Developing a Data Management Guide for Researchers

John Borghi, Stephen Abrams, John Chodacki, Daniella Lowenberg, Stephanie Simms
University of California Curation Center (UC3)
California Digital Library

Developing our Guide

Researchers are faced with an evolving array of expectations related to how they manage and share data. Academic libraries are well positioned to provide support related to research data management (RDM) owing to their extensive expertise in preserving and curating information¹. Unfortunately, researchers and RDM service providers think and talk about data in very different ways and RDM assessment tools are often several steps removed from a researcher's day-to-day practice.

To address these difficulties, we are developing a customizable set of tools that describe data-related practices using language and terminology familiar to researchers.

This project consists of two related elements, a customizable rubric intended to enable researchers to self-assess their data-related practices and a series of one page guides designed to provide actionable information regarding requirements, standards, and best practices.

RDM Rubric: Building upon maturity-based assessment tools², the rubric is intended to help researchers self-assess how they manage their data over the course of a research project.

Guides: Drawing from the data curation literature, the guides are intended to provide researchers concrete information about how to advance their data-related practices.

The Structure of our Guide I have detailed plans related to my data. I have some plans My plans are When it comes to my I have plans related Planning related to my data, iterative and any data, I have an to my data that are but any related informal "way of related optimized to Planning for Data documentation is doing things", but not streamline sharing or documentation written before or a documented plan. publishing. informs my after the fact. subsequent research process. **Organizing Data** Collection Ad hoc, Non-reproducible Optimized, Standardized Saving Data Planning for Data plan detailing how you'll manage your data, code, and other research material There is a lot more to saving data than ensuring you have appropriate backups. How and where you save your data depends on a number of factors including the size and (including physical samples) over the course of a project will help your research Preparing Data nalysis onsider when saving your data When possible, save multiple copies of your data across a variety of torage mediums. Hard drives, cloud storage, and other options have fferent levels of reliability, but all will eventually fail or become document called a Data Management Plan (DMP) that provides details about the type o data to be collected and managed within a research project as well as the individuals Saving data takes time, but losing data wastes time. Backing up data Þ should be a regular part of your research practice, but you should also The DMPTool (https://dmptool.org/) is a free tool that provides guidance for creating a have a plan for how data will be saved after your research is concluded data management plan. The RDM team at your institution also provides DMP-related Analyzing Data volve saving data in open file formats, using version control, or simply toring your data alongside the documentation and other research · Plans should identify the data you intend to collect, as well as how you plan to example, if your data contains participant names or other personally identifying information, it will need to either be de-identified or stored in a secure system like A plan is really only useful if people can follow it. Be sure your plans are • Even if you do not have a Data Management Plan (DMP), you may have a The characteristics of your data determine how much flexibility you will have about where it can be saved. If you have large quantities of data or data document that describes your . This could be included in a lab protocol, an IRE Sharing Data While planning for data you may encounter unfamiliar terms or familiar terms code, etc) needed to make sense of or use that data. used in unfamiliar ways. If you have any questions, do not hesitate to reach out

- 1. Phases of the research project: Terminology drawn from a review of research data lifecycle models.
- 2. Activities within each phase: Terminology drawn from an informal survey of how researchers actually describe their work.
- 3. Declarative statements: Intended to allow researchers to self-assess the maturity of their data-related practices.
 - 4. Each row of the RDM rubric is tied to a guide that provides information about requirements, standards, and best practices.

A Call to Action

As we develop our tools, we are seeking feedback from (and encouraging collaboration with) librarians, researchers, and other research data management stakeholders.

If you would like to receive information about the development of our tools, provide feedback, or discuss collaborating on institution or discipline-specific versions, please get in touch with John Borghi at John.Borghi@UCOP.edu.

Works Cited

- 1. Cox, A. M., & Verbaan, E. (2016). How academic librarians, IT staff, and research administrators perceive and relate to research. *Library & Information Science Research*, *38*(4), 319-326.
- 2. Crowston, K., & Qin, J. (2011). A capability maturity model for scientific data management: Evidence from the literature. *Proceedings of the American Society for Information Science and Technology, 48*(1), 1-9.



