Provenance, Journal of the Society of Georgia Archivists

Volume 30 | Number 1 Article 2

January 2012

Digital Forensics Meets the Archivist (And They Seem to Like Each Other)

Christopher A. Lee
University of North Carolina Chapel Hill

Digital Forensics Meets the Archivist (And They Seem to Like Each Other) *

Christopher A. Lee

Materials with archival value are now predominantly "born digital." Archivists have unprecedented opportunities to acquire and preserve traces of human and associated machine activity. Seizing these opportunities will require archivists to extract digital materials from their storage or transfer media in ways that reflect the metadata and ensure the integrity of the materials. They must also support and mediate appropriate access: allowing users to make sense of materials and their context, while also preventing inadvertent disclosure of sensitive data.

There are a variety of methods, strategies and applications from the field of digital forensics that archivists are beginning to incorporate into their workflows. The application of digital forensics to their collections allows archivists to advance the fundamental concepts of provenance, original order and chain of custody.

Digital records can be considered and encountered at multiple levels of representation, ranging from aggregations of records down to bits as physically inscribed on a storage medium; each level of representation can provide distinct contributions to the information and evidential value of records. There is a substantial body of information within the underlying data

* ***Note from the editor:*** The Society of Georgia Archivists was honored to have Cal Lee as the keynote speaker for the 2012 Annual Meeting. His keynote about digital records and digital forensics was based on his previous writings and presentations. His contribution to *Provenance* is a summary of his presentation with a bibliography for further reading.

structures of computer systems that often can be discovered or recovered, revealing new types of records or essential metadata associated with existing record types.

Archives can incorporate a variety of forensics practices and methods by treating disk images – rather than individual files or packaged directories – as basic units of acquisition. A disk image is a complete copy of every storage sector from a drive, which captures many forms of information that can be lost in a simple file copy. Using write blockers, creating full disk images and extracting data associated with files can all be essential to ensuring provenance, original order and chain of custody. Incorporation of digital forensics methods also will be essential to the sustainability of archives as stewards of personally identifying information; the same tools that are used to expose sensitive information can be used to identify, flag and redact or restrict access to it.

Digital forensics offers valuable methods that can advance the archival goals of maintaining authenticity, describing born-digital records and providing responsible access. However, most digital forensics tools were not designed with archival objectives in mind. The BitCurator project is attempting to bridge this gap through engagement with digital forensics, library and archives professionals, as well as dissemination of tools and documentation that are appropriate to the needs of memory institutions. Funded by the Andrew W. Mellon Foundation, BitCurator is a joint effort – led by the School of Information and Library Science at the University of North Carolina, Chapel Hill (SILS) and Maryland Institute for Technology in the Humanities (MITH), and involving contributors from several other institutions—to develop a system for librarians and archivists that incorporates the functionality of many digital forensics tools. Much of the BitCurator activity is translation and adaptation work, based on the belief that archivists will benefit from tools that are presented in ways that use familiar language and run on platforms that archivists can support.

Two groups of external partners are contributing to BitCurator: a Professional Expert Panel (PEP) of individuals who are at various stages of implementing digital forensics tools and methods in their collecting institution contexts, and a Development

Advisory Group (DAG) of individuals who have significant experience with development of software. Input from the PEP and DAG have helped us to refine the project's requirements and clarify the goals and expectations of working professionals.

BitCurator is packaging, adapting and disseminating a variety of open-source applications. Rather than developing everything from scratch, BitCurator is able to benefit from numerous existing open-source tools, many of which are now quite mature. The goal is to provide a set of tools that can be used together to perform archival tasks but can also be used in combination with many other existing and emerging applications.

For Further Reading:

AIMS Working Group. "AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship." 2012.

BitCurator Project. http://bitcurator.net

Forensics Wiki. http://www.forensicswiki.org/

Garfinkel, Simson and David Cox, "Finding and Archiving the Internet Footprint," Paper presented at the First Digital Lives Research Conference: Personal Digital Archives for the 21st Century, London, UK, February 9-11, 2009.

Gengenbach, Martin J. "'The Way We Do it Here': Mapping Digital Forensics Workflows in Collecting Institutions." A Master's Paper for the M.S. in L.S degree. August, 2012.

Kirschenbaum, Matthew G., Richard Ovenden, and Gabriela Redwine. "Digital Forensics and Born-Digital Content in Cultural Heritage Collections." Washington, DC: Council on Library and Information Resources, 2010.

John, Jeremy Leighton. "Adapting Existing Technologies for Digitally Archiving Personal Lives: Digital Forensics, Ancestral Computing, and Evolutionary Perspectives and Tools." Paper presented at iPRES 2008: The Fifth International Conference on Preservation of Digital Objects, London, UK, September 29-30, 2008.

Lee, Christopher A. "Digital Curation as Communication Mediation," In Handbook of Technical Communication, edited by Alexander Mehler, Laurent Romary, and Dafydd Gibbon, 507-530. Berlin: Mouton De Gruyter, 2012.

Lee, Christopher A., Matthew Kirschenbaum, Alexandra Chassanoff, Porter Olsen, and Kam Woods. "BitCurator: Tools and Techniques for Digital Forensics in Collecting Institutions." D-Lib Magazine 18, No. 5/6 (May/June 2012). http://www.dlib.org/dlib/may12/lee/05lee.html

Ross, Seamus and Ann Gow. "Digital Archaeology: Rescuing Neglected and Damaged Data Resources." London: British Library, 1999.

Woods, Kam and Geoffrey Brown. "Creating Virtual CD-ROM Collections." International Journal of Digital Curation 4, no. 2 (2009): 184-198

Woods, Kam and Geoffrey Brown. "From Imaging to Access – Effective Preservation of Legacy Removable Media." In Proceedings of Archiving 2009, 213-18. Springfield, VA: Society for Imaging Science and Technology, 2009.

Woods, Kam and Christopher A. Lee. “Acquisition and Processing of Disk Images to Further Archival Goals." In Proceedings of Archiving 2012 (Springfield, VA: Society for Imaging Science and Technology, 2012), 147-152.

Woods, Kam, Christopher A. Lee, and Simson Garfinkel. "Extending Digital Repository Architectures to Support Disk Image Preservation and Access." In JCDL '11: Proceeding of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries, 57-66. New York, NY: ACM Press, 2011.

Christopher (Cal) Lee is Associate Professor at the School of Information and Library Science at the University of North Carolina, Chapel Hill. His primary area of research is the long-term curation of digital collections. He is particularly interested in the professionalization of this work and the diffusion of existing tools and methods into professional practice. Lee edited and provided several chapters to *I, Digital: Personal Collections in the Digital Era*. He is Principal Investigator of the BitCurator project, which is developing and disseminating open-source digital forensics tools for use by archivists and librarians.