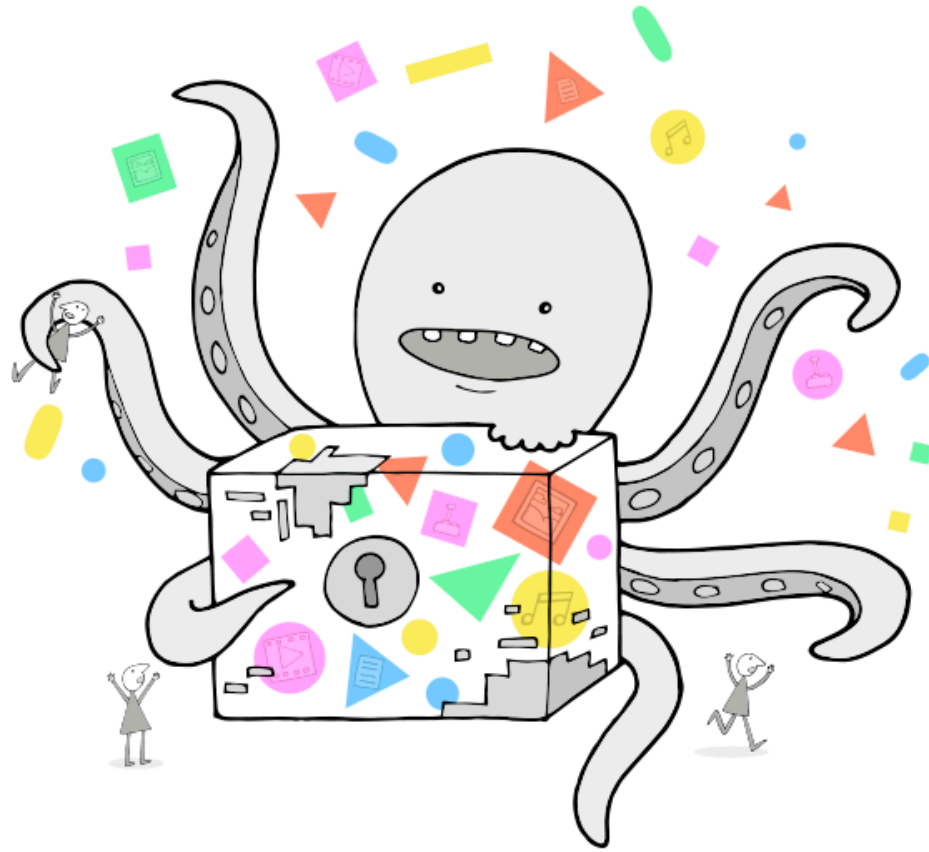


BUILDING ON COMMON GROUND: EXPLORING THE INTERSECTION OF ARCHIVES AND DATA CURATION

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?

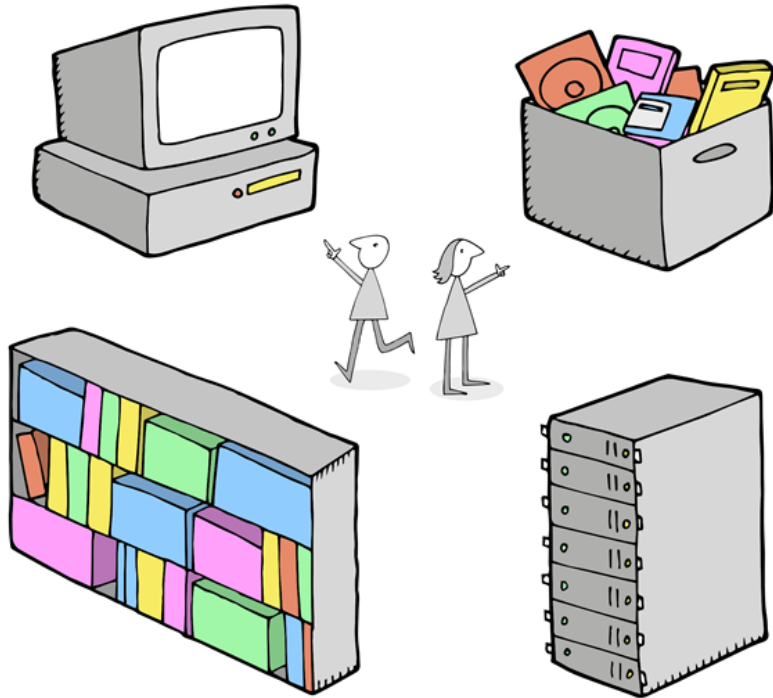
GREAT MINDS THINK ALIKE



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

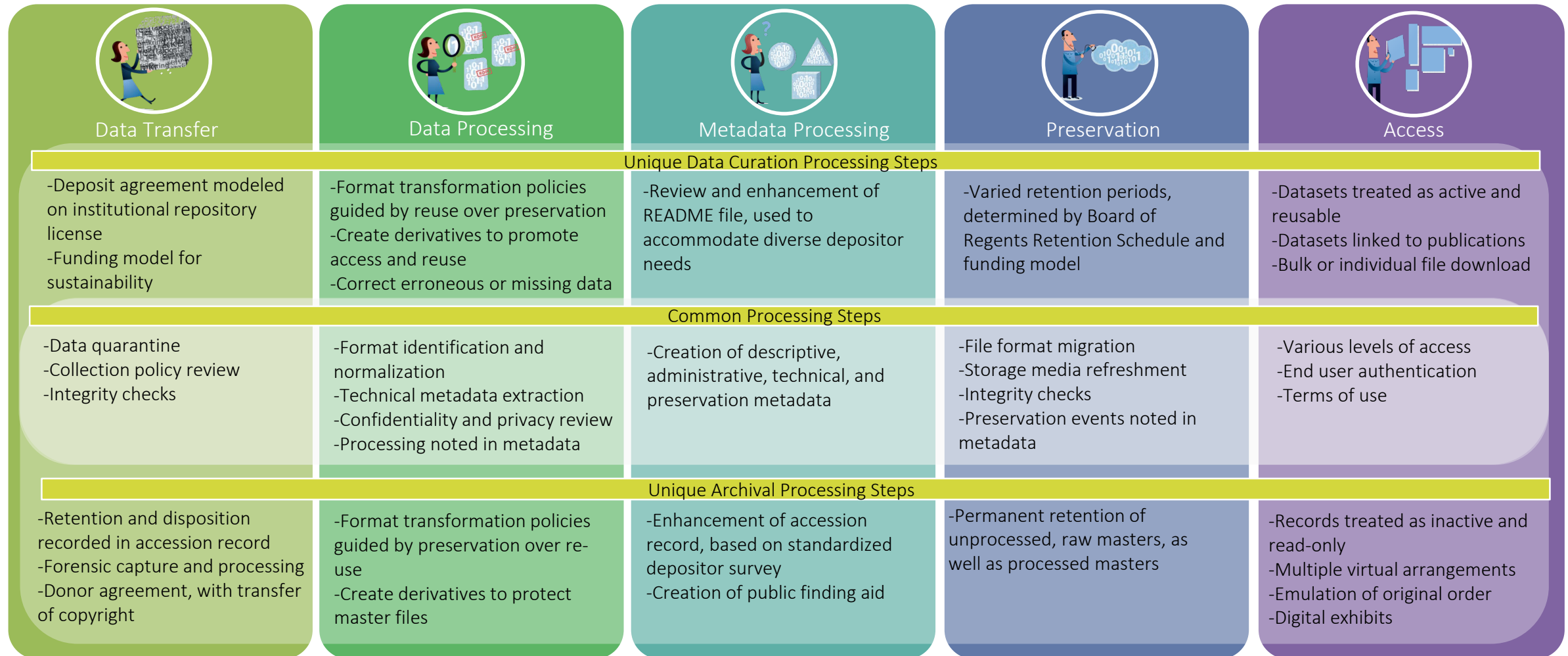


Diagram from Rolando, Hagenmaier, & Wells Parham, 2014.



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 3rd 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

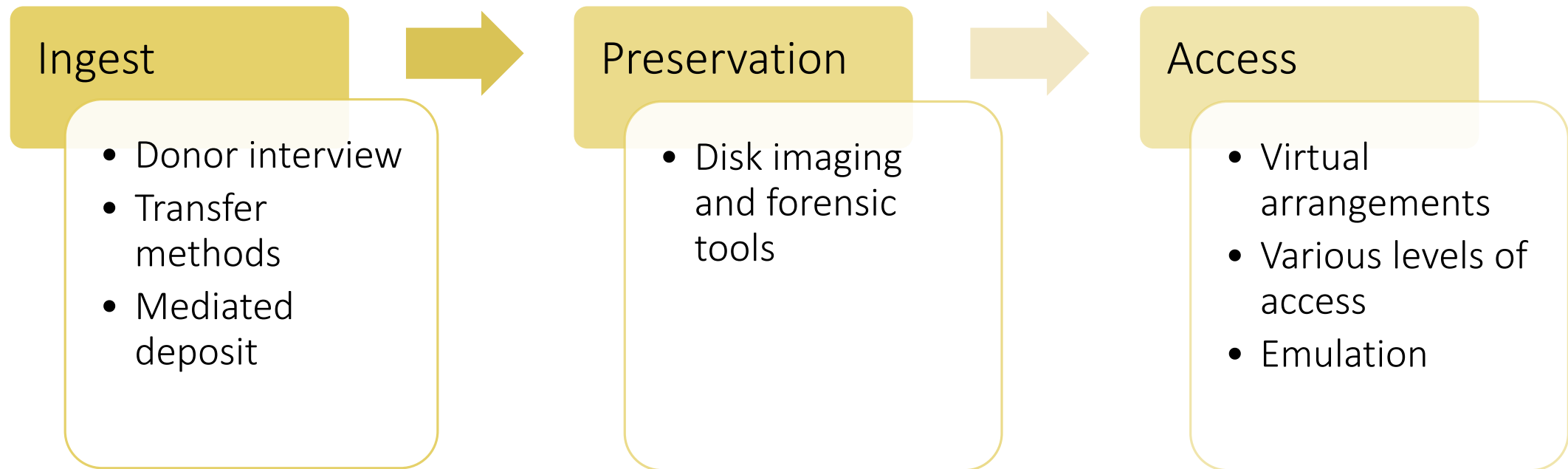
- Need for documented processing levels and levels of commitment
- More Product Less Process?

Archives

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

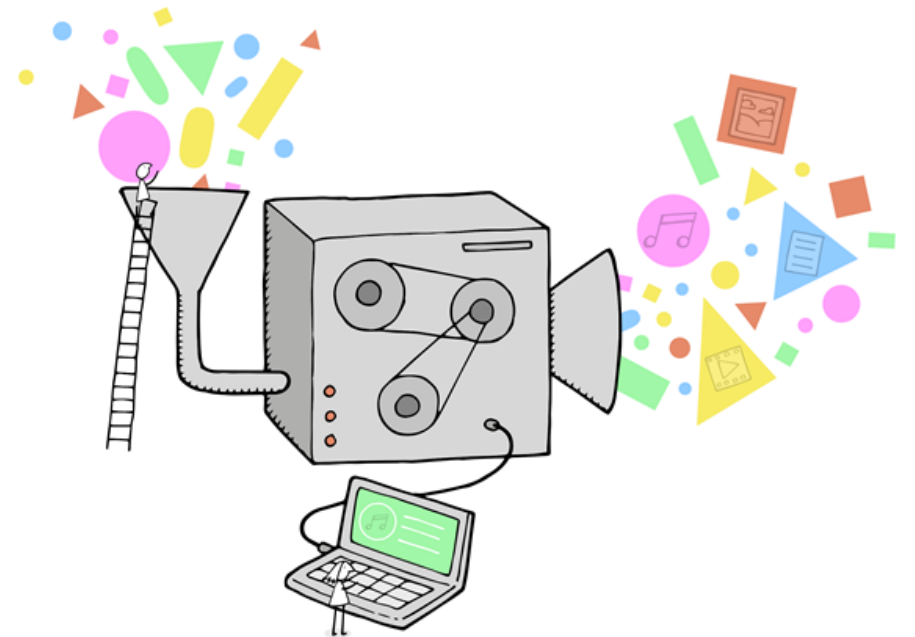


WORKFLOWS



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

http://wiki.dpconline.org/index.php?title=SPRUCE_Digital_Preservation_Illustrations.

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.

http://www.dcc.ac.uk/sites/default/files/documents/IDCC14/175Rolando_Hagenmaier_Parham_idcc_14.pdf

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

