Developing a Data Management Plan (DMP) in the Cognitive Sciences

Jessie Cass
Simmons College, Graduate School of Library & Information Science

Follow this and additional works at: https://escholarship.umassmed.edu/escience_symposium

Part of the Cognitive Psychology Commons, Communication Sciences and Disorders Commons, and the Scholarly Communication Commons

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 License.

Repository Citation

Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 License. This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
OBJECTIVE: To experience the process of using principles of scientific research data management (SRDM) to work with a researcher to create a data management plan (DMP). SRDM is an area where research in the traditional sciences intersects with information science. SRDM guides researchers through all stages of the data life cycle. A DMP is a document explaining how a study will progress through the data life cycle that is increasingly required by research funders. This project was undertaken as part of a class on SRDM through the Simmons College School of Library and Information Science. METHODS: After corresponding via email with a researcher studying the cognitive and linguistic skills of deaf children with autism, a set of questions was created based on an interview instrument developed by the Digital Curation Centre and a Skype interview was conducted. Using the information gathered during the interview and in follow-up emails, as well as knowledge of SRDM principles learned in class and through independent research, a DMP (following National Science Foundation guidelines) was created. Additionally, aspects of the researcher’s study which proved challenging when creating a DMP were identified. RESULTS: A seven-part DMP was created. Challenging aspects were identified as a set of teaching points. These included: data being collected via video camera; children as subjects; subject IDs; repository requirements. CONCLUSIONS: This project was successful in teaching both this author and the interviewed researcher about SRDM and DMPs. This will improve the cognitive science community’s understanding of the principles and importance of SRDM.

Challenges Throughout the Data Cycle

All of the subjects in the study were children. This complicated the researcher’s ability to share gathered data. The researcher had the parents of subjects sign consent forms, however these did not give the researcher permission to submit the data to a repository. If the researcher had been looking at the whole life cycle of the data from the beginning, the wording of the consent form could have been different.

Preserving Data

One issue the researcher had early in the data life cycle was with the naming of files and folders. To maintain confidentiality, the researcher needed to use IDs, rather than names, when referencing subjects and naming files. In this case the researcher used the sex of the subject, the relevant diagnosis of the subject, and the chronological number of the subject to make an ID. This was problematic because it would be possible for someone to figure out which subject went with which ID, and therefore which data, as long as they knew the order in which data were taken and the sex and diagnosis of the subjects. A common alternative solution is to use a number randomizer to create IDs that are unrelated to any aspect of the subject. These IDs are written in a protected spreadsheet in order to keep track of which subject has which ID.

Pictured above is a version of the Data Life Cycle. This is a concept used in data management to show what happens to data from creation forward. A DMP addresses how the researcher will deal with their data throughout each of the steps. This requires the researcher to give some thought to their data and its management, which usually results in data of better quality.