# New and Improved: Refining the CURATE(D) model and developing online modules

- 🜟 Briana Ezray Wham, PhD // Penn State
- 🜟 Mikala Narlock // Data Curation Network @ UMinnesota

## Agenda

- Context
  - What is the Data Curation Network
- CURATE(D) Steps
- Online modules
- Feedback / Q&A

Ethical. Reusable. Better.

**DATA CURATION NETWORK** 

datacurationnetwork.org

We have a very short time together today, so we will move at a brisk pace through these topics. Our contact information will be on the last slide, but we also whole-heartedly invite you to chat with us after the presentation!



I'm excited to share a little bit about the Data Curation Network with you. Before I get started, I want to say that the work I'll be discussing today, though I'll say words like "we" and "our" is the result of years worth of labor from countless individuals.

The Data Curation Network, or DCN, is a community-led network of curators advancing open research by making data more ethical, reusable, and understandable. We are an active community of data stewards and curation practitioners who share knowledge and time to collaboratively curate research outputs and advance the data curation profession. The DCN has been active since 2016, and officially launched as a member-funded organization in July 2021

The DCN collaborates to curate datasets, shares expertise and resources, offers education and professional development opportunities to the wider research community, and advocates for the role of data curators and information professionals. In this way, we help ensure that data shared by researchers at our institutions is more findable, accessible, and reusable.

### The CURATE(D) Steps

C Check files and read documentation.

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

Request missing information or changes.

A Augment metadata for findability.

**Transform** file formats for reuse.

**Evaluate** for FAIRness.

**D)** Document your curation activities



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

Ethical. Reusable. Better.

**DATA CURATION NETWORK** 

datacurationnetwork.org

One of the things the DCN is known for is our CURATE(D) steps
This workflow is primarily a teaching tool, to help data curators, researchers, or
students understand the various steps that go into the curation process
Originally developed at the first All Hands Meeting in 2018, these have served as an
onboarding tool for new curators

Under each step, there was a description of the action, the key steps, and then a checklist of the actions that was not meant to be exhaustive but more explanatory. These steps became the foundation for a two day workshop the DCN hosts in which attendees are invited to a deep dive into the checklist and apply the steps to real data.

**Centering Racial Equity in Data Curation** 



**Report to the Data Curation Network** 

Written By: Fay Cobb Payton, PhD

April 2021

- FATE Fairness, Accountability, Transparency & Ethics in Al
- 2. **FAIR** Findable, Accessible, Interoperable & Reusable
- CARE Collective Benefit, Authority of Control, Responsibility & Ethics

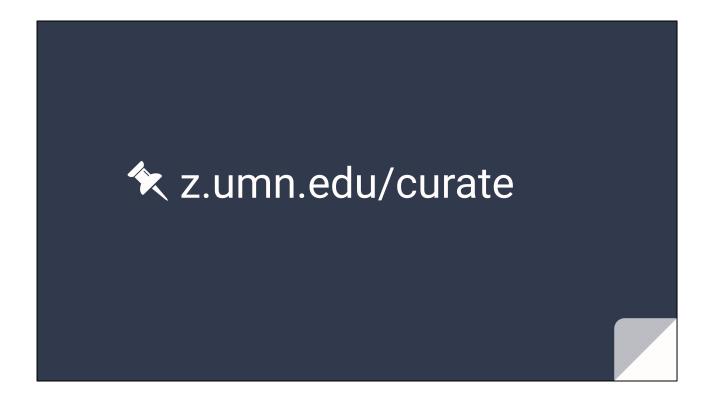


DATA CURATION NETWORK

datacurationnetwork.org

During Winter 2021, specifically March, a small group of curators dug into the DCN CURATE(D) steps and discuss how our model might better incorporate the actions that data curators often take to ensure data are shared in ethical, equitable ways. They applied real data sets when discussing the FATE, FAIR, CARE Principles

One key element we've identified as part of our IDEA action plan is to refine our CURATE(D) model to better reflect how curators consider, respect, and value the people within the data. So, a special interest group was launched to tackle this, to incorporate ethical considerations into every step. This group started back in 2020, but carried through most of 2021



So now let's look at the newly revised steps, which you can find at z dot umn dot edu slash curate

#### The CURATE(D) Steps

#### Augment the dataset

In this step we ensure metadata conforms to repository and/or appropriate discipline standards; adjust metadata to improve findability and accessibility; and improve documentation to make data more understandable, interoperable and reusable. Common AUGMENT steps include:

- Enhance metadata to best facilitate discoverability, such as by ensuring datasets have a persistent identifier.
- Create and apply metadata for the data record, including descriptive keywords
- When appropriate, structure and present metadata in domain-specific schemas to facilitate interoperability with other systems
- Implement any other agreed-on enhancements to metadata or documentation following discussion with researcher

#### **Key Ethical Considerations**

- Make sure bibliographic information reflects correct author attribution.
- Ensure any augmentation by the depositor to resolve ethical questions from previous steps is completed.

Ethical. Reusable. Better.

DATA CURATION NETWORK

datacurationnetwork.org

And they look something like this! Each step has a definition of the action, key ethical considerations, and essential tasks. In all sections, we've not only tightened up the language for consistency, but we've also integrated how curators consider, respect, and value the people within the data. We also collectively drafted a preamble to the CURATED steps, which describes how the steps are to be used and who the target audience is.

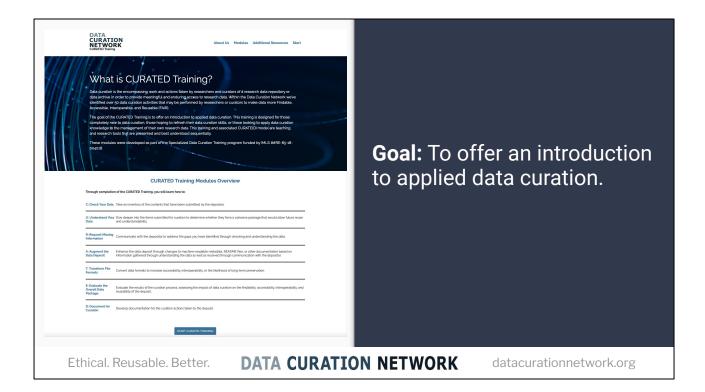
# **CURATED Online Learning Modules**



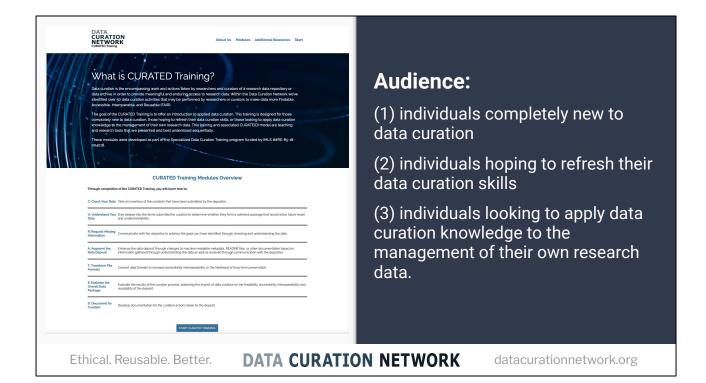
datacurationnetwork.github.io/CURATED/

IMLS ##RE-85-18-004018

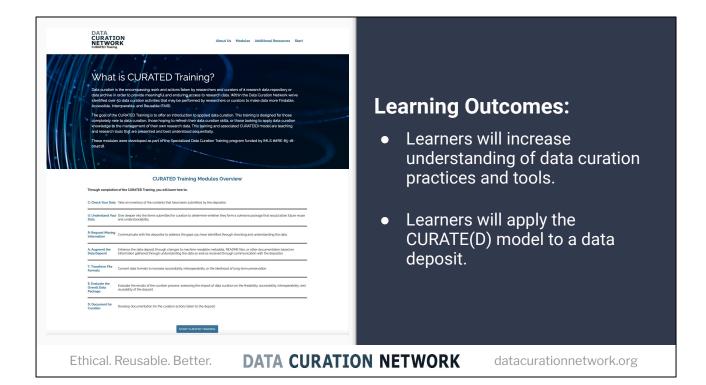
To further support the work the DCN is doing to both expand the network of data curators as well as enhance the capacity for librarians to robustly curate heterogeneous datasets created on their campuses as part of the IMLS Building the Digital Curation Workforce: Advancing Specialized Data Curation, we built the CURATED Online Learning Modules to offer an online, hands-on experience to learn how to apply the CURATED steps



The goal of the CURATED online training modules is to offer an introduction to applied data curation.



The training is designed for those completely new to data curation, those hoping to refresh their data curation skills, or those looking to apply data curation knowledge to the management of their own research data.



And through taking this online module, learners can expect to increase their understanding of data curation practices and tools as well as have an opportunity to apply the CURATE(D) model to a real-world data deposit.

	<b>CURATED Training Modules Overview</b>	
Through completio	of the CURATED Training, you will learn how to:	
C: Check Your Data	Take an inventory of the contents that have been submitted by the depositor.	
U: Understand Your Data	Dive deeper into the items submitted for curation to determine whether they form a cohesive package that would allow future reuse and understandability.	
R: Request Missing Information	Communicate with the depositor to address the gaps you have identified through checking and understanding the data.	-
A: Augment the Data Deposit	Enhance the data deposit through changes to machine-readable metadata, README files, or other documentation based on information gathered through understanding the data as well as received through communication with the depositor.	
T: Transform File Formats	Convert data formats to increase accessibility, interoperability, or the likelihood of long-term preservation.	
E: Evaluate the Overall Data Package	Evaluate the results of the curation process, assessing the impact of data curation on the findability, accessibility, interoperability, and reusability of the deposit.	
D: Document for Curation	Develop documentation for the curation actions taken to the deposit.	-
	START CURATED TRAINING	
ıl. Reusable. Better	DATA CURATION NETWORK datacurat	ionnetwork.o

The online modules are structured following the CURATE(D) model with lessons for each step of the model: C: Check Your Data, U: Understand Your Data, R: Request Missing Information, A: Augment the Data Deposit, T: Transform File Formats, E: Evaluate the Overall Data Package, and D: Document for Curation. As the lessons follow the CURATED model they are presented and best understood sequentially, but we have formatted the site so that individuals who are looking to learn about a particular step can jump to that lesson.

## **Data Deposit**

Throughout the following lessons, a single dataset will be used for all activities. This dataset was deposited by an entomologist to their local data repository; it has been slightly simplified for the purposes of these lessons. The dataset contains experimental measurements documenting how honey bee colonies exposed to high temperatures maintain and meet the water needs of the colony to survive. We would like to thank the dataset authors for allowing us to use their work.

From the link below you can access the initial deposit which consists of a metadata form and a tabular file.

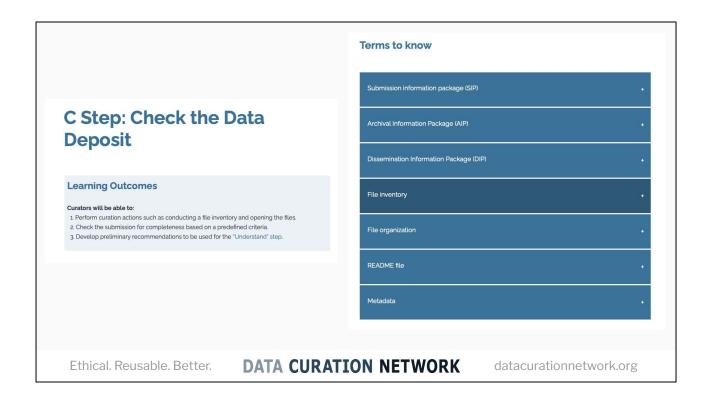
ACCESS THE DATA DEPOSIT HERE

Ethical. Reusable. Better.

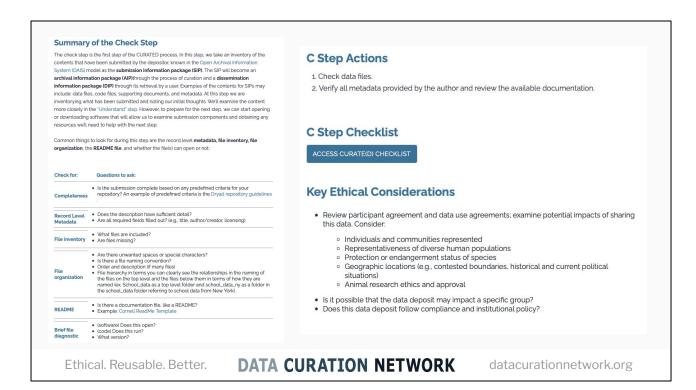
**DATA CURATION NETWORK** 

datacurationnetwork.org

As I mentioned, the CURATED modules give learners the opportunity to work with a real-world data deposit, taking it through each step of the CURATED model. We opted to use a single dataset for all module activities so that learners could see how each step is necessary and can be applied when curating a data deposit. However, throughout the modules we have also included opportunities for learners to find additional datasets to work with based on their expertise or interests.



Within each lesson, a general template was followed so that learners can anticipate and easily navigate through the content. Each lesson includes learning outcomes, terms to know,



A step summary, step actions, a link to the checklist Mikala showed, the key ethical considerations.

Activ	7		
Materials	Needed		
1. Data d	eposit.		
Direction	s		
	ivity, using the checklist below, you will perform the Check on the data	donosi	Once
		Contracting the Contracting	
ou have	completed this activity, feel free to run the C step on another dataset o	f your c	hoosin
2 64			
	C Step	Yes/	No/N
Number	C Step Files open as expected?	Yes/	
Number C1	•		
Number C1 C2	Files open as expected?	0.0	
Number C1 C2 C3	Files open as expected?  Code runs as expected?  Metadata has all required fields filled out such as the title, author, and		
C Step Number C1 C2 C3 C4 C5	Files open as expected?  Code runs as expected?  Metadata has all required fields filled out such as the title, author, and licensing information.		

And an activity to solidify the knowledge gained.

# CURATED Online Learning Modules Feedback Form





https://forms.office.com/r/CPZeB7t2BR

IMLS ##RE-85-18-004018

As these modules are a new output for us, we are hoping to collect feedback from initial users. We are collecting feedback via an open-ended form which is linked here as well as at the end of the modules.

## **CURATED Online Modules Development Team**

Subohi Agarwal

Mara Blake Joel Herndon Mikala Narlock Amanda Albert

Susan Borda Lisa Johnston **Dorris Scott** 

Jake Carlson Marley Kalt Cynthia Hudson Vitale

Briana Ezray Wham Jennifer Darragh Wendy Kozlowski

David Fearon Sophia Lafferty Hess Sarah Wright

Punya Nayan Shah Jennifer Moore Hannah Hadley

**DATA CURATION NETWORK** Ethical. Reusable. Better.

datacurationnetwork.org

# Questions?

