

eScholarship@UMassChan

Curating Audiovisual Data in Data Repositories

Item Type	Presentation
Authors	Grace, Madina;Phegley, Lauren;Valade, Meg
DOI	10.13028/1254-bd50
Publisher	eScholarship@UMassChan
Rights	Copyright © 2023 Grace, Phegley and Valade. This is an open-access document distributed under the terms of the Creative Commons Attribution 4.0 License (CC BY 4.0). The use, distribution or reproduction in other forums is permitted, provided the original author(s) are credited.
Download date	2025-05-23 07:35:21
Item License	https://creativecommons.org/licenses/by/4.0/
Link to Item	https://hdl.handle.net/20.500.14038/52783

Curating Audiovisual Data in Data Repositories

Presented by Madina Grace, Lauren Phegley, Meg Valade
NIRD 2023

Our Team



Madina Grace

Repository Services Analyst

Duke University



Lauren Phegley

Research Data Engineer

University of Pennsylvania



Meg Valade

Data Project Assistant

ICPSR at the University of Michigan

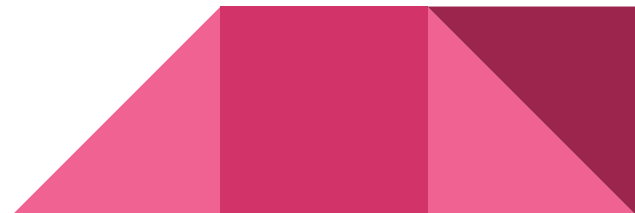
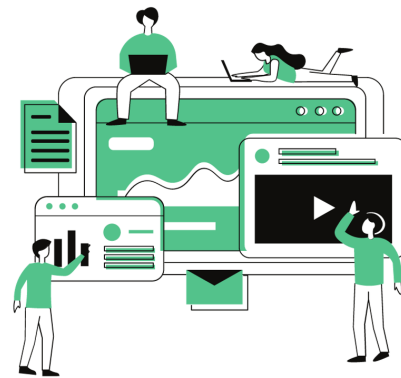
Data Curation Network

- Membership organization of data repositories working together to advance open research
- Facilitates a shared-curation workflow which shares curatorial time and expertise
- Participated in the October 2022 Specialized Data Curation Workshop in Minneapolis, Minnesota
- Picked audiovisual data as our primer topic

**DATA
CURATION
NETWORK**

Setting the Stage

- Realized audiovisual data was not an easy topic!
- New to audiovisual data curation
- Conducted informational interviews



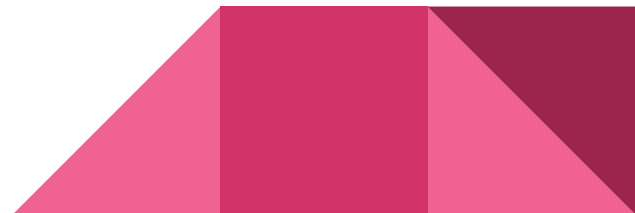
The State of Audiovisual Data Sharing

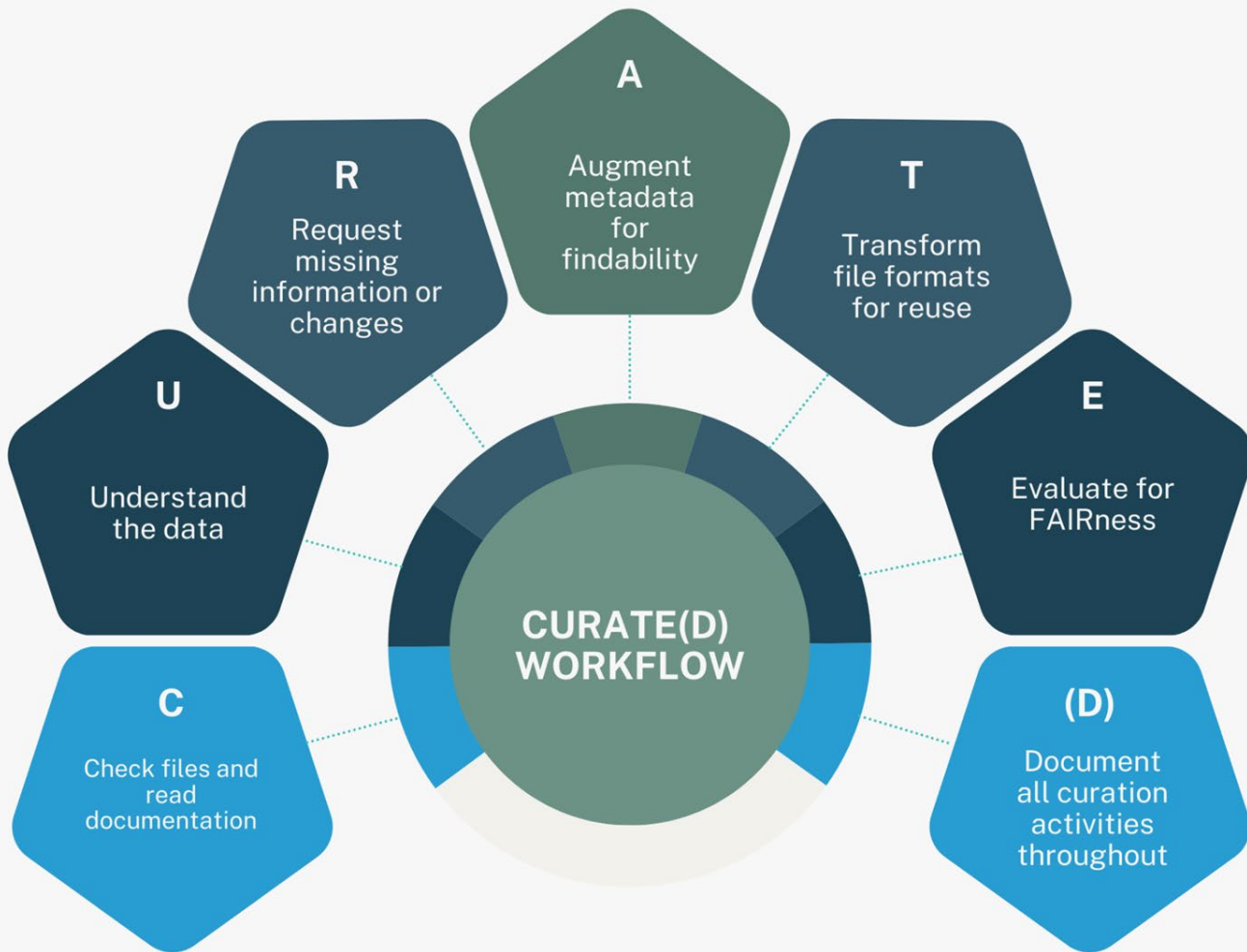


- Rarely shared
- More common in fields like behavioral psychology, anthropology, [oral history](#)
 - Discipline specific repositories that specialize in these formats
- Continued expectation for data sharing will increase the prevalence of sharing this data format
- Our Scope - previously digitized or born digital materials

Whose Role is it Anyways?

- Curating research data - why we do it
- Institutional repositories vs. Data Repositories
 - Different expectations and norms
- Bandwidth, Expertise, Software





CURATE(D) Steps

Step	Tasks for Audiovisual Data Curation
C heck files and read documentation	<ul style="list-style-type: none">• Review if your repository is able to handle the type/volume/required preservation level of the data• Check to see if files open
U nderstand the data	<ul style="list-style-type: none">• Check to see if audiovisual materials have obvious technical issues• Review the video payload information to check formats, codecs, etc.• Determine if there are ethical issues that need to be addressed
R equest missing information or changes	<ul style="list-style-type: none">• Are the changes that need completion within the abilities of the depositor?
A ugment metadata for findability	<ul style="list-style-type: none">• Appropriately describe content of the video, such as with keywords• Link to existing publications, project information, and other datasets
T ransform file formats for reuse	<ul style="list-style-type: none">• Transform into a well-established open access file format. If you compress the file, use lossless compression.• Retain original format file
E valuate for FAIRness	<ul style="list-style-type: none">• Review your actions thus far to see how you can improve the dataset's FAIRness
D ocument all curation activities throughout the process	<ul style="list-style-type: none">• Record actions from metadata changes to depositor communications in your Curators Log

Ethical Issues - Human Participants

- Human participants are common in video data, especially when dealing with the social sciences
- Personal Identifiable Information (PII)
- Level of reuse availability: restricted, public-use, combination
- Conduct disclosure risk review, potential masking or remediation



Ethical Issues - What to consider

Documentation

- IRB Approval _____
- Data transfer agreement _____
- Data Use Agreement (DUA) _____
- Informed Consent _____

Questions to ask yourself

- What does the IRB allow for dissemination? Is there an expectation of privacy?
- How will the data be utilized? Will different types of DUAs be needed for different use types?
- What kinds of data has the researcher been given permission to share and at what level (e.g., transcript-only, audio-only files, audio-video files; at a public or restricted level)?
- Have subjects provided informed consent that is in line with what the researcher wishes to release?

Technical concerns

- Staff expertise
- Software
- Storage space

Questions to ask:

- Is there sufficient storage in your network?
- Is there a way to transfer files from a depositor?
- Are there any edits to the original files?
- What level of curation is needed?
- Are files available to end-users through the repository's user interface




Documentation

Submitted by a depositor:

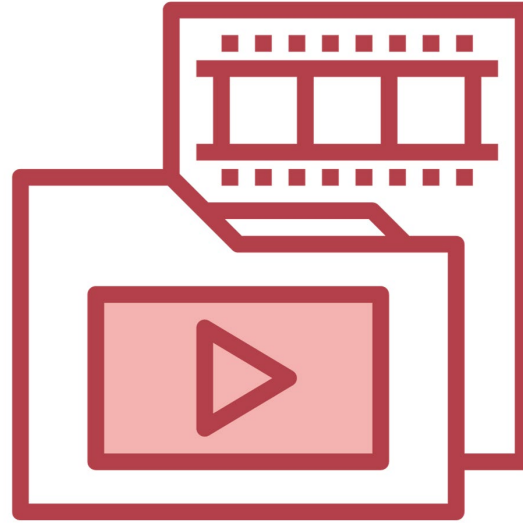
- README file
- Human Subjects Documentation (blank informed consent, etc.)

Internal documentation:

- Data Use Agreements (DUA) for any human subject audiovisual data provided by a depositor.
 - Curation log documenting any edits and additions
 - User agreements for any legal access requirements
 - Manifests
 - Any preservation documents, e.g. checksum logs
- 

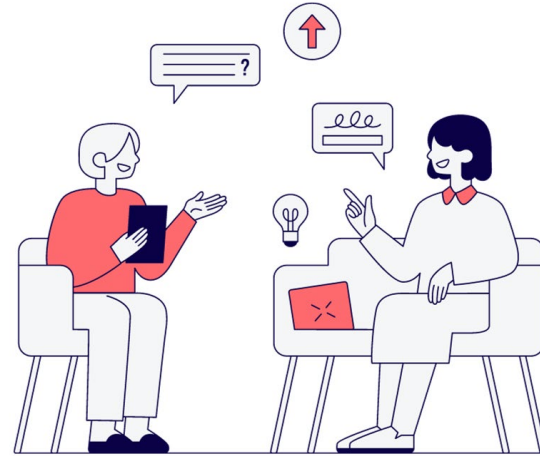
Suggested metadata standards

- DublinCore
- PREMIS
- METS
- AudioMD and VideoMD



Special considerations

- Communication with a depositor
- Preservation actions
- Files evaluation



References:

- Darragh, J., Hofelich Mohr, A., Hunt, S., Woodbrook, R., Fearon, D., Moore, J., & Hadley, H. (2020, October 6). *Human Participants Data Essentials Data Curation Primer*. GitHub. <https://github.com/DataCurationNetwork/data-primers/blob/master/Human%20Participants%20Data%20Essentials%20Data%20Curation%20Primer/human-participants-data-essentials-data-curation-primer.md>
- Grace, Madina; Jerrild, Meg; Phegley, Lauren. (2023). *Audiovisual Data Curation Primer*. Data Curation Network. GitHub. <https://github.com/mikalaraethelib/data-primers/blob/master/Audiovisual%20Data%20Curation%20Primer/AV-data-curation-primer.md>
- Pryse, J. A., Harp, M., Mannheimer, S., Marsolek, W., & Cowles, W. (2022, December 2). *Oral History Interview Data Curation Primer*. GitHub. <https://github.com/DataCurationNetwork/data-primers/blob/c6ec438e76fea49eaaf2806bc79ec2c8c12de7f3/Oral%20History%20Interviews%20Data%20Curation%20Primer/oral-history-interviews-data-curation-primer.md> (Original work published 2018)



Thank you!

Contact us:

Meg Valade - mjerrild@umich.edu

Madina Grace - madina.grace@duke.edu

Lauren Phegley - lphegley@upenn.edu