LAS 6292: METADATA

updated: 2023-02-28

Pre-Class Preparation (Instructor):

Send in an email to students:

- content of any pre-class emails.

Bring to Class:

- Snacks
- Flip charts and markers
- Dry write markers
- Tent cards for student names

Objectives and Competencies:

- Be able to explain what metadata are and why they are important
- Identify & list the types of information typically included in metadata records
- Identify reasons metadata is of value to data users, creators, and organizations
- Be able to identify metadata standards relevant to different disciplines
- Explain the elements of good metadata

Pre-class Preparation (Students):

Readings

1. Michener, W.K., et al . 1997. Non-geospatial metadata for the ecological sciences. Ecological Applications 7: 330–342.
   [read online] [download pdf]
2. Pp 446-450 in Bernard, H.R. and Bernard, H.R., 2013. Social research methods: Qualitative and quantitative approaches. Sage. [[download pdf]]

Online Lectures:

- Metadata

Class Outline

Topic 1 Overview: ——- (10 min)

Topic 2 Overview: Topic (10 min)

Metadata Review

Data repositories such as Dryad and ICPSR are designed to permanently store the data thatused in research so it is available to future scholars. To use the data, it is important to have good metadata….but how good are the metadata, really?

Go to either Dryad or ICPSR and download 2-3 datasets. Now review the data and metadata. Based on the information provided, could you explain what the abbreviations are? How the data were collected? What the values represent? Could you recreate the analysis? Is there anything missing or that stands out?

- Here is an example from Dryad: The Page "Data from: Resilient networks of ant-plant mutualists in Amazonian forest fragments" includes an overview of the project and dataset, along with some other information. If you click "Download Dataset" you will get a zip file with {{% staticref "course-materials/class-sessions/07-metadata/sample_metadata/Ant.Plant.Mutualists.Fragmentation.csv" "newtab" %}}[the data in .csv format]{{% /staticref %}} and {{% staticref "course-materials/class-sessions/07-metadata//sample_metadata/Passmore_etat_README.txt" "newtab" %}}[the Metadata in .txt format]{{% /staticref %}}.

Breakout & Return Results

Breakout (15 min):

1. Review the Metadata of Others. Did they include the 5 metadata descriptors?

Returning results & Take-home message (35 min)

1. *Take-home message:* Metadata are hard to write but invaluable

anything before the break? (10 min)

Breakout 2: After the Break (10 min), In-class Assignment (45 min)

Metadata Templates

Today's session is an opportunity to start drafting the metadata for your project. Although there are links in the notes for today's sesssion to tools that will build your metadata in machine-readable XML schema, for this class (and maybe even in most cases) a .txt or .Rmd file with information on the relevant Class Descriptors (*sensu* Michener *et al.* 1997) is all you need sufficient. To save you time, I have created metadata templates based on information from ICSPR (for social sciences) and Michener *et al.* (for biophysical sciences) that you can download and edit; you can add more fields or delete any that are not relevant. Note that Table 1 in Michener

et al. is much more comprehensive and provides additional guidance on how to make sure the metadata are useful.

1. Download the following templates: Click the link for your preferred format (.xtx or .Rmd) and save the file in the RStudio project you've created for your Course Project. The .txt version can be opened and edited in any word processor, a text editor, or in R. The .Rmd file is an R Markdown Document.

a. **Metadata Template for Social Sciences:** {{% staticref "course-materials/class-sessions/07-metadata/ICPSR_Metadata.txt" "newtab" %}}[.txt format]{{% /staticref %}} or {{% staticref "course-materials/class-sessions/07-metadata/ICPSR_Metadata.Rmd" "newtab" %}}[.Rmd format]{{% /staticref %}} (from ICSPR; see note below for additional info on Qualitative Data).

b. **Metadata template for Biophysical Sciences**: {{% staticref "course-materials/class-sessions/07-metadata/Michener_etal_1997_Table1.txt" "newtab" %}}[.txt format]{{% /staticref %}} or {{% staticref "course-materials/class-sessions/07-metadata/Michener_etal_1997_Table1 "newtab" %}}[.Rmd format]{{% /staticref %}}(Table 1 from Michener et al. 1997).

Note for researchers in the Humanities or those working primarily with Qualitative Data: The metadata required often depend on the type of material with which you work (e.g., oral history, photos, digital, printed). If your data is in this domain, you can download templates here: Template No.1 is from UF's Samuel Proctor Oral History Project, Template #2 is a more general one from the UF Humanities Archives. You can also review the metadata required by the Qualitative Data Repository.

2. Choose the template that is most appropriate for your discipline, then review both templates. Is there metadata from the other one would be useful to include in yours? If so, copy the items over and save the revised file with a new (correctly styled) name.

3. Start filling out the metadata requested in the template. You might want to begin by making notations on the ones for which you will have to present the range of possible values, units, the names / brands / models of equipment used to make or record measurments, etc.

4. **Submission:** NONE. This is a component of the final project, so the goal for today is to jump-start your work and to realize that preparing a good metadata file takes longer than anyone anticipates.

Note:Be sure to check the Notes for todays topic - they include excellent resources for preparing metadata.

Free Time

There are 30 min remaining that can be used to —–

Sources for Today's Session

1. ICPSR *Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle (6th Edition)*
2. DataONE Community Engagement & Outreach Working Group (2017) "Metadata Management". Accessed through the Data Management Skillbuilding Hub at https://dataoneorg.github.io/Education/lessons/07_metadata/index on Aug 31, 2020

Resources

Codebooks

This will help you ensure that including all the relevant information in your codebooks.

1. ICPSR: What is a Codebook?

Discipline-Specific Metadata Standards

It is worth looking in these catalogs to see if you can find metadata standards for your discipline. They will provide suggestions on not only what to include, but the standard vocabulary for your discipline.

1. RDA Catalog of metadata standards for different disciplines
2. UK Digital Curation Center Directory of metadata standards for different disciplines
3. Ecological data: Ecological Metadata Language
4. Museum Specimens: Darwin Core
5. Geography Markup Language (GML): Emphasis on geographic features (roads, highways, bridges)
6. Humanities: UF Digital Collections (UFDC) key metadata fields used for non-published items such as posters, archival materials, artists' files, field notebooks, etc. Includes a link to a template you can download. See also Template No.1 from UF's Samuel Proctor Oral History Project, Template #2, which is a more general one from the UF Humanities Archives, and the metadata required by the Qualitative Data Repository.

Tools for Creating machine-readable metadata

I include these here in case you want to try using them. It's not required, but it could definitely make your life easier if there is a standard tool for your discipline (e.g., MORPHO if you are working with ecological data).

1. giant list from the RDA of tools for creating standardized metadata for different disciplines
2. USGS Metadata Wizard

3. TKME
4. CatMDEdit
5. GRIIDC

Metadata Dictionaries

1. USGS
2. Global Change Master Dictionary
3. USGS Geographic Names
4. Getty Thesaurus of Geographic Names

Organizations Developing Metadata Standards and Schema

1. The Research Data Alliance (RDA)"has the goal goal of building the social and technical infrastructure to enable open sharing and re-use of data."
2. DDI Alliance: "Established in 2003, the Data Documentation Initiative Alliance (DDI Alliance) is an international collaboration dedicated to establishing metadata standards and semantic products for describing social science data, data covering human activity, and other data based on observational methods."
3. The Dublin Core Metadata Initiative is "an organization supporting innovation in metadata design and best practices across the metadata ecology".
4. "Best Practices in Creating Social Science Metadata." p.32 in the ICPSR *Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle (6th Edition)*.