1-1-2017

A Case Study for Organizing and Documenting Research Data

Qianjin Zhang
University of Iowa

Sara Scheib
University of Iowa

Zhang, Q. & Scheib, S. (2017). A Case Study for Organizing and Documenting Research Data. In L. Johnson (Ed.), *Curating Research Data, Volume Two: A Handbook of Current Practice* (pp. 18-20). Chicago, Illinois: Association of College and Research Libraries.

Pre-Deposit Interview in Action

Qianjin Zhang and Sara Scheib present a case study on how academic librarians conducted in-depth interviews with faculty researchers to better understand their data management needs.

A Case Study for Organizing and Documenting Research Data

*Qianjin (Marina) Zhang and Sara Scheib**

PROJECT SUMMARY

The University of Iowa Libraries, like many other academic libraries, is dedicated to the development of research data services. From 2012 to 2014, we ran an environmental scan on campus data management resources that assessed research data needs and identified gaps in existing research data services.[3] Based on the outcomes of the environmental scan and assessment, we formed a working group to create a guide for researchers and campus service providers on data management resources in August 2014. To explore good practices in research data services and establish a long-term partnership with researchers, faculty, and students, two subject specialist librarians from the group started a case study for organizing and documenting research data.

In consideration of the researcher's specific data generation, collection, and analysis methods, we developed a data organization scheme based on the ISA-Tab metadata standard and applied it as the data was migrated to secure storage. As a result, the organized data is not only more accessible to the researcher, her students, and collaborators but also easily understood by newcomers to their laboratory.

CHALLENGE AND PROBLEM

In this case study, we worked with a researcher in the field of chemical and biochemical engineering and her graduate students to help them organize and document their research data resulting from an NSF-funded project. We conducted

in-depth interviews to better understand their data management needs. The researcher's data was stored on a variety of digital media, including personal computers, laboratory computers, commercial cloud-based storage, removable flash drives, and hard drives. Recently, our university's information technology unit began providing access to secure storage via remote server access for researchers, faculty, and students. The researcher contacted us to help prepare their data for migration to the secure storage and therefore organize and document their data in the process. In preparation for the first interview, we asked to review a sample of their data and the data management plan as submitted to the NSF.

In November 2014, we conducted our first in-person interview with the researcher and her graduate students. This conversation helped us understand her research project and their data process throughout the data life cycle. The interview questions were based on a questionnaire from the 2014 University of Iowa Libraries' report on campus research data needs.[4] The results of our first interview suggested that they needed a data organization scheme and an appropriate metadata standard for their data.

APPROACH AND SOLUTION

After the first interview, we analyzed how the researcher and her graduate students did research. Their research project included multiple small projects. Each small project employed several types of instruments and technologies. We investigated existing metadata standards listed on Digital Curation Centre (DCC) and found the ISA-Tab metadata standard. The ISA-Tab metadata standard stands for the Investigation/Study/Assay tab-delimited format metadata standard.[5] In this standard, the investigation, study, and assay are organized in a hierarchical structure. This structure fits well with their data flow: the investigation corresponded to their research project, the study corresponded to a small project, and the assay corresponded to an instrument or technology employed.

A few weeks later, we conducted a second interview with the researcher and her graduate students to ensure that the ISA-Tab metadata standard could be adapted to meet their data organization and documentation expectations. In consideration of their specific laboratory equipment, materials, and measurements, we selected particular elements from the ISA-Tab metadata standard, customized these elements, and developed a data organization scheme in a hierarchical outline format as well as a visual map format (see figure 1.3).

Over the next two months, the researcher and her graduate students created their folder structure (see appendix 1.0 A on page 45) and sorted their files based on the data organization scheme. We also recommended some best practices on file-naming conventions, file version control, and data documentation. We answered clarifying questions about the data organization scheme via e-mail.

After completing this process, the researcher and her graduate students gave us positive feedback that they were satisfied with the data organization scheme and best practices on data documentation. In future projects, it may be wise to conduct a final wrap-up meeting with all participants at the conclusion of the project to solicit more detailed feedback.

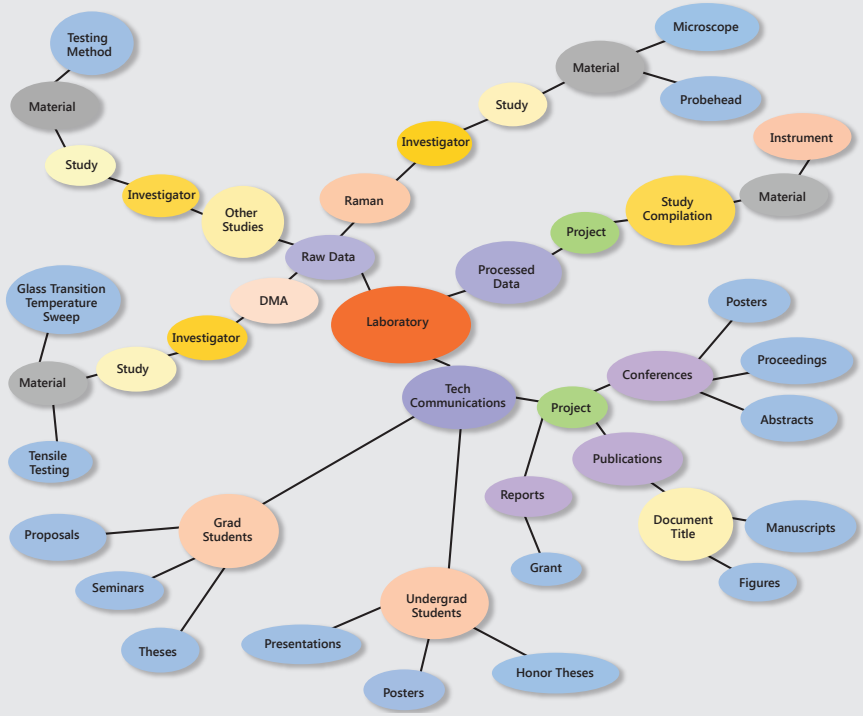

FIGURE 1.3
Data organization scheme.

FUTURE WORK

We plan to share the data organization scheme and best practices on data documentation with other researchers in the same field and apply them to other fields. This case study can be used as an example to help researchers use best practices to manage their data throughout the data life cycle.