

# What are we even doing here?

## Building a community of people working with data

**Sophia Lafferty-Hess, Wanda Marsolek, Jennifer Moore,  
Neggin Keshavarzian, Mikala Narlock, Seth Erickson**

*This project was made possible in part by the Institute of Museum and Library Services  
(RE-252343-OLS-22).*



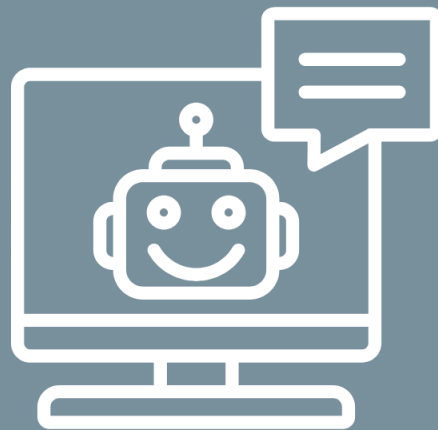
Ethical. Reusable. Better.

**DATA CURATION NETWORK**

[datacurationnetwork.org](http://datacurationnetwork.org)

# Data Are People are Data

# AI is useful.... Just ask....



Created by Happy Girl  
from Noun Project

Conversation with OpenAI's GPT-4.5 model, an advanced iteration in the GPT series on March 19, 2025

## What is data curation?

“Data curation refers to the process of organizing, managing, and maintaining data to ensure it is accurate, accessible, and usable over time.”

“The goal of data curation is to ensure that data is high-quality, well-organized, and can be used effectively for analysis, decision-making, or research purposes.”

## Are you good at it?

“**I can definitely help** with various aspects of data curation, like organizing and cleaning data, suggesting ways to structure it for easy use, and helping with analysis or sharing strategies.

**While I don't physically manage or store data myself**, I can guide you through the process and provide the tools, tips, and formulas you need to handle it efficiently. If you have a specific data curation task in mind, feel free to share it, and I can assist you with that!”

## Are you the best resource for it?

“**I can be a very useful resource for data curation tasks**”

“However, when it comes to actual hands-on tasks like physically storing or managing large datasets, or dealing with very complex data systems, tools designed specifically for data management (like specialized software or a dedicated data team) would be more appropriate.

**In short, I'm a great resource for understanding, analyzing, and organizing data, but for large-scale, physical data management, you may need additional tools or human expertise.** Let me know what you need, and I'll do my best to help!”

## Are you good at creating curricula for data curation?

Yes, **I can assist** in developing curricula for data curation. Data curation involves organizing, managing, and preserving data throughout its lifecycle to ensure its quality and accessibility.

For reference, the **Data Curation Network offers a CURATE(D)** workshop curriculum that has been effective in training professionals across various institutions. **This curriculum is structured around the CURATE(D) Steps and has been utilized to enhance data curation competencies.** These resources can serve as valuable references in designing a comprehensive data curation curriculum tailored to specific educational objectives and audience needs.

## What is the value of a human data curator?

“A **human data curator plays a pivotal role** in managing and enhancing data assets within an organization.”

“In summary, **human data curators add significant value** by ensuring that data is accurate, accessible, integrated, preserved, and ethically managed. Their expertise transforms raw data into a strategic asset, driving informed decisions and fostering a data-driven culture within organizations.”

## What is the value in human created data curation training curricula?

“**Human-created data curation training curricula offer significant value by tailoring educational content to the nuanced needs of specific disciplines and data types.**”

“In summary, human-designed data curation training curricula provide tailored, up-to-date, and practical education that aligns with the evolving demands of data stewardship, ensuring that professionals are well-prepared to manage and curate data effectively.”

# How we may use AI

**C** - familiarization with formats in a data package

**U** - familiarization with a new topic; Can use small bits of code BUT

**R** - get help writing a message

**A** - find metadata standards for a given discipline area

**T** - ask what open file types might be appropriate

**E** - thank for your help, but no thanks

**D** - leveraging project management tools with AI functionality

# How we will not use AI

## Ethical:

- we cannot feed datasets into AI;
  - it's not our data. It's not its purpose.
  - FAIR and CARE

## Reusable:

- Even when we take an assist, we will evaluate it with human eyes
  - Make reusable for humans and machines

## Better:

- Community development through community engagement and community learning

# People are Essential

# DATA CURATION NETWORK

## Mission

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

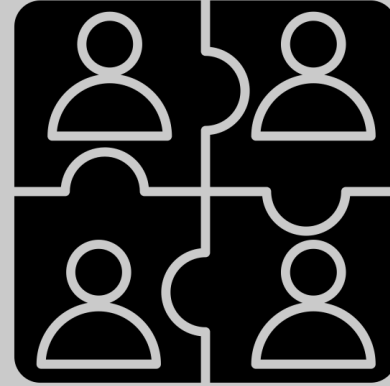
Ethical. Reusable. Better.



# Why collaboration matters



Created by Ifanicon  
from Noun Project



Created by parnowowo  
from Noun Project



# People-Centric Program that Cares for Data and Leverages Appropriate Technology



## 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



## Interest Groups



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

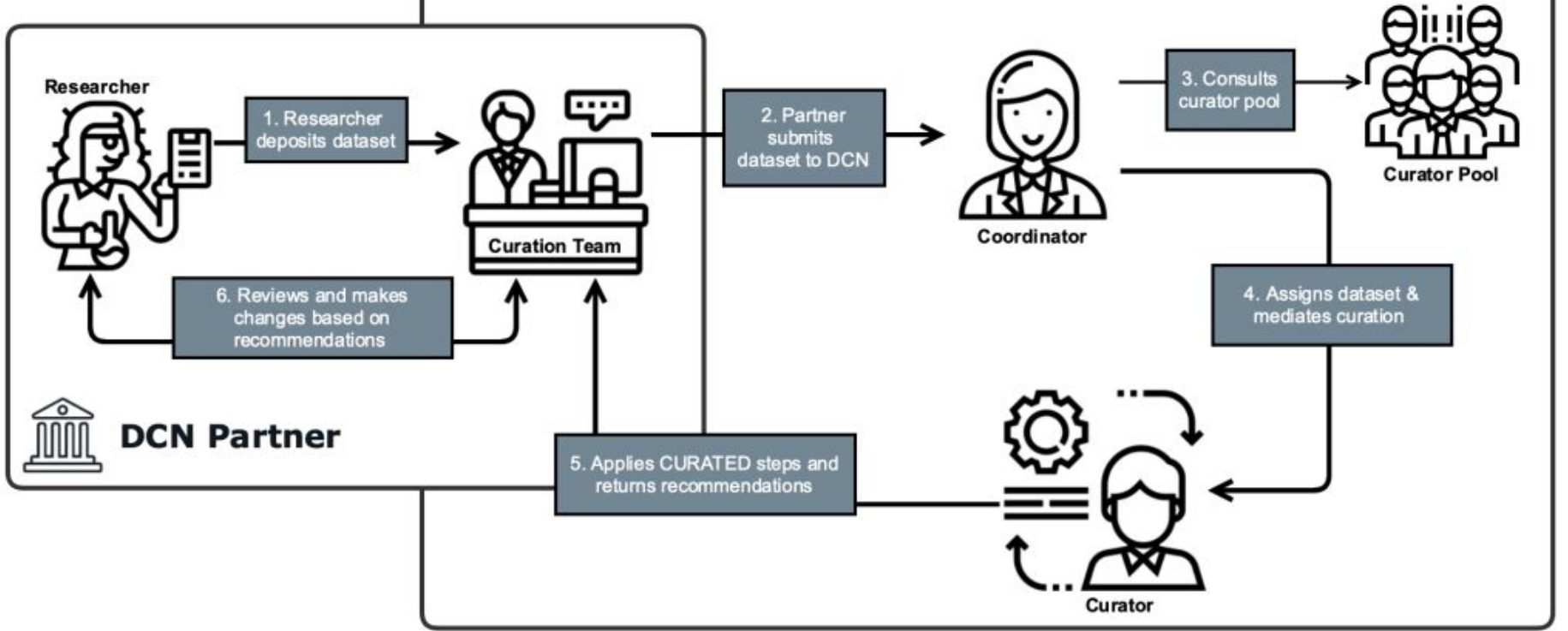


## Community



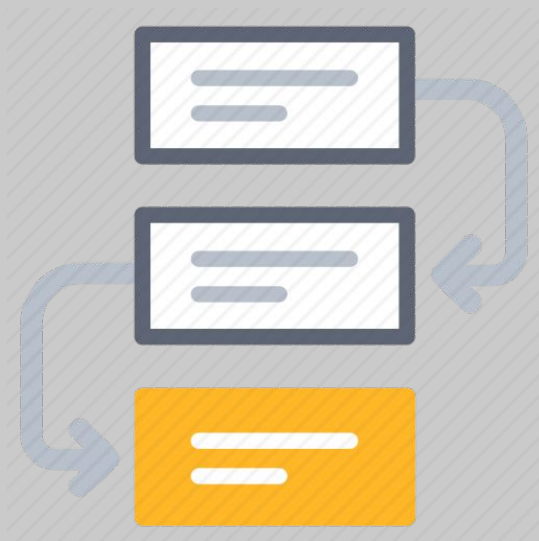
Build community at annual events for discussion, training, and networking.

# DATA CURATION NETWORK



# 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



<https://datacurationnetwork.org/resources/workflows/>

# Data Curation Primers

- How-to manuals
- Peer reviewed
- Topic / format background
- Openly available under CC-BY

## Accessibility Data Primer

Authors: Emily Oxford and Rachel Woodbrook ([woodbr@umich.edu](mailto:woodbr@umich.edu))

DCN Mentors: Wendy Kozlowski, Cornell University

## Executive Summary

### ArcGIS Pro Project Package (PPKX) Data Curation Primer

Top

Authors: Jaime Martindale, Aditya Ranganath, Amy Work, Melinda Kernik

Mentor: Melinda Kernik

Affiliate Contributors (peer reviewers): Erich Purpur, Leighton Christiansen, and Keith Jenkins

Key

revi

Suggested Citation: Martindale, Jaime; Ranganath, Aditya; Work, Amy; Kernik, Melinda. (2023). ArcGIS Pro Package Data Curation Primer. Data Curation Network [GitHub Repository](#).

## Table of Contents

[Format Overview](#)

[Introduction](#)

[Description of format](#)

[Examples](#)

[Key questions to ask yourself](#)

[Key clarifications to get from researcher](#)

[Software and resources for reviewing data](#)

# Phase 1 (2017 - present): CURATED Fundamentals Trainings

- 1.5 day and 3 hour workshops
  - In person and virtual
  - More than 250 individuals trained!



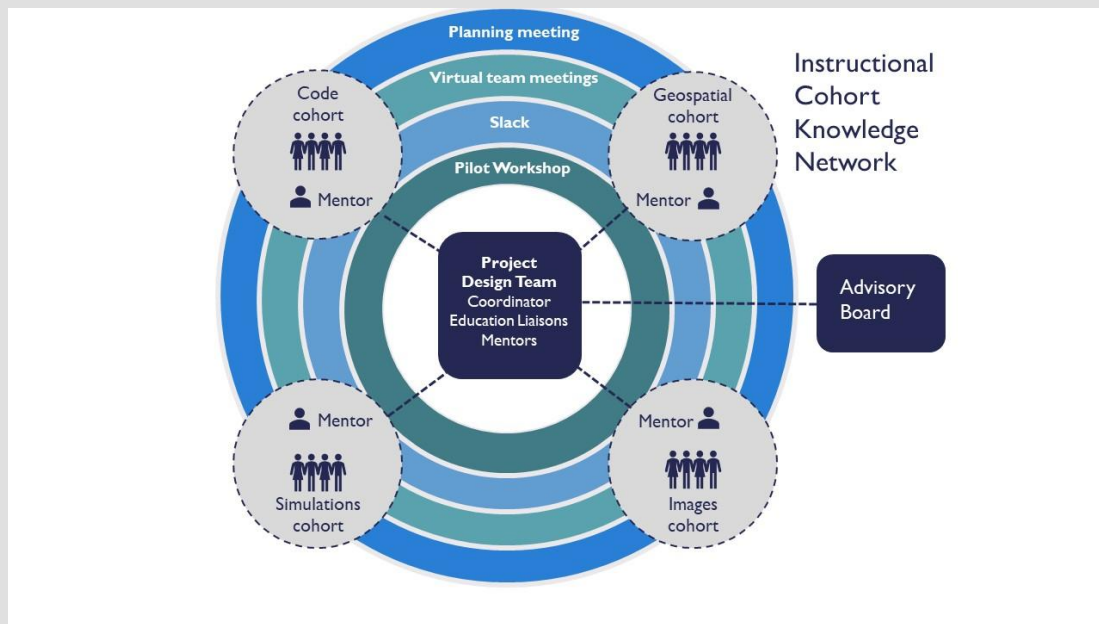
# Phase 2 (2022-2024): Expand curriculum for specialized data types

- Develop curriculum materials for four specialized data types:
  - Geospatial data
  - Scientific images
  - Code
  - Simulations
- Consider ethical, equitable, and accessible considerations for curation of that data type



# Bringing people together for a shared goal

We employed a community-driven “cohort model” to facilitate knowledge sharing and bring together intersectional voices.



# Instructional Cohorts

## Simulation data

**Mentor: Wanda Marsolek**, Data Curation Librarian, University of Minnesota

**L. Wynholds**, University of Utah

**Heather Shimon**, Science & Engineering Librarian, University of Wisconsin-Madison

**Fernando Rios**, Research Data Management Specialist, University of Arizona

**Girmaye Misgna**, Mapping & Geospatial Data Librarian, University of Pennsylvania

## Code

**Mentor: Seth Erickson**, Data Services Librarian, University of California at Santa Barbara

**Greg Janée**, Director, Library Research Data Services, University of California at Santa Barbara

**Nick Ruhs**, Research Data Management Librarian, Florida State University

**Talya Cooper**, Software Curation Specialist, New York University

**Kay P. Maye**, Scholarly Engagement Librarian for Social Sciences and Data, Tulane

## Scientific images

**Mentor: Neggin Keshavarzian**, Research Data Management Specialist, Princeton University

**Mariah Kenney**, Data Curator and Metadata Librarian, Brain Image Library

**Amy Schuler**, Director, Information Services & Library, Gary Institute of Ecosystem Studies

**Sarah Wright**, Research Data and Life Sciences Librarian, Cornell University

**Paul Gignac**, Associate Professor, Director of Global Graduate Programs, University of Arizona College of Medicine

## Geospatial data

**Mentor: Jennifer Moore**, Head of Data Services, Washington University in Saint Louis

**Leighton I Christiansen**, Data Curator, Center for Disease Control

**Kelly Grove**, GIS and Earth Sciences Librarian, Florida State University

**Tim Norris**, Data Scientist, University of Miami Libraries

**Melinda Kernik**, Spatial Data Analyst and Curator, University of Minnesota

# Our approach to curriculum development

- Kick-off meeting**
- Establish relationships and a shared understanding of goals
- Peer review**
- Each data type was reviewed by one member from each cohort
  - Allowed for cross-learning between cohorts
- Pilot workshop  
(Fall 2023)**
- 35 participants including members of the archives/library community and cohort team members
  - Each cohort had 2.5 hours to test materials of their choosing
  - Refined materials based on feedback from participants

# Our approach to continuous learning

Centered feedback as a mechanism for the iterating and improving content as well as reflecting on what worked well and what can be improved in the future

- Peer review of module materials performed by each cohort in Summer 2023
- Individual paper surveys distributed after each data type module during the pilot workshop and shared back with the individual cohorts
- Final survey sent to all pilot workshop participants (cohort members and applicants)
- Feedback survey about their experience sent to cohort team members

# Our approach to continuous learning

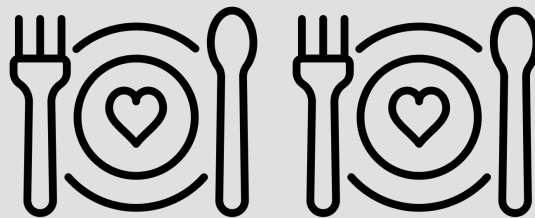
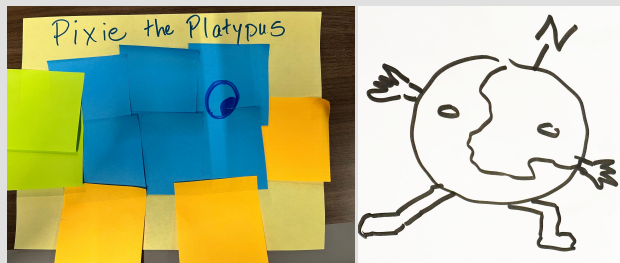
The depth of reflection, iterations, and learning came from the qualitative data.

“For me, this workshop showcased the DCN's unique position and ability to bring experienced practitioners and novice curators together in a joint learning experience, and to advance the tools, techniques, and principles of data curation. Thank you for doing this!”

“There simply wasn't enough time. Each topic was given 2.5hrs, but I'm thinking some of the topics would have benefited from having 4hrs, or even 5hrs. 5hrs (with suitable breaks) would have given the chance for participants to engage more deeply and more hands-on with a topics, especially one as complex as geospatial.”

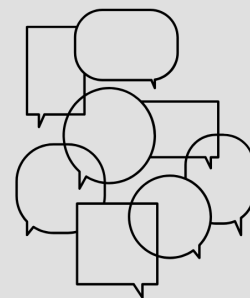
“Over all this was a wonderful experience. There were times that the communication could have been better, just some confusion about documents and how to structure some things. Would love to participate in something like this again.”

# Our approach to community building

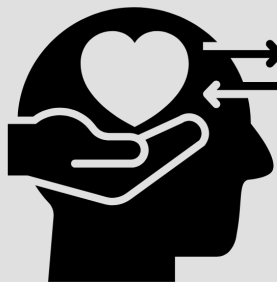


Created by Iqbal Widianto  
from Noun Project

Created by Iqbal Widianto  
from Noun Project



Created by Jooyun Lee  
from Noun Project



Created by Ali Nur Rohman  
from Noun Project



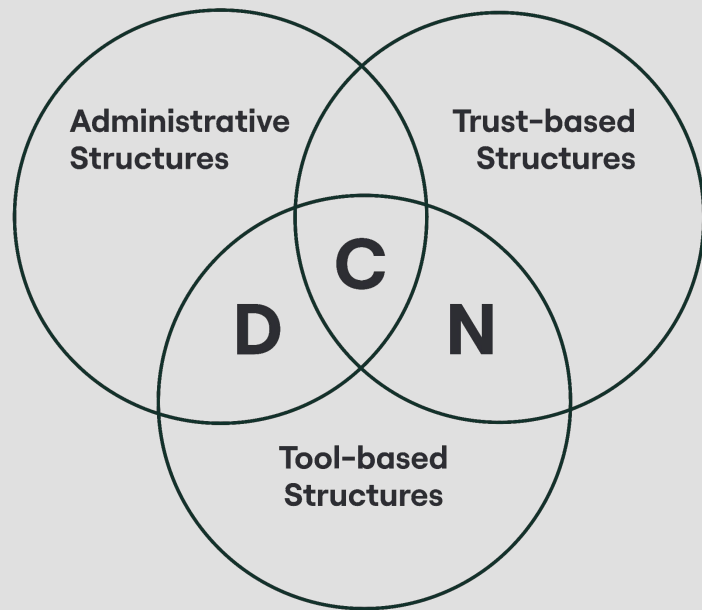
Created by kholifah  
from Noun Project



Created by bsd studio  
from Noun Project

# Adaptation and Extension

“The labor of building and sustaining an active community is an artform that requires **humility, intuition, and vulnerability**. In reflecting on the work and success of the DCN, a key theme that emerged was the DCN’s ability to **center humanity**: in our curators, in our workflows, and in our leadership, particularly acknowledging and leaning into our feelings of vulnerability.”



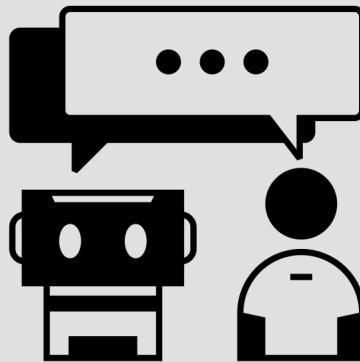
The Art, Science, and Magic of the Data Curation Network: A Retrospective on Cross-Institutional Collaboration.  
<http://doi.org/10.3998/mpub.12782791>

# Impact of bringing together people & technology

- Spring-Summer training series in partnership with NIH
- All education materials openly available - see the [DCN Education Page](#)
- Growth of the network over time
  - From 6 original institutions to now 25 institutions
- Growth and development of primers
  - Almost 50 primers published since 2017
  - AI data curation primer group formed at last NIH CURATED Fundamentals workshop!

# Conclusion

“**human data curators add significant value** by ensuring that data is accurate, accessible, integrated, preserved, and ethically managed.”



Created by erifqizeicon  
from Noun Project

“**human-designed data curation training curricula provide tailored, up-to-date, and practical education that aligns with the evolving demands** of data stewardship, ensuring that professionals are well-prepared to manage and curate data effectively.”

# References

Carlson, J., Narlock, M., Imker, H. J., Blake, M., Herndon, J., & Luong, H. (2023). The Art, Science, and Magic of the Data Curation Network: A Retrospective on Cross-Institutional Collaboration.

<http://doi.org/10.3998/mpub.12782791>

Conversation with OpenAI's GPT-4.5 model, an advanced iteration in the GPT series on March 19, 2025

Christiansen, Leighton L.; Grove, Kelly; Kernik, Melinda; Norris, Timothy; Moore, Jennifer. (2024).

Geospatial Curation Curriculum. Retrieved from the University Digital Conservancy,

<https://doi.org/10.24926/2024.265767>.

Cooper, Talya; Janée, Greg; Maye, Kay P.; Ruhs, Nick; Erickson, Seth. (2024). Code Curation

Curriculum. Retrieved from the University Digital Conservancy,

<https://doi.org/10.24926/2024.265764>.

D'Ignazio, C., & Klein, L. F. (2020). Data feminism. The MIT Press.

Gignac, Paul; Wright, Sarah; Kenney, Mariah; Schuler, Amy; Keshavarzian, Neggin. (2024). Scientific

Images Curation Curriculum. Retrieved from the University Digital Conservancy,

<https://doi.org/10.24926/2024.265766>.

Wynholds, L.; Shimon, Heather; Rios, Fernando; Misgna, Girmaye; Marsolek, Wanda. (2024).

Simulations Curation Curriculum. Retrieved from the University Digital Conservancy,

<https://doi.org/10.24926/2024.265765>.

# Acknowledgments

The Data Curation Network is supported by our current members, the University of Minnesota, and grants from the Alfred P. Sloan Foundation (Primary Award: G-2018-10072; Planning Award: G-2016-7044), the Institute of Museum and Library Services (Phase 1 CURATED Workshops: RE-85-18-0040-18 and Phase 2 Specialized Workshops: RE-252343-OLS-22), and the National Institutes of Health Office of Data Science Strategy