

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

DATA CURATION NETWORK



Mission

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

Ethical. Reusable. Better.

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 NETWORK

Report to the Data Curation Network

Written By: Fay Cobb Payton, PhD

April 2021



1. **FATE** – Fairness, Accountability, Transparency & Ethics in AI
2. **FAIR** – Findable, Accessible, Interoperable & Reusable
3. **CARE** – Collective Benefit, Authority of Control, Responsibility & Ethics

 z.umn.edu/curate



The CURATE(D) Steps

C
U
R
A
T
E
(D)

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.

CURATED Online Learning Modules



datacurationnetwork.github.io/CURATED/

What is CURATED Training?

Data curation is the encompassing work and actions taken by researchers and curators of a research data repository or data archive in order to provide meaningful and enduring access to research data. Within the Data Curation Network we've identified over 50 data curation activities that may be performed by researchers or curators to make data more Findable, Accessible, Interoperable, and Reusable (FAIR).

The goal of the CURATED Training is to offer an introduction to applied data curation. This 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. This training and associated CURATE(ID) model are teaching and research tools that are presented and best understood sequentially.

These modules were developed as part of the Specialized Data Curation Training program funded by IMLS #RE-85-18-004018.

CURATED Training Modules Overview

Through completion 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](#)

Goal: To offer an introduction to applied data curation.

What is CURATED Training?

Data curation is the encompassing work and actions taken by researchers and curators of a research data repository or data archive in order to provide meaningful and enduring access to research data. Within the Data Curation Network we've identified over 50 data curation activities that may be performed by researchers or curators to make data more Findable, Accessible, Interoperable, and Reusable (FAIR).

The goal of the CURATED Training is to offer an introduction to applied data curation. This 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. This training and associated CURATE(D) model are teaching and research tools that are presented and best understood sequentially.

These modules were developed as part of the Specialized Data Curation Training program funded by IMLS ##RE-85-18-004018.

CURATED Training Modules Overview

Through completion 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

Audience:

(1) individuals completely new to data curation

(2) individuals hoping to refresh their data curation skills

(3) individuals looking to apply data curation knowledge to the management of their own research data.

What is CURATED Training?

Data curation is the encompassing work and actions taken by researchers and curators of a research data repository or data archive in order to provide meaningful and enduring access to research data. Within the Data Curation Network we've identified over 50 data curation activities that may be performed by researchers or curators to make data more Findable, Accessible, Interoperable, and Reusable (FAIR).

The goal of the CURATED Training is to offer an introduction to applied data curation. This 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. This training and associated CURATE(D) model are teaching and research tools that are presented and best understood sequentially.

These modules were developed as part of the Specialized Data Curation Training program funded by IMLS #RE-85-18-004018.

CURATED Training Modules Overview

Through completion 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

Learning Outcomes:

- Learners will increase understanding of data curation practices and tools.
- Learners will apply the CURATE(D) model to a data deposit.

CURATED Training Modules Overview

Through completion 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

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](#)

C Step: Check the Data Deposit

Learning Outcomes

Curators will be able to:

1. Perform curation actions such as conducting a file inventory and opening the files.
2. Check the submission for completeness based on a predefined criteria.
3. Develop preliminary recommendations to be used for the "Understand" step.

Terms to know

Submission information package (SIP) +

Archival Information Package (AIP) +

Dissemination Information Package (DIP) +

File inventory +

File organization +

README file +

Metadata +

Summary of the Check Step

The check step is the first step of the CURATED process. In this step, we take an inventory of the contents that have been submitted by the depositor, known in the [Open Archival Information System \(OAIS\)](#) model as the **submission information package (SIP)**. The SIP will become an **archival information package (AIP)** through the process of curation and a **dissemination information package (DIP)** through its retrieval by a user. Examples of the contents for SIPs may include: data files, code files, supporting documents, and metadata. At this step we are inventorying what has been submitted and noting our initial thoughts. We'll examine the content more closely in the "Understand" step. However, to prepare for the next step, we can start opening or downloading software that will allow us to examine submission components and obtaining any resources we'll need to help with the next step.

Common things to look for during this step are the record level **metadata**, **file inventory**, **file organization**, the **README file**, and whether the file(s) can open or not:

Check for: Questions to ask:

Completeness	<ul style="list-style-type: none">Is the submission complete based on any predefined criteria for your repository? An example of predefined criteria is the Dryad repository guidelines.
Record Level Metadata	<ul style="list-style-type: none">Does the description have sufficient detail?Are all required fields filled out? (e.g., title, author/creator, licensing)
File inventory	<ul style="list-style-type: none">What files are included?Are files missing?
File organization	<ul style="list-style-type: none">Are there unwanted spaces or special characters?Is there a file naming convention?Order and description (if many files)File hierarchy in terms you can clearly see the relationships in the naming of the files on the top level and the files below them in terms of how they are named (ex. School_data as a top level folder and school_data_ny as a folder in the school_data folder referring to school data from New York).
README	<ul style="list-style-type: none">Is there a documentation file, like a README?Example: Cornell ReadMe Template
Brief file diagnostic	<ul style="list-style-type: none">(software) Does this open?(code) Does this run?What version?

C Step Actions

1. Check data files.
2. Verify all metadata provided by the author and review the available documentation.

C Step Checklist

ACCESS CURATE(D) CHECKLIST

Key Ethical Considerations

- Review participant agreement and data use agreements; examine potential impacts of sharing this data. Consider:
 - Individuals and communities represented
 - Representativeness of diverse human populations
 - Protection or endangerment status of species
 - Geographic locations (e.g., contested boundaries, historical and current political situations)
 - Animal research ethics and approval
- Is it possible that the data deposit may impact a specific group?
- Does this data deposit follow compliance and institutional policy?

Activity

Materials Needed

1. Data deposit.

Directions

In this activity, using the checklist below, you will perform the Check on the [data deposit](#). Once you have completed this activity, feel free to run the C step on another dataset of your choosing.

C Step Number	C Step	Yes/No/NA
C1	Files open as expected?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
C2	Code runs as expected?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
C3	Metadata has all required fields filled out such as the title, author, and licensing information.	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
C4	Is there any documentation?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
C5	Are there human participant data present?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
C6	Do the file names have unwanted spaces or characters?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

CURATED Online Learning Modules Feedback Form



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

CURATED Online Modules Development Team

Content Creators

Mara Blake

Joel Herndon

Susan Borda

Lisa Johnston

Jake Carlson

Marley Kalt

Jennifer Darragh

Wendy Kozlowski

David Fearon

Sophia Lafferty Hess

Hannah Hadley

Jennifer Moore

Mikala Narlock

Dorris Scott

Cynthia Hudson Vitale

Briana Ezray Wham

Sarah Wright

Instructional Designer

Amanda Albert

Web Developers

Subohi Agarwal

Punya Nayan Shah

Questions?



★ Briana Ezray Wham, PhD // bde125@psu.edu

★ Mikala Narlock // mnarlock@umn.edu