

Big 10 Academic Alliance

Data Curation Network

*A Collective Approach to Providing Research Data
Curation Services*

*Lisa Johnston
University of Minnesota*

May 16, 2017

Data Curation Network

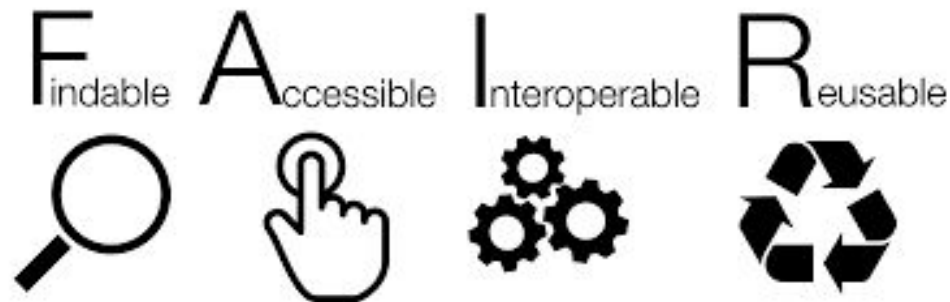
Rise of the Data Sharing Culture

Researchers are increasingly required/incentivised to share data

- Funder data sharing mandates
- Journal data sharing policies
- Disciplinary practices → emphasis on transparency and reproducibility

But! It's not enough to just share the files, **well-curated data** are more valuable!

Goal of data curation ⇒ Prepare and maintain research data in ways that make it findable, accessible, interoperable and reusable.



Data Repository for Univ of Minnesota (DRUM)

LIBRARIES
digital conservancy

Search Browse Help Sign in

Data Repository for U of M

Search the Data Repository **Q 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 (up to 2GB each) and use the upload form to fill out metadata about your data.

Features

Flexible Access Options

Choose to make your data immediately accessible to everyone, or moderate access to your data upon request.

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.



The Center of Pressure Data from "The Rim and Ancient Mariner: The Nautical Horizon Affects Postural Sway in Older Adults"

Munafo, Justin G; Wade, Michael G; Stoffregen, Thomas A; Stergiou, Nicholas (2016)

Submission
under
curatorial
review

Published Date

2016-11-24

Author Contact

Stoffregen, Thomas A (tas@umn.edu)

Type

Dataset

Abstract

This dataset contains the raw center of pressure data collected on the Enrichment Voyage (www.semesteratsea.org) on an AMTI (Advanced Mechanical Technology, Inc.) force plate. The data was collected across two days from 18 participants. There were two conditions (the near condition and the far condition) split evenly across six trials in a randomized order. In the trials of the near condition, participants stood on the force plate with their hands comfortably on their sides with their shoes on. They maintained their gaze on a tripod located 50 cm from their heel for the duration of the trial. In the trials of the far condition, everything remains the same, except that the tripod was removed, and participants were instructed to look at the horizon. Trials were 60 seconds long.

Suggested Citation

Munafo, Justin G; Wade, Michael G; Stoffregen, Thomas A; Stergiou, Nicholas. (2016). The Center of Pressure Data from "The Rim and Ancient Mariner: The Nautical Horizon Affects Postural Sway in Older Adults". Retrieved from the Data Repository for the University of Minnesota, <http://doi.org/10.13020/D6MG63>.

Show full item record

Persistent link to this item

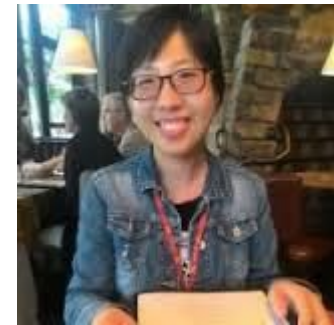
<http://doi.org/10.13020/D6MG63>
<http://hdl.handle.net/11299/183102>

Services

Full Metadata (xml)
View Usage Statistics

View/Download file

File View/Open	Description	Size	Format
----------------	-------------	------	--------



Data Management Curation Lead (30%)

GIS/Spatial Data Curator (10%)

Scientific Data Curator (Grad Ras, 50%)



Team

Health Sciences/ Public Health Data Curator (10%)

Social Sciences Data Curator (10%)



CLA Data Management Specialist (n/a)

Digital Arts & Humanities Specialist (10%)



sea_for_mfdfa.csv

100%

Home Layout Tables Charts SmartArt Formulas Data Review

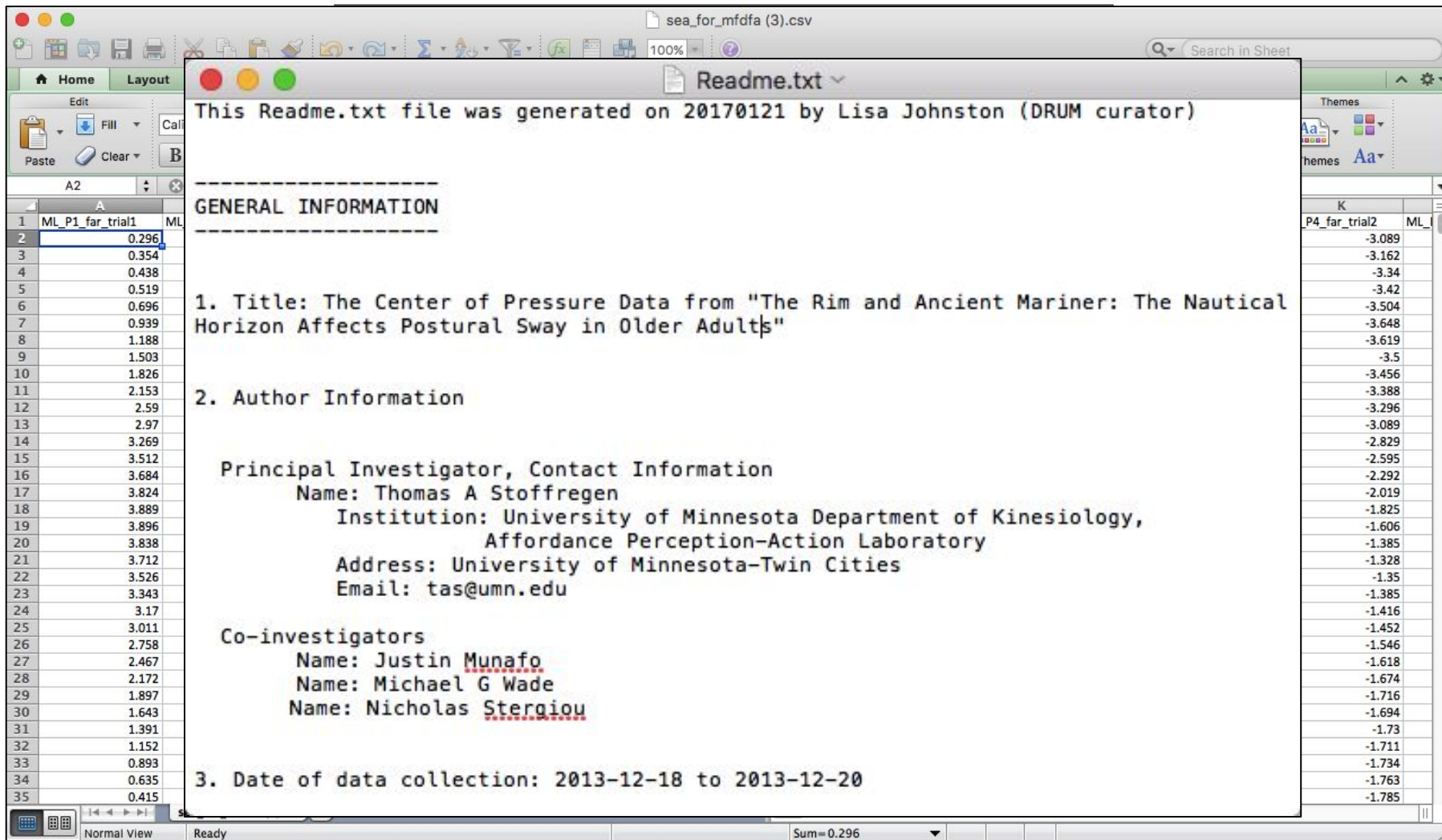
Edit Font Alignment Number

Fill Calibri (Body) 12 A A abc Wrap Text General

Paste Clear B I U Merge % .0 .00

A1

A	B	C	D	E	F	G	H	I	J	K	L	M
0.296	-0.55	-0.243	-3.04	-2.885	0.071	1.847	-0.521	0.107	-3.528	-3.089	2.45	-
0.354	-0.487	-0.193	-2.974	-2.933	-0.029	1.951	-0.506	0.061	-3.643	-3.162	2.131	-1.
0.438	-0.449	-0.099	-2.979	-2.901	-0.09	2.051	-0.501	-0.005	-3.647	-3.34	1.713	-1.
0.519	-0.431	-0.018	-3.042	-2.831	-0.13	2.107	-0.452	-0.027	-3.616	-3.42	1.322	-1.
0.696	-0.37	0.023	-3.065	-2.832	-0.187	2.202	-0.415	-0.072	-3.648	-3.504	0.831	-1.
0.939	-0.332	0.083	-3.089	-2.815	-0.233	2.314	-0.342	-0.119	-3.603	-3.648	0.38	-1.
1.188	-0.295	0.171	-3.08	-2.753	-0.295	2.431	-0.197	-0.186	-3.598	-3.619	-0.07	-1.
1.503	-0.284	0.279	-3.129	-2.746	-0.363	2.52	-0.116	-0.298	-3.487	-3.5	-0.608	-1.
1.826	-0.288	0.36	-3.183	-2.743	-0.496	2.59	-0.012	-0.316	-3.318	-3.456	-0.989	-1.
2.153	-0.289	0.369	-3.162	-2.632	-0.615	2.653	0.197	-0.345	-3.249	-3.388	-1.33	-0.
2.59	-0.244	0.359	-3.205	-2.51	-0.761	2.761	0.412	-0.416	-3.204	-3.296	-1.58	-0.
2.97	-0.196	0.319	-3.218	-2.463	-0.944	2.933	0.643	-0.421	-3.143	-3.089	-1.746	-0.
3.269	-0.222	0.297	-3.148	-2.454	-1.045	3.051	0.904	-0.356	-2.983	-2.829	-1.813	-0.
3.512	-0.266	0.274	-3.157	-2.429	-1.147	3.119	1.116	-0.286	-2.783	-2.595	-1.927	-0.
3.684	-0.271	0.289	-3.214	-2.396	-1.255	3.052	1.222	-0.227	-2.627	-2.292	-2.081	-0.
3.824	-0.275	0.233	-3.289	-2.4	-1.262	2.996	1.39	-0.16	-2.475	-2.019	-2.286	-0.
3.889	-0.294	0.186	-3.295	-2.303	-1.306	2.961	1.545	-0.083	-2.293	-1.825	-2.461	-0.
3.896	-0.295	0.158	-3.289	-2.266	-1.383	2.93	1.645	-0.095	-2.195	-1.606	-2.573	-0.
3.838	-0.283	0.152	-3.286	-2.273	-1.352	2.876	1.615	-0.074	-2.086	-1.385	-2.672	-0.
3.712	-0.338	0.139	-3.328	-2.23	-1.302	2.778	1.637	-0.007	-1.971	-1.328	-2.759	0.
3.526	-0.363	0.125	-3.387	-2.198	-1.275	2.604	1.624	-0.019	-1.814	-1.35	-2.817	0.



The screenshot shows a web browser window displaying a data repository page. A spreadsheet titled 'sea_for_mfdfa (3).csv' is open, showing a column of numerical data. A 'Readme.txt' file is overlaid on the spreadsheet, providing metadata for the data. The README file contains the following information:

This Readme.txt file was generated on 20170121 by Lisa Johnston (DRUM curator)

GENERAL INFORMATION

- Title:** The Center of Pressure Data from "The Rim and Ancient Mariner: The Nautical Horizon Affects Postural Sway in Older Adults"
- Author Information**

Principal Investigator, Contact Information
 Name: Thomas A Stoffregen
 Institution: University of Minnesota Department of Kinesiology,
 Affordance Perception-Action Laboratory
 Address: University of Minnesota-Twin Cities
 Email: tas@umn.edu

Co-investigators
 Name: Justin Munafò
 Name: Michael G Wade
 Name: Nicholas Stergiou
- Date of data collection:** 2013-12-18 to 2013-12-20

The spreadsheet in the background shows a column of numerical values, with the first cell containing 'ML_P1_far_trial1' and the first data point being 0.296. The status bar at the bottom of the spreadsheet shows 'Sum= 0.296'.

Challenges for Data Curation Services



- How to scale local data curation services across all disciplines?
- How many data curation experts are needed?
 - Types: GIS, spreadsheet/tabular, statistical/survey, software code, video/audio...
 - Disciplines: genomic sequence, chemical spectra, bioinformatics...
- Are there ways to more efficiently curate rare or infrequently generated data types?
- Might our institution specialize in curation skills and represent our academic expertise?

Data Curation Network

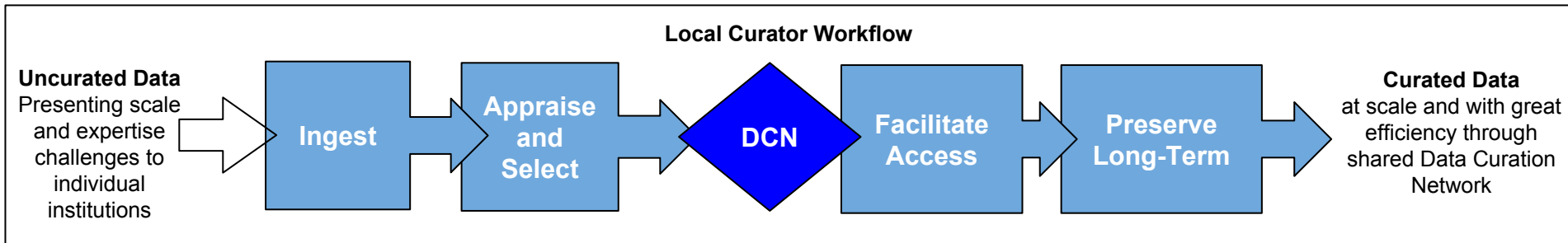
A network of expertise model for curating research data in digital repositories

The Data Curation Network will enable academic institutions to better support researchers that are faced with a growing number of requirements to ethically share their research data.

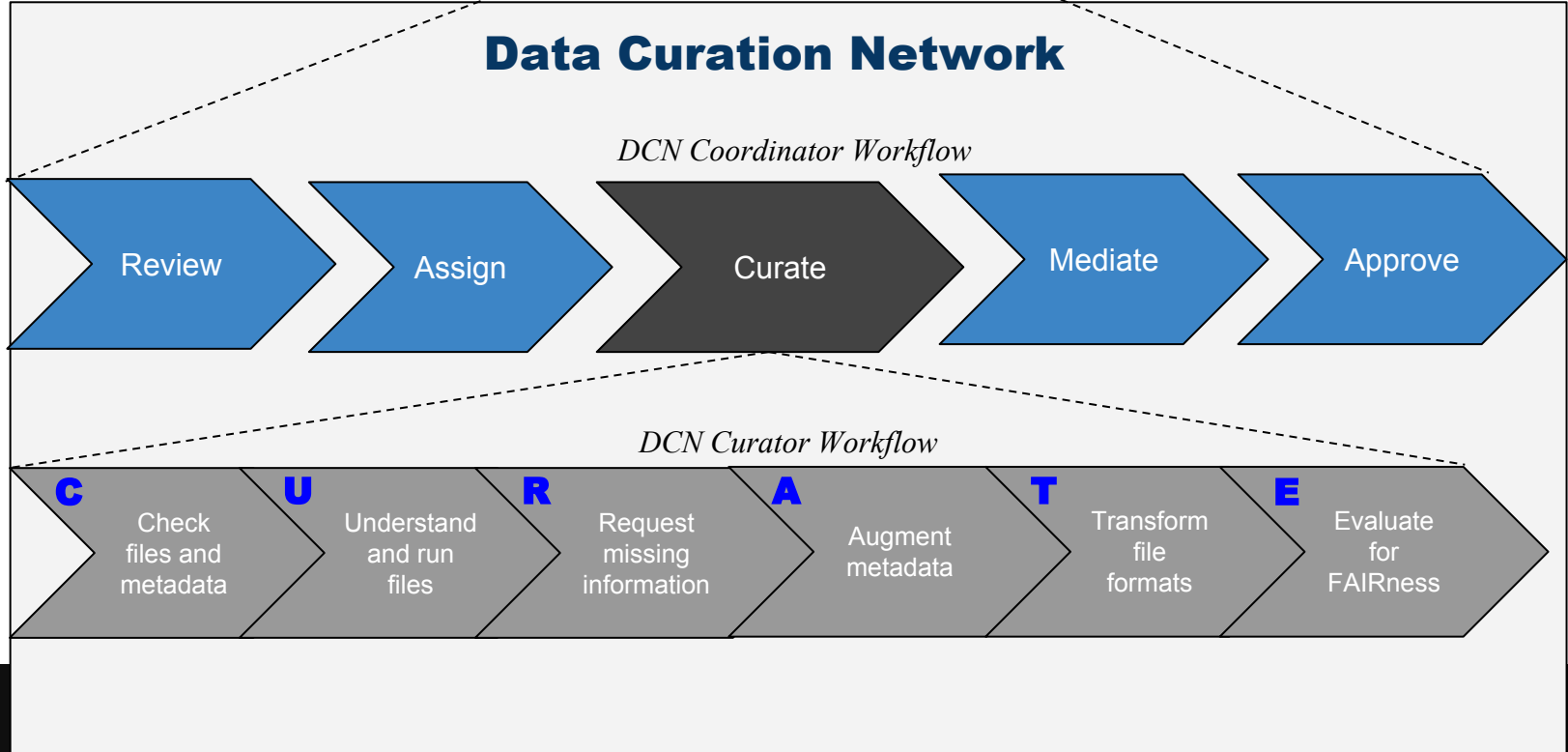
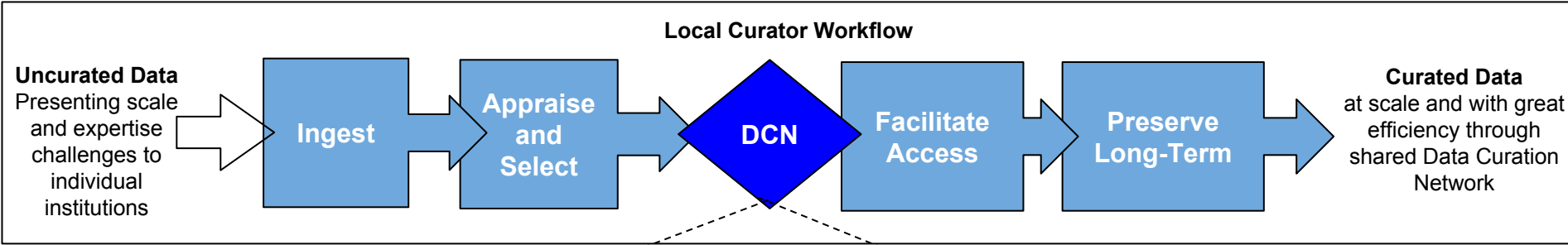
DCN Planning Phase Partners



Model for the Data Curation Network



Model for the Data Curation Network





**Lessons learned from building
collaborative research data
curation services...**

Collab tips that worked for the Data Curation Network

1. Develop a shared vision

Mission of the DCN

The Data Curation Network will enable academic institutions to better support researchers that are faced with a growing number of requirements to ethically share their research data.

In the next 3-5 years we will...

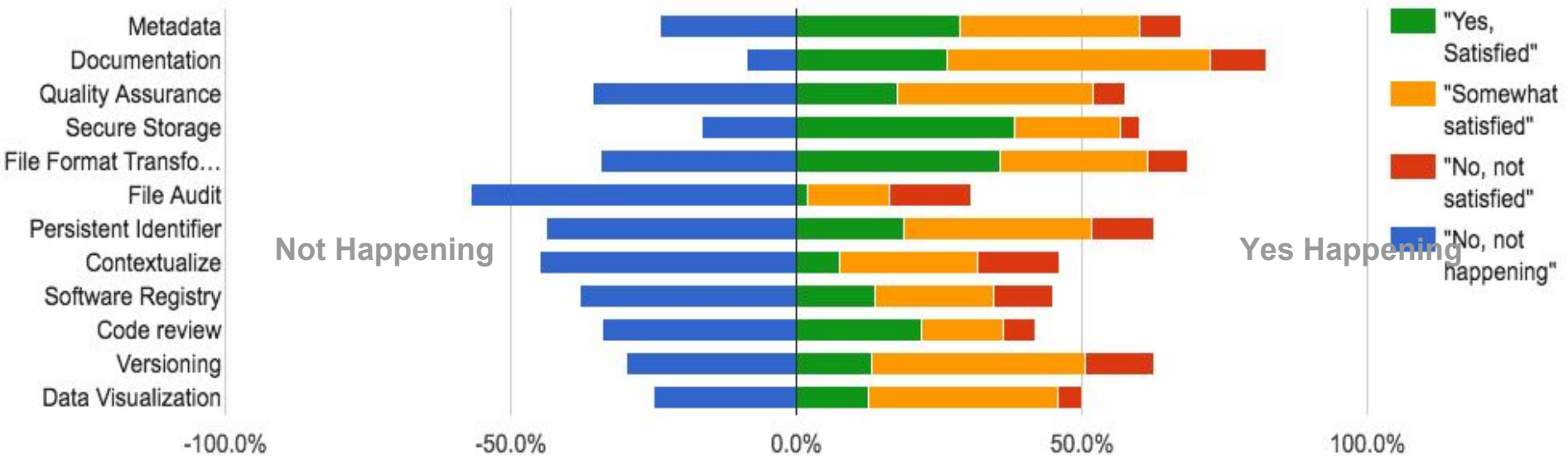
1. Develop standards-driven data curation techniques for all types of repository workflows and infrastructure.
2. Expand into a sustainable entity that grows beyond our initial six partner institutions.
3. Datasets curated by the Data Curation Network will be used to advance research and education in ways that are measurably of greater reuse value than non-curated data.
4. Build an innovative community that enriches capacities for data curation writ large.

Collab tips that worked for the Data Curation Network

1. Develop a shared vision
2. Define the need with evidence (e.g. track demand)

Important but not satisfactorily happening (n=91 researchers)

Does this happen for your data and if so, are you satisfied?



Not Happening (Left) and Level of Satisfaction when Happening (Right)

Gaps: Very important but not happening >50%

Most Important Activities (4 out of 5)

- **(Create) Documentation (4.6)**
- **Secure Storage (4.4)**
- **Quality Assurance (4.3)**
- **Persistent Identifier (4.3)**
- **Software Registry (4.1)**
- **Data Visualization (4.0)**
- **File Audit (4.0)**
- **(Create) Metadata (4.0)**
- **Versioning (3.9)**
- **Contextualization (3.9)**
- **Code Review (3.9)**
- **File Format Transformations (3.9)**

Not Happening for Majority of Researchers

- **Persistent Identifier (37% happens)**
- **Software Registry (41% happens)**
- **File Audit (16% happens)**
- **Contextualization (38% happens)**
- **Code Review (38% happens)**

Happening, but not satisfactorily

- **Documentation (26% satisfied),**
- **Secure storage (38% satisfied),**
- **Quality Assurance (14% satisfied),**
- **Data Visualization (12.5% satisfied),**
- **Metadata (29% satisfied)**
- **Versioning (13% Satisfied)**
- **File Format Transformations (29% satisfied)**

Gaps: Happening but not in a satisfactory way

Most Important Activities (4 out of 5)

- **(Create) Documentation (4.6)**
- **Secure Storage (4.4)**
- **Quality Assurance (4.3)**
- **Persistent Identifier (4.3)**
- **Software Registry (4.1)**
- **Data Visualization (4.0)**
- **File Audit (4.0)**
- **(Create) Metadata (4.0)**
- **Versioning (3.9)**
- **Contextualization (3.9)**
- **Code Review (3.9)**
- **File Format Transformations (3.9)**

Not Happening for Majority of Researchers

- **Persistent Identifier (37% happens)**
- **Software Registry (41% happens)**
- **File Audit (16% happens)**
- **Contextualization (38% happens)**
- **Code Review (38% happens)**

Happening, but not satisfactorily

- **Documentation (26% satisfied),**
- **Secure storage (38% satisfied),**
- **Quality Assurance (14% satisfied),**
- **Data Visualization (12.5% satisfied),**
- **Metadata (29% satisfied)**
- **Versioning (13% Satisfied)**
- **File Format Transformations (29% satisfied)**

Collab tips that worked for the Data Curation Network

1. Develop a shared vision
2. Define the need with evidence (e.g. track demand)
3. Get to know one another, *really* (perception vs practice)

Baseline Assessments and Pilots

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

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 to accept or reject	Auto Accept	Go Live Here	As needed	Review metadata only	Review files and metadata	Add DOI
Minnesota	X			X	X		X			X	X
Cornell	X		X*	X		X	X			X	X*
Illinois	X			X		X	X			X*	X
Michigan	X			X		X	X			X*	X*
Penn State	X			X		X	X				
Wash U	X		X	X		X	X			X	X

* On request

Published as: Johnston, Lisa R., Jake R. Carlson, Patricia Hswe, Cynthia Hudson-Vitale, Heidi Imker, Wendy Kozlowski, Robert K. Olendorf, and Claire Stewart. 2017. "Data Curation Network: How Do We Compare? A Snapshot of Six Academic Library Institutions' Data Repository and Curation Services." *Journal of eScience Librarianship* 6(1): e1102. <https://doi.org/10.7191/jeslib.2017.1102>.

Collab tips that worked for the Data Curation Network

1. Develop a shared vision
2. Define the need with evidence (e.g. track demand)
3. Get to know one another, *really* (perception vs practice)
4. Dedicate time and staff to the project

DCN Project Team



Lisa Johnston (PI)
Research Data Management &
Curation Lead,
University of Minnesota



Jake Carlson
Research Data Services
Manager,
University of Michigan



Robert Olendorf
Science Data
Librarian,
Penn State University



Cynthia Hudson-Vitale
Digital Data Outreach
Librarian,
Washington University in
St. Louis



Heidi Imker
Director of the Research
Data Service,
University of Illinois



Wendy Kozlowski
Data Curation Specialist,
Cornell University



Claire Stewart
Associate University Librarian for
Research and Learning,
University of Minnesota

Collab tips that worked for the Data Curation Network

1. Develop a shared vision
2. Define the need with evidence (e.g. track demand)
3. Get to know one another, *really* (perception vs practice)
4. Dedicate time and staff to the project
5. Learn from others!

Peer Engagement

Shared staffing models



Large-scale data repositories



Collaborative digital libraries



Governmental data sharing



Emerging disciplinary data curation groups



Planning the Data Curation Network

To Do List:

- ✓ **Compare local practice** of data repository workflows, policies, and staffing.
- ✓ **Track baseline metrics** and monitor the demand for (types, disciplines, frequency) and effort (e.g., cost, time, expertise) involved with curating data.
- ✓ **Seek input from researchers** to better understand how data curation services fit into their research workflow.
- ✓ **Pilot data curation** workflows and determine staff training needs.
- ✓ **Engage our peers** through interviews and surveys.
- ✓ **Develop a model** for shared data curation services detailing the projected staffing, costs, skills sets, and demand necessary for implementation.
- ✓ **Establish a financial model** for sustainable continuation and growth.

(Future) **Implementation Phase** will

- Test the model across our partner institutions
- Grow and sustain the Network beyond original partners

Thanks!

Web: <https://sites.google.com/site/DataCurationNetwork>

Twitter #DataCurationNetwork