DATA CURATION NETWORK

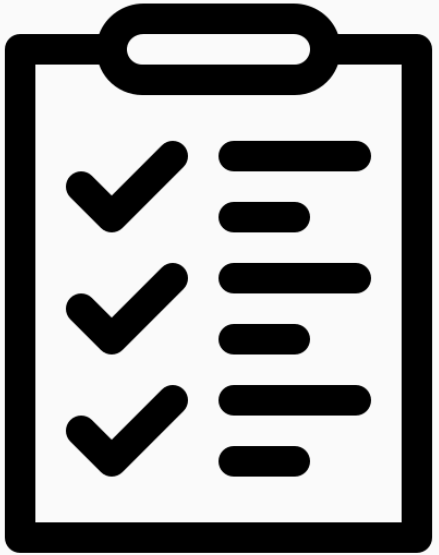
Curation + Community: About the Data Curation Network

Mikala Narlock, Director, Data Curation Network

Presentation to
Landesinitiative für Forschungsdatenmanagement fdm.nrw
May 2023

Agenda

- Why data curation?
- Overview of the DCN
  - How we were started
  - How we're structured
  - How we're funded



DATA CURATION NETWORK

Mission

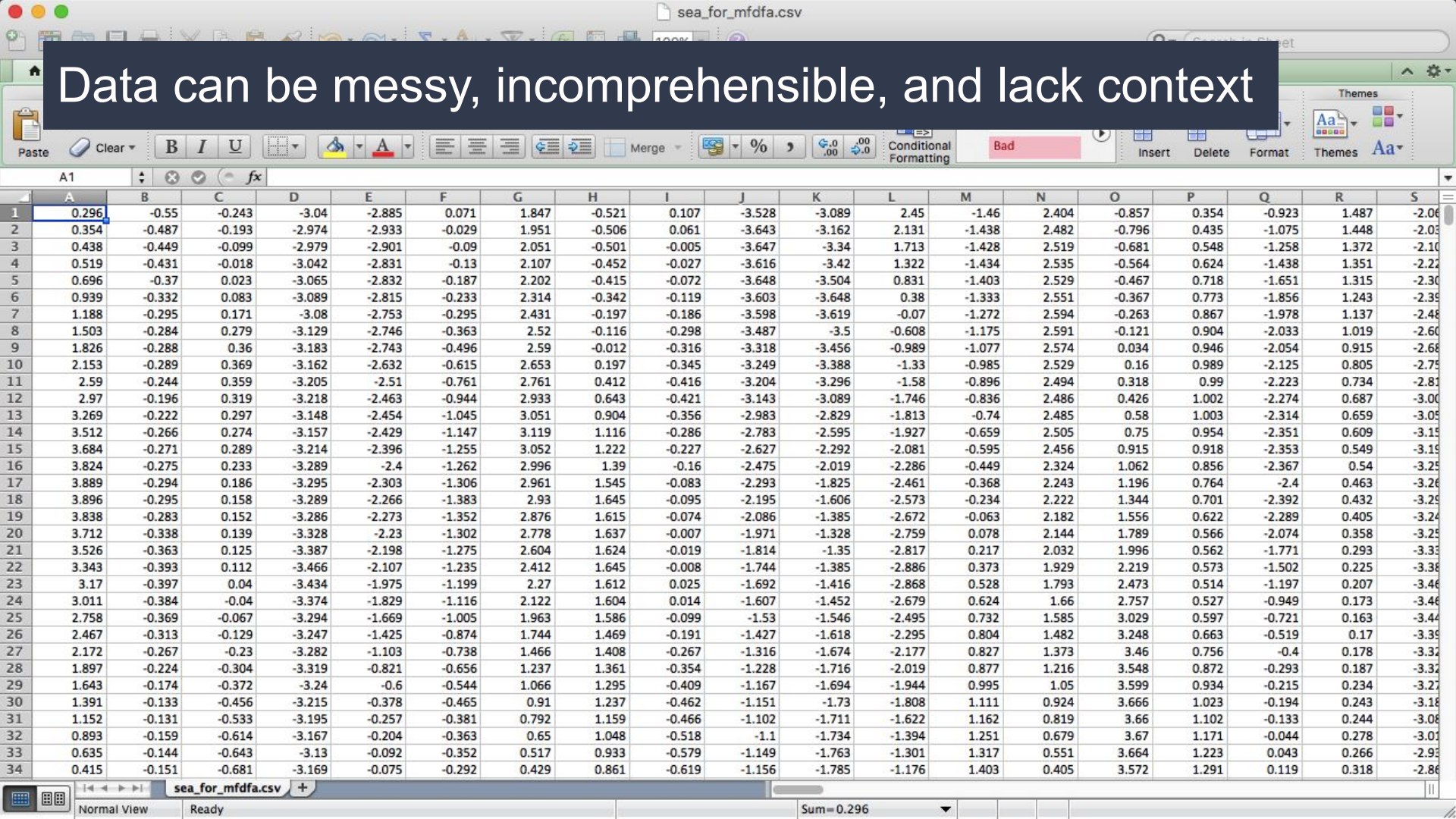
Trusted, community-led network of curators advancing open research by making data

Ethical. Reusable. Better.

Why data curation?

Research data have value beyond their original purpose.

But….


sea_for_mfdfa.csv
Data can be messy, incomprehensible, and lack context
Paste
Clear
Merge
Conditional Formatting
Bad
Insert
Delete
Format
Themes
A1
sea_for_mfdfa.csv
Normal View
Ready
Sum=0.296

Digital file formats are constantly at risk

What is data curation?

“Data curation is the active and ongoing management of data through its lifecycle and interest and usefulness to scholarship, science, and education; curation activities enable data discovery and retrieval, maintain quality, add value, and provide for re-use over time.”

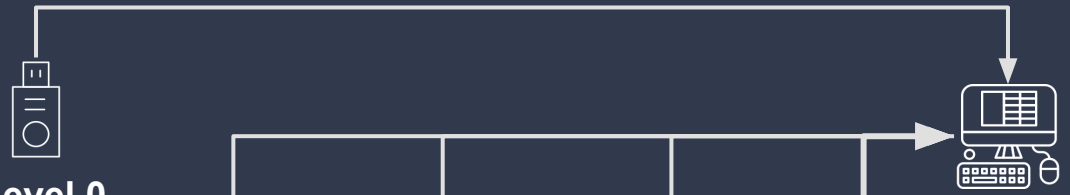
- University of Illinois’ Graduate School of Library and Information Science

What is data curation?

“Data curation is akin to work performed by an art or museum curator. Through the curation process, data are organized, described, cleaned, enhanced, and preserved for public use, much like the work done on paintings or rare books to make the works accessible to the public now and in the future.”

- ICPSR (Inter-university Consortium for Political and Social Research)

Levels of Curation



Distributed as deposited

Level 0

Record Level Curation
perform brief metadata checks for increased **findability (F)**

Level 1

File Level Curation
review files arrangement and perform file format conversions for increased **accessibility (A)**

Level 2

Documentation Level Curation
review documentation and request/add missing information for increased **reusability (R)**

Level 3

Data Level Curation
open files and review data contents and may annotate /edit the data for accuracy or **interoperability (I)**

Level 4

Who assists in curation?

Data curation involves

- Automated curation actions taken by **repositories**
- Decisions and actions taken by the **curator**
- Engagement with the **researcher**

Definitions of Data Curation Activities used by the DCN

Overview of the DCN

Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

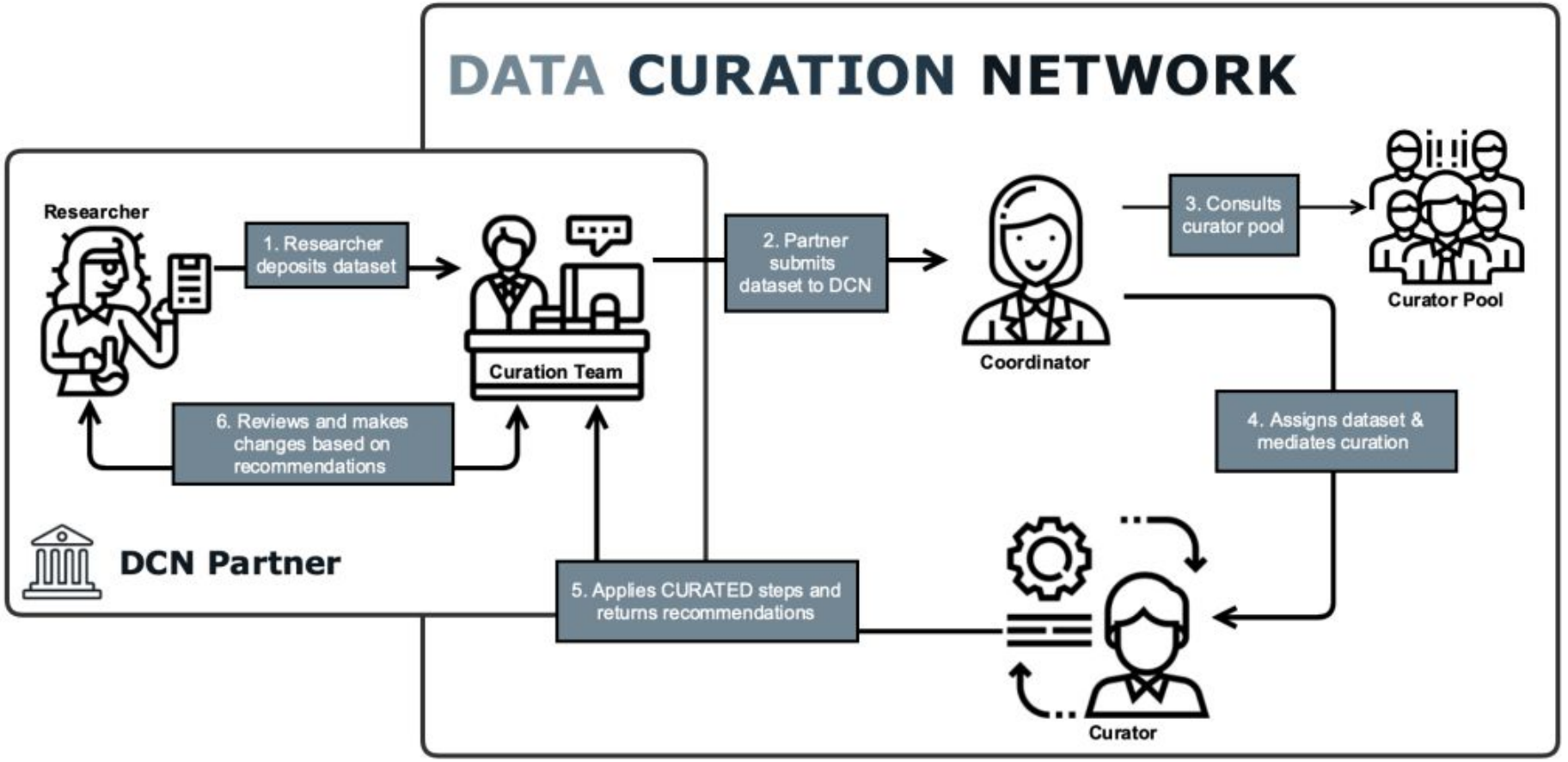
Curation → Curate data as a cross-institutional network of nearly 50 individual experts

Education → Offer professional development opportunities for an emerging data curator professional community

Primers → Create and openly share data curation best practices

Research → Support informal research teams addressing a specific topic: big data, human subjects, racial justice...

Community → Build community at annual events for discussion, training, and networking.


DATA CURATION NETWORK
Researcher
1. Researcher deposits dataset
Curation Team
2. Partner submits dataset to DCN
Coordinator
3. Consults curator pool
Curator Pool
4. Assigns dataset & mediates curation
Curator
5. Applies CURATED steps and returns recommendations
6. Reviews and makes changes based on recommendations
DCN Partner

The CURATE(D) Steps

C **Check** files and read documentation.

U **Understand** the data (or try to), if not...

R **Request** missing information or changes.

A **Augment** metadata for findability.

T **Transform** file formats for reuse.

E **Evaluate** for FAIRness.

(D) **Document** your curation activities

z.umn.edu/curate

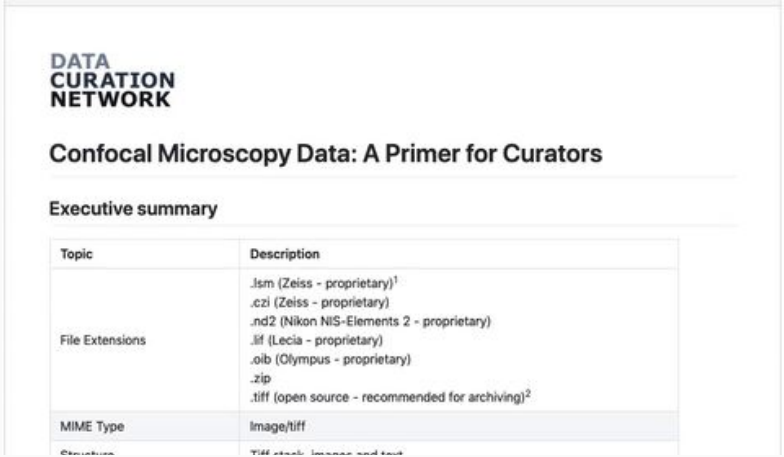
Data Curation Primers

- Concise and iterative resources
- 30 primers released so far on Github!

Ivey, Susan; Koshoffer, Amy; Sneff, Gretchen; Wang, Huajin. (2019). Confocal Microscopy Images Data Curation Primer. Data Curation Network GitHub Repository.

INSTITUTE of Museum and Library SERVICES

Grant # RE-85-18-004018



DATA CURATION NETWORK

Confocal Microscopy Data: A Primer for Curators

Executive summary

Topic	Description
File Extensions	.lsm (Zeiss - proprietary)[1] .czi (Zeiss - proprietary) .nd2 (Nikon NIS-Elements 2 - proprietary) .lif (Lecia - proprietary) .oib (Olympus - proprietary) .zip .tiff (open source - recommended for archiving)[2]
MIME Type	Image/tiff

Free and open!

Peer-reviewed concise resources to assist the data curator with recommendations for curating specific formats or curation topics!

Radical interdependence is more than cooperation –

- It's depending on others **outside of your organization** to "do the work."
- It's about trust, vulnerability, and accountability, not from a punitive perspective, but from an intrinsic desire to show up for and serve one another based on **clear expectations, commitments, and the shared goal of advancing the field.**
- It's about pooling our resources, our knowledge, and our interests, to better accomplish our work and **advocate for the role of data curators**.
- It's about recognizing the **interconnectedness** of the work we are all doing.

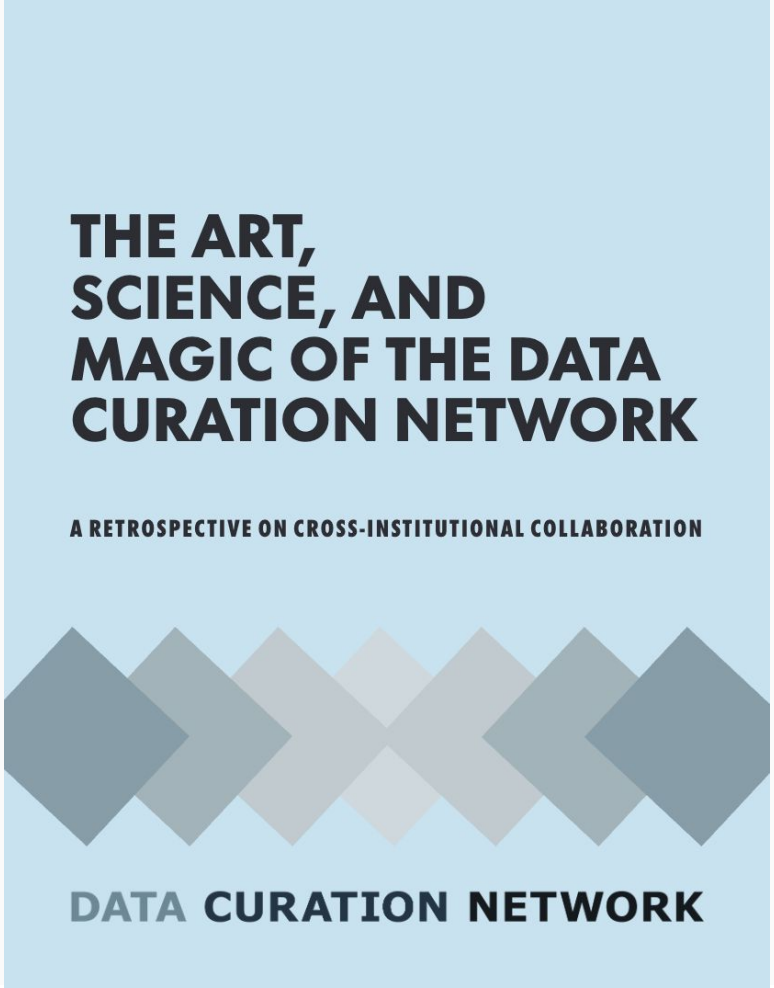
Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

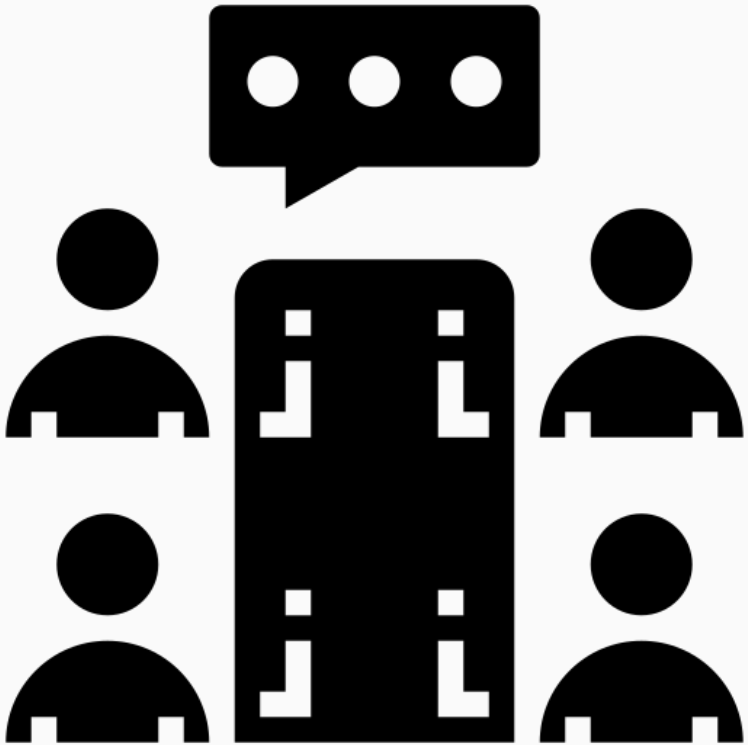
How we started

doi.org/10.3998/mpub.12782791

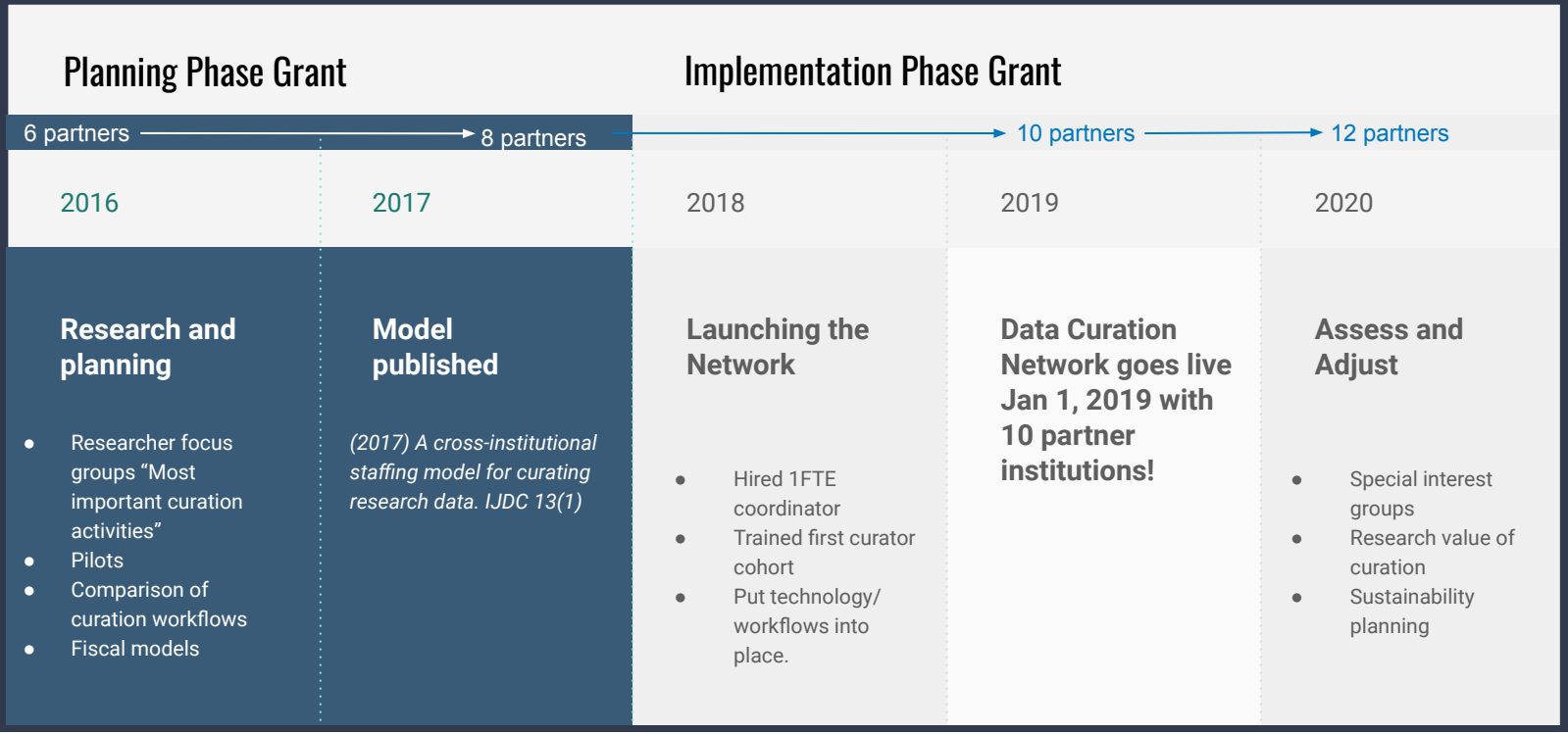


What if…

- Six individuals at six institutions
- Similar challenges in launching curation services
- They wondered – what if we build a shared workflow? What if we could design our services to they depended on each other?



Development Timeline



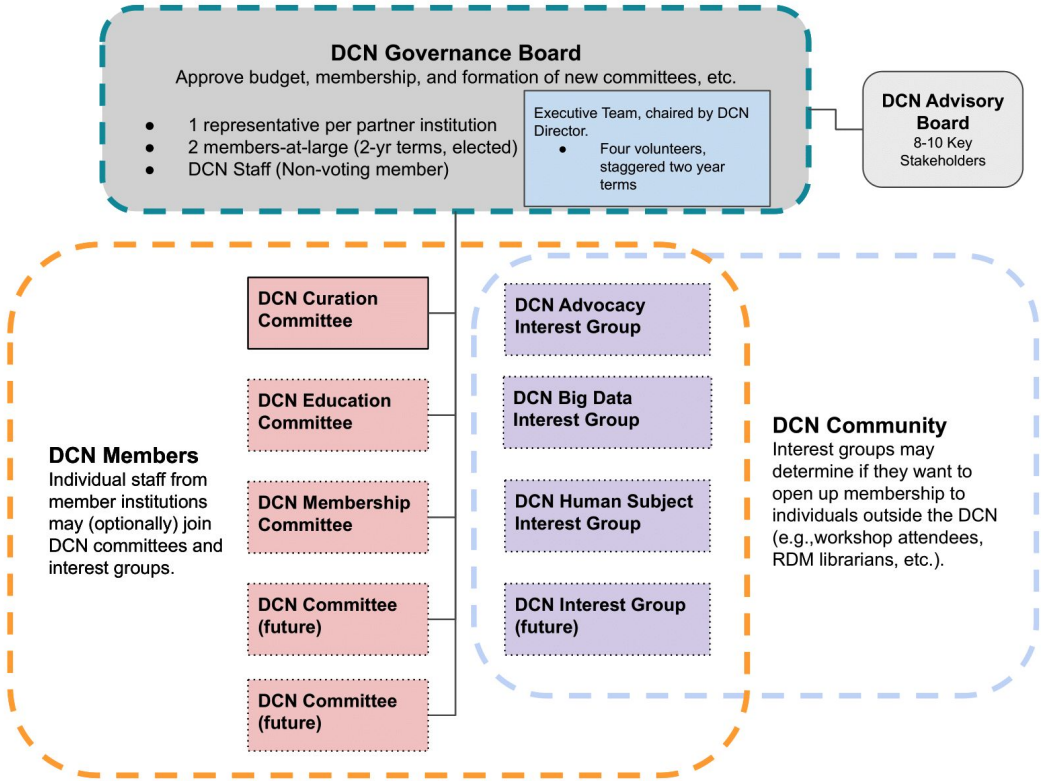
Important notes

- Started with shared challenge and opportunity
- Full time project coordinator and staff member– essential
- Slow, thoughtful growth to ensure community connections and trust

How we’re structured

DATA CURATION NETWORK

Governance Model version 2

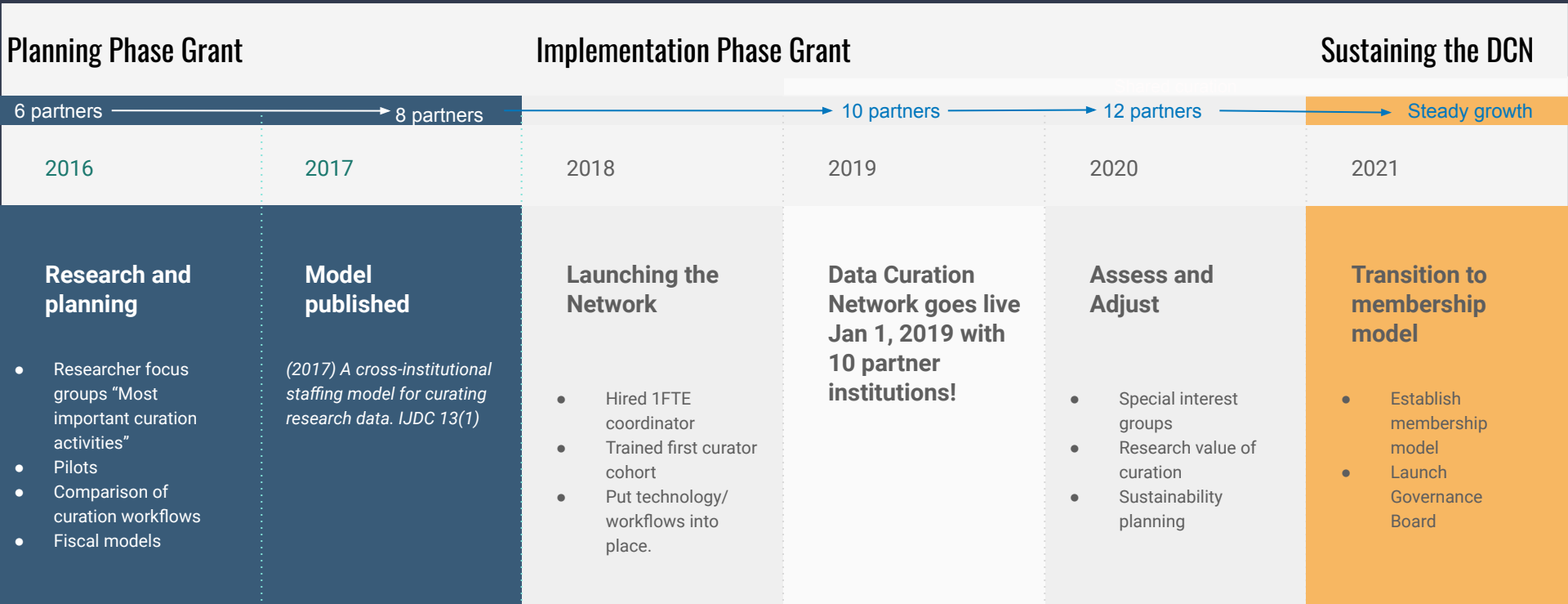


Strategic Framework

- Collaboratively developed
- Goals for 2022-2025
- Balancing continued growth and development with available resources
- https://hdl.handle.net/11299/226182

How we’re funded

Towards Sustainability



UNIVERSITY OF MINNESOTA
NYU
Duke
UNIVERSITY
UNIVERSITY OF MICHIGAN
JOHNS HOPKINS
UNIVERSITY
ILLINOIS
PRINCETON
UNIVERSITY
DRYAD
THE MICHAEL J. FOX FOUNDATION
FOR PARKINSON'S RESEARCH
PennState
UNIVERSITY LIBRARIES
VIRGINIA TECH
Washington
University in St. Louis
UNIVERSITY LIBRARIES
UNIVERSITY of NEBRASKA
LINCOLN
University of Colorado Boulder
CORNELL UNIVERSITY
Library
UNIVERSITY of VIRGINIA

DCN's Membership Model

Sustainer Institution ($10,000)

This is the highest level of participation. Sustainer institutions contribute curator staff to the Network and have full participation in governance.

- Radical interdependence: Participate in shared curation model (up to 200 hours)
- Representation in DCN Governance
- Travel support ($2k) to annual All Hands Meeting (AHM)
- Logo on website

Ambassador ($2,500)

Ambassadors invite DCN instructors to teach our 2-day in-person CURATED workshop for local/regional staff.

- Host a DCN workshop
- DCN community of practice
- Invited to present virtually at AHM
- Logo on website

Some thoughts...

- Community is essential
- Community and trust must be nurtured and grown
- Slow growth can be essential

Please reach out with any questions!

mnarlock@umn.edu