Title

Preservation Metadata for Digital Forensics. Report of the ALCTS PARS Preservation Metadata Interest Group, American Library Association Annual Meeting, Las Vegas, June 2014.

Authors

Sarah Potvin and Chelcie Juliet Rowell

Body Text

At the 2014 ALA Annual Meeting in Las Vegas, the Preservation Metadata Interest Group meeting centered on the use of the BitCurator software environment to generate metadata in support of preservation functions. The program paired an overview of BitCurator by Principal Investigator (PI) Cal Lee with two lightning talks from Princeton University and Rice University, where archivists are integrating BitCurator into their processing workflows. The meeting drew approximately 46 attendees.

The session opened with a brief business meeting. Co-Chair Chelcie Juliet Rowell (Wake Forest University) announced that at the January 2014 ALA Midwinter Meeting in Philadelphia, at the request of the co-chairs, the PARS Executive Committee voted unanimously to change the name of this interest group from the "Intellectual Access to Preservation Metadata Interest Group" to the "Preservation Metadata Interest Group." Rowell then welcomed new Preservation Metadata Interest Group Co-Chair Drew Krewer (University of Houston) and thanked outgoing Co-Chair Sarah Potvin (Texas A&M University) for her service.

As libraries, archives, and other collecting institutions increasingly receive receive "computer storage media (and sometimes entire computers) as part of their acquisition of 'papers' from contemporary artists, writers, [...] and other public figures," the need for digital forensics methods has emerged within technical services workflows (Kirschenbaum et al, 2010, pp. 1–2). Three presentations during the meeting of the Preservation Metadata Interest Group provided perspectives on the BitCurator environment from both the software's developers and some of its current users. The aims of the BitCurator project are to "build, test, and analyze systems and software for incorporating digital forensics methods into the workflows of a variety of collecting institutions" to aid in the acquisition of born-digital content stored on source media, such as personal computers, external hard drives, floppy disks, and Zip disks" (BitCurator, n.d., "About: The Project").

Many of these tools are adapted from the world of law enforcement and re-purposed by BitCurator for the needs of libraries, archives, and museums that are acquiring born-digital content. As Kirschenbaum, Ovenden, and Redwine (2010) write, "The same forensic software that indexes a criminal suspect's hard drive allows the archivist to prepare a comprehensive manifest of the electronic files a donor has turned over for accession" (p. 2). BitCurator addresses two fundamental needs of collecting institutions that are not typically recognized by the digital forensics industry (Lee, 2014):

- incorporation into the ingest workflows and collection management environments of libraries, archives, and museums
- provision of public access to the data captured

Lee reminded the audience that cultural heritage institutions have the same goals when acquiring born-digital materials as when acquiring analog materials: they seek to ensure the materials' integrity, preserve the context of the materials (with the understanding that this will help make sense of the materials), and prevent the inadvertent disclosure of sensitive information such as credit card and Social Security numbers. In recognition of the gap between the law enforcement context in which digital forensics tools were originally developed and the cultural heritage context in which they are also needed, BitCurator was designed to support the following functions:

- acquisition — extracting digital content from source media
- reporting — characterizing directory structures, user activity, file similarity, actions performed during processing, etc.
- redaction — removing sensitive or personally identifying information
- metadata export — structuring the output of reports into the format required by the collecting institution, such as the Metadata Encoding and Transmission Standard (METS), the Metadata Object Description Schema (MODS), or Encoded Archival Description (EAD)

In designing the software environment, the developers of BitCurator thought very carefully about how preservation metadata provides evidence to support assertions regarding digital materials' authenticity, provenance, and chain of custody. BitCurator's primary workflows are structured around capturing and analyzing disk images. Unlike copying individual files, capturing a disk image provides a "perfect capture" of both the contents and structure of an entire storage device—e.g., a floppy disk, hard drive, or CD-ROM.[1] The BitCurator developers advise the acquisition of born-digital materials as disk images in order to ensure integrity and proof of custody as well as for preserving provenance information about a file system: "Even if the image is not retained, it allows the [information professional] to store information about how files were acquired, where and when they came from, under what circumstances they were transferred, and relevant environmental data concerning their production" (BitCurator, n.d., "FAQ"). While disk imaging ensures that this kind of information is preserved, other tools bundled into the BitCurator environment extract this information in the form of a wide range of reports and package report output into established metadata formats, such as PREMIS events for each forensics tool that is applied to a disk image.

Given the opportunities for forensic analysis bundled into the BitCurator environment, how are particular institutions integrating the tool into their workflows?

As Rebecca Russell (Archivist/Special Collections Librarian) and Amanda Focke (Assistant Head of Special Collections), both from the Woodson Research Center in Rice's Fondren Library, explained in their lightning talk, "Implementing BitCurator at Rice University: Baby Steps to Digital Preservation Glory," BitCurator played a vital role in workflows developed as part of their larger effort to establish a digital preservation program over the course of 2014. This effort also involved developing staff skills and augmenting institutional support, policies, workflows,

[1] In their lightning talk to the Preservation Metadata Interest Group, Rebecca Rusell and Amanda Focke define disk imaging as "the process of extracting unaltered bitstreams from digital media; it creates a perfect capture of a device's file structure and all contents (including hidden files and fragments) into one file, the 'disk image.'"

software and hardware.[2] Rice staff identified BitCurator as a resource that could aid in their implementation of an OAIS-compliant system for preserving and accessing digital materials in archival collections.

Rice now uses BitCurator to create disk images of digital materials stored on source media (Russell and Focke, 2014). They also rely on BitCurator reporting tools to extract information that may comprise different components of an information package as described by the Open Archival Information System (OAIS)—that is, content information, preservation description information, packaging information, and descriptive information. Together with metadata supplied by Rice staff, the metadata extracted from the disk images, which BitCurator exports as Digital Forensics XML and crosswalks into other formats, populate the preservation description information component of OAIS-compliant information packages. Rice chose a simple deployment of BitCurator, operating its environment as an optical disc (like a CD or a DVD), rather than as a standalone operating system or a virtual machine.

In his lightning talk, "It Takes a System to Curate a Bit: Preservation Planning at the Princeton University Archives," Digital Archivist Jarrett M. Drake described the digital curation program established in the Princeton University Archives in April 2014. Princeton University Archives does not receive digital accessions on external storage media, but instead has established processes for university offices, departments, or student organizations to transfer digital records via the university network. Before incorporating BitCurator into their workflows for processing born-digital materials, Princeton University Archives captured preservation metadata as unstructured values in user-defined fields in Archivist's Toolkit. Now, by making use of the reporting and metadata export functions supported by the BitCurator environment, Princeton University Archives is taking steps toward making their preservation metadata more actionable by generating PREMIS metadata and packaging it within METS. Drake emphasized a functional approach to preservation metadata, observing, "There's no point in recording checksums if you don't have a mechanism and process to later verify that those checksums are still the same as they were [before transfer]" (Drake, 2014).

In order to continue supporting collecting institutions' stewardship of born-digital materials, the BitCurator project must ensure that is community and its software will persist into the future. During BitCurator's start-up phase, the project has been led by personnel at partner institutions—the School of Information and Library Science at the University of North Carolina at Chapel Hill and the Maryland Institute for Technology in the Humanities—with support from the Andrew W. Mellon Foundation (2011–2014). Now, having developed a strong community of practice around the software, BitCurator is transitioning to its long-term model of stewardship and governance. Going forward, it will be overseen by the BitCurator Consortium, with an administrative base in the Educopia Institute and funding based on institutional membership dues (BitCurator, n.d., "BitCurator Consortium"). There are two categories of membership in the consortium: general and charter. Among other benefits, general members receive help desk support, service opportunities, and voting rights within the consortium. Charter members (a membership option available through December 31, 2014) receive all the benefits of general membership, as well as the opportunity to engage in higher level leadership of the consortium during its early days.

More information on BitCurator is available on the project's website, including the white paper titled "From Bitstreams to Heritage: Putting Digital Forensics into Practice in Collecting Institutions" (Lee et al, 2013).

[2] See Rice Fondren Library (2014) for their publicly posted digital preservation policy.

References

BitCurator. (n.d.) “About: The Project.” Retrieved August 11, 2014, from BitCurator website: http://www.bitcurator.net/aboutbc/#project

BitCurator. (n.d.) “BitCurator Consortium.” Retrieved August 11, 2014, from BitCurator website: http://www.bitcurator.net/bitcurator-consortium/

BitCurator. (n.d.) “FAQ.” Retrieved August 11, 2014, from Bitcurator wiki: http://wiki.bitcurator.net/index.php?title=FAQ

Drake, J.M. (2014, June 27) “It Takes a System to Curate a Bit: Preservation Planning at the Princeton University Archives.” Presentation to the Preservation Metadata Interest Group at the American Library Association Annual Conference. Las Vegas, Nevada.

Kirschenbaum, M.G., Ovenden, R., & Redwine, G. (2010, December) “Digital Forensics and Born-Digital Content in Cultural Heritage Collections” [PDF document]. CLIR Report 149. Washington, DC: Council on Library and Information Resources. Retrieved from http://www.clir.org/pubs/reports/reports/pub149/pub149.pdf

Lee, C.A., Woods, K., Kirschenbaum, M.G., & Chassanoff, A. (2013, September 30) “From Bitstreams to Heritage: Putting Digital Forensics into Practice in Collecting Institutions” [PDF document]. Retrieved from BitCurator website: http://www.bitcurator.net/wp-content/uploads/2013/11/From-Bitstream-to-Heritage-S.pdf

Lee, C.A. (2014, June 27) “An Introduction to the BitCurator Environment.” Presentation to the Preservation Metadata Interest Group at the American Library Association Annual Conference. Las Vegas, Nevada.

Russell, R. & Focke, A. (2014, June 27) “Implementing BitCurator at Rice University: Baby Steps to Digital Preservation Glory.” Presentation to the Preservation Metadata Interest Group at the American Library Association Annual Conference. Las Vegas, Nevada.

Rice Fondren Library. (2014, June) “Woodson Research Center Digital Preservation Policy” [PDF document]. Retrieved from http://library.rice.edu/collections/WRC/WRCDigitalPreservationPolicyrevJune2014.pdf