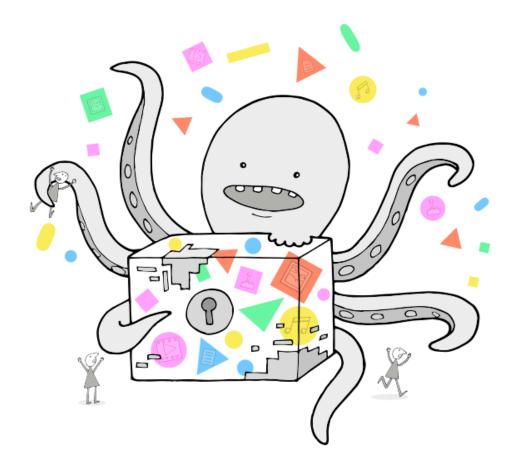
### BUILDING ON COMMON GROUND: EXPLORING THE INTERSECTION OF ARCHIVES AND DATA CURATION

Lizzy Rolando & Wendy Hagenmaier



# THE MOTIVATION



- Digital preservation is really hard, regardless of material or content type.
- What best practices, theories, tools, etc., can we adopt from other communities of practice to improve how we curate data?

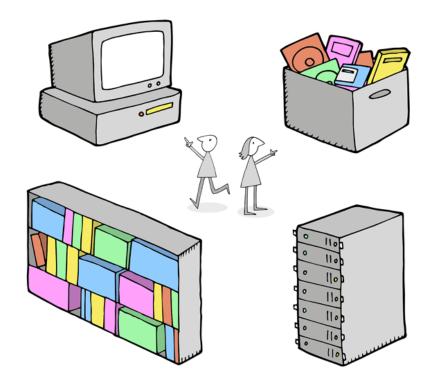
# **GREAT MINDS THINK ALIKE**



Important concepts and practices from Archives that are relevant to data curation:

- Selection and Appraisal
- Description and Arrangement
- Context
- Evidence
- And many more...

## METHODS



Using the same collection(s)

- Develop a processing plan based on archival practices and theory
- Develop a processing plan based on data curation practices

Compare outcomes to uncover areas of convergence or divergence

## **INITIAL OUTCOMES**

Data Transfer	Data Processing	Metadata Processing	Preservation	Access		
Unique Data Curation Processing Steps						
-Deposit agreement modeled on institutional repository license -Funding model for sustainability	<ul> <li>-Format transformation policies guided by reuse over preservation</li> <li>-Create derivatives to promote access and reuse</li> <li>-Correct erroneous or missing data</li> </ul>	-Review and enhancement of README file, used to accommodate diverse depositor needs	-Varied retention periods, determined by Board of Regents Retention Schedule and funding model	-Datasets treated as active and reusable -Datasets linked to publications -Bulk or individual file download		
Common Processing Steps						
-Data quarantine -Collection policy review -Integrity checks	-Format identification and normalization -Technical metadata extraction -Confidentiality and privacy review -Processing noted in metadata	-Creation of descriptive, administrative, technical, and preservation metadata	-File format migration -Storage media refreshment -Integrity checks -Preservation events noted in metadata	-Various levels of access -End user authentication -Terms of use		
Unique Archival Processing Steps						
-Retention and disposition recorded in accession record -Forensic capture and processing -Donor agreement, with transfer of copyright	-Format transformation policies guided by preservation over re- use -Create derivatives to protect master files	-Enhancement of accession record, based on standardized depositor survey -Creation of public finding aid	-Permanent retention of unprocessed, raw masters, as well as processed masters	<ul> <li>-Records treated as inactive and read-only</li> <li>-Multiple virtual arrangements</li> <li>-Emulation of original order</li> <li>-Digital exhibits</li> </ul>		

Ē

# WHERE THE RUBBER MEETS THE ROAD

- What changes do we need to make now, to improve our Archives and Data Curation programs?
- What changes should we plan to make in the future?
- Can we leverage existing resources to address the needs of both Archives and Data Curation?

# THEORY

Data Curation

- Prioritize use over preservation
- Traces and context of materials are valuable for informing use of original materials

#### Archives

- Prioritize preservation over use
- Traces and context of materials are valuable for their own uses

## **POLICIES - RETENTION SCHEDULES**

#### Data Curation

 Scholarly products are usually retained "permanently"  Materials subject to State Retention Schedules Archives

- Archivists decide which permanent, inactive records are accessioned
- Records Management will dispose of materials, in accordance with retention schedule

### POLICIES - OWNERSHIP, COPYRIGHT, AND LICENSING

#### Data Curation

- Encourage use of CC-0 license
   over
   transferring
   copyright to
   Institution
- Licenses emphasize the rights of the Institution to take actions to preserve the materials and provide access in accordance with Fair Use
- Need to accommodate cases where materials include 3<sup>rd</sup> party copyright or licenses
- Unclear ownership

Archives

 Encourage transfer of copyright to institution or CC-BY and little use of CC-0 license

## **POLICIES - COLLECTION DEVELOPMENT**

#### Data Curation

- Beholden to acquire certain materials because of federal agency requirements
- Materials that meet content policies will almost always be accepted

- Need for documented processing levels and levels of commitment
- More Product Less Process?
- Records

   Management may
   acquire certain
   materials because of
   institutional or state
   requirements
- Archives will acquire materials based on historical or archival value

Archives

## WORKFLOWS

#### Ingest

Ē

- Donor interview
- Transfer methods
- Mediated deposit

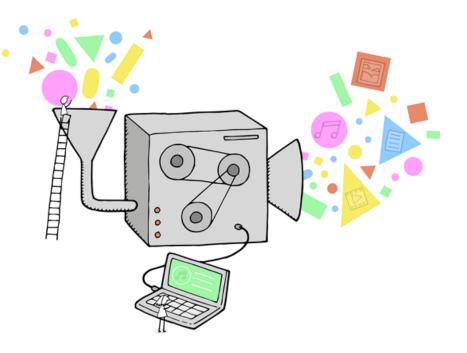
#### Preservation

 Disk imaging and forensic tools

А	ccess
	<ul> <li>Virtual arrangements</li> <li>Various levels of access</li> <li>Emulation</li> </ul>

# TOOLS

- Hydra/Fedora repository infrastructure
- Archivematica format normalization, file arrangement
- Bit Curator forensic capture and analysis
- EPADD -natural language processing and redaction



# FINAL THOUGHTS AND FUTURE RESEARCH

This comparison challenges us to consider where there may be gaps in our professional practice.

Question status quo

Strengthen local workflows and practices

How do our patrons' communities of practices shape our theoretical frameworks and practices?

# **QUESTIONS?**

#### Lizzy Rolando

lizzy.rolando@library.gatech.edu

Wendy Hagenmaier

wendy.hagenmaier@library.gatech.edu

# CREDITS

Images are available under a CC-BY-NC license. They were created by Tom Woolley for the Jisc funded Digital Preservation Business Case Toolkit, and they are available at <a href="http://wiki.dpconline.org/index.php?title=SPRUCE\_Digital\_Preservation\_Illustrations">http://wiki.dpconline.org/index.php?title=SPRUCE\_Digital\_Preservation\_Illustrations</a>.

Rolando, E., Hagenmaier, W., & Wells Parham, S. (2014). Repurposing Archival Theory in the Practice of Data Curation. Poster presented at IDCC 2014. San Francisco, CA. <u>http://www.dcc.ac.uk/sites/default/files/documents/IDCC14/175Rolando\_Hagenmaier\_Parham\_idcc\_14.pdf</u>

This work is licensed under a Creative Commons Attribution 4.0 International License.