

# Data Curation Network

*A network of expertise model for curating research data in academic digital repositories*

**Project Name:** Planning the Data Curation Network (Alfred P. Sloan Grant Number: 2016-7044)

**Partner Institutions:** University of Minnesota (lead), Cornell University, Penn State University, University of Illinois at Urbana-Champaign, University of Michigan, and Washington University in St. Louis.

**Duration:** May 2016 - April 2017

**Goals:** This multi-institution planning grant will develop a “Data Curation Network” model that aims to enable academic institutions to better support researchers that are faced with a growing number of requirements to ethically share their research data. Data curation services are currently provided by expert staff at each of our institutions to prepare digital research data for open access and reuse. Our goal is to demonstrate that a ‘network of expertise’ model for data curation services will enable academic libraries to collectively, and more effectively, curate a wider variety of data types (e.g., discipline, file format, etc.) that expands beyond what any single institution might offer alone. With input from academic researchers from a wide range of disciplines our aim is to determine how this model might be effectively implemented, assessed, and sustained.

**Need:** Researchers are required by many federal and private funders and publishers to make the digital data underlying their research openly available for sharing and reuse. However, in order for data to be fully publicly accessible to search, retrieve, and analyze, most data require that specialized curatorial actions be taken to best prepare them for reuse including:

- quality assurance,
- file integrity checks,
- documentation review,
- metadata creation for discoverability, and
- file transformations into archival formats.

Due to the heterogeneous and multidisciplinary nature of research data generated in our nation's academic institutions, the skills and expertise required to curate data for optimal reuse cannot reasonably be provided by the few experts that may reside at each institution. Rather, University of Minnesota research<sup>1</sup> has shown that multiple data curation experts are needed to effectively curate the diverse data types a university typically creates. Local implementation of a small-scale distributed data curation staffing model<sup>2</sup> for the Data Repository at the University of Minnesota has proven to be a workable solution.

**Approach:** Supporting researchers with data curation is an important role that academic

---

<sup>1</sup> The full report and recommendation from this pilot is published as “[A Workflow Model for Curating Research Data in the University of Minnesota Libraries.](#)”

<sup>2</sup> This model was successfully instituted in November 2014 and includes 5 data curation experts each covering data curation needs for data in the physical sciences, health sciences, social sciences, spatial data, and digital humanities data. See “[The Supporting Documentation for Implementing the Data Repository for the University of Minnesota \(DRUM\): A Business Model, Functional Requirements, and Metadata Schema.](#)”

# Data Curation Network

*A network of expertise model for curating research data in academic digital repositories*

research libraries aspire to fill as we transform our workforce to assume greater digital stewardship responsibilities in the academy. Libraries are experts at identifying, selecting, organizing, describing, preserving, and providing access to information materials, print and digital. And as a cornerstone of the academic institution, libraries are persistent, with a demonstrated and sustainable model for providing services such as collection management, preservation, and access to a broad variety of information. While discipline-specific data repositories (e.g., ICPSR, GenBank) manage largely homogeneous data sets, our general-purpose repositories for data, called institutional repositories (IRs), can accommodate a wider range of data formats and metadata from many disciplines. Given limited resources, it is unrealistic however to expect that every academic library can hire a data curator for every data type (e.g., GIS, spreadsheet/tabular, statistical/survey, video/audio, computer code) or discipline-specific data set (genomic sequence, chemical spectra, biological image) that one might encounter. Similarly, each type of data curation expertise might only be needed occasionally depending on the disciplinary makeup at each institution. This project will conceptualize and develop a “network of expertise” model<sup>3</sup> for U.S. academic libraries to collectively provide data curation services to support digital research data deposit into repositories for open access and reuse. The appeal of a shared expertise model for delivering unique library services has been expressed through recent research<sup>4</sup> and demonstrated in successful projects.<sup>5</sup>

**Outcomes:** This project will deliver a model describing how a Data Curation Network spanning a consortium of U.S. research libraries can be implemented, assessed, and sustained.

1. The model will include an implementation plan for a clear, well-coordinated system that addresses the procedures necessary to handle the curation workload for a wide-variety of data types and formats as well as the challenges of managing a geographically- and institutionally-distributed staff.
2. Baseline measures will be established to forecast the potential workload and staffing needs involved with providing data curation services, using our institutions as a gauge. Understanding the metrics for individually providing data curation services will allow us to assess the efficiency, impact, and the “value add” of a shared data curation service once implemented. Example metrics include: How many data sets are deposited into the local data repository? What types of data are encountered? What skills and expertise are needed to curate these data sets? How much time is involved?

---

<sup>3</sup> [Kirchner et al. \(2015\)](#) define networks of expertise as “... a way to implement and sustain new information services for research.... Through this method, existing organizations will start to change as they integrate experts more fully into the daily work and as a greater number of information professionals share knowledge” (p.16).

<sup>4</sup> For example, R. Erway calls for a collaborative expert network for handling the variety of born digital media in the nation’s libraries in “[Swatting the Long Tail of Digital Media: A Call for Collaboration](#).”

<sup>5</sup> The [2CUL project](#), [DPN](#), [DuraSpace](#), for example.

# Data Curation Network

*A network of expertise model for curating research data in academic digital repositories*

3. During the process, the team will gather input from researchers at each of our institutions, through informal engagement and discussion, to refine our model. Participants at each campus will be recruited to represent a variety of data types and research cultures.
4. Our assessment plan will define ways to assess the cost-effectiveness, efficiency, and demand (both in data variety and skills utilized) for a Data Curation Network model. Key measures of success, such as how the Network will demonstrate value (e.g., reuse metrics of curated vs. uncurated data) will be addressed.
5. A sustainability plan will recommend levels of support needed to sustain the Data Curation Network post-implementation and provide the necessary incentives to grow the network beyond these initial partners (e.g., a membership model that allows for new institutions to join the Network, for example).

**Personnel:** The project team brings together the perspectives of research data librarians, academic library administration, and data curation subject experts from six major academic institutions that each, separately, provide repository and curation services to their campus constituents. Project members include: [Lisa R. Johnston](#) (PI), Research Data Management/Curation Lead, University of Minnesota; [Jake Carlson](#), Research Data Services Manager, University of Michigan; [Robert Olendorf](#), the Science Data Librarian at Penn State University; [Cynthia Hudson-Vitale](#), Data & GIS Services Librarian, Washington University in St. Louis; [Heidi Imker](#), director of the Research Data Service, University of Illinois at Urbana-Champaign; [Wendy Kozlowski](#), Data Curation Specialist, Cornell University; and [Claire Stewart](#), Associate University Librarian for Research and Learning, University of Minnesota. Meanwhile project personnel are expected to seek input from their local experts in data curation.

**Impact:** The institutional partners plan to use this model to implement a fixed-term pilot of the Data Curation Network using the staffing and resources across our six institutions. Participation in the grant may also allow partners to make informed decisions around data curation related activities (e.g., technology, policy, etc.). Additionally, by sharing the results of this project broadly, academic institutions will be better informed to collectively and effectively provide data curation support for researchers.

**Budget:** The Alfred P. Sloan Foundation awarded \$88,725 to this one-year planning grant (Grant Number: 2016-7044) to cover 0.4FTE of the PI's time to serve as project coordinator, honoraria for two invited speakers from successful shared staffing models and one team collaboration expert; travel expenses for two in-person team meetings; travel and event costs to facilitate researcher engagement activities at each institution; and conference travel support to share project findings at relevant conferences.