

IASSIST 2017

Data Curation

Perception and Practice

Lisa Johnston, Jake Carlson, Cynthia Hudson-Vitale, and Wendy Kozlowski

5-24-2017

Data Curation Network

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.

Our Mission

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.

<https://sites.google.com/site/datacurationnetwork/>

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
Data Services Coordinator,
Washington University in
St. Louis



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

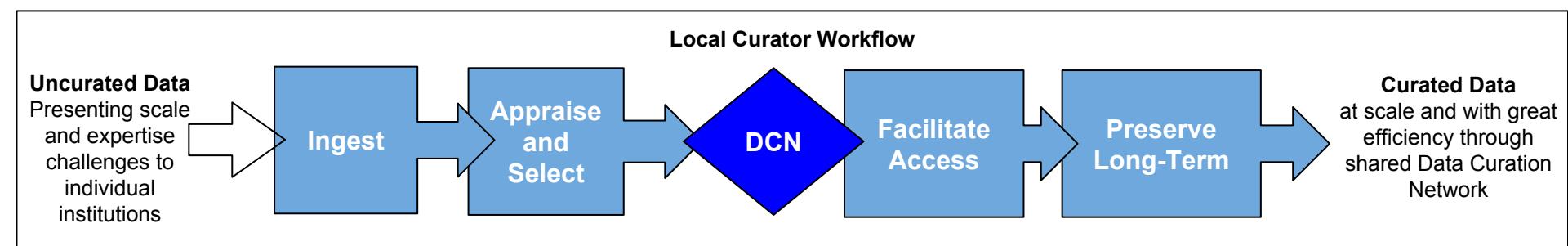


Wendy Kozlowski
Data Curation Specialist,
Cornell University

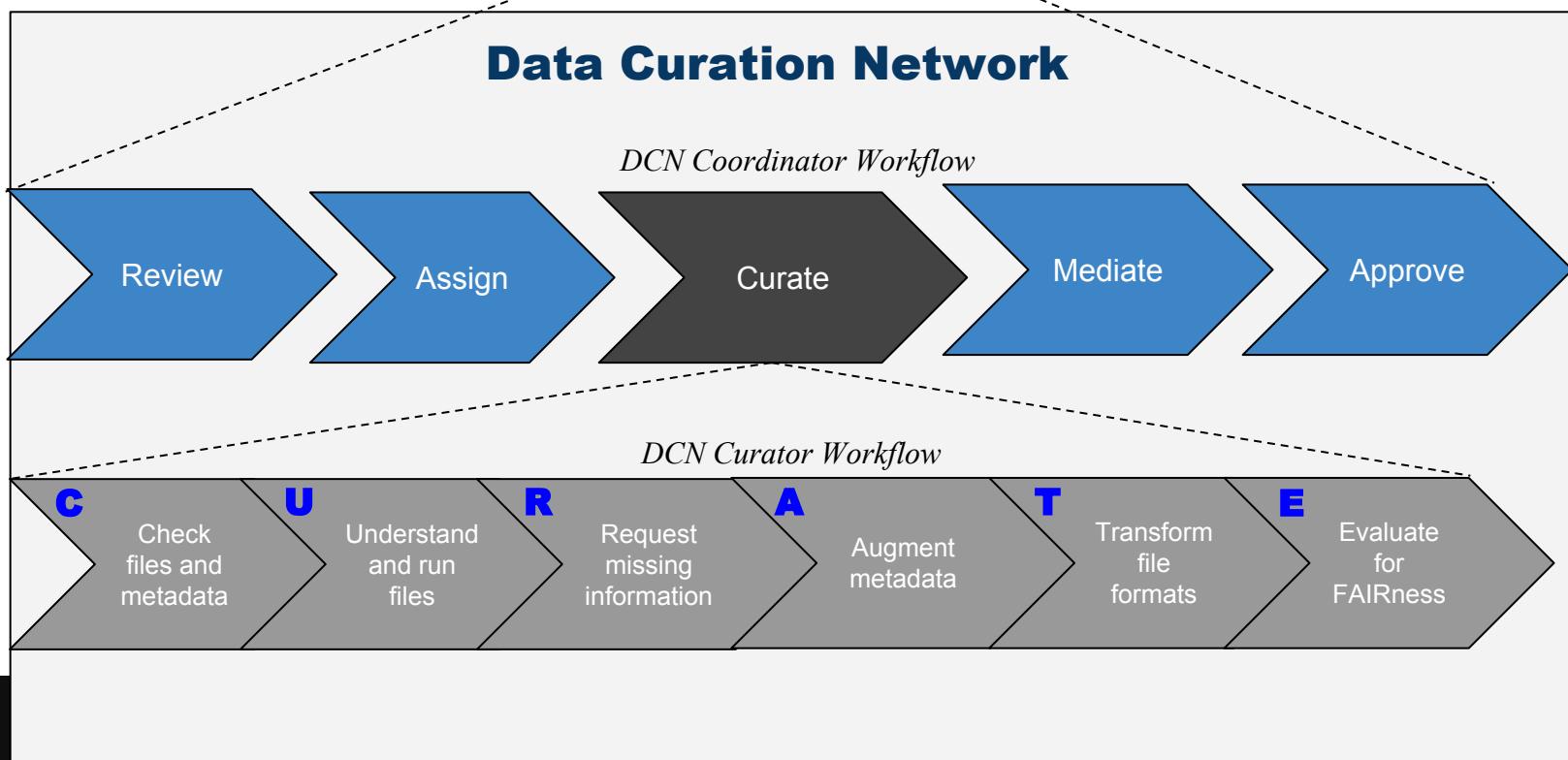
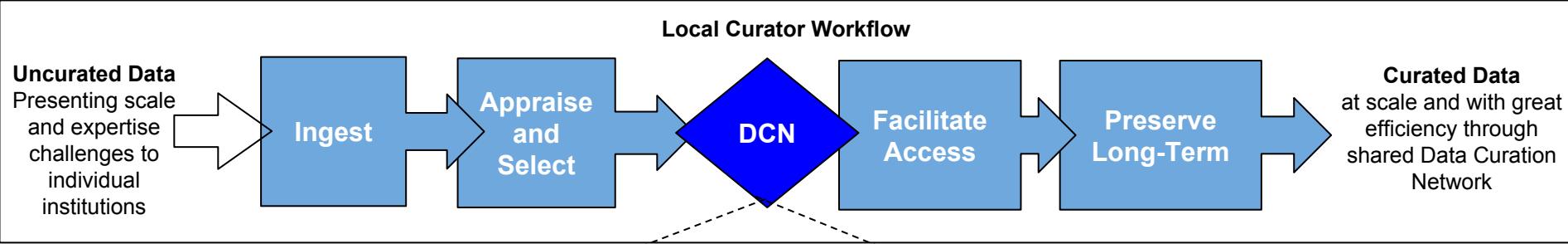


Claire Stewart
Associate University Librarian for
Research and Learning,
University of Minnesota

Model for the Data Curation Network



Model for the Data Curation Network



Perception vs Practice for Data Curation

Focus for our panel today....

- *Perception*
 - Perceived importance of data curation activities by **researchers**
 - Perceived importance of data curation activities by **libraries**
- *Practice*
 - Panel: Four snapshots of institutional data curation services in practice
 - Results of piloting the Data Curation Network
- Unpacking the Findings
- Discussion with the Audience



Perception

Results of Engagement with Researchers and Librarians



Perceptions of Data Curation



Source: Flickr - <https://flic.kr/p/ccmWs>

- What data curation activities are important to Researchers?
- What data curation activities are important to Librarians?
- Are there gaps between the perceptions of researchers and librarians?

Data Curation Activities

Data Curation Activity	Data Curation Network Definition
Documentation	Information describing any necessary information to use and understand the data. Documentation may be structured (e.g., a code book) or unstructured (e.g., a plain text Readme file).
Chain of custody	Intentional recording of provenance metadata of the files (e.g., metadata about who created the file, when it was last edited, etc.) in order to preserve file authenticity when data are transferred to third-parties.
Secure Storage	Data files are properly stored in a well-configured (in terms of hardware and software) storage environment that is routinely backed-up and physically protected. Perform routine fixity checks (to detect degradation or loss) and provide recovery services as needed.

Full list of Activities: <http://bit.ly/DCNcurationActivities>

Poll: Importance of Data Curation Activities

**** Participation implies consent for us to share the deidentified results**

POLL: The first IASSIST conference was held in 1971 in Daytona Beach, FL with 29 representatives from Canada and the US. How many IASSIST conferences have you attended?

This is my first

2-3

3-5

5-10

>10

Start the presentation to activate live content

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

DOCUMENTATION: Information describing necessary information to use and understand the data. Documentation may be structured (e.g., a code book) or unstructured (e.g., a plain text "Readme" file).

1 – This activity is not important

2

3

4

Start the presentation to activate live content
activity is very important

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

POSITIONING: Provide mechanisms to ingest multiple versions of the data overtime that includes metadata describing the version history and any changes made for each version.

1: This activity is not important

2
3
4

activity is very important

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

RESTRICTED ACCESS (Data Enclave): In order to maintain the privacy of research subjects without losing integral components of the data, some data access will be protected and/or mediated to individuals that meet predefined criteria.

1: This activity is not important

2

3

4

Start the presentation to activate live content

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

CONTENT IDENTIFIER: A URL (or Uniform Resource Locator) monitored by an authority to ensure a stable web location for consistent citation and long-term discoverability. Provides redirect when necessary (e.g., DOI or handle).

1: This activity is not important

2

3

4

5: This activity is very important

Start the presentation to activate live content

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

DE REVIEW: Run and validate computer code, look for missing files and/or errors) in order to find mistakes overlooked in the initial development phase, improving the overall quality of software.

1: This activity is not important

2

3

4

Start the presentation to activate live content

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

Researcher Engagement Events



Source:

<https://unsplash.com/search/photos?photo=PJzc7LOt2Ig>

Focus groups (Oct-Nov 2016) at each of the 6 DCN partner institutions asked:

- What data curation activities are important to you?
- What data curation activities are currently being done by you or a 3rd party?
- If the data curation activity is being performed, how satisfied are you with the results?

Data Curation Activities

Each institution selected Data Curation Activities from our list of 35 possibilities.

Metadata

Information about a data set that is structured (often in machine-readable format) for purposes of search and retrieval. Metadata elements may include basic information (e.g. title, author, date created, etc.) and/or specific elements inherent to datasets (e.g., spatial coverage, time periods).

Rate how important this activity is to you.

(Write a number 1-5 with 5 = highest importance, 1 = not important)

Round 1	Round 2	Round 3	Round 4
---------	---------	---------	---------

Full list of Activities: <http://bit.ly/DCNcurationActivities>

Importance of Curation Activities

Rank	Curation Activity	CU	WU	IL	PSU	MN	MI	# Inst.	Total Responses	1-5 Avg.
#1	Documentation	✓	✓	✓	✓	✓	✓	6	91	4.6
#2	Secure Storage	✓	✓		✓		✓	4	60	4.4
#3	Quality Assurance	✓	✓	✓	✓	✓		5	73	4.3
#4	Persistent Identifier									
		✓	✓	✓	✓	✓	✓	6	91	4.3
#5	Software Registry		✓				✓	2	29	4.1
#6	Data Visualization		✓			✓		2	24	4.0

For activities rated by more than one institution. Ranking on a 1 to 5 scale with 5 = Most Important
 (CU = Cornell, PSU = Penn State, IL = Illinois, WU = Washington University, MI = Michigan, MN = Minnesota).

Researcher Engagement Events Results

Rank	Curation Activity	CU	WU	IL	PSU	MN	MI	# Inst.	Total Responses	1-5 Avg.
#7	File Audit	✓		✓	✓			3	49	4.0
#8	Metadata	✓		✓	✓	✓	✓	5	80	4.0
#9	Versioning	✓	✓	✓	✓	✓	✓	6	91	3.9
#10	Contextualize	✓	✓	✓	✓	✓	✓	6	91	3.9
#11	Code Review	✓	✓	✓	✓	✓	✓	6	29	3.9
#12	File Format Transformations	✓	✓	✓	✓	✓		5	73	3.8

For activities rated by more than one institution. Ranking on a 1 to 5 scale with 5 = Most Important
 (CU = Cornell, PSU = Penn State, IL = Illinois, WU = Washington University, MI = Michigan, MN = Minnesota).

Researcher Engagement Events Results

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)**

Researcher Engagement Events Results

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)

Poll: Data Curation Activities Taking Place

**** Participation implies consent for us to share the deidentified results**

CUMENTATION: Information describing necessary information to use and understand the data. Documentation may be structured (e.g., a code book) or unstructured (e.g., a plain text "Readme" file).

No, this is not happening for my data
Yes, this is happening but I'm not satisfied with it
Yes, this is happening and I'm somewhat satisfied with it

Yes, this is happening and I'm satisfied with it

I don't know!

Start the presentation to activate live content

N/A If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

POSITIONING: Provide mechanisms to ingest multiple versions of the data overtime that includes metadata describing the version history and any changes made for each version.

No, this is not happening for my data

Yes, this is happening but I'm not satisfied with it

Yes, this is happening and I'm somewhat satisfied with it

Yes, this is happening and I'm satisfied with it

I don't know!

N/A for me or my data

Start the presentation to activate live content

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

RESTRICTED ACCESS (Data Enclave): In order to maintain the privacy of research subjects without losing integral components of the data, some data access will be protected and/or mediated to individuals that meet predefined criteria.

No, this is not happening for my data
Yes, this is happening but I'm not satisfied with it

Yes, this is happening and I'm somewhat satisfied with it

Yes, this is happening and I'm satisfied with it

I don't know!

Start the presentation to activate live content

N/A If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

PERSISTENT IDENTIFIER: A URL (or Uniform Resource Locator) that is monitored by an authority to ensure a stable web location for consistent citation and long-term discoverability. Provides redirection when necessary (e.g., DOI or handle).

No, this is not happening for my data

Yes, this is happening but I'm not satisfied with it

Yes, this is happening and I'm somewhat satisfied with it

Yes, this is happening and I'm satisfied with it

Start the presentation to activate live content

I don't know!

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

N/A for me or my data

DE REVIEW: Run and validate computer code, look for missing files and/or errors) in order to find mistakes overlooked in the initial development phase, improving the overall quality of software.

No, this is not happening for my data
Yes, this is happening but I'm not satisfied with it
Yes, this is happening and I'm somewhat satisfied with it

Yes, this is happening and I'm satisfied with it

1

2

3

4

5

6

I don't know!
Start the presentation to activate live content

N/A If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

ARL SPEC Kit: Data Curation



Survey of the 124 members of the Association of Research Libraries

Distributed in January 2017

80 libraries responded (65%)

- 51 currently provide services
- 13 are developing services

<http://publications.arl.org/Data-Curation-SPEC-Kit-354/>

Data Curation @ ARL Libraries

Rschr. Rank	Activity	Offer	Plan to Offer	Percentage of Respondents
#1	Documentation	36	3	80%
#2	Secure Storage	39	2	84%
#3	Quality Assurance	22	1	47%
#4	Persistent Identifier	40	2	86%
#5	Software Registry	4	2	12%
#6	Data Visualization	14	4	37%

Number of ARL institutions = 49

Data Curation @ ARL Libraries

Rschr. Rank	Activity	Offer	Plan to Offer	Percentage of Respondents
#7	File Audit	21	7	57%
#8	Metadata	43	1	90%
#9	Versioning	24	3	55%
#10	Contextualization	28	4	65%
#11	Code Review	4	1	10%
#12	File Format Transformation	25	5	61%

Number of ARL institutions = 49

Libraries Plan to Support ... (or do we?)

Rsch. Rank	Data Curation Activity	Currently providing	Plan or would like to provide	No Interest	Unsure
#31	Repository Certification	3	30	5	10
#11	Code Review	4	29	10	6
#32	Emulation	1	26	14	7
#23	Peer Review	1	22	20	5
#5	Software Registry	4	23	12	9
#30	Deidentification	8	25	11	5
#17	Interoperability	11	28	5	4



Practice

Institutional Data Curation Service

Case Studies



DCN Baseline Assessment

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

	Pre-ingest Curation?		Mediated vs Self-deposit?		Accept/Reject Stage?		Public	Post-ingest curation			
Workflow Steps by Institution	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

Journal of eScience Librarianship 6(1): e1102.
<https://doi.org/10.7191/jeslib.2017.1102>

Local Practice: Cornell University

eCommons
digital repository
ecommons.cornell.edu

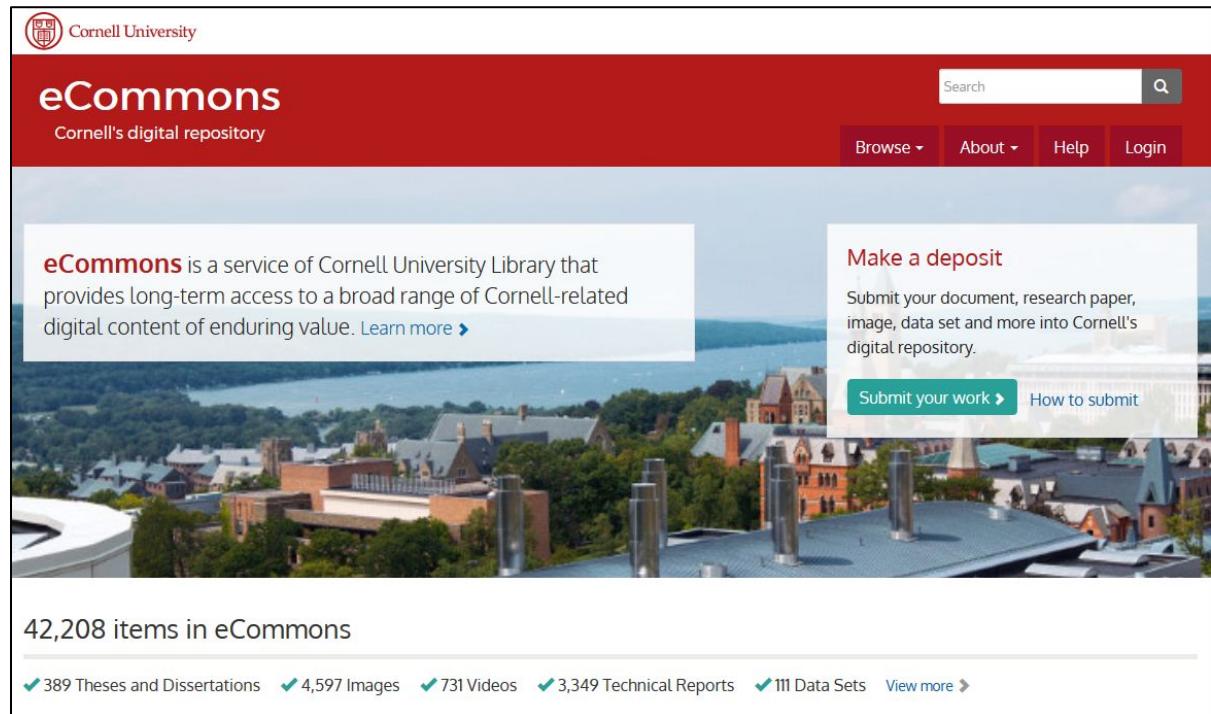
Launched: 2002

First dataset: 2003 / 2005

Datasets published: 111

Features:

- DSpace 5.5
(Dublin Core)
- Curation encouraged
- Versioning encouraged
- Handles for all items, and DataCite DOI's upon request
- “Embargos” for delayed publication needs
- Links to associated publications



The screenshot shows the homepage of the eCommons digital repository. At the top, there is a header with the Cornell University logo and the text "eCommons Cornell's digital repository". A search bar and navigation links for "Browse", "About", "Help", and "Login" are also present. Below the header, a large image of the Cornell University campus is displayed. On the left side of the main content area, there is a box containing text about eCommons and a "Learn more" link. On the right side, there is a box titled "Make a deposit" with a "Submit your work" button and a "How to submit" link.

eCommons is a service of Cornell University Library that provides long-term access to a broad range of Cornell-related digital content of enduring value. [Learn more](#)

42,208 items in eCommons

✓ 389 Theses and Dissertations ✓ 4,597 Images ✓ 731 Videos ✓ 3,349 Technical Reports ✓ 111 Data Sets [View more](#)

Local Practice: Cornell University

Staffing

IR Administration and Development



Gail Steinhart
Repository
Manager



Chloe McLaren
Admin and Policy
Development



Mira Basara
Admin and Batch
Upload Support



George Kozak
System Maintenance
and Development

Data-specific Support and Curation



Wendy Kozlowski
Data Curation
Specialist



Erica Johns
Research Data
and Environmental
Sciences



Data expertise from across
Cornell - available for
consultation on specific issues

Local Practice: University of Michigan

Deep Blue Data

<https://deepblue.lib.umich.edu/data>

Soft Launch: Feb 2016

Official Launch: Sept 2016

96 entries

- 60 Works
- 36 Collections

Features:

- Hydra Fedora (Dublin Core)
- DataCite DOIs
- Extension of Deep Blue (UM's IR)

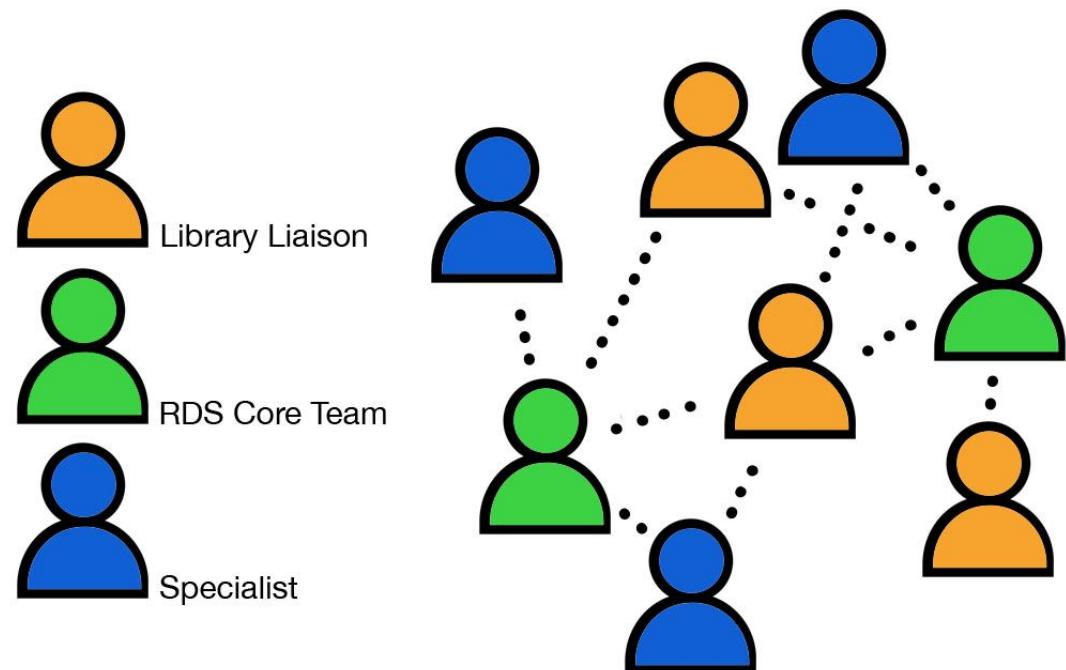
The screenshot shows the homepage of the Deep Blue Data repository. At the top, there is a navigation bar with links for Home, About, Help, Contact, a search bar, and a login link. Below the navigation bar, a main banner area contains text about the repository's purpose and two calls-to-action: 'Browse' and 'Deposit Your Work'. The central part of the page is titled 'Featured Works' and displays three items, each represented by a blue cube icon and a title. The first item is 'Neighborhood Effects...', the second is 'The Observation and...', and the third is 'Citations to Open and...'. Each item also includes a thumbnail image, the depositor's name, and their email address.

Work Title	Deposited by	Email
Neighborhood Effects...	tveinot@umich.edu	
The Observation and...	dcsoztrk@umich.edu	
Citations to Open and...	hellpop@umich.edu	

Local Practice: University of Michigan

Roles in Research Data Services

- **RDS Core Team**
Oversee services including Deep Blue Data and the curation process
- **Liaisons**
The front line in offering RDS. Consult on data deposit and curation
- **Specialists**
Apply expertise as needed



University of Minnesota Data Repository for U of M (DRUM)

Launched: March 2015
124 data sets published
Curation: Required

Features:

- DSpace 6.0 (Dublin Core)
- DataCite DOIs
- Versioning
- Web of Science Indexing
- Documentation assistance (e.g., readme.txt template)



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.

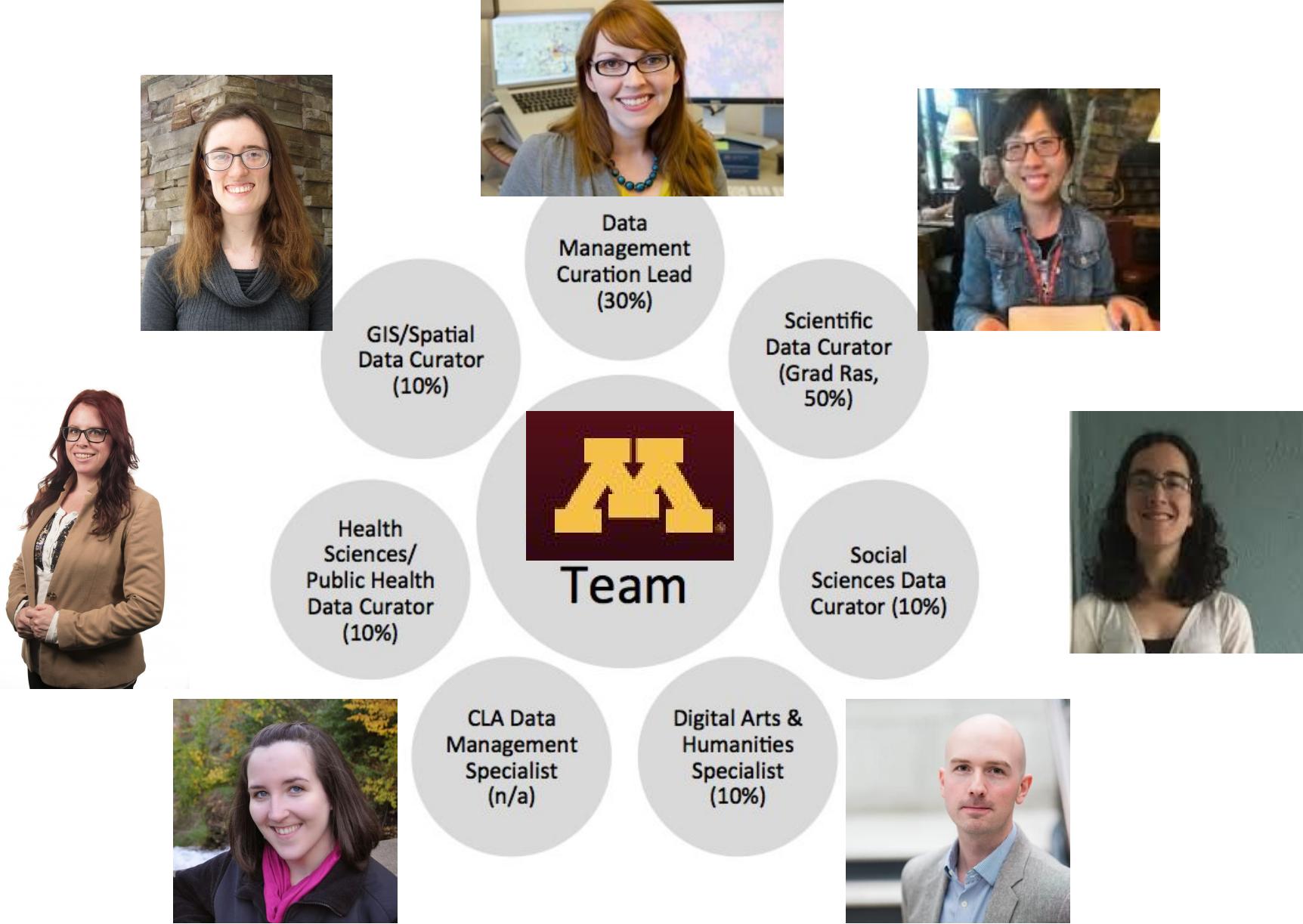
Our Services

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.



Local Practice: Washington University

Digital Research Materials Repository

Launched -
Soft launch 2015

Features -

- BePress Platform
- DataCite DOI's
- Restricted access
- 10 year

preservation then review by subject liaison

The screenshot shows the Washington University Open Scholarship homepage. On the left, there is a sidebar with search and browse functions. The main content area features a large graphic of three interlocking gears labeled 'D', 'R', and 'RM'. To the right of the graphic, the text 'DIGITAL RESEARCH MATERIALS (DATA & SUPPLEMENTAL FILES)' is displayed. Below the graphic, the text 'Digital Research Materials Repository' is written. At the bottom, there is a paragraph about the purpose of the repository and a list of benefits.

WASHINGTON UNIVERSITY
OPEN SCHOLARSHIP

Home About FAQ My Account

Enter search terms:
 Search
 Advanced Search
Notify me via email or RSS

Links
Data Management Research Guide

Browse
Collections
Disciplines
Authors

Author Corner
Author FAQ
Policies
Submission Guidelines
Submit Research

Home > DATA

DIGITAL RESEARCH MATERIALS (DATA & SUPPLEMENTAL FILES)

Digital Research Materials Repository

This Digital Research Materials series is a place for WUSTL affiliates, including faculty, students, and affiliated researchers, to share and publish digital data and supplemental files for long-term access and future use. While some academic disciplines have established research data repositories, many fields of research do not have easily available options for archiving and online access.

Benefits include:

- **Flexible Access Options:** Make your data immediately accessible to all, or moderate access to your data upon request.
- **Long-term Access:** Persistent links and [identifiers](#) (DOIs) make it easy for others to cite your data.
- **Analytics:** Track how often your data are viewed and downloaded.
- **Meet Grant Requirements:** Comply with federal mandates for data management planning (DMP)

Local Practice: Washington University

Staffing



Jennifer
Moore,
GIS & Data
Projects
Manager



Lauren Todd,
Engineering
Librarian



Daria
Carson-Dussan,
Romance
Languages
Librarian



Micah Zeller,
Copyright
Specialist



Emily Stenberg,
Repository
Manager



Cynthia
Hudson-Vitale,
Data Services
Coordinator



Practice

Data Curation Pilots

DCN Curation Pilots

Fall of 2016 → conducted two rounds of controlled data curation pilots among 16 institutional data curation staff to:

1. Identify actual and individual curation practices taken at partner DCN institutions (compare).
2. Establish training needs of DCN curators.
3. Identify any issues, misaligned expectations, and/or conflicts with the goals of the project.

	Minnesota	Cornell	Penn State	Illinois	Michigan	Wash U
Curators	3	2	2	3	3	3
Example Expertise	Phys Sci, genomics	Earth & Enviro	Software code, GIS	Bio/chem, statistical	Poly Sci, clinical	Health, GIS

DCN Curation Pilots

Rsch. Rank	Data Curation Activities*	Round 1 (n=6)	Round 2 (n=9)	Total (n=15)	% Did This
#3	Quality Assurance	6	8	14	93%
#1	(Create) Documentation	6	5	11	73%
#34	Correspondence (with author)	6	5	11	73%
#12	File Format Transformations	6	2	8	53%
#8	(Create) Metadata	4	4	8	53%
#10	Contextualization	3	4	7	47%
#28	File Inventory or Manifest	2	4	6	40%
#21	Risk Management	4	1	5	33%

*Not ranked by researchers, curators also Inspected Files, Inspected Metadata and (33%) created a working copy.

DCN Curation Pilots

- **Recommendation 1:** Assignments to curators prioritize file format and software expertise over discipline when necessary.
- **Recommendation 3:** Centralize all DCN correspondence and perform routine checks on all submissions before assigning to DCN curator.
- **Recommendation 5:** Create levels of curator criteria for curators to aim for rather than allowing curators to fall into the “never ending” quest for high standards.
- **Recommendation 7:** Data curation activities taken should differentiate between the role of the local repository curators versus the role of the Data Curation Network curator.

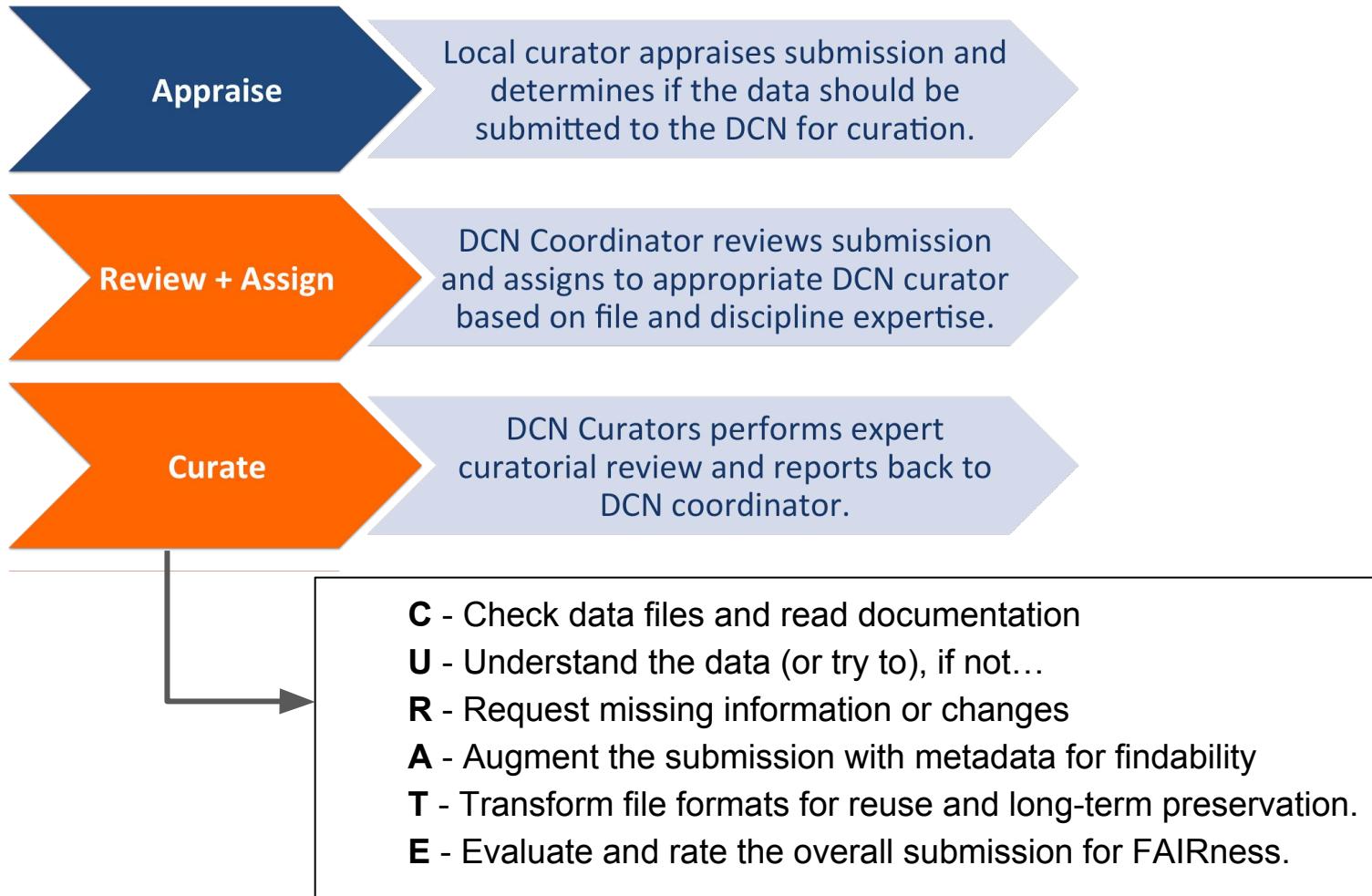
<https://sites.google.com/site/datacurationnetwork/>



Our Findings

Data Curation Network

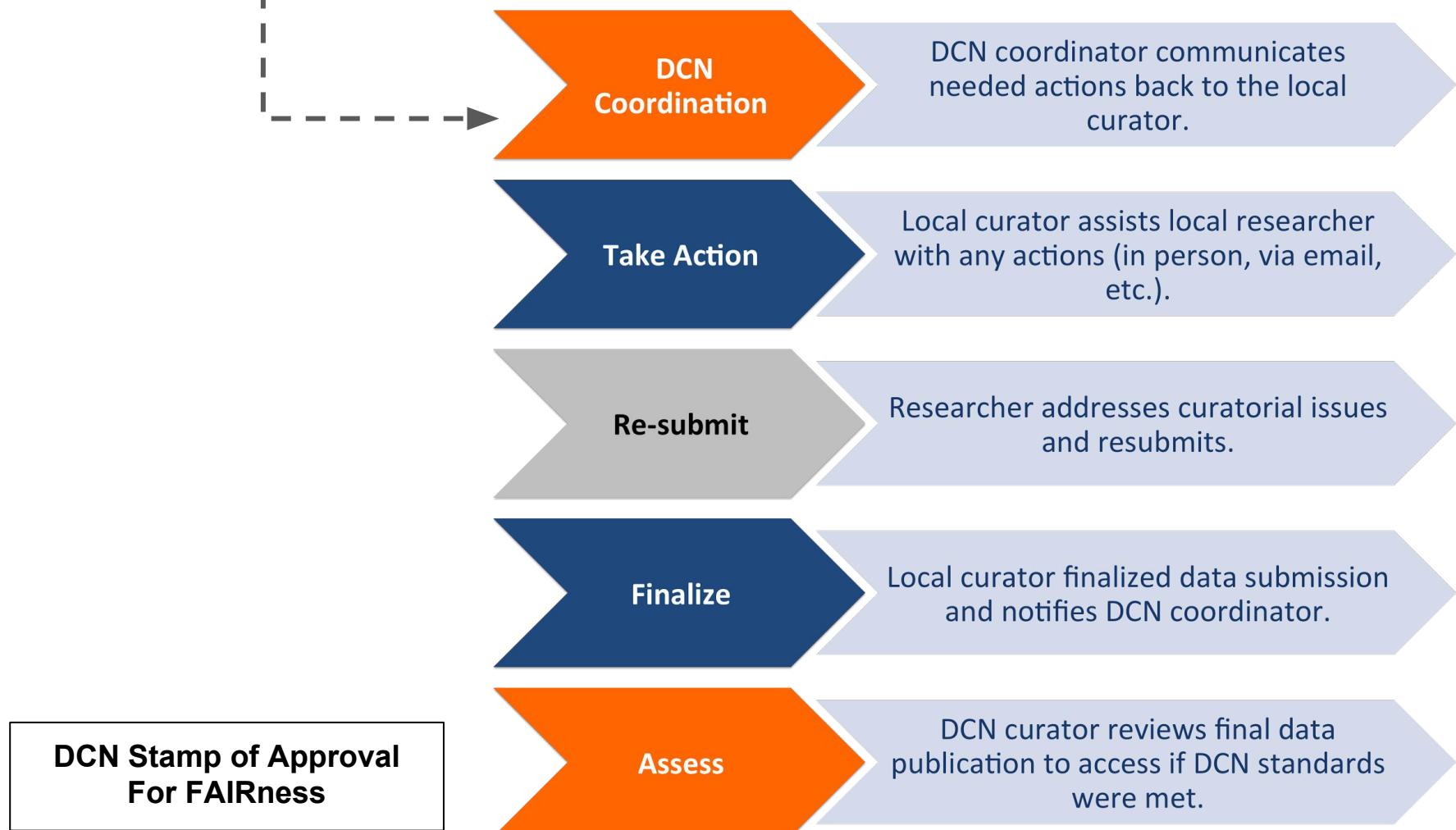
CURATE Steps



DCN Draft Procedures and Checklist

CURATE Actions	Curation Checklist
<p>Check data files and read documentation</p> <ul style="list-style-type: none">• Review the content of the data files (e.g., open and run the files or code).• Verify all metadata provided by the author and review the available documentation.	<ul style="list-style-type: none"><input type="checkbox"/> Files open as expected<ul style="list-style-type: none"><input type="checkbox"/> Issues _____<input type="checkbox"/> Code runs as expected<ul style="list-style-type: none"><input type="checkbox"/> Produces minor errors<input type="checkbox"/> Does not run and/or produces many errors<input type="checkbox"/> Metadata quality is rich, accurate, and complete<ul style="list-style-type: none"><input type="checkbox"/> Metadata has issues _____<input type="checkbox"/> Documentation Type (circle) Readme / Codebook / Data Dictionary / Other: _____<input type="checkbox"/> Missing/None

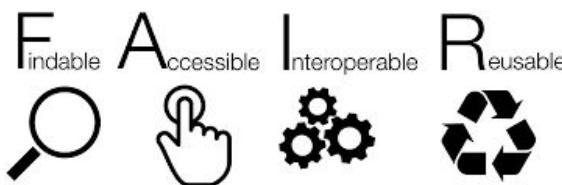
Local Control



DCN Draft Procedures and Checklist

Evaluate and the overall data record for FAIRness.

- Score the dataset and recommend ways to increase the FAIRness of the data and become "DCN approved."



Findable -

- Metadata exceeds author/ title/ date,
- Unique PID (DOI, Handle, PURL, etc.)
- Discoverable via web search engines like Google

Accessible -

- Retrievable via a standard protocol (e.g., HTTP)
- Free, open download.

Interoperable -

- Metadata formatted in a standard schema (e.g., Dublin Core)
- Metadata provided in machine-readable format (OAI feed)

Reusable -

- Data include sufficient metadata about the data characteristics to reuse without the direct assistance of the author
- Clear indicators of who created, owns, and stewards the data help ensure trust
- Data are released with clear data usage

Data Curation Network Outcomes

- 1. Standards-driven data curation techniques for all types of repository workflows and infrastructure.**
- 2. 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. An innovative community that enriches capacities for data curation writ large.**

Results

Poll 1: Curation Activity Importance

Activity	Average @ DCN Engagement Events	Average @ IASSIST 2017
Documentation	4.6	4.7
Persistent Identifier	4.3	4.3
Versioning	3.9	4.1
Code Review	3.9	3.6
Restricted Access	2.6	4.0

Results (partial)

Poll 2: Curation Activity Engagement and Satisfaction

Activity	Happening @ DCN Events	Happening @ IASSIST 2017	Satisfied / Somewhat Satisfied @ DCN events	Satisfied / Somewhat Satisfied @ IASSIST 2017
Documentation	80%	89%	26 / 46 (72%)	15 / 54 (69%)
Persistent Identifier	37%	74%	19 / 33 (52%)	8 / 15 (23%)
Versioning	56%	68%	13 / 37 (50%)	25 / 20 (45%)
Code Review	39%	31%	22 / 14 (36%)	10 / 13 (23%)
Restricted Access	38%	67%	21 / 4 (25%)	12 / 26 (38%)

Discussion

Data Curation Perception vs Practice

Discussion Question

- What do we value in terms of curation and how does that align with our approach?
- How do we measure the the value/impact of data curation?
- How do you known when data are “well curated” ? What does “done” look like? Is it a matter of time spend, effort made, perfection?

Thanks!

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

Twitter #DataCurationNetwork

Data Curation Network



Planning a network of expertise model for curating research
data in academic libraries

2016-2017

Planning the Data Curation Network project is supported by a grant from the
ALFRED P. SLOAN FOUNDATION.