such as packaging, format, and documentation.
- We agreed that a good metadata review requires looking at the data to ensure that the metadata represents it correctly.
- USGS policy requires two reviews (metadata and data) and not two separate reviewers, but this is a minimum standard. Science center directors can require additional reviews before they approve the data release. Our group agreed that two different reviewers looking at a data release would be a good thing for ensuring quality, and for high-profile datasets more than two might be good – we might want to think about defining levels of review. The Alaska Science Center has a scientist peer review the data content and a data manager review technical aspects including metadata.

Observation: As USGS implements the new policy with unprepared reviewers, it's almost inevitable that some "horror story" data will be released that will be embarrassing. It would be good for us to keep our ears to the ground – who needs help in reviewing data and metadata?

Issue: Metadata writing and reviewing are a significant time investment. How can we help our scientists and managers plan realistically?

- We could "pass around notes" about how long it takes, producing a community estimate that could be shared more widely.
- Data management plans will be helpful when they are required.

- Metadata reviewers tend to also become metadata counselors, helping new metadata writers avoid difficult and time-consuming approaches, and even providing training.
- Another way of helping research projects get started with metadata is to provide templates customized with appropriate contacts and disclaimers, which simplifies the project's work, helps standardize their metadata, and makes review easier.
- We agreed to enlarge the scope of our community to include metadata counseling, training, and resources.

Future community meetings.

- We agreed to meet monthly at 2:00 Eastern Time on the first or third Mondays of the month. (This is the same time of the week as the CDI Data Management Working Group and the Science Center Points of Contact for the new policies, but different weeks.) If we meet on the first week, the third week might be used for subgroup meetings.
- We agreed to have a session on keywords.
- We can post possible discussion topics on the confluence forum.

Future community activities.

- Similar to the library of recommended disclaimers, we could recommend wording that can be used in metadata records for referring to information that is documented somewhere else, for example in data dictionaries, "techniques and methods" publications, or NWIS documentation.
- We could revise the checklists for data reviewers and metadata reviewers on the USGS, incorporating the separate list that VeeAnn and Peter provided as part of their training.
- We will start a confluence forum topic on recommended resources, and start it off with the "green workbook" which several people recommend.

A question was raised about metadata for a geodatabase that includes multiple data sets. The discussion was diverted to one about acceptable data release formats for GIS data. SDTS has been withdrawn with no replacement, geodatabases are proprietary, shapefiles are said to have problems with spatial reproducibility. The discussion will continue on the forum page on our confluence site. The larger question is how we as reviewers should advise authors about data distribution packaging (convenience, clarity, longevity).