Data Curation Network: How Do We Compare? A Snapshot of Six Academic Library Institutions' Data Repository and Curation Services

Lisa R. Johnston
University of Minnesota - Twin Cities

Et al.

Let us know how access to this document benefits you.

Journal of eScience Librarianship
putting the pieces together: theory and practice

eScience in Action

Data Curation Network: How Do We Compare? A Snapshot of Six Academic Library Institutions' Data Repository and Curation Services

Lisa R. Johnston[1], Jake Carlson[2], Patricia Hswe,[3] Cynthia Hudson Vitale[4],
Heidi Imker[5], Wendy Kozlowski[6], Rob Olendorf[7], Claire Stewart[1]

[1]University of Minnesota, Minneapolis, MN, USA
[2]University of Michigan, Ann Arbor, MN, USA
[3]Andrew W. Mellon Foundation, New York, NY, USA
[4]Washington University in St. Louis, St. Louis, MO, USA
[5]University of Illinois, Urbana, IL, USA
[6]Cornell University, Ithaca, NY, USA
[7]Penn State University, State College, PA, USA

Abstract

Objective: Many academic and research institutions are exploring opportunities to better support researchers in sharing their data. As partners in the Data Curation Network project, our six institutions developed a comparison of the current levels of support provided for researchers to meet their data sharing goals through library-based data repository and curation services.

Methods: Each institutional lead provided a written summary of their services based on a previously developed structure, followed by group discussion and refinement of descriptions. Service areas assessed include the repository services for data, technologies used, policies, and staffing in place.

Conclusions: Through this process we aim to better define the current levels of support offered by our institutions as a first step toward meeting our project's overarching goal to develop a shared staffing model for data curation across multiple institutions.

Introduction and Background

Funder requirements, institutional and journal data sharing policies, and new trends in research reproducibility signal that academic research will become increasingly more open in the coming years. We, and others,[1] believe that data curation is critical to ensuring that this movement is fully actualized. Our six institutions are beginning to dedicate some level of resources towards data curation services. In doing so we are interested in leveraging our individual progress to contribute to the greater data curation community. The six academic library-run repository services compared here are participants in the Alfred P. Sloan Foundation-funded Data Curation Network project (https://sites.google.com/site/datacurationnetwork). The goal of the Data Curation Network project is to bring together institutions individually providing local support for data repository deposit and curation in order to plan a shared, cross-institutional staffing model for applying expert-level human curation across disciplines than any one institution could offer alone. This assessment captures our current institutional support, which will continue to grow and evolve. This comparison will help the Data Curation Network team design a shared service that fits within the existing scope of our institutions' capacities, yet broadens our ability to curate a wider variety of digital data for researchers than would be available to any individual institution. This assessment is also intended to help others who are at the beginning stages of developing data curation services and are scanning for examples of what peer institutions have implemented. It is not intended to be a scientific comparison or a comprehensive representation of existing data repository and curation services in the field.

Methods

Data curation is a term that is often used to describe a wide range of activities, and the term itself may have different meanings depending on the context and environment in which it is used. In the Data Curation Network, our understanding of data curation is based on the FAIR guiding principles: to prepare and maintain research data in ways that make it findable, accessible, interoperable and reusable.[2] Under this definition, data curation services could include a wide range of possible activities including developing metadata, associating documentation, providing access, or supporting preservation. Data curation services are often provisioned through a data repository as is the case for the current members of the Data Curation Network.

To understand the baseline levels of service currently provided for data repository and curation services, the following six repositories were examined: the Data Repository for the University of Minnesota (DRUM), the eCommons at Cornell University, the Illinois Data Bank at the University of Illinois Urbana-Champaign, Deep Blue Data at the University of Michigan, ScholarSphere at Penn State University, and the Digital Research Materials Repository (DRMR) at Washington University in St. Louis. This is a sample of convenience based on the institutions' involvement with the Data Curation Network project. A project team member from each institution (author) was asked to write a summary report and address specific questions

---

1   See for example, Helena Karasti, Karen S. Baker, and Eija Halkola. "Enriching the notion of data curation in e-science: data managing and information infrastructuring in the long term ecological research (LTER) network," Computer Supported Cooperative Work (CSCW) 15, no. 4 (2006): 321-358; and Principle #7 of the 2016 UK Concordat on Open Data, http://www.rcuk.ac.uk/documents/documents/concordatonopenresearchdata-pdf.
2   Force11, "The Fair Data Principles," accessed January 26, 2017, https://www.force11.org/group/fairgroup/fairprinciples.

(presented here as tables) based on their own knowledge and experience.[3] Following the self-reporting exercise, each team member gave a 20-minute webinar presentation to the project team to further clarify responses. The results of this exercise were captured and described in this report for sharing with peer institutions. This review is a snapshot in time — the six institutional service offerings represented here will change and grow in the future. For the sake of developing a baseline understanding of their practices, this report describes each institution's repository technologies. However, the Data Curation Network is a staffing-focused effort and does not intend to dictate specific technologies or practices taken at our partner institutions. Our goal is to develop a model in which Network curators can work effectively across a variety of similar, but not identical, services. Therefore, this report focuses primarily on which repository and curation services are offered as well as their policy and staffing parameters. Issues around the mechanics of data curation and specific steps taken to prepare data for sharing and preservation, will be addressed in greater depth in future reports by the Data Curation Network project.

Comparisons of Our Six Institutions

The following four sections describe and compare our data repository and curation services.

Section 1.0: overviews for data repository and curation services at each institution are presented along with our workflows and a comparison for how we track curation activity.

Section 2.0: presents and compares the repository technologies used at each institution.

Section 3.0: focuses on policy related to our services.

Section 4.0: assess our staffing, organizational approaches, and provides samples of our position descriptions.

1.0 Services Overview

Each of the six institutions currently provides data repository and curation services and tracks their holdings as either the number of data files or data records (which may hold multiple related files). They do so either as a service of the traditional institutional repository or IR (Minnesota, Cornell, Penn State, WUSL) or via a dedicated data repository (Illinois, Michigan). Although the underlying software and infrastructure may be identical, the service is described as an institutional repository if it is used to collect a variety of research output types, and as a data repository if its scope is limited to data. The intention is to draw focus to the specific needs and demands of a data curation service, rather than to focus on repository practice or services more broadly. All of the repositories make content available on an open access basis, meaning the data housed in these repositories are publicly accessible for search, retrieval, and download.

---

3   See team member bios at Data Curation Network home page, "Who is involved," accessed January 26, 2017, https://sites.google.com/site/datacurationnetwork/people.

Data Repository for University of Minnesota (DRUM)

Institution: University of Minnesota, Twin Cities (Minneapolis, Minnesota)
URL: http://hdl.handle.net/11299/166578
Launched: November 15, 2014
Data Holdings: 103 data records as of January 9, 2017

The University of Minnesota (U of M) Libraries has been providing research data management services for a number of years, including support for writing data management plans, educational training and workshops, and consultation (see http://lib.umn.edu/datamanagement). The Libraries launched the Data Repository for the University of Minnesota (DRUM) in 2014 for U of M researchers to self-deposit their data for long-term open access and reuse when no other discipline-appropriate data repository exists. DRUM resides within the existing institutional repository service, the University Digital Conservancy, as a sub-collection with a custom metadata schema and submission workflow. An example dataset in DRUM is shown in Figure 1. All data submitted to DRUM undergo curatorial review by a data curator who collaborates with the data author to ensure that the data are in a format and structure that meet our policies and best facilitate reuse.

eCommons at Cornell University

Institution: Cornell University (Ithaca, New York)
URL: http://ecommons.cornell.edu
Launched: Fall 2002
Data holdings: 110 data records as of January 9, 2017

The purpose of eCommons is to provide stable, long-term public access to digital content produced by members of the Cornell University community and its sponsored associates. Because policies and submission processes are the same for datasets as other content, our approach to providing open and persistent access to research results is to accept all forms of "scholarly output" in Cornell Library's institutional repository. We encourage use of eCommons for data, particularly when there are no appropriate discipline-based repositories available, or when a researcher doesn't wish to incur a cost associated with their deposit. Data submitted to eCommons are assigned a type "dataset" for discovery purposes, and can be added to the organizational collection of the submitter's choice. Since 2015, datasets must undergo a discovery metadata review, and some receive an additional curation of science metadata and data file format and structure. Most science metadata are submitted as readme files, but standardized metadata are accepted as item files. If a researcher rejects suggestions of the curator, data are still accepted to the repository. eCommons at Cornell University launched in the fall of 2002, and the first dataset was deposited in 2005. An example dataset from eCommons at Cornell is displayed in Figure 1.

ScholarSphere

Institution: Penn State University (State College, Pennsylvania)
URL: https://scholarsphere.psu.edu
Launched: Fall 2012
Data Holdings: 802 public data files as of January 9, 2017

Figure 1: Example dataset in DRUM (left, http://dx.doi.org/10.13020/D6PK5C) and in Cornell eCommons (right, http://hdl.handle.net/1813/43783) which both use the DSpace software.

ScholarSphere is a self-deposit repository service through which faculty, students, and staff at Penn State are able to share their work, including research data sets, on a worldwide scale and be assured of its long-term preservation and thus access. The main impetus behind designing ScholarSphere was to help researchers comply with research data management requirements, as well as with increasing requirements from publishers to link research articles to the data sets associated with them. At the same time, until ScholarSphere, Penn State did not have an institutional repository capturing the scholarly record of its faculty, students, and staff for preservation and access purposes. (There has been an electronic thesis and dissertation service since the mid-2000s, but the University perceived a need for a service to accept a broader array of scholarship — hence, the decision for ScholarSphere to take in both data sets and conventional scholarly publications.) The University also has a stand-alone, mediated-deposit data repository, DataCommons,[4] more specifically geared toward earth and environmental sciences, including geosciences. We connect our researchers to data repositories beyond Penn State as needed via consultation and via a LibGuide for research data management services (http://psu.libguides.com/rdm), which points users to re3data,[5] an online index of data repositories, and to repository services known to accept data sets, such as figshare[6] and Zenodo.[7] Users with deposits in ScholarSphere may access graph visualizations depicting the number of pageviews and downloads for their deposits. Data submitted to ScholarSphere do not undergo any curatorial review, apart from an automatic audit of the files

---

4   DataCommons homepage, accessed August 16, 2016, http://datacommons.psu.edu.
5   Re3data.org homepage, accessed August 16, 2016, http://www.re3data.org.
6   Figshare homepage, accessed August 16, 2016, https://figshare.com.
7   Zenodo homepage, accessed August 16, 2016, https://zenodo.org.

for preservation purposes. However, in some cases researchers request this service. We are also expecting to implement curatorial review for datasets in the future to improve the quality of the data ingested. An example data record in ScholarSphere is shown in Figure 2.

Deep Blue Data

Institution: University of Michigan (Ann Arbor, Michigan)
URL: https://deepblue.lib.umich.edu/data
Launched: February 29, 2016 (soft launch), September 20, 2016 (official launch)
Data Holdings: 50 data records as of January 9, 2017

Deep Blue Data is a repository offered by the University of Michigan Library that provides access and preservation services for digital research data that were developed or used in the support of research activities at U-M. Deep Blue Data is a component of a suite of services provided by the U-M Library designed to broadly disseminate the intellectual contributions in research, teaching and creativity made by the University of Michigan community and to ensure its longevity. It is a companion repository to Deep Blue, which serves to provide access to papers, presentations, reports and other human readable scholarship from the University of Michigan. Our primary goal in providing research data services is to connect researchers to the resources that are best suited to support their specific needs for their data. In cases where subject-based data repositories and services are available that meet a researcher's needs we will consult with the researcher and the repository to assist with the submission process as appropriate. However, researchers in many fields do not yet have a data repository devoted to their needs, or in some situations the disciplinary repository is not a viable option. The Deep Blue Data repository was developed to provide these researchers with the means to satisfy

Figure 2: Example dataset in ScholarSphere (left, https://scholarsphere.psu.edu/files/m900nt50p) and Deep Blue Data (right, https://deepblue.lib.umich.edu/data/concern/generic_works/rf55z7781) both using Hydra (https://projecthydra.org) with Fedora (http://fedorarepository.org).

requirements and take advantage of the benefits that sharing and curating data affords. As we continue to develop the capabilities of Deep Blue Data our intent is to go beyond providing a place to put data and create more of a platform for others to interact with the data in ways that add value. An example data record from Deep Blue Data is shown in Figure 2.

Illinois Data Bank

> Institution: University of Illinois at Urbana Champaign (Illinois)
> URL: https://databank.illinois.edu
> Launched: May 16, 2016
> Data Holdings: 33 data records as of January 9, 2017

The Illinois Data Bank's mission is to centralize, preserve, and provide persistent and reliable access to the research data created by affiliates of the University of Illinois at Urbana-Champaign, such as its faculty, academic staff, and graduate students. The Illinois Data Bank is intended to be responsive to the Illinois research community, is supported by the University of Illinois at Urbana-Champaign, and endeavors to be both durable and sustainable. The Illinois Data Bank is a platform for making datasets created from research projects by University of Illinois at Urbana-Champaign researchers publicly accessible by seeing that the research data is both widely discoverable and linked to associated works, such as journal articles, source code, or data deposited elsewhere. During consultations we may point to alternative repositories and encourage depositors to reconsider if a more appropriate repository is available. We elected to go with development of a web application that interacts directly with our preservation system in order to leverage that system's functionality, and allows us to focus our long-term efforts on centralizing our preservation efforts. Depositing research data into the Illinois Data Bank is voluntary. An example data record in Illinois Data Bank is shown in Figure 3.

Digital Research Materials Repository (DRMR)

> Institution: Washington University in St. Louis (Missouri)
> URL: http://openscholarship.wustl.edu/data
> Launched: January 5, 2015
> Data Holdings: 3 data records as of January 9, 2017

The purpose of the digital research materials repository (DRMR) is to provide a long-term, institutional home for research data and supplemental materials produced at Washington University in St. Louis (WUSTL). A free service of the University Libraries, DRMR curates data and the supporting documentation used to verify or support research, including any analysis scripts, data dictionaries, and domain metadata. The DRMR at WUSTL is a companion collection within our institutional repository, Open Scholarship, which serves to provide access to dissertations, theses, and other scholarly output of the university. DRMR provides a data archiving solution for anyone in the WUSTL community who does not have an appropriate discipline or domain repository available to them, or does not want to incur the costs of deposit. Once submitted to DRMR, datasets and submitted materials undergo archival processing and curation treatments. Curators work directly with WUSTL researchers to enhance records and documentation for reuse and accessibility. An example data record in DRMR is shown in Figure 3.

Figure 3: Example dataset in the Illinois Data Bank (left, https://doi.org/10.13012/J8PN93H8) using a custom-build Ruby on Rails application and the Digital Research Materials Repository (right, https://doi.org/10.7936/K7J67F60) using Digital Commons by BePress.

1.1 Data Curation Workflows

The comparison of curation workflows (illustrated in Table 1) demonstrate how a "dataset" typically flows through the curation process prior, during, and post-ingest to the local repository and curation services offered by the six institutions. Some columns in Table 1 were not used by any of our institutions but are included here as alternative or contrasting approaches. Each institution commonly defines data sets as: Facts, measurements, recordings, records, or observations about the world collected by scientists and others, with a minimum of contextual interpretation. Data may be any format or medium (e.g., numbers, symbols, text, images, films, video, sound recordings, drawings, designs or other graphical representations, procedural manuals, forms, data processing algorithms, or statistical records (the definition is based on the Research Data Alliance definition of data, http://smw-rda.esc.rzg.mpg.de/index.php/Data).

Our comparison found that each curation workflow is based on a self-submission model allowing researchers to deposit their data at will. All but one repository (Minnesota) automatically accepts the data once deposited. All but one repository (Penn State) provides post-ingest curatorial review of the deposited files and metadata. Persistent identifiers in the form of a digital object identifier (DOI) are added in various ways. These similarities are encouraging and may allow our model to scale data curation work across the institutions in a similar post-ingest manner. Four institutions provided illustrative diagrams that depict this curation workflow process and they appear as Figure 4 (Minnesota), Figure 5 (Cornell), Figure 6 (Illinois), and Figure 7 (WUSL).

Table 1: Comparison of the data curation workflows at the six institutions
*On request

| Service Workflow Steps by Institution	Pre-ingest Curation?		Mediated vs. Self-deposit?		Accept/Reject Stage?		Public	Post-ingest curation?			
	Consult only	Staging Area for deposit	Mediated deposit	Self-deposit	Approval required to accept or reject	Auto Accept	Go Live Here	As needed	Review metadata only	Review files and metadata	Add DOI
University of Minnesota	X			X	X		X			X	X
Cornell University	X		X*	X		X	X			X*	X*
University of Illinois	X			X		X	X			X*	X
University of Michigan	X			X		X	X			X	X*
Penn State University	X			X		X	X				
Washington University in St. Louis	X		X	X		X	X			X	X
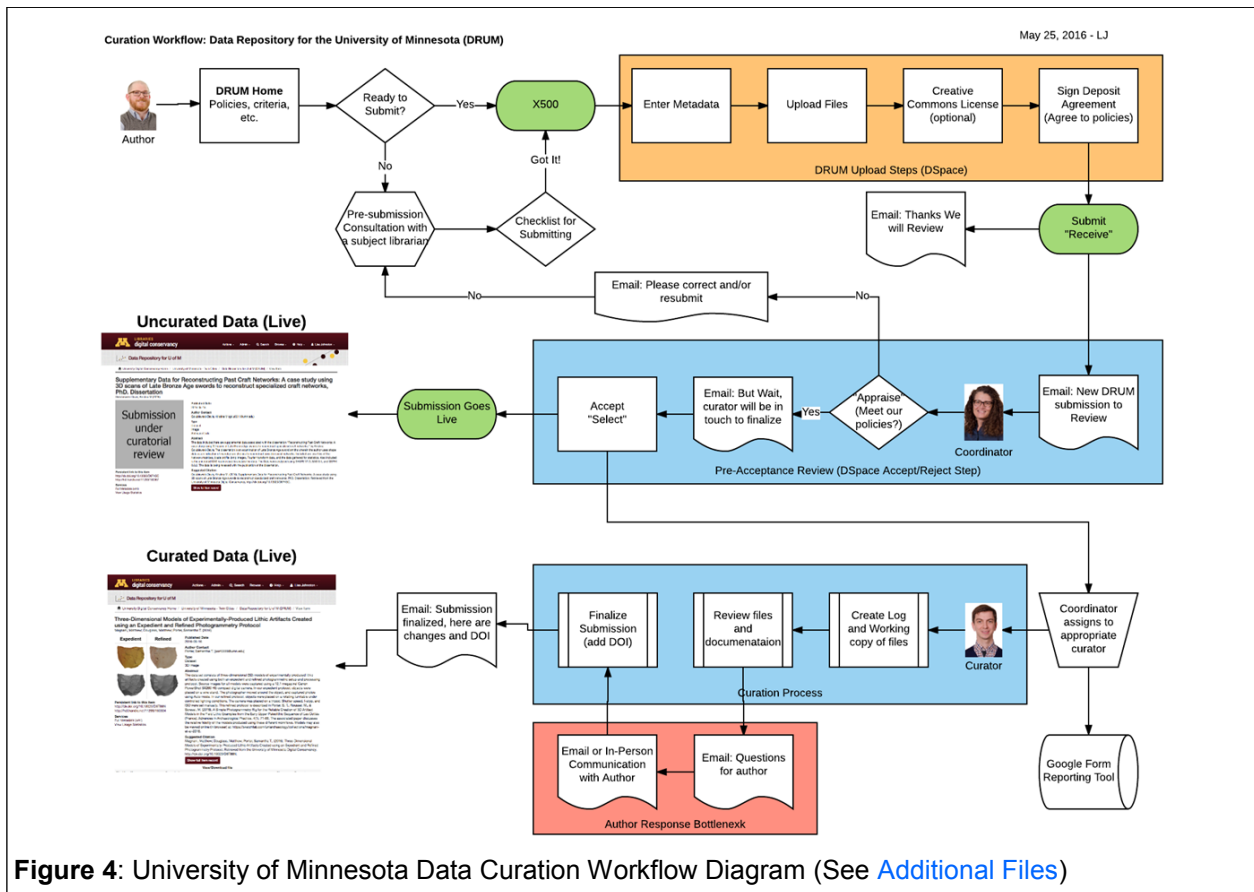
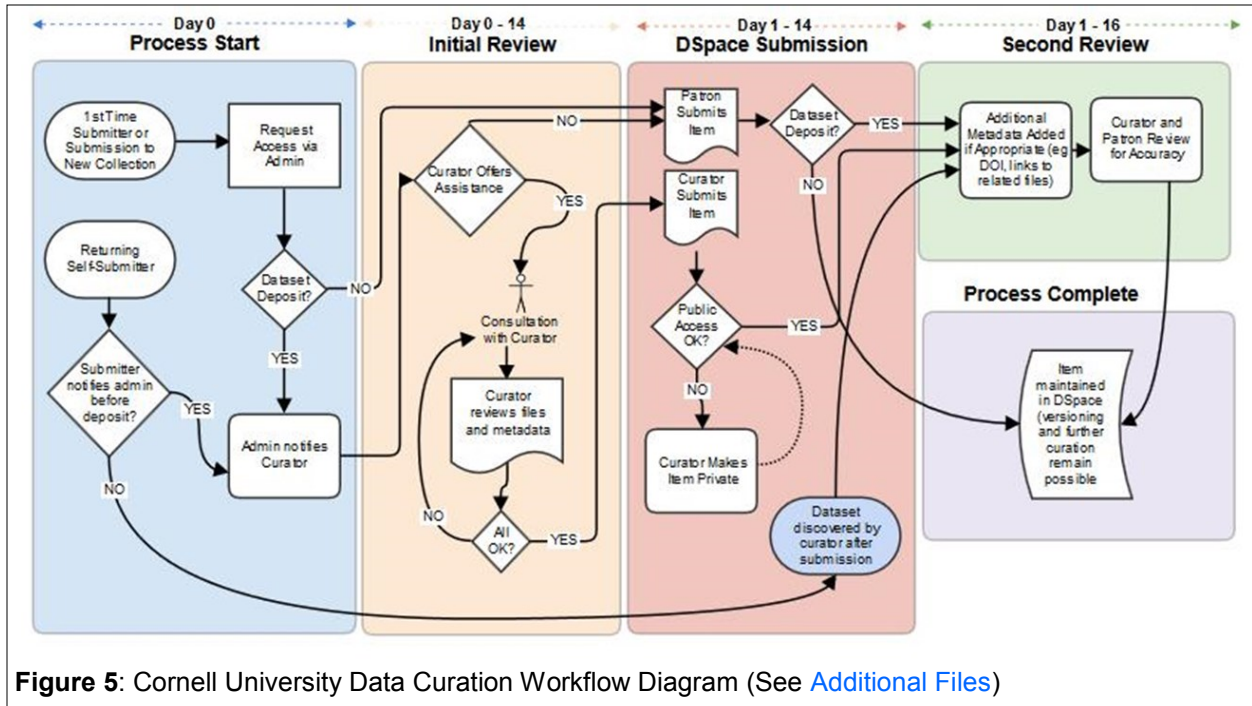**Figure 4**: University of Minnesota Data Curation Workflow Diagram (See Additional Files)

Figure 5: Cornell University Data Curation Workflow Diagram (See Additional Files)
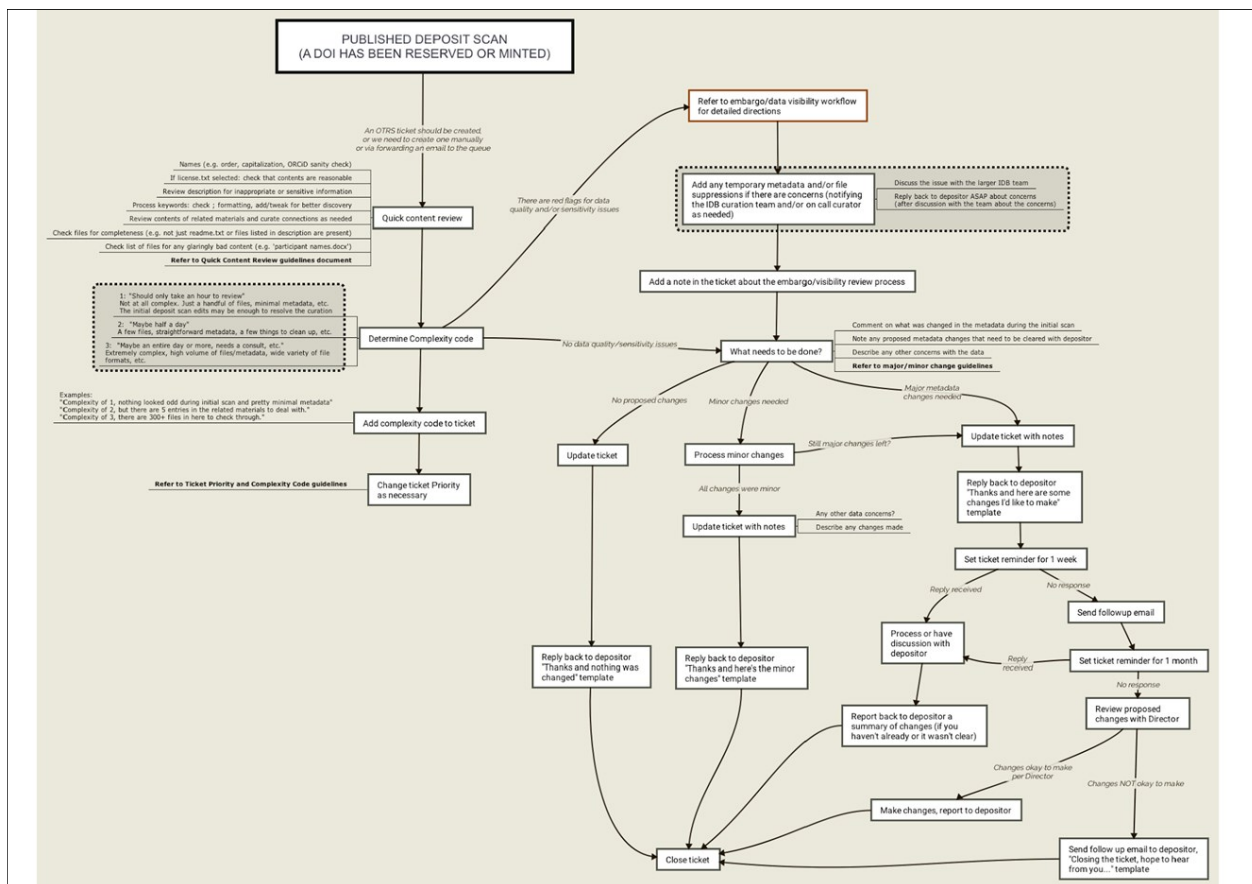
Figure 6: University of Illinois Data Curation Workflow Diagram (See Additional Files)
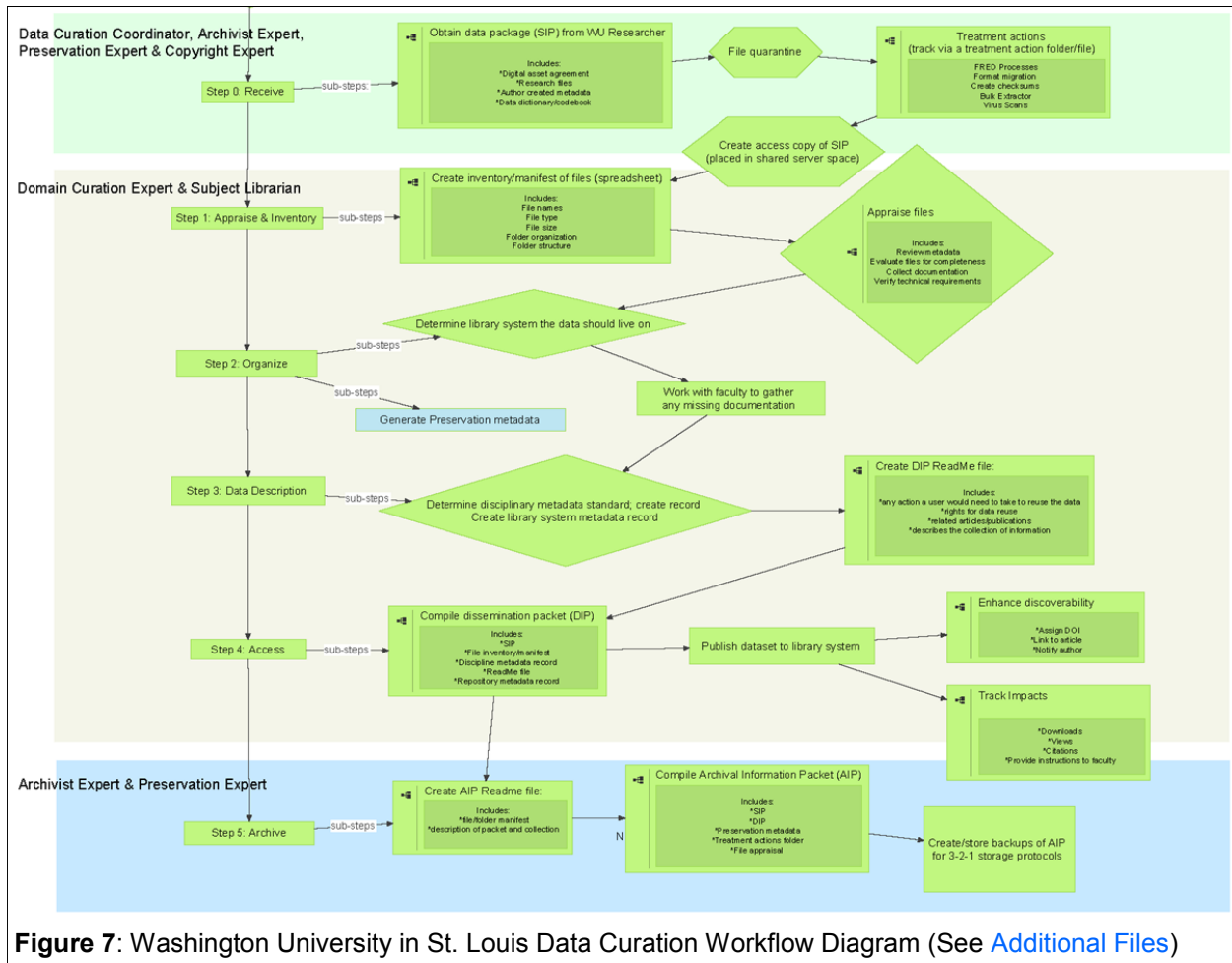
Figure 7: Washington University in St. Louis Data Curation Workflow Diagram (See Additional Files)

1.2 Tracking Data Curation Activities

Data curation services may also involve augmentation to the metadata, file format transformations (e.g., preservation friendly file formats), and documentation added to the record. Each repository tracks these changes to the data deposit in a variety of ways.

- **University of Minnesota**: Before making any changes, curators create a working copy of the submission and store the original files and metadata as a back-up copy, in case reversion is needed. During the curation process our staff keep a text-based curator's log file detailing all changes made during the curation process. The curators also (manually) capture all relevant correspondence with author (e.g., email exchanges) regarding the changes made and save with the log. This log file is archived with the dataset in DRUM but not made publicly available.

- **Cornell University**: Prior to submission, the curator documents all interactions, either in person, or via email, on an internal wiki; no strict format/standard yet in place. Once submitted changes are tracked by DSpace in a basic provenance record (date, time, user), and the curator logs any additional, relevant information to both the discovery and science metadata.

- **Penn State University**: Depositors with valid Penn State access account IDs may log into ScholarSphere any time to edit metadata on their files. Versions are automatically tracked in ScholarSphere, so if there are metadata changes, the system is monitoring these. Depositors can backtrack to the earlier version(s) as needed and select the one(s) they would like to make public. There is no notification to the repository service manager when deposits are made to ScholarSphere.

- **University of Illinois**: We've implemented a ticketing system (OTRS[10]). All deposits automatically create a ticket. After the curation review, depositors get an email documenting changes (even if none) or asking questions as needed. Metadata changes available as changelog; file changes would occur as versioned datasets.

- **University of Michigan**: Research Data Services staff are sent an email notifying us of a new deposit. The Data Curation Librarian contacts the appropriate subject liaison and they review the submission. Any issues, questions or suggestions for improvements with the deposit are documented and depositor is contacted via email and asked if action should be taken. Interactions with depositors are currently tracked in a spreadsheet but we are moving towards adopting ticketing system.

- **Washington University in St**. **Louis**: Treatment action text files are created and email exchanges documented and stored with the AIP along with an image of the original submission, Bitcurator[11] reports, and checksums generated. Additionally, a spreadsheet tracks the processing of the submission as it travels along the curation workflow.

Tracking data curation activities will be a key aspect of the resulting Data Curation Network model in order to measure the levels of curation staffing needs for particular disciplines, to monitor the time involved, and to demonstrate efficiencies gained by each Network participant.

2.0 Repository Technologies

Each of the repositories uses software to manage the digital assets in their data repository service. Two systems use DSpace[12] (Minnesota, Cornell), two use or intend to use Sufia[13] running on a Hydra/Fedora platform (Michigan, Penn State), Illinois runs a custom Ruby on Rails solution with a preservation back-end known as Medusa,[14] and Washington University in St. Louis uses Digital Commons by BePress.[15] The specific software versions, upload limitations, features, metadata schemas, and support for external services are compared in Table 2. As network of shared staffing, it will be critical for curators in the Data Curation Network to be able to work across a variety of technology solutions and this cross-section provides an excellent base from which to build on.

10   OTRS, "Simple Service Management," accessed January 25, 2017, https://www.otrs.com.

11   Bitcurator is a digital forensic software environment available for free download from their homepage, accessed January 25, 2017, https://www.bitcurator.net.

12   DSpace is open source repository software offered by the DuraSpace organization, accessed August 16, 2016 at https://wiki.duraspace.org/display/DSDOC5x/Release+Notes.

13   Sufia is an open source repository front-end application for Hydra and Fedora, accessed August 16, 2016, http://sufia.io.

14   Medusa is a digital preservation repository developed at the University of Illinois at Urbana-Champaign, accessed August 16, 2016, https://wiki.cites.illinois.edu/wiki/display/LibraryDigitalPreservation/Medusa+FAQ.

15   Digital Commons is a fee-based hosted digital repository solution provided by BePress, accessed August 16, 2016, http://digitalcommons.bepress.com.

Table 2: Comparison of the technology, features, and limitations for the six institutions' repositories for data

	DRUM	eCommons	Illinois Data Bank	Deep Blue Data	ScholarSphere	DRMR
Technology Platform	DSpace 5.5	DSpace 5.5	Custom-built Ruby on Rails web app as a microservice to Medusa, a local preservation repository	Hydra/Fedora Sufia 7	Hydra/Fedora (soon to be Sufia 7)	Digital Commons
Upload limits	Self-deposit up to 2 GB per file. Larger files must be mediated (up to 100GB per collection).	Self-deposit up to 2 GB per file. Larger files must be mediated. Total size per project per year is 10GB.	Self-deposit up to 15 GB via Box.[16] Larger files may be ingested via a mediated mechanism.	Self-deposit up to 2 GB per file. Larger files must be mediated. No defined limits. Exploring capability & capacity to handle large data sets.	Self-deposit up to 500 MB per file. Larger files via Dropbox (1.9 GB) or Box (5 GB). Up to 100 files and totaling less than 1 GB in size.	Self-deposit up to recommended 2 GB per file (not a hard limit - up to 10-20 GB).
Features	- Open Access - Versioning (mediated) - Related material linking - API - OAI/PMH feed	- Open Access - Versioning (mediated) - Related material linking - OAI/PMH feed	- Open Access - Versioning (mediated) - Related material linking (mediated) - Descriptive metadata editing	- Open Access - Related material linking	- Open Access (default) with option of Penn State only and Private - Versioning - Descriptive metadata editing - Basic analytics and data visualization about files	- Open Access - Versioning (mediated) - OAI/PMH feed - Related material linking (mediated)
Service/ Software Add-ons	- DataCite DOI[17]	- DataCite DOI (mediated and only upon request)	- DataCite DOI - Box integration - ORCID[18] integration	- Data Cite DOI (upon request)	- Dropbox[19] - Box integration	- DataCite DOI
Discovery Services	Web indexing: full-text and metadata - Data Citation Index (WoS)[20] - Datacite.org - re3data.org - SHARE[21]	- Web indexing: full-text and metadata	- Web indexing: metadata only - Datacite.org - re3data.org - SHARE (forthcoming)	- Web indexing: metadata only - Datacite.org	- Web indexing: full-text and metadata - re3data.org - SHARE	-Web indexing -Datacite.org -SHARE
Descriptive Metadata Schema	Dublin Core[22]	Dublin Core	Compatible with DataCite Metadata Schema 3.1[23]	Dublin Core	Dublin Core (PCDM[24] soon)	Dublin Core
Published Schema	Published online[25]	Not yet published	Published online[26]	Not yet published	Not yet published	Not yet published

16   Box, "Secure File Sharing, and Collaboration," accessed January 25, 2017, https://www.box.com/home.
17   DataCite, "Assign DOIs," accessed January 25, 2017, https://www.datacite.org/dois.html.
18   ORCID, "Connecting Research and Researchers," accessed January 25, 2017, https://orcid.org.
19   Dropbox homepage, accessed January 25, 2017, https://www.dropbox.com.
20   Web of Science, "Data Citation Index, Clarivate Analytics," accessed January 25, 2017, http://wokinfo.com/products_tools/multidisciplinary/dci.
21   SHARE homepage, accessed January 25, 2017, https://share.osf.io.
22   Dublin Core Metadata Initiative homepage, accessed January 25, 2017, http://dublincore.org.
23   DataCite Schema, "DataCite Metadata Schema 3.1, Released October 16, 2014," accessed January 25, 2017, https://schema.datacite.org/meta/kernel-3.
24   Portland Common Data Model homepage, accessed January 25, 2017, http://pcdm.org/2016/04/18/models.
25   University of Minnesota Libraries, "The Supporting Documentation for Implementing the Data Repository for the University of Minnesota (DRUM): A Business Model, Functional Requirements, and Metadata Schema," Issued April 2015, http://hdl.handle.net/11299/171761.
26   Elise Dunham and Stein, Ayla, "Illinois Data Bank Metadata Documentation v1.0," Released August 4, 2016, http://hdl.handle.net/2142/91020.

3.0 Policy Comparison

Policy development is a critical component of developing data repository and curation services. The institutions all have publically viewable policies for deposit, access, documentation, and preservation (compared in Table 3). However several challenging policy limitations and themes emerged from our discussions.

- **Undefined documentation requirements**: Several institutions (Minnesota, WUSL, Michigan, Cornell, Penn State) described their policies for what constitutes adequate documentation for a data deposit to be vague. Our partner at Michigan said that "Currently, the expected documentation is only loosely defined in our policy." and our Penn State partner said "We could define our documentation requirements, period."

- **Difficulty in determining who can deposit**: All six institutions require at least one author be an institutional affiliate to deposit their data. However, our Illinois team member reports, "There are lots of collaborations and infrastructure projects at our university, so some asking to allow data deposit where an Illinois affiliate is not always an author. Similarly, some centers and projects want to be labeled at the data author or the long-term contact (e.g. organization as author)."

- **Sensitive data concerns**: None of the repositories allow data deposit that contains private data. Our Illinois partner mentioned "Lots of issues around sensitive data, third party data and Data Use Agreements (DUAs)." While our Michigan partner said "We do encounter researchers with sensitive data issues who would like guidance on how to share their data. We are still learning how we can respond effectively."

- **Overlapping or competing data repositories**: If the institution houses other data repositories, scope can become an issue. Our Minnesota partner reported, "We have a large medical school with separate clinical data repository and a do-it-ourselves approach limits our outreach in this side of campus." Our Penn State member reported, "There are two other repository services at Penn State, in addition to ScholarSphere. These are DataCommons and Penn State Law eLibrary. Depositors would benefit from a clearer, more explicit expression of our policies, particularly around the scope of our collections." Penn State is currently working to further define the scope of ScholarSphere in relation to other repositories to help users better understand which repository is appropriate.

- **Access control**: Some institutions provide authors the ability to embargo or temporarily restrict access to their data deposits (Minnesota, Cornell, Illinois, WUSTL). Our team member at Cornell said, "We do get submitters who want to control access (either to Cornell community, or only "upon request")."

The Data Curation Network must consider conflicting policy issues, build a shared understanding (e.g., memorandum of agreement), and create a governance model that addresses the unique needs and restrictions in place at each institution.

Table 3: Comparison of policies for data repository and curation service at the six institutions

Link to Deposit License Agreement	U. Minnesota	https://conservancy.umn.edu/pages/drum/policies/#deposit-license
	Cornell	https://ecommons.cornell.edu/page/policy#license
	U. Illinois	https://databank.illinois.edu/policies#deposit_agreement
	U. Michigan	https://deepblue.lib.umich.edu/data/agreement
	Penn State	https://scholarsphere.psu.edu/agreement
	Wash U. St. Louis	http://openscholarship.wustl.edu/data/policies.html
Data Type/ Collection Scope	U. Minnesota	General, all data
	Cornell	General, all data (but recommend disciplinary repositories as relevant)
	U. Illinois	General, from either research project or with an expectation to be used for research purposes (but recommend disciplinary repositories as relevant)
	U. Michigan	General, all data that conform to our collections policy (we assist researchers in identifying disciplinary repositories if desired)
	Penn State	General, all data (but recommend disciplinary repositories as relevant)
	Wash U. St. Louis	General, all data
Deposit	U. Minnesota	University of Minnesota affiliates (Shib log n, no authorization required)
	Cornell	Cornell University affiliates (Shib log in or upon request to admins after justification)
	U. Illinois	University of Illinois affiliates (login - restricted to fac, grad student and staff groups)
	U. Michigan	University of Michigan affiliates (log in)
	Penn State	Penn State University affiliates (Shib login, no authorization required)
	Wash U. St. Louis	WUSL affiliates (log in)
Private data (PII)	U. Minnesota	Not accepted, run Identity Finder[27] to be sure
	Cornell	Not accepted
	U. Illinois	Not accepted
	U. Michigan	Not accepted
	Penn State	Not accepted
	Wash U. St. Louis	Not accepted
Access	U. Minnesota	Default open access. Submitters may mediate access to just the files for up to 2 years.
	Cornell	Default open access. Submitters may request delayed access until related publication has been released (not advertised as an "embargo").
	U. Illinois	Default open access. Submitters may embargo either the entire dataset or just the files for up to 1 year.
	U. Michigan	Open access. No embargos at this time (though we are considering it for the future).
	Penn State	Open Access. No embargos at this time.
	Wash U. St. Louis	Default open access. Submitters may embargo just the files for up to 2 years.
Data Restrictions	U. Minnesota	No private, confidential, or other legally protected information.
	Cornell	No confidential information, proprietary information of others or export controlled information.
	U. Illinois	No private, confidential, or other legally protected information.
	U. Michigan	No private, confidential, or other legally protected information. Data must have research value.
	Penn State	N/A
	Wash U. St. Louis	No private, confidential, or other legally protected information.

27   Identity Finder software released by Spirion, "Identity Finder Data Sheet: Sensitive Data Manager," accessed January 25, 2017, http://info.identityfinder.com.

Table 3 *(continued)*: Comparison of policies for data repository and curation service at the six institutions

Documentation Restrictions	U. Minnesota	Data must include "adequate documentation describing the nature of the data at an appropriate level for purposes of reuse and discovery."
	Cornell	None required but strongly encouraged (and assistance offered).
	U. Illinois	None required but strongly encouraged to deposit metadata files that meet minimum standards as outlined in the Dataset Documentation Help section.
	U. Michigan	None required (outside of some basic metadata) though "A detailed description of a data's origins, purpose, and use" is strongly encouraged.
	Penn State	None required (outside of some basic metadata)
	Wash U. St. Louis	"adequate documentation for reuse."
Preservation Commitment	U. Minnesota	Files preserved at least 10 years via Rosetta.[28]
	Cornell	"...committed to preserving the binary form of the digital object..."; no commitment to format migration.
	U. Illinois	Minimum of 5 years via the preservation repository (Medusa).
	U. Michigan	Minimum of 10 years. 3 tiers of commitment depending on format.
	Penn State	For long-term preservation & access (no finite number of years expressed).
	Wash U. St. Louis	Minimum 10 years followed by collection review in IR
Creative Commons License	U. Minnesota	Optional, author-specified via submission form: CC0, CC BY, CC BY NC. Other licenses on request (mediated).
	Cornell	Optional, author-specified: CC0, CC-BY, CC-BY-ND, CC-BY-SA, CC- BY-NC, CC-BY-NC-ND, CC-By-Nc-Sa; Other licenses (eg. ODC) can be added via metadata)
	U. Illinois	CC0 and CC BY encouraged, licence.txt allowed
	U. Michigan	Required, author-specified CC0, CC BY, CC BY NC (other licenses on request)
	Penn State	Default is CC BY-NC-ND but depositor may change to any of the following: CC BY; CC BY-SA; CC BY-NC; CC BY-ND; CC0; All rights reserved.
	Wash U. St. Louis	Optional, CC-BY, CC-BY-SA, CC-BY-NC, CC-BY-NC-SA, CC0
End-User Terms of Use	U. Minnesota	The user not make any use of data to identify or otherwise infringe the privacy or confidentiality rights of individuals discovered inadvertently or intentionally in the data. The user will give appropriate attribution to the author(s) of the data in any publication that employs resources provided by the Data Repository. If your use or publication requires permission, you must contact the authors directly; administrators of the Data Repository cannot respond to requests for permission.
	Cornell	n/a
	U. Illinois	Datasets published in the Illinois Data Bank are discoverable and openly available to anyone with access to the World Wide Web. Data Files and Metadata Files are provided at least in the original format deposited. When appropriate, items in proprietary formats may be con-verted to formats that can be opened and read using freely available software. When Data Files and/or Metadata Files in a Dataset are made available in a converted format, Research Data Service staff will document the conversion in the Dataset's associated Descriptive Metadata and/or Metadata File(s).
	U. Michigan	You agree that Deep Blue repositories and its administrator, the University of Michigan, shall have no liability for any consequential, indirect, punitive, special or incidental damages, whether foreseeable or unforeseeable (including, but not limited to, claims for defamation, errors, loss of data, or interruption in availability of data), arising out of or relating to your use of Deep Blue repositories or any resource that you access through Deep Blue repositories.
	Penn State	n/a
	Wash U. St. Louis	n/a
Disclaimer on data quality	U. Minnesota	"Data are offered with no warranty or claim of fitness for any purpose. In no event shall the University be liable for any actual, incidental or consequential damages arising from use of these files."
	Cornell	N/A
	U. Illinois	... does not attempt to judge the scholarly quality of the Dataset. ... Thus a determination of the research quality is at the discretion of, and also the responsibility of, the Long-Term Contact Person...
	U. Michigan	Deep Blue services and content therein are provided "as is" without warranty of any kind… including... fitness for a particular purpose. Use of Deep Blue Data is at your own risk.
	Penn State	N/A
	Wash U. St. Louis	N/A

28   Rosetta homepage, accessed November 28, 2016, http://www.exlibrisgroup.com/category/RosettaOverview.

4.0 Staffing for Data Repository and Curation Services

Of the six institutions' reported staffing levels, one commonality was the heavy reliance on partial or shared staff that dedicates only a percentage of their time to data repository and curation services. In fact, for the six institutions, this was the case for each of our positions. Table 4 describes the levels of staffing for the six services and is followed by a brief description of the organizational oversight and staffing structure in each case. The implications for this baseline metric are key for the Data Curation Network. A shared staffing model across the Network will provide each of our services with an infusion of expert staff that will increase the collective capacities for offering data curation services and allow our individual services to scale.

Table 4: Comparison of staffing levels for data repository and curation services

	DRUM	**eCommons**	**Illinois Data Bank**	**Deep Blue Data**	**ScholarSphere**	**DRMR**
# Full-Time Employees (FTE)	0 FTE	O FTE	1 FTE (developer)	1 FTE (librarian)	0 FTE	0 FTE
Approximate Shared Employee time	~1.5 FTE librarian and curator time and 0.5 FTE developer time	~1 FTE librarian and curator time and ~0.5 FTE developer time	~1.4 FTE librarian and curator time	~1 FTE librarian and curator time 1.75 FTE developer time	~0.75 FTE librarian and curator time and ~1.5 FTE developer time	~0.65FTE librarian and curator time
List any non-library staff	0.1 FTE curator time (volunteer from the College of Liberal Arts)		0.05 FTE developer time (volunteer from central IT for user experience)			

4.1 Organizational Approaches to Data Repository and Curation Services

Each institution has a unique approach to how data curation services fit within the broader campus landscape. Understanding these relationships will aid in developing clear incentives for joining the Data Curation Network that reaches stakeholders both within and external to the library. Each of the six services were assessed for:

1. **University Oversight**: The campus-wide body or policy that governs data management-related decisions.

2. **Library Oversight**: The group or individuals that sponsors and oversees the data repository and curation services provided by the library.

3. **Organizational Structures**: The management and reporting structure for the key personnel providing these services.

4. **Committee Structures**: The related library and non-library groups and committees that participate in providing the services.

<u>University of Minnesota</u>

1. University Oversight: Defined in the campus Data Management Policy[30] as shared between the Vice President for Research, the CIO of the Office of Information Technology, and the University Librarian.

2. Library Oversight: Associate University Librarian (AUL) for Research and Learning and Libraries Cabinet

3. Organizational Structures: Launched as a "library initiative" reporting directly to three AULs, the Data Management and Curation Initiative primarily sits under the Research and Learning AUL (reference and subject liaisons group).

    a. Lead/Director for Data Curation Services: Director of DRUM (30%)

    b. Direct Report Curation Staff: Scientific Data Curator, a graduate research assistant at 50% time (Fall/spring only)

    c. Non-Direct Report Curation Staff:

        i. DRUM Coordinator (Repository Archivist, 10%) → reports to University Archivist

        ii. Public Health/human subjects data curator (10%) → reports to Dir of health sciences unit

        iii. Social sciences data curator (10%) → reports to Dir of social sciences unit

        iv. Spatial/GIS data curator (10%) → reports to Dir of Map Library

        v. College of Liberal Arts (CLA) Data Management Specialist (10%) → Reports to CLA Information Technology unit

4. Committee Structures:

    a. Library groups: Research Data Services Team (outreach and training); DSpace Management Team (technical development/road mapping); University Digital Conservancy (institutional repository) Management Team (governance)

    b. Non-library groups: Data policy implementation team (based in the Office for the Vice President for Research), informal Community of Practice for Research Data Management

<u>Cornell University</u>

Org Chart Link:
https://www.library.cornell.edu/sites/default/files/CULibrary_20160600_OrgChart_Kenney.pdf

1. University Oversight: n/a

---

29   University of Minnesota, "UMN Policy: Research Data Management: Archiving, Ownership, Retention, Security, Storage, and Transfer," accessed November 28, 2016, http://policy.umn.edu/research/researchdata.

2. Library Oversight: Coordinated by Cornell University Library's Scholarly Communication Librarian, who reports to Associate University Librarian (AUL) for Scholarly Resources and Preservation Services, who is part of the Library Executive Group (University Librarian+AULs)

3. Organizational Structures:

   a. Lead for Data Curation Services: Data Curation Specialist (based in the Scholarly Communication Services, a Unit of Digital Scholarship and Preservation Services)

   b. Direct Report Curation Staff: 0

   c. Non-Direct Report Curation Staff: 0.2 (collection-specific data curator/subject librarian)

4. Committee Structures:

   a. Related library groups: Metadata Services (Unit of Library Technical Services); eCommons Advisory Board

   b. Related non-library groups: n/a

Penn State University

Org Chart Link: https://libraries.psu.edu/file/psul-organizational-chart

1. University Oversight: n/a

2. Library Oversight: Associate Dean for Technology and Digital Strategies

3. Organizational Structures: Based in the library's Technology and Digital Strategies unit and the Research, Collections, and Scholarly Communications unit are:

   a. Lead/Director for Data Curation Services: (still being determined)

   b. Direct Report Curation Staff: Science Data Librarian, Geospatial Data Services Librarian

   c. Non-Direct Report Curation Staff: n/a

4. Committee Structures:

   a. Related library groups: ScholarSphere Service Team, Data Services Action Working Group (strategic implementation group), Research Data Management Team

   b. Related non-library groups: Penn State Institute for CyberScience, Research IT Advisory Council, Data Governance Working Group, Data Commons Team.

University of Illinois

Org Chart Link: http://cms.library.illinois.edu/cms/staff/orgchart

1. University Oversight: Office of the Vice Chancellor for Research (for research data services specifically - otherwise it's the Provost)

2. Library Oversight: AUL for Research

3. Organizational Structure: Based in the library's Office of Research, the unit includes:

    a. Lead/Director for Data Curation Services: Research Data Services Director (33%)

    b. Direct Report Curation Staff: 2 data curators

    c. Non-Direct Report Curation Staff: n/a

4. Committee Structures

    a. Related library groups: Repository Team, Digital Preservation, Archives, Scholarly Commons (data discovery and use), Subject Specialists, Cataloging and Metadata

Related non-library groups: Research IT (e.g. storage and security + campus IT professionals), Institutional Review Board, Research Ethics, Technology Transfer, Graduate College, National Center for Supercomputing Applications (supercomputing institute).

<u>University of Michigan</u>

1. University Oversight: Provost (loosely)

2. Library Oversight: AUL for Research & AUL for Library Information Technology

3. Organizational Structures: Based in the library's Research Unit, Library Information Technology are

    a. Lead/Director(s) for Data Curation Services:

        i. Head of Sci/Eng/SAND Libraries and Director of Research Data Services

        ii. Head, Architecture and Engineering, Library Information Technology

    b. Direct Report Curation Staff: Research Data Services (RDS) Manager

    c. Non-Direct Report Curation Staff: Data Curation Librarian (reports to RDS Manager)

4. Committee Structures

    a. Related library groups: Learning & Teaching, Technical Services, Preservation, University of Michigan Publishing, Taubman Health Sciences Library.

    b. Related non-library groups: Advanced Research Computing, institutional review board, Office of Research and Sponsored Programs, MIDAS (Data Science Initiative), Consulting for Statistics, Computing and Analytics Research.

<u>Washington University in St. Louis</u>

Org Chart Link: http://library.wustl.edu/about/orgchart

1. University Oversight: Provost

2. Library Oversight: AUL; Director of Scholarly Services

3. Organizational Structure: Based in the Library's Data & GIS Services Unit are

    a. Lead/Director for Data Curation Services: Data Services Coordinator

    b. Direct Report Curation Staff: n/a

    c. Non-Direct Report Curation Staff: GIS Projects Manager (reports to Director of Scholarly Services), subject specialists, repository librarian, copyright librarian

4. Committee Structures:

    a. Related library groups: Scholarly Publishing, Special Collections (Archives), Subject Specialists

    b. Related non-library groups: research computing infrastructure (HPC, storage, and data transfer), Research Office, Center for Biomedical Informatics

4.2 Position Descriptions and Job Duties

By reviewing position descriptions for research data curation staff and other library staff with data repository and curation responsibilities we aim to better understand the skills needed and the encompassing roles already expected from the staff that our Network model is aimed toward. Here are some experts from the partner institutions' position descriptions. See also the recent report[30] from the Joint Task force on Librarians' Competencies in Support of E-Research and Scholarly Communication.

<u>Lead/Director for Data Curation Services</u>. Example duties include:

- Collect, manage, curate, provide access to and assist in the discovery of research data; refer researchers to disciplinary repositories as appropriate.

- Provide consultation services for researchers and liaisons to enhance the ability of others to manage, preserve, and conduct new research using digital data collections.

- Develop innovative methods for data discovery to enhance the library's delivery and discovery environment.

- Work with faculty, graduate and post-doctoral students, academic and administrative units, and research centers to enable them to better manage, describe, archive, preserve, and make available university research data.

- Work with researchers to identify, recruit, ingest and deposit data into repositories, including the library's digital repositories, adhering to local policies and national and international standards and best practices for data management, public access and preservation.

---

30   Birgit Schmidt and Kathleen Shearer, "Librarians' Competencies Profile for Research Data Management," June 2016, https://www.coar-repositories.org/files/Competencies-for-RDM_June-2016.pdf

<u>Data Curators</u>. Example duties include:

- Serve as primary expert contact for new users inquiring to submit content to the data repository; authorizes new submitters, and answers questions to assist during the upload process for distributed content providers.

- Process submissions for deposit and archive datasets in the digital repository; research data-related repository activities, workflows, and policies.

- Collect, manage, curate, provide access to and assist in the analysis of research data related to [specific subject discipline]; refer researchers to disciplinary repositories as appropriate.

- Engage with [disciplinary] data producers at the University, as well as at the state and local government levels, to acquire and build a corpus of digital spatial data for access and preservation.

- Perform data curation actions for [disciplinary] data contributed to the data repository or other appropriate repositories.

- Apply data management and data curation techniques for a variety of digital formats (text, code, images, video, etc.).

<u>Library Staff/Subject liaison</u>. Example duties that related to data repository and curation services include:

- Work closely with faculty and students in [subject area] to understand and respond to their changing workflows and patterns of research, research dissemination, and management and preservation of research data.

- Educate and inform faculty, students, and campus administrators about scholarly communication issues such as author's rights agreements, open access publishing models, and discipline repositories for publications and data.

Discussion

The data repository and curation services at our six institutions represent a snapshot-in-time for library-based activities in this area. By comparing side-by-side services, policies, technology, and staffing levels, our Data Curation Network team holds a better understanding of the similarities and contrasting approaches underway so that we may move forward in our goal of developing a shared staffing model for providing data curation services across our institutions. For example, throughout our assessment it became clear that many of our service goals were well aligned and the basis for our model began to form. Based on the similarities that most of our services featured, including self-deposit submission workflows, post-ingest curation, DOI minting services, and closely aligned metadata requirements, we now envision a model for shared staffing that delineates the "local" curator role from the "Network" curator role. A possible outcome is envisioned in Table 5.

Table 5: Possible roles and activities of curators in the Data Curation Network
*Note: The local curator may prefer to be the primary contact with the data submitter rather than someone from the Network.

Local Curator: The data curator at the institution where the data submission originated.	Network Curator: The subject-expert curator in a non-local Data Curation Network institution that is assigned the submission to review.
• receiving data and appropriate metadata • appraisal and selection (e.g., initial review of the submission to determine if it meets local policy) • assigning persistent identifiers (e.g., DOI) • providing access and dissemination • Providing digital preservation of the files (e.g, fixity check, validation, providing checksums, etc.)	• assigning submission to the appropriate curation expert • reviewing data files and providing quality assurance • reviewing documentation (e.g, readme files) • communicating with the data submitter for missing information and files* • transforming file formats (e.g, proprietary file formats to non-proprietary)

Another finding of this assessment were the perceived similarities in our institutional policies, thus alleviating concerns that a future shared-service model might face an uphill battle to avoid conflicts with policy. Differences in the repository policies were not described as fundamental divergences, but rather, as policy gaps that should have or will be addressed. It was common to hear a team member say, "No, our policy does not say that, but it probably should." This process of comparing policies in our assessment and review allowed team members to deeply engage with other institutions' policies in order to benchmark and compare to their own. As a result, team members could detect gaps in their own process and fill in any gaps in local policy where needed.

Additionally, in our parallel yet separate implementations of repository technology, each using a variation of multiple software approaches, we found much common ground in the workflows and design of how data interacted with the service. For example, one possible workflow in our Data Curation Network model will be review datasets post-ingest when they are already publically available, rather than needing special-access permissions for non-local curators. These technology and workflow commonalities are thanks, in large part, to the institutional repository model that each of our systems are either based on or emulating for the use case of research data.

Finally, the staffing models had strong similarities, even though the lack of stable full-time staff was the underlying theme. Yet, as the primary goal of the Data Curation Network is to approach a shared staffing model for data curation services, it is this lack of staffing resources that fuels our project. By pooling our staffing resources, we hope to have a stronger and more diversified portfolio of skills and expertise to draw from in our data curation service efforts at home.

Conclusion

Data-specific curation activities are relatively new to academic libraries and based on the assessment presented here it is clear to us that we, individually, have much to learn. The Data

Curation Network serves as a way for us to learn from each other about how to best curate datasets. However, moving forward we hope the Network will begin to enable the community to pragmatically and effectively provide added value to published datasets. The next phase of the project will develop a model for how the Data Curation Network will function, including how data will enter and flow through the service in ways that match our shared expectations, as well as how the Network will be administered and sustained. Most importantly, by intentionally structuring our efforts to coordinate as a Network that can grow and incorporate new institutions over time, we hope to play a role in engaging and empowering the larger data curation community through sharing experiences and providing a platform for continued dialog and discussion in this area.

Supplemental Content

An online supplement to this article can be found at http://dx.doi.org/10.7191/jeslib.2017.1102 under "Additional Files".

Funding Statement

Disclosure

The authors report no conflict of interest.

References

BePress. 2016. "Digital Commons." Accessed August 16. http://digitalcommons.bepress.com

Bitcurator. 2017. "BitCurator." Accessed January 25. https://www.bitcurator.net

Box. 2017. "Secure File Sharing, and Collaboration." Accessed January 25. https://www.box.com/home

Center for Open Science. 2017. "SHARE." Accessed January 25. https://share.osf.io

Concordat Working Group. 2016. "Concordat on Open Data." July 28.
http://www.rcuk.ac.uk/documents/documents/concordatonopenresearchdata-pdf

DataCite. 2017. "Assign DOIs." Accessed January 25. https://www.datacite.org/dois.html

DataCite Schema. 2017. "DataCite Metadata Schema 3.1, Released October 16, 2014." Accessed January 25.
https://schema.datacite.org/meta/kernel-3

DataCommons. 2016. "Penn State Data Commons - Penn State's source for data sharing, discovery, and archiving." Accessed August 16. http://datacommons.psu.edu
Data Curation Network. 2017. "Who is involved." Accessed January 26.
https://sites.google.com/site/datacurationnetwork/people

Dropbox. 2017. "Dropbox." Accessed January 25. https://www.dropbox.com

DSpace. 2017. "DSpace is a turnkey institutional repository platform." Accessed January 25. http://www.dspace.org

Dublin Core Metadata Initiative. 2017. "DCMI Home: Dublin Core® Metadata Initiative (DCMI)." Accessed January 25. http://dublincore.org

Dunham, Elise and Ayla Stein. 2016. "Illinois Data Bank Metadata Documentation v1.0." *Illinois Data Bank*.
http://hdl.handle.net/2142/91020

Duraspace. 2016. "DSpace 5.x Release Notes." Accessed August 16.
https://wiki.duraspace.org/display/DSDOC5x/Release+Notes

Fedora Repository. 2017. "Home - Fedora Repository." Accessed January 25. http://fedorarepository.org

Figshare. 2016. "figshare - credit for all your research." Accessed August 16. https://figshare.com

Hydra Project. 2017. "Hydra Project." Accessed January 25. https://projecthydra.org

Karasti, Helena, Karen S. Baker, and Eija Halkola. 2006. "Enriching the notion of data curation in e-science: data managing and information infrastructuring in the long term ecological research (LTER) network." *Computer Supported Cooperative Work* 15(4): 321-358. http://www.dx.doi.org/10.1007/s10606-006-9023-2

ORCID. 2017. "Connecting Research and Researchers." Accessed January 25. https://orcid.org

OTRS. 2017. "Simple Service Management." Accessed January 25. https://www.otrs.com

Portland Common Data Model. 2017. "Portland Common Data Model." Accessed January 25.
http://pcdm.org/2016/04/18/models

Re3data.org. 2016. "Home - re3data.org." Accessed August 16. http://www.re3data.org

Rosetta. 2016. "Rosetta - Digital Management and Preservation." Accessed November 28.
http://www.exlibrisgroup.com/category/RosettaOverview

Schmidt, Birgit and Kathleen Shearer. 2016. "Librarians' Competencies Profile for Research Data Management."
https://www.coar-repositories.org/files/Competencies-for-RDM_June-2016.pdf

Spirion. 2017. "Identity Finder Data Sheet: Sensitive Data Manager." Accessed January 25.
http://info.identityfinder.com

Sufia. 2016. "Sufia: a community-supported repository front-end." Accessed August 16. http://sufia.io

University of Illinois. 2016. "Medusa." Accessed August 16.
https://wiki.cites.illinois.edu/wiki/display/LibraryDigitalPreservation/Medusa+FAQ

University of Minnesota. 2016. "UMN Policy: Research Data Management: Archiving, Ownership, Retention, Security, Storage, and Transfer." Accessed November 28. http://policy.umn.edu/research/researchdata

University of Minnesota Libraries. 2015. "The Supporting Documentation for Implementing the Data Repository for the University of Minnesota (DRUM): A Business Model, Functional Requirements, and Metadata Schema." *University of Minnesota Digital Conservancy*. http://hdl.handle.net/11299/171761

Web of Science. 2017. "Data Citation Index, Clarivate Analytics." Accessed January 25.
http://wokinfo.com/products_tools/multidisciplinary/dci

Zenodo. 2016. "Zenodo - Research. Shared." Accessed August 16. https://zenodo.org