Position Statement

July 26, 2019

Behind the Scenes:

Metadata and Data Visualization

Jo Klein

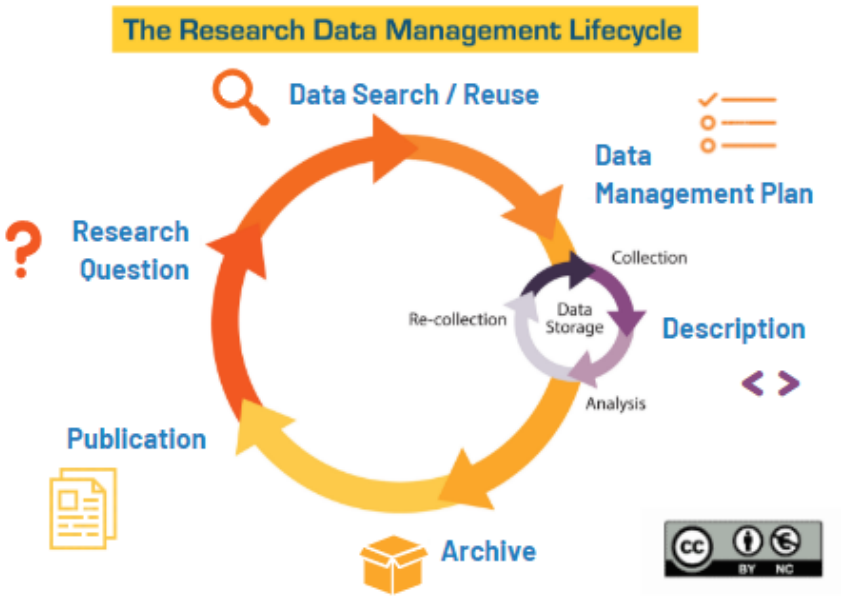
Geospatial and Data Visualization Librarian
UNC Greensboro University Libraries

Background

Creating a good quality data visualization requires a well-thought-out design process. The steps of that design process are debatable, and vary depending on who you ask or what blog or guide you're referring to, but they usually include a few **key guidelines**:

- Know your question, audience, and data
- Select a visualization/chart type and development tool
- Pick out the key message using color, shape, scale, and other visual indicators
- Publish and share your final product

These steps can take place at multiple points during the research process (illustrated by the Research Data Management (RDM) Lifecycle diagram to the right). While most would decide that a certain type of chart best illustrates already-collected data during analysis or publication, you can also plan to collect or find data for a specific type of visualization during the research question and planning stages.



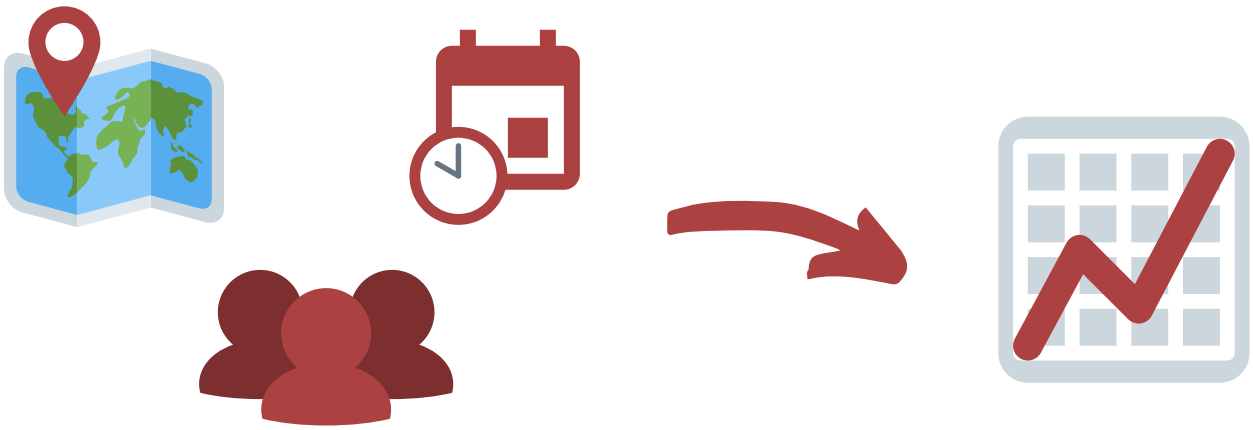
Adapted from: The University of California, Santa Cruz, Data Management LibGuide, Research Data Management Lifecycle, diagram, viewed 26th July 2019 <http://guides.library.ucsc.edu/datamanagement>

Where does metadata fit in all this?

The answer is **everywhere and anywhere**.

In addition to being required by many research-funding agencies, metadata is essential to making data Findable, Accessible, Interoperable, and Reusable (FAIR) and promoting sharing and open use of existing data across the research lifecycle (1,2). This extends to the visualizations we make using that data.

Most importantly, metadata adds **context** to data. Creating a good quality visualization requires an understanding of the data being used; i.e. where did the data come from, when was it collected and under what circumstances, what is its format, what geographic regions does it encompass, are there copyrights or other rules for reuse, who do we contact for more information, etc. Good quality metadata can tell us this information and more (3).



Where do libraries come in?

In order to use metadata, it has to be there. Producers of data and data-derivative works (which includes data visualizations) need to capture accurate and adequate metadata to provide context to their work. Libraries, by nature of being "in the business of preserving, providing access, and supporting faculty research," are at the perfect place to both support and get buy-in from data-producers through instruction and outreach and their role as central hubs of information and social activity (4). Many libraries, especially those at research/academic institutions like UNC Greensboro (UNCG), already offer support and services for research data management, data discovery and storage/archiving, and scholarly communications, all of which involve metadata in some form (4).

Research Areas (and what we've found so far)

Challenges and Areas of Need

What challenges do researchers and data visualization creators/users face involving metadata? What areas need the most support?

Data visualizations share characteristics of (and common metadata fields with) data, images, and digital objects, but don't completely fit into any one category. This is helpful in some ways, as existing standards and learning resources (such as information literacy, data literacy, and visual literacy frameworks) can be easily adapted for data visualization, but there are drawbacks.
Few, if any, metadata standards or extensions have been developed specifically for data visualizations and with the needs of the data visualization creator or user in mind. Having such a standard would help creators and users create, share, find, attribute, and use visualizations, but with discussion of the following challenges that should be addressed:

- What information fields a researcher versus libraries/repositories and down-stream users consider to be valuable can differ, leading to missing vital information or redundant information after the data has been shared (5). In addition, different metadata creators have different ideas of future use and needs, and will have different ideas about what needs to be recorded (6).
- Metadata can bring extra challenges, as observed previously with visual/image objects (7):
  - Maintaining links between metadata and objects can be difficult
  - Identical metadata for different objects and other quality issues can complicate the search/find process
  - Will researchers/creators feel too much time is being spent on metadata capture?
- Metadata requirements will change depending on where data is shared and stored; researchers will need a thorough research data management plan to determine what metadata to capture over the research process.
  - Cloud storage (such as Google Drive and Dropbox), hard drives or CDs, and other non-repository methods of sharing data don't allow for as much embedded metadata as repositories with required schemas and extensions do. Metadata shared this way usually has to be moved with the data file (or visualization file/object) as a separate "readme" file and can be easily unlinked and lost. Some creators share associated metadata (including code) using associated

GitHub sites, such as the visualizations and apps in the Shiny from R Studio gallery.

- Out of thirty UNCG faculty and research staff that indicated they shared data through cloud methods, email, CDs, and other methods, only one used an "online data deposit" or repository (8) that allows for more detailed description using metadata. It's unknown how these respondents shared associated metadata, or whether they did at all.
- The example data visualization above contains information about the source of the data, but doesn't have any other associated metadata, for example when the visualization was created, who created it, or why. For a visualization created for and published in a position statement, this may be sufficient. For a visualization shared through a repository like Figshare and created using open-source code with a Creative Commons Attribution License, it may not.

Metadata Standards, Schemas, and Tools

Are there any standards and schemas that already exist that can be applied to data visualization? What repositories accept data and data visualization, and what metadata fields do they record?

- Some data repositories and other research communication services support and have established metadata schemas for data (including geospatial data) and visual resources like data visualizations and maps, including Figshare, Esri Living Atlas, and UNC Dataverse.
- There are many existing metadata schemas for field-specific data description, but few for geospatial data and digital objects. There are three directly-associated standards in the RDA Metadata Standards Directory.
  - ISO19115 for geographic information and services
  - Common Information Model (CIM) for climate data and the models and software that are used to produce and visualize them.
  - PREMIS for the preservation of digital objects

- Some groups, such as the OCLC Research Library Partnership Metadata Managers Focus Group, the Big Ten Alliance (BTAA) Geospatial Data Project, and government agencies like the USGS are working on standardized metadata schemas and issues with geospatial visualization, as well as overall challenges involving metadata.

3 Instruction, Services, and Advocacy

Are there existing instructional materials or guides that cover metadata, data visualization, or both? What gaps exist and how can they be bridged?

- In 2017 the OCLC Research Library Partnership Metadata Managers Focus Group published a page on metadata advocacy, which contains a list of strategies and topics to use to advocate for metadata and research data management support and services.
- In addition, finding motivators for visualization creators/learners to learn metadata and RDM through needs assessments and surveys will be helpful: for example, incorporating "real-world case studies and authentic project involvement" into instruction (9).
- Instruction and services on metadata, RDM, and data visualization (including geospatial data) as separate topics are already available from some university libraries and library-adjacent organizations like the Association of College and Research Libraries (ACRL), visualization tools/companies like Esri, Tableau, and Google, and standard-setting authorities and committees. For example:
  - The Content Standard for Digital Geospatial Metadata Workbook from the Federal Geographic Data Committee
  - The Learn Data Viz guide and research Guide to Data Visualization from UNC Chapel Hill University Libraries
  - The Data Science and Visualization Institute at NC State.

What's Next?

Literature Review & Environmental Scan

Conduct a literature review and environmental scan on the intersection of metadata and data visualization, especially in relation to library services.

Needs Assessment - Fall 2019

Invite faculty, researcher staff, and students to give feedback on metadata and data visualization services at UNCG and in the broader Visualizing the Future Symposium community.

Metadata Development - Ongoing

Develop and advocate for adoption of metadata standards/schemas for data visualization, through outreach to standards authorities and groups offering repository and sharing services, such as Figshare and institutional repositories.

Instruction & Outreach - Ongoing

Create instruction and outreach materials (including infographics, video tutorials, and guides) and provide instruction sessions (such as workshops and webinars). Utilize existing resources and supplement as needed.

Follow-Up

Conduct follow-up surveys of workshop attendees and previously surveyed groups to evaluate implementation of findings.

Bibliography

1. FAIR Principles & Digital Objects: The role of METADATA | Agricultural Information Management Standards (AIMS). (n.d.). Retrieved July 26, 2019, from http://aims.fao.org/activity/blog/fair-principles-digital-objects-role-metadata
2. Hagstrom, S. (2014, September 3). The FAIR Data Principles. Retrieved July 23, 2019, from FORCE11 website: https://www.force11.org/group/fairgroup/fairprinciples
3. Smith-Yoshimura, K. (2016, April 18). Metadata for research data management. Retrieved July 23, 2019, from Hanging Together website: https://hangingtogether.org/?p=5616
4. Craft, A., & Kellam, Lynda. (2017, February). Under Construction: Building Research Data Services at UNCG. Presented at the Empirical Librarians Conference, NC A&T State University, Greensboro, NC.
5. Data Justice Research | Detroit Community Technology Project. (n.d.). Retrieved July 23, 2019, from https://detroitcommunitytech.org/datajustice
6. Dangerfield, M.-C., & Kalshoven, L. (2015). Report and Recommendations from the Task Force on Metadata Quality (p. 54). Retrieved from Europeana website: https://pro.europeana.eu/files/Europeana_Professional/Europeana_Network/metadata-quality-report.pdf
7. Smith-Yoshimura, K. (2015, April 9). Managing Metadata for Image Collections. Retrieved July 25, 2019, from Hanging Together website: http://hangingtogether.org/?p=5130
8. Crowe, K., & Crumpton, M. (2014). Defining the Libraries' Role in Research: A Needs Assessment; A Case Study. Proceedings of the 2014 Library Assessment Conference.
9. Qin, J., & D'ignazio, J. (2010). The Central Role of Metadata in a Science Data Literacy Course. Journal of Library Metadata, 10(2–3), 188–204.