Bringing Values to the Bitstream: A Framework for Digitally-Aware Professional Ethics of Curation

Christopher (Cal) Lee
School of Information and Library Science
University of North Carolina, Chapel Hill

Society of American Archivists Research Forum
August 11, 2009
Austin, TX

When acquiring, managing and providing access to materials, professionals in collecting institutions must consider various norms, laws, codes of ethics, policies, procedures and personal values.

As they address curation of digital collections, they will increasingly discover that established sources of guidance suffer from “latent ambiguities.”*

*Lessig, Lawrence. *Code and Other Laws of Cyberspace*. New York: Basic Books, 1999.

Digital resources are composed of interacting components that can be considered and accessed at different levels of representation, e.g.:

- Bitstream as it resides on a physical medium
- Bitstream as read through input/output hardware & software
- Sub-file data structure (e.g. field or record in a database)
- File accessed through a filesystem
- Document as rendered through specific applications
- Record composed of multiple files

Because of the possibility of interacting with at different levels, there is no single, canonical representation of digital data.

To ensure integrity and future use, archivists and other information professionals must make decisions regarding treatment of materials at multiple levels of representation.

Digital curation professionals are faced with many new decisions, which require an **understanding** of underlying digital representations, in order to appropriately enact professional values.

Some Examples

When acquiring a disk as part of a collection, should an archivist create a bit-level **image of the disk**, in order to ensure the potential to recreate not only the "payload" data of files but also various forms of information within and below the filesystem?

How should an archivist deal with “hidden” data in a Word document?

Retain the Word document but provide a surrogate to patrons?

Keep the Word document closed for a given period of time?

Convert to a “clean” format upon acquisition in order to reflect the text that the author intended to present?

If an archivist acquires a Microsoft Outlook .pst file (including saved and sent messages, calendar items, draft and deleted messages, address book, and viruses), should she retain the whole .pst file or extract messages and attachments that were sent and received?

Someone cropped a set of images in order to remove sensitive parts, but the images still have pixel information & embedded thumbnail reflecting the “removed” parts.

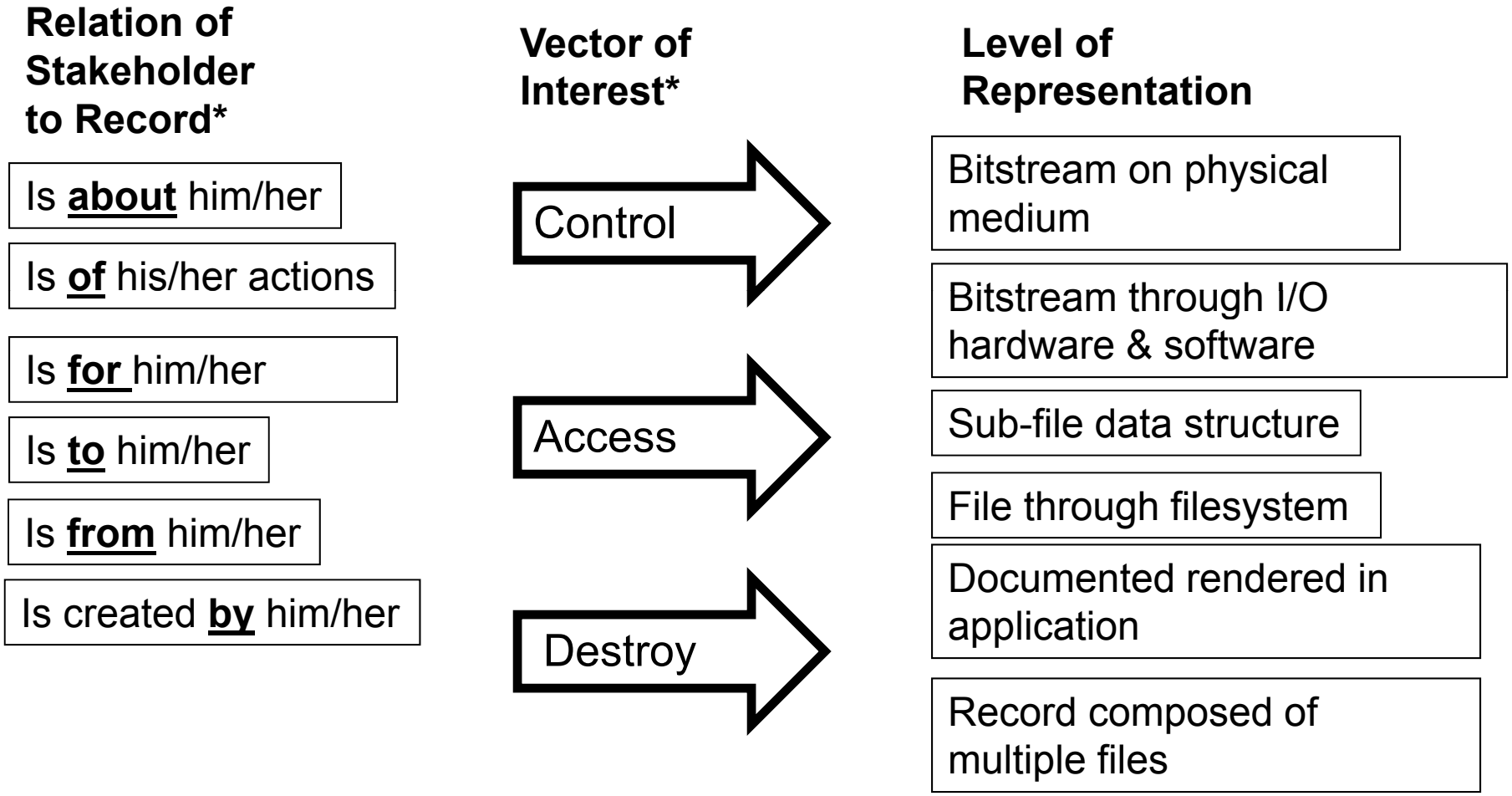
How should the archivist approach the management of the images?

How should archivists deal with materials that are encrypted in the following ways?

- At file level, encrypted by applications
- At file level, encrypted by operating system
- Applied to entire storage volume

For discussion of levels of encryption: Carrier, Brian. *File System Forensic Analysis*. Boston, MA: Addison-Wesley, 2005.

A Framework for Considering and Teaching Ethics of Digital Curation



*Not mutually exclusive

Application of Framework in a Course

- Present & describe a case involving specific set of digital objects
- Focus on specific “transition point”* in life of digital objects
- Group elaborates a workflow based on rules to implement
- Justify rules in terms of above framework

*Transition Point in the Information Continuum (Dimension 6 of Matrix of Digital Curation Knowledge and Competencies)
http://ils.unc.edu/digccurr/digccurr-transition-point.html