

DATA **CURATION** **NETWORK**

A Curator's Perspective

IDCC 2019

**Debra Fagan, Erin Clary and
Lisa Johnston**

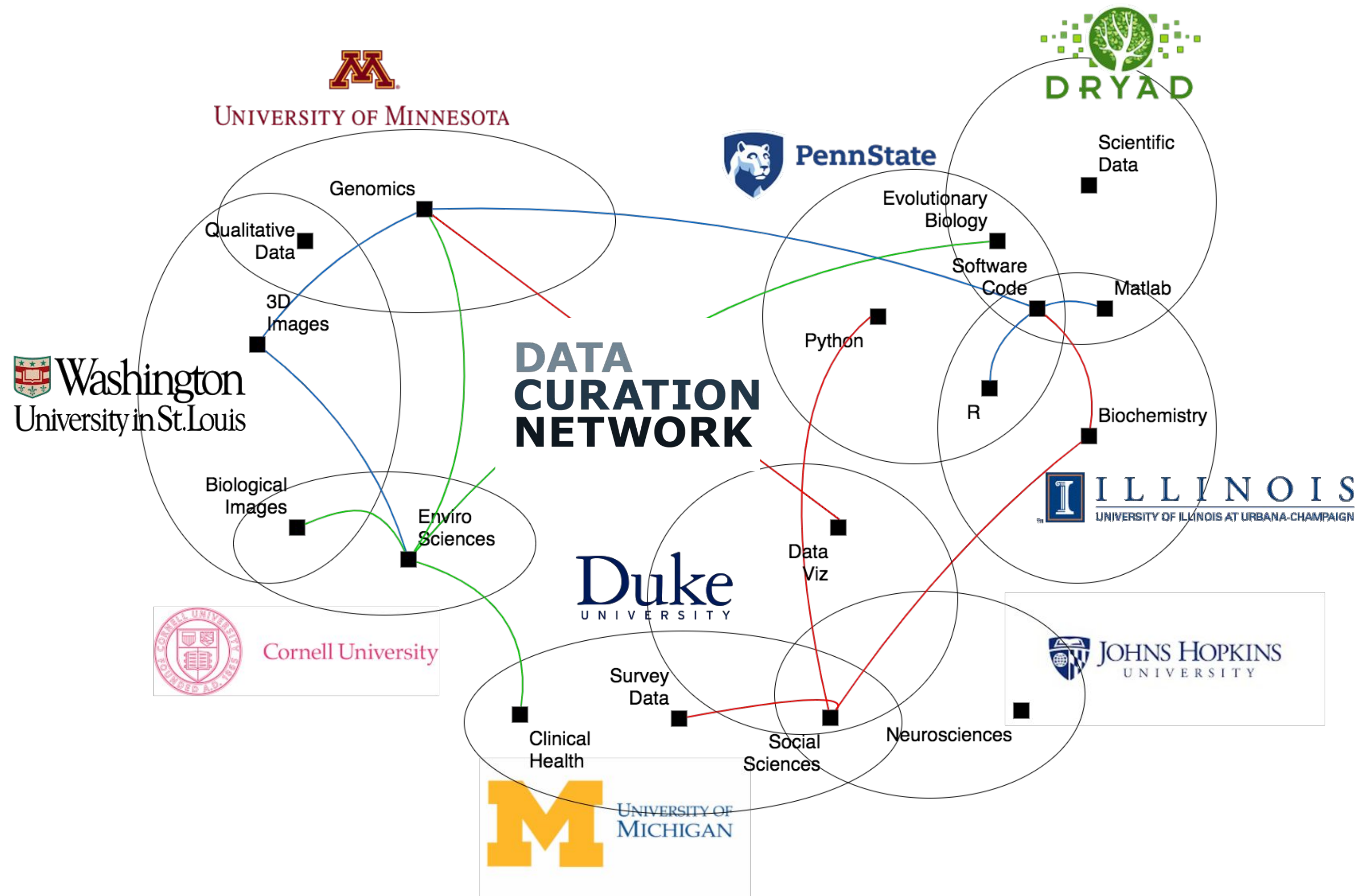
DATA CURATION NETWORK

Our mission

- Develop standards-driven curation techniques
- Enhance curation capacity of digital data repositories
- Build a curation community
- Provide training and other resources

Data Curation Network

- Collaborative staffing model
- Match data to curation experts
- Alfred P. Sloan Foundation
- Implementation phase (2018-2021)
 - 8 partner institutions
 - 32 individuals

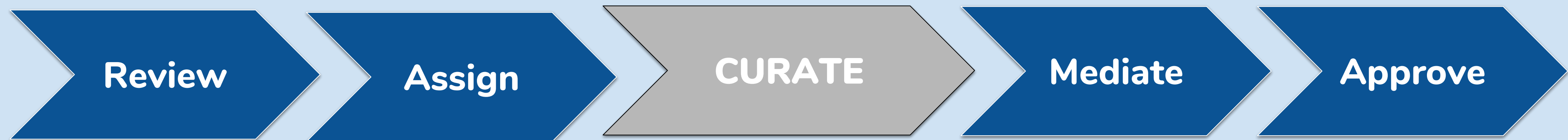


Workflow and C-U-R-A-T-E Steps



Data Curation Network

DCN Coordinator Workflow



DCN Curator Workflow





**Curated repository for
data underlying
scientific publications**

<https://datadryad.org>

DRYAD THEN AND NOW

**Ecology and
evolutionary biology
publication data**



**General purpose
repository for
scientific data**

DRYAD CURATION CHALLENGES

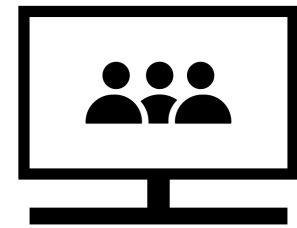
Skills and expertise needed for a wide variety of data

- Numerous file formats
- Varied community standards
- Licensing
 - Data reuse
 - Software / code

Data packages	25379
Data files	80697
Curated per week	>100
Journals	778
Integrated journals	137

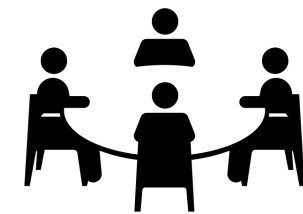
Turns out, we are not the only ones facing these challenges!

PREPARATION FOR DCN LAUNCH



Webinars

- Introduction to DCN
- “Piloting the Data Curation Network”



All Hands Meeting

- Minneapolis - summer
- Learn basic curation
 - Identified challenges
 - Shared resources
 - Met DCN members



Training

- Train the trainer
- Jira



DCN PILOT

- Jira board up and running
- Weekly standup established
- "Practice" submissions completed
- First "real" submission completed

Data Curation N...
Software project

DCN Pilots Board

Kanban board

Reports

Releases

Issues and filters

Pages

Components

CURATE Steps D...

DCN Pilots

Kanban board

Release

Quick filters

QUEUE	COORDINATOR R...	IN CURATION	LOCAL REVIEW	LOCAL SIGNOFF	CLOSED
	Counts Minneapolis None 14/Dec/18 Unassigned DCPN-4	with "Direct Effects of CO2 None 21/Dec/18 Nathan Piekielek DCPN-7	Shear Wave Speed Digital Phantoms - None 14/Dec/18 Joel Herndon DCPN-2	Pilot - CT scan data None 31/Dec/18 Rachel Woodbr... DCPN-1	Counts Minneapolis None 21/Dec/18 Lisa Johnston DCPN-14
	Comparisons in Mouse Self- Stimulation_UMich None 14/Dec/18	Comparisons in Mouse Self- Stimulation_UMich None 01/Feb/19	Neuroscience dataset from Dryad / eLife None 21/Dec/18	Nitrogen NEXAFS Spectra Dataset - Cornell None 14/Dec/18	Minnesota Bird Counts Minneapolis None 14/Dec/18

PROCESS

Checklist of CURATE Steps and FAIRness Scorecard

(published in the [DCN Planning Phase Report \(2017\)](#))

- C** **Check** files and read documentation (risk mitigation, file inventory, appraisal/selection)
- U** **Understand** the data (or try to), if not... (run files/environment, QA/QC issues, readmes)
- R** **Request** missing information or changes (tracking provenance of any changes and why)
- A** **Augment** metadata for findability (DOIs, metadata standards, discoverability)
- T** **Transform** file formats for reuse (data preservation, conversion tools, data viz)
- E** **Evaluate** for FAIRness (licences, responsibility standards, metrics for tracking use)

Notes

CHECK Step

CURATE Action	Curator Checklist
Check data files and read documentation <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. 	<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"/> Did not try to run code <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: <ul style="list-style-type: none"> <input type="checkbox"/> Missing/None <input type="checkbox"/> Needs work <input type="checkbox"/> Human subjects data present <ul style="list-style-type: none"> <input type="checkbox"/> Request consent form / participation agreement

Notes:

There are a number of tools that will help inventory a dataset (e.g., number of files; histogram of file format types, etc.). [View tools list here.](#)

- Assigned in Jira
- Updated status on Kanban board
- Had weekly stand up for brief update

MOVING FORWARD: AN ITERATIVE PROCESS

- Determine how to handle things internally on an institutional level
 - When and how to submit to DCN
 - How to handle/prioritize curation assignments
- Continue developing shared resources (including each other)
- Refine and add procedures as needed
- Handle juggling expectations/requirements of local institutions
- Ensure consistency

DCN VALUE FOR CURATORS

- Not a one-way street / not just volunteering time to do curation
- Provides networking / collaboration opportunities
- Leads to development of standards
- Opportunity to contribute to something bigger

Thanks!

Any questions?

DATA
CURATION
NETWORK

<https://datacurationnetwork.org/>

CREDITS

- Presentation template designed by [Slidesmash](#)
- Vector Icons by [Matthew Skiles](#)
- Additional Icons:
 - online meeting by Ismael Ruiz from the Noun Project
 - Meeting by Scott Lewis from the Noun Project
 - Share by Adrien Coquet from the Noun Project