## Repurposing Archival Theory in the Practice of Data Curation Georgialnstitute of Technology Elizabeth Rolando | Wendy Hagenmaier | Susan Wells Parham

# Introduction

- Expansion of data curation and digital archiving services at the Georgia Tech Library and Archives.
- How do data curation and archival science intersect?
- How can comparing data curation and archival science lead to improvements in local workflows and practices?

Data Trans	for	Deta Procesing	Metadata Processing	Preservation
		Data Processing	Metauata Processing	FIESEIVALION
		U	nique Data Curation Processing Steps	
-Deposit agreement institutional reposito -Funding model for s	ory license guid ustainability -Crea acce	mat transformation policies ed by reuse over preservation ate derivatives to promote ss and re-use rect erroneous or missing data	-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
			Common Processing Steps	
-Data quarantine -Collection policy rev -Integrity checks	view norm -Tech -Con	mat identification and nalization nnical metadata extraction fidentiality and privacy review cessing noted in metadata	-Creation of descriptive, administrative, technical, and preservation metadata	-File format migration -Storage media refreshment -Integrity checks -Preservation events noted in metadata
			Unique Archival Processing Steps	
<ul> <li>-Retention and dispondecided upon and reaccession record</li> <li>-Forensic capture and</li> <li>-Donor agreement, word of copyright</li> </ul>	corded inguide-Creationd processingfiles	nat transformation policies ed by preservation over re-use ate derivatives to protect master	<ul> <li>-Enhancement of accession record, based on standardized depositor survey</li> <li>-Creation of public finding aid</li> </ul>	-Permanent retention of unprocessed, raw masters, as well as processed masters

Figure 1: Highlights from comparison between archival and data curation processing plans. The first row lists those elements of the data curation processing plans. The first row lists those elements of the data curation processing plan that were unique, while the bottom row lists those steps in the archival processing plan that were distinct. The middle row identifies those elements of the processing plans that were common between the two. Images used in the diagram were created by Jørgen Stamp (www.digitalbevaring.dk) and are published under a Creative Commons Attribution 2.5 Denmark license.

# Findings

What data curation might learn from archival science and processing:

- Forensic capture and processing may be valuable for certain data sets
- Existing repository license agreement models might not work for digital data sets
- Retention and disposition should be planned at the point of data transfer
- Creating virtual arrangements that emulate the data creator's original environment could be valuable
- Data curators might question how much should be done to correct data in order to  $\bullet$ facilitate re-use--how much effort is enough?

### References

<sup>1</sup> Consultative Committee for Space Data Systems. (2012). Reference model for an Open Archival Information System (OAIS) (Magenta Book CCSDS 650.0-B-1). Retrieved from <a href="http://public.ccsds.org/publications/archive/650x0m2.pdf">http://public.ccsds.org/publications/archive/650x0m2.pdf</a>. <sup>2</sup> Inter-university Consortium for Political and Social Research. (n.d.). A Case Study in Repository Management. Retrieved from <u>http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/lifecycle/index.html</u>. <sup>3</sup> UK Data Archive. (2014). *How We Curate Data*. Retrieved from http://www.data-archive.ac.uk/curate. <sup>4</sup> Society of American Archivists, Glossary of Archival and Records Terminology: <u>http://www2.archivists.org/glossary</u>

# Methodology

- Data curation processing informed by OAIS Reference Model<sup>1</sup>, ICPSR workflow<sup>2</sup>, and UK Data Archive workflow<sup>3</sup>
- Archival processing informed by concepts, such as appraisal, respect des fonds, original order, and archival value<sup>4</sup>, as well documented practices at peer institutions
- Compare processing plans to discover areas of agreement and areas of conflict

#### What archival science and processing might learn from data curation:

- Establish a balance between supporting future access and use and maintaining the integrity of the record--do disk images support future access?
- Existing donor agreement and copyright transfer models might not work for digital archives acquisitions
- Funding model should be planned at the point of record transfer
- Processed records may not be "inactive"; the life of the record continues through re- $\bullet$ use, which enriches the record and should be documented in the record itself

Process the same digital collection, once by data curator, once by digital archivist

	Access
	<ul> <li>-Datasets treated as active and reusable</li> <li>-Datasets linked to publications</li> <li>-Bulk or individual file download</li> </ul>
	-Various levels of access -End user authentication -Terms of use
as	<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>