

DATA CURATION NETWORK

"What a long, strange trip it's been": A Retrospective of the Data Curation Network

Jake Carlson, University of Michigan
Lisa Johnston, University of Wisconsin - Madison
Mikala Narlock, University of Minnesota

Please note that we're recording and will share the slides.

The early days...

Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

Jake

Remember 2010?



(ii) **Plans for Data Management and Sharing of the Products of Research.** Proposals must include a document of no more than two pages uploaded under "Data Management Plan" in the supplementary documentation section of Research.gov. This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results (see [Chapter XI.D.4](#)), and may include:

1. the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
2. the standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);
3. policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
4. policies and provisions for re-use, re-distribution, and the production of derivatives; and
5. plans for archiving data, samples, and other research products, and for preservation of access to them.

Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

The twilight series, Lady Gaga's meat dress at the MTV Music Video Awards, Angry Birds as the game that we were all playing...

And the NSF released their Data Management Plan requirement. Researchers applying for grants from the NSF now had to include a 2 page document (Data Management Plan) describing how they would conform to the NSF's policy on the dissemination and sharing of research data. In particular, researchers had to describe their data, the standards that would be used in documenting and formatting the data, how people would be able to access the data, what people could do with it once they had the data and how access to the data would be preserved.

This caused a lot of consternation and confusion among researchers as many of them had not had to consider these types of questions before. Libraries saw an opportunity to apply our knowledge, skills and expertise with managing and sharing information towards working with research data to provide services to address researcher concerns.

Response from Libraries

- A proliferation of guides on how to create a Data Management Plan.
- A recognition that we needed to know more about how researchers managed, shared and preserved their data.

What Should Your Data Management Plan Address?

A Guide for Purdue University Researchers

This guide was created by Purdue University Libraries in accordance with the National Science Foundation (NSF) data management planning [requirements](#). This information is also available as a DMP template on [DMPTool.org](#). Instructions for accessing the DMPTool template, as well as other data management resources, are available on the Purdue University Research Repository (PURR) [website](#).

YOUR DMP SHOULD INCLUDE

- 1. Data Types**
The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project
- 2. Metadata**
The standards to be used for data and metadata format and content
- 3. Access and Sharing**
Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements
- 4. Re-use**
Policies and provisions for re-use, re-distribution, and the production of derivatives
- 5. Archiving**
Plans for archiving data, samples, and other research products, and for preservation of access to them

The NSF announcement was referred to as the announcement that launched 1k LibGuides. This is the guide that I had a hand in creating back when I was at Purdue University. But we quickly recognized that just guiding researchers through creating a DMP was not going to be sufficient. Researchers needed to be able to follow through on the promises that they were making in their DMPs, which would require a greater level of understanding about the data and researcher needs.

Data management itself being ill defined concept at the time, there was not much written about data management as a topic in the literature and so it was difficult to know where or how to start.

Data Curation Profiles

DCP Toolkit:
[https://docs.lib.purdue.edu/
dcptoolkit/](https://docs.lib.purdue.edu/dcptoolkit/)

DCP Directory:
[https://docs.lib.purdue.edu/
dcp/](https://docs.lib.purdue.edu/dcp/)

Data Curation Profiles Toolkit



[About](#) [News](#) [Download The Toolkit](#) [Submit a Profile](#) [Completed Profiles](#) [Workshops](#) [Forums](#) [Resources](#)

The Data Curation Profile (DCP) is a structured interview tool for librarians to capture:

- Information about a particular data set
- Current practices in managing the data set
- The unmet needs of the researcher.
 - i.e. What would the researcher like to do with the data set, but is not currently for whatever reason.

PURDUE
UNIVERSITY
LIBRARIES

Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

One of the first things that I worked on at Purdue was something called the Data Curation Profile Toolkit. The DCP Toolkit was designed to be used by librarians as a means to gather the information they might need to really understand a researchers situation in making a data set of theirs more widely available. is essentially a reference interview on steroids.

Research Data Management Group at Cornell



RESEARCH DATA MANAGEMENT SERVICE GROUP

Comprehensive Data Management Planning & Services

The Research Data Management Service Group (RDMSG) is a collaborative, campus-wide organization that assists with creating and implementing data management plans, applying best practices for managing data, and finding data management services at any stage of the research process.

CORNELL SERVICES

- Collaboration tools
- Data collection and analysis
- Data curation
- Data sharing
- High performance computing
- Intellectual property and copyright
- Metadata
- Security, privacy, and confidentiality
- Storage, backup, and recovery
 - Data Storage Finder
- Support for researchers in the Life Sciences

Ethical. Reusable. Better.

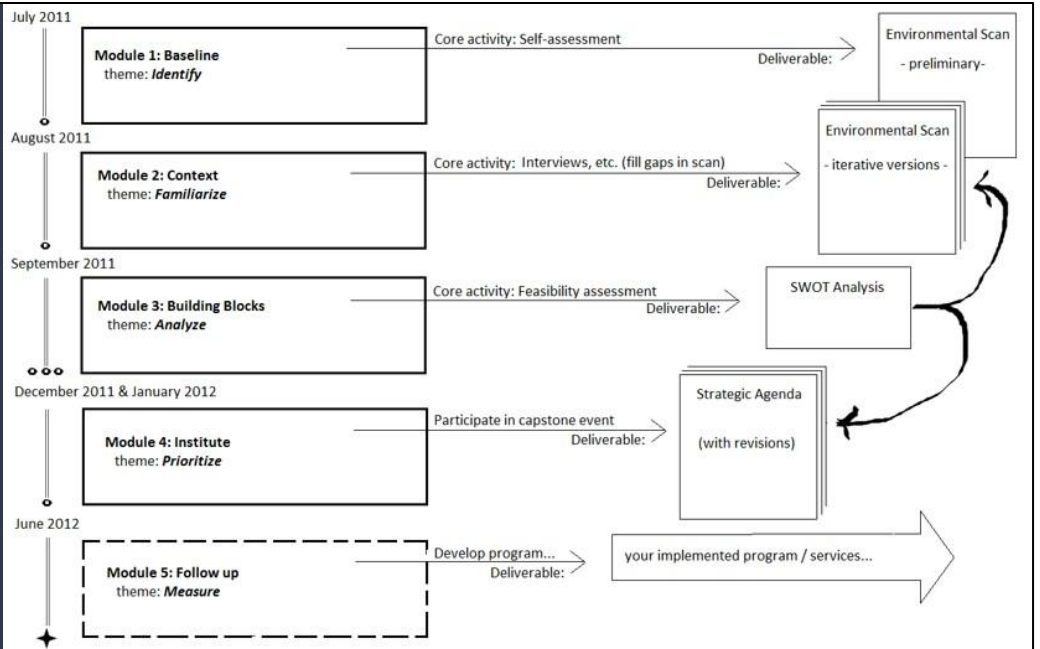
DATA CURATION NETWORK

datacurationnetwork.org

Another early effort by libraries to address researcher needs around data sharing was the Research Data Management Group at Cornell. Managing, Sharing and working with data requires different types of services and resources, more than any one department could provide on their own. The Cornell Libraries formed working partnerships with the IT units, Office of Research and others to form their university wide Research Data Management Group. They engaged in research early on to better understand how researchers were responding to the DMP requirement, but have continued on and developed an incredibly useful guide to building readme files and created the Data Storage Finder tool to help researchers navigate through the many options they had for storing their data. The Data Storage Tool has been adopted by many other universities, including Michigan.

I'll also say that Cornell's Research Data Management Group has served as the model for Michigan's Research Data Stewardship Initiative and other campus wide communities at other universities.

ARL / DLF E-Science Institute 2011-12



Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

The ARL-DLF E-Science institute was another initiative designed to help prepare academic libraries to understand and address the emerging needs for data services. Here too, we recognized the need to work collaboratively with our peers given the scope and complexity of research data.

70 institutions went through the 1st iteration of the e-science institute and many more went through the second and third offerings.

Five stages:

- Environmental scanning - what services relevant to managing, sharing and preserving research data were offered by the institution and by whom? (The Library and others)
- Interviews of researchers - what data were they generating, what were they doing with it and what were their needs for it?
- SWOT analysis - analyzing what was learned
- Develop a strategic agenda to identify and prioritize areas for investment at an individual institution
- Measure the progress made (6 month follow up report out)

LIBRARIES digital conservancy

Search Browse Collections Help Sign in

Data Repositories

Data Repository for U of M

Search the Data Repository **Go**

The Data Repository for University of Minnesota (DRUM)

DRUM is a publicly available collection of digital research data generated by U of M researchers, students, and staff. Anyone can search and download the data housed in the repository, instantly or by request.

The Data Repository accepts submissions from University affiliates for digital archiving and access. [Learn more](#) about depositing to the Data Repository and other services to manage your data.

Upload to the Data Repository >

*U of M affiliates only | [How to submit](#)

How to Upload

- 1. Prepare Data**
Data should be free of identifying or sensitive information and include adequate documentation. Not sure? Contact us for help!
- 2. Upload**
Have your files ready and use the upload form to fill out metadata about your data.

Features

- Access Options**
Choose to make your data immediately accessible to everyone, or restrict access to your data for up to two years.
- Meet Grant Requirements**
Comply with federal mandates for data management planning (DMP) and sharing. [Read more.](#)

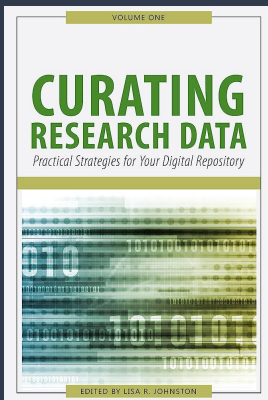
Our Services

- Data Management Plan Assistance**
We offer personalized assistance for drafting your next grant's Data Management Plan. Contact us for assistance during your planning process.
- Metadata Consultation**
We can help structure your data using disciplinary best practices to ensure the best organization of your data.

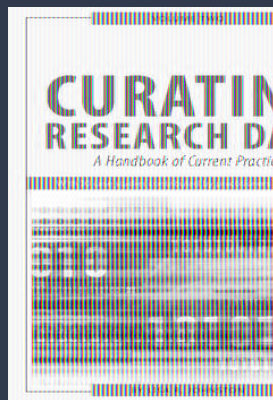
Libraries also saw their institutional repositories being used as a way to share the data in ways that would satisfy federal agency requirements, and in some cases Libraries developed stand alone data repositories such as DRUM at the University of Minnesota.

However, we did not want these repositories to be dumping grounds for poorly organized or described data sets. If we invested into data IRs we wanted them to be useful not just for people producing data who needed to satisfy the data sharing requirement but for consumers of data as well. We want the data that we host to be understandable, trusted and reusable as possible. This requires that we curate the data, that we invest staff time and energy into reviewing the data, the documentation and the connection between files to ensure that the data is as valuable as it could be in our repository.

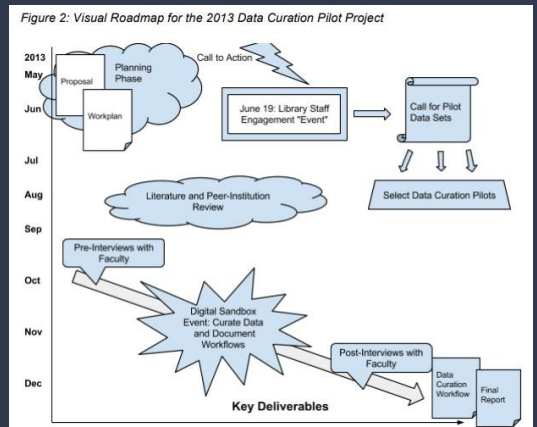
What is Data Curation



World Cat listing
<https://www.worldcat.org/title/69352704>



World Cat listing
<https://www.worldcat.org/title/66989181>



Johnston, L. (2014). A Workflow Model for Curating Research Data in the University of Minnesota Libraries: Report from the 2013 Data Curation Pilot. University of Minnesota Digital Conservancy. <http://hdl.handle.net/11299/162338>

Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

However, The concept of data management was not well formed back in 2010 let alone data curation. We needed not just to define data curation but to operationalize it and make it something that could be operationalized and actionable by librarians.

Lisa did a masterful job in researching data curation as a concept and in recruiting experts in the field to define the current practice of data curation and to provide practical strategies for how it could be done.

She then took it a step further and tried an experiment with librarians at the University of Minnesota. She give librarians access to several data sets for them to review and then told them to try and curate the data as best they could. This study provided some important insights about our preparedness to do data curation work and the nature of the time and labor involved in doing it.

Sum up the challenges

- Librarians have relevant skills and expertise for digital preservation and sharing, but we needed to develop an understanding as to how to translate and apply them to research data
 - Leverage library / archival best practices
 - Understand specialized research data types and researcher needs for them
 - Make curation work actionable (not just conceptual)
- Lots of work happening, but in isolation at institutions. Given the scope and the scale of research data, we needed to transcend our institutional boundaries.
 - Break down barriers across institutions
 - Developing shared common practices and infrastructure

Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

Data Curation Network

Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

Lisa start

In light of these challenges, a group of six librarians at six institutions came together and asked, why not try to work together to address these as a collective instead of individuals operating in silos?

Forming the Data Curation Network



Radical Independence:

“Why not work together to address our common challenges? – And not just to share information, but to actually do the work of curating the data that ingest into our repositories?”

Five Key Areas

- build a shared curation workflow between the DCN and the data curators at each member institution;
- define a staffing and governance model for our network;
- craft a 6-year implementation plan to serve as a roadmap in growing curation offerings and membership;
- develop a financial plan focused on the long-term sustainability of the DCN
- assess our networked approach to curating research data

Ethical. Reusable. Better.

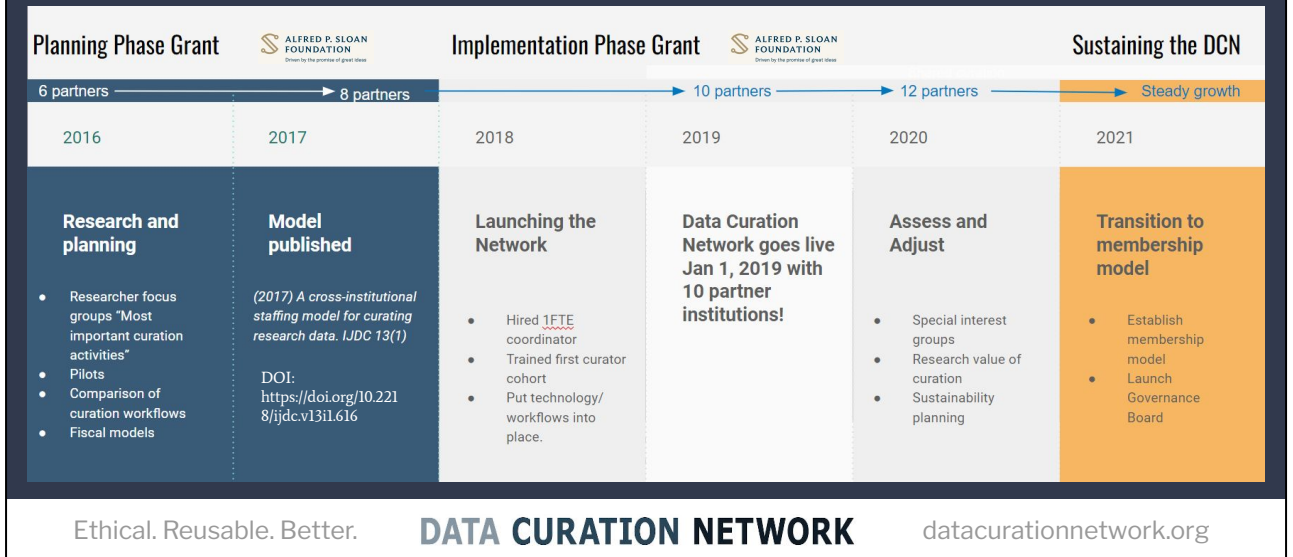
DATA CURATION NETWORK

datacurationnetwork.org

This discussion led to the six PIs submitting for and being awarded a one year planning grant from the Alfred P. Sloan Foundation to better investigate what a cross-institutional data curation network would look like. This included understanding what a shared curation workflow would look like, how it would be staffed and governed, what financial sustainability would look like, and developing a roadmap to grow the network and its curation offerings.

This work, from the outset, was rooted in radical interdependence, which is something that we've actively worked to maintain in the DCN. Radical interdependence is about relying on one other to work together to common challenges– to go beyond information sharing and to actually do the work of curating research data ingested into repositories. It's about being able to say I don't know can you help, and trusting that your colleague will show up for you and complete their work.

Data Curation Network Timeline



This is just a high-level timeline of the development of the DCN. After the one year planning grant, the Alfred P. Sloan foundation awarded the team a three year implementation grant, which also provided enough funding to hire a full time project coordinator. This full time staff person was essential for developing and launching the shared curation workflow, which officially launched January 2019. During this implementation phase, the team put into practice what was learned during the research and planning phase. In 2021, the DCN transitioned to be a fully member funded project, rolling off the grant funding into our sustainability and growth model.

I wanted to highlight that the community started with six partners and six librarians in 2016, growing to 8 partners in 2017, 10 partners in 2019, 12 partners in 2020, 14 partners in 2021, and 17 partners in 2022. Starting in 2018, the community also grew beyond just the PIs from each institution, as partners started adding curators to the community. This means that we started to grow both in terms of number of participants and in cohesion.

Data Curation Network - 19 members and growing!



New in 2023!



Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

Mikala Start

And this has been essential as we've continued to grow and onboard both new institutions and new curators into our network – we're now up to 19 member institutions with the University of Wisconsin and University of Pennsylvania joining us in FY 23.

Mission

Trusted community-led network
that enables researchers to openly
share data in ways that are

Ethical. Reusable. Better.

DATA CURATION NETWORK



Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org  Alfred P. Sloan
FOUNDATION

Now that we've been in operation for more than 7 years, what is the DCN now? We are a community-led network of curators advancing open research by making data more ethical, reusable, and understandable. We are an active community of data stewards and curation practitioners who share knowledge and time to collaboratively curate research outputs and advance the data curation profession.



Curation



Access curation expertise across our cross-institutional network



Education



Offer professional development opportunities for an emerging data curator professional community



Primers



Contribute to openly shared data curation best practices



Research



Join a DCN special interest group addressing a specific topic: big data, human subjects, racial justice, + more



Community



Be a part of the virtual community for discussion, training, networking events, more

Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

Some of the things the DCN provides and facilitates are listed here, but this includes Curation: from 17 partner institutions, soon to be 19, we have nearly 50 individual curation experts who are the foundation of our cross-institutional curation network. The DCN also serves as an educational resource for data curators, through workshops and the data curation primers, as well as a community of practice that offers interest groups, professional development opportunities, and networking.

Why a retrospective?

- Project debrief
- Visioning and future direction
- Onboarding / Staff transitions

Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

The DCN has been active in these areas, and more, since 2016. In March 2022, we conducted a project retrospective and brought together many project personnel, to better understand the process of planning and implementing the DCN. We were at an interesting point in time: we'd transitioned to be member funded, we had launched our governance model and process, I had recently been hired, and Lisa Johnston, who had served as the PI and leader for the grants, was preparing to transition to a new position outside of the DCN. So, we decided to host this project retrospective to better understand: what were the successes and challenges we faced in building, launching, and sustaining the DCN? Where are we headed for the next phase of our work? And, how can we ensure that this transition in personnel is as smooth as possible?

Project Retrospective Process

- Structured and unstructured time
- Different activities and prompts
- Celebrating the good, acknowledging the challenges



Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

Lisa and I collaborated on a retrospective meeting agenda that, over the course of 2.5 days, set out to accomplish all of those things. We think this effort was successful, and wanted to quickly share how we organized this meeting. We balanced structured and unstructured time and conversations— sometimes we had very specific questions and conversations, and other times took a broad prompt and had a wide ranging discussion.

We also brought in different activities and prompts to make sure we were engaging folks in mechanisms they were most comfortable with— whether that was free writing time, jamboard activities, or just conversation.

And, we incorporated both celebrating the accomplishments of our community while acknowledging the challenges and shortcomings.

Retrospective Report

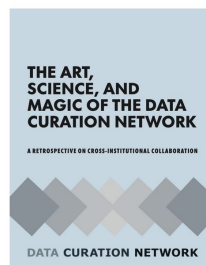
**DATA
CURATION
NETWORK**

Ethical. Reusable. Better.

[Home](#) [About](#) [Resources](#)

doi.org/10.3998/mpub.12782791

Jake Carlson, Mikala Narlock, Mara Blake, Joel Herndon, Heidi Imker, Lisa Johnston, Wendy Kozlowski, Cindy Xuying Xin, Sophia Lafferty-Hess, Hoa Luong, Wanda Marsolek, Jennifer Moore, Dorris Scott, Cynthia Hudson Vitale, Briana Ezray Wham, & Sarah Wright
2023 Michigan Publishing Services



We are excited to share that our open access retrospective report is now accessible through the University of Michigan Press! You can read "[The Art, Science, and Magic of the Data Curation Network: A Retrospective on Cross-Institutional Collaboration](#)" now. The open access version is available to use now. Physical copies will be made available for purchase later this spring.

In this report, we provide insight into the formation of the DCN. In particular, we provide additional context into how the DCN was envisioned and launched, in particular the key areas of focus for planning the work of the network. After briefly

Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

We captured our discussions through notes, and as a team, authors a retrospective report published through the University of Michigan Press in March 2023, which you can access at this DOI. And now I'll turn it over to Lisa!

Some of our findings

Ethical. Reusable. Better.

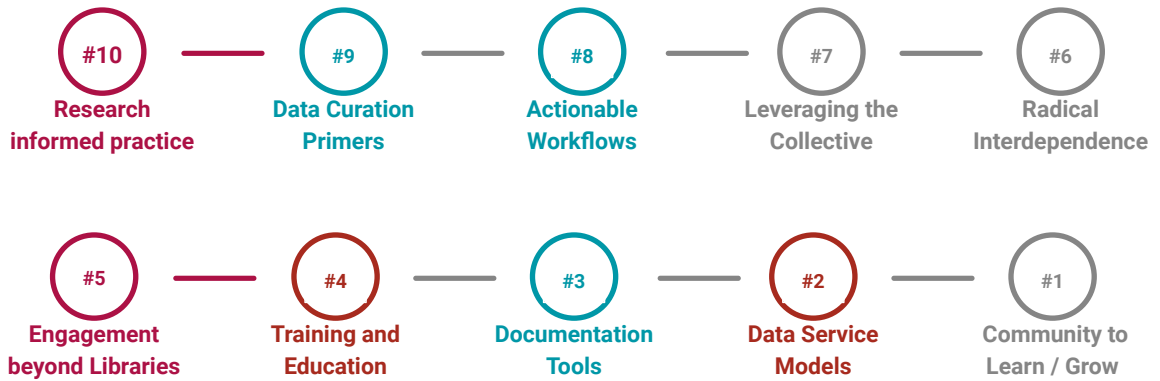
DATA CURATION NETWORK

datacurationnetwork.org

Lisa Start

Top 10 Successes

- Administrative structures
- Tool-based structures
- Trust-Based structures



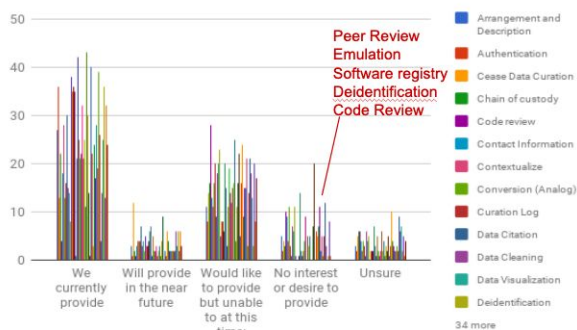
Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

Success: Research informed practice

ARL Spec Kit Data Curation 2017 (n=80)



SPEC Kit #354: Data Curation. Association of Research Libraries (ARL), May 2017.
<http://publications.arl.org/Data-Curation-SPEC-Kit-354/--FreeAttachments/Data-Curation-SPEC-Kit-354.pdf>

Focus Group Results (n=91 across 6 institutions)

Top 12 Most Important Curation Activities (4 out of 5)

1. (Create) Documentation (4.6)
2. Secure Storage (4.4)
3. Quality Assurance (4.3)
4. Persistent Identifier (4.3)
5. Software Registry (4.1)
6. Data Visualization (4.0)
7. File Audit (4.0)
8. (Create) Metadata (4.0)
9. Versioning (3.9)
10. Contextualization (3.9)
11. Code Review (3.9)
12. File Format Transformations (3.9)

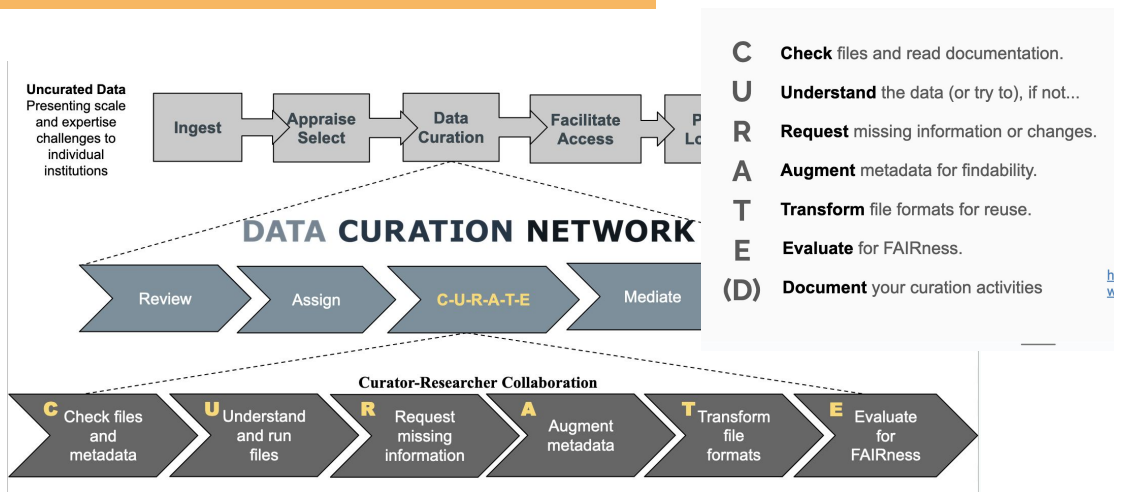
Data Curation Network. (2018). How Important is Data Curation? Gaps and Opportunities for Academic Libraries. *Journal of Librarianship and Scholarly Communication*, 6 (1), e2198. <http://doi.org/10.7710/2162-3309.2198>

Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

Success: Data Curation Service Model

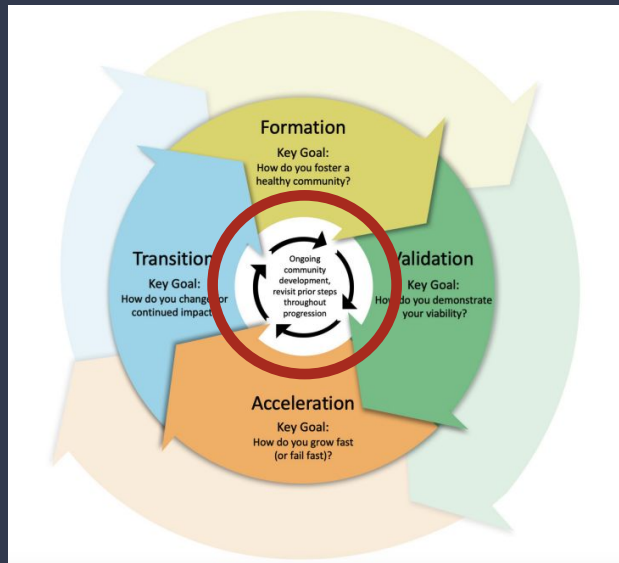


Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

Continued Efforts



Source: Educopia, "Community Cultivation Field Guide"

Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

Mikala start:

Thanks Lisa! The DCN has accomplished some amazing things over the years, but this is not to suggest that there haven't been challenges in developing, launching, and sustaining the DCN. In fact, there remains work to be done. We're in the ongoing development phase, highlighted in the Community Cultivation Lifecycle here, which requires a continued focus on fostering a healthy community, demonstrating our sustainability, as well as ensuring we are growing to meet the changing needs of our community. But, we are in a fortunate position to not only be financially stable, but also have an active and involved community that will help us address some challenges on the horizon.

Future challenges

- Homogeneity of DCN Members
- Balance interests and goals with capacity
- Maintaining trust and community

Ethical. Reusable. Better.

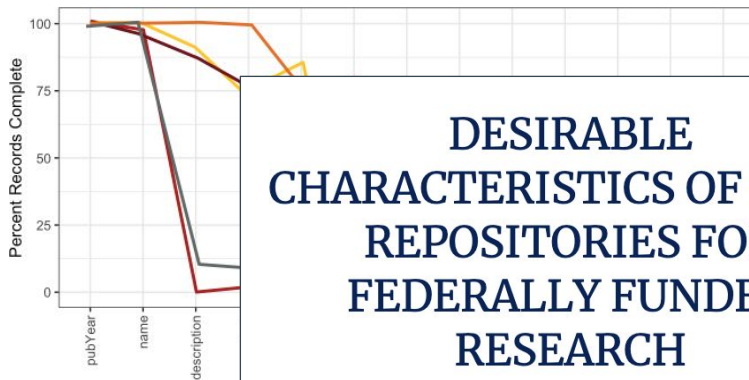
DATA CURATION NETWORK

datacurationnetwork.org

One future challenge we see is the homogeneity of our members. We are primarily R1s and all well-resourced institutions. We want to be sure that the DCN is welcoming to all institutions, regardless of their type. We are here to support curators looking for a community. In the coming years, we will continue to find ways to serve as a welcoming and inclusive community for curators— and remember that your institution doesn't have to be a sustaining member to participate in the DCN.

Another challenge will be balancing our interests, our goals, our research projects with our very real and limited capacity. I am the only full time staff member of the DCN— everyone else participates on top of their day jobs. So while there are many exciting opportunities and interests, we have to be sure that we are moving strategically so as not to burn ourselves out.

And lastly, as we grow, we're reflecting on our radical interdependence at the fabric of the DCN. We know that the trust between members and the community is one of the key benefits of the DCN. As we continue to onboard new members and institutions, we will need to be thoughtful and intentional about building and maintaining trust and community.



eraging Data Communities to
Open Science

Workshop Series

DESIRABLE CHARACTERISTICS OF DATA REPOSITORIES FOR FEDERALLY FUNDED RESEARCH

Guidance by the
SUBCOMMITTEE ON OPEN SCIENCE

of the
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

Ethical. Reusable. Better.

DATA CURATION NETWORK ITHAKA S+R

Looking more broadly at the data curation profession, there are exciting new challenges and opportunities on the horizon and happening now

- Supporting Data Communities– Recently the DCN collaborated with Ithaka S+R on an NSF funded project to better understand the data management and sharing needs of data communities. Data communities are “formal or informal groups of scholars who share a certain type of data with each other, regardless of disciplinary or institutional boundaries,” and the results of this workshops demonstrate that there is a lot for room for supporting these communities.
- Another interesting and challenging opportunity is collaboration across IRs with specific regards to metadata application. Recent work through the NSF funded Realities of Academic Data Sharing Initiative, based at the association of research libraries. They have recently done some research into metadata completeness in participating IRs, and learned that many institutions are applying metadata in slightly different ways, which can make it hard to compare apples and apples. Plus, by the time the metadata gets to an aggregator level, like DataCite, the records are less complete than they are in the host IR.

These are just two challenges that we see on the horizon, but we’re excited about the possibilities of working with others in the RDM community to address these shared opportunities.

Q&A / Discussion

jakecar@umich.edu
lrjohnston@wisc.edu
mnarlock@umn.edu

Key Takeaways:

- Successful cross-institutional collaboration takes trust, reciprocity, and resource investment (time and money)
- Retrospective process helps large, multi-year projects look back, celebrate successes, and prioritize for the future

Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

So, to wrap up this long, strange journey, some of our key takeaways are that successful cross-institutional collaboration requires trust and vulnerability, and especially resources. The retrospective process can provide a structured reflection time to celebrate successes and prioritize for the future. And, many of us are facing similar challenges— and we don't need to go through those challenges alone.

And with that, we thank you sincerely for your time and would love to open the floor for questions and comments!