Archival Application of Digital Forensics Methods for Authenticity, Description and Access Provision

Christopher A. Lee, University of North Carolina at Chapel Hill

Summary

When acquiring born-digital materials, archivists often must extract digital materials from media in ways that reflect the rich metadata associated with records and ensure records' integrity. They must also allow users to make sense of materials and understand their context, while preventing inadvertent disclosure of sensitive data. There are methods and strategies from the field of digital forensics that can aid this work. This paper discusses the development and application of digital forensics tools to improve the acquisition, management and access functions of archives. It reports on the BitCurator project, which is identifying current and desirable workflows of several archival institutions, as well as developing and testing tools to support the workflows. Incorporation of digital forensics methods also will be essential to the sustainability of archives as stewards of personally identifying information. There are a variety of potential changes within the archival profession that are associated with adopting digital forensics tools and practices.

Introduction

Materials with archival value are now predominantly "born digital," and archivists have unprecedented opportunities to acquire and preserve traces of human and associated machine activity. In order to seize these opportunities, archivists must be able 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 understand 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 can aid this work.

Applying Archival Principles to Born-Digital Acquisitions

Three fundamental archival concepts are provenance, original order and chain of custody. The provenance of a record is its "life history." Although the concept had precedents in the archival writing and practices of several countries of the seventeenth, eighteenth, and early nineteenth centuries,[1] the first widely recognized articulation of the principle of provenance—called *respect des fonds*—was in France in 1841.[2]

For purposes of describing archival collections, one of the most important aspects of provenance is the identification of one or more origins or sources of a record (for example, the person who wrote a diary entry or the specific business transaction that generated a receipt). However, provenance more broadly "consists of

the social and technical processes of the records' inscription, transmission, contextualization, and interpretation which account for its existence, characteristics, and continuing history."[3] According to the principle of provenance, records from a common origin or source should be managed together as an aggregate unit and should not be arbitrarily intermingled with records from other origins or sources.

The etymology of the term provenance is from the Old French, meaning origin or cause.[4] There are many different interactions with records that are important to document, in order to understand the records' origins and "life history" (e.g., those who influenced the creation of the records, those who received them, custodians who transformed them over time), not simply one isolated moment of creation. These considerations illustrate the importance of provenance not only as the source of a record but also as a "history of the ownership . . . used as a guide to authenticity or quality" and "a documented record of this" history.[5] Given the complex and evolving relationships between entities (e.g., people, agencies) and records, provenance is not simply a matter of identifying the one person who created a record at a point in time but instead "relate[s] a multitude of contextual entities to a multitude of recordkeeping entities in a multitude of ways."[6] In digital environments, it can be important to consider provenance at levels of granularity finer than an entire record, such as why a specific data element appears within a dataset and where specifically the data element was generated; [7] and to include additional technical components in one's notion of provenance, such as system configuration information.[8]

Closely related to provenance is the principle of original order, which indicates that archivists should organize and manage records in ways that reflect their arrangement within the creation environment. The concept of original order had some precedent in archival writings of late-nineteenth-century Italy, [9] but it was most strongly influenced by Prussian archival practice in the late nineteenth century.[10] Its most widely recognized articulation within the context of archival description was the Manual for the Arrangement and Description of Archives—known as the "Dutch Manual"—which was originally published in 1898 and first published in English translation in 1940.[11] For personal records, the principle of original order implies that archivists should carry forward (either by perpetuating or attempting to reconstruct) the peculiar ways in which individuals label and organize their own records.

There are two primary motivations for retaining original order:

- It "preserves existing relationships and evidential significance that can be inferred from the context of the records." [12] This supports what Hugh Taylor calls "authentic pattern recognition."[13]
- It "exploits the record creator's mechanisms to access the records, saving the archives the work of creating new access tools."[14]

A compelling argument for retaining original order in a digital environment is that—even if that order is messy and idiosyncratic—it conveys meaningful information about the recordkeeping context, and additional

layers of description can be laid on top of that order to facilitate various forms of navigation and access.[15] However, rather than simply "freezing or restoring one particular past arrangement as 'the' original order," [16] original order is most usefully understood within the context of a larger, ongoing chain of custody.

The chain of custody is the "succession of offices or persons who have held materials from the moment they were created."[17] For purposes of legal compliance, authenticity, evidential integrity, and legal admissibility, the ideal recordkeeping system would provide "an unblemished line of responsible custody"[18] through control, documentation, and accounting for all states of a record and changes of state (e.g., movement from one storage environment to another, transformation from one file format to another) throughout its existence—from the point of creation to each instance of use and (when appropriate) destruction.

The reality of contemporary information management is rarely consistent with the recordkeeping ideal. In most cases, the best that an information professional can do is to capture or create limited documentation of the portion of the chain of custody that occurred before he/she first encountered the records, and then attempt to provide much more detailed chain of custody control and documentation from that point forward. For example, an archivist acquiring a floppy disk containing records from a donor often will not know with certainty what the states and transitions of the records were before they were last saved onto that disk, but she can use various forms of information (e.g., other records, discussions with the donor) to make inferences about earlier points in the "life" of the records. Tom Nesmith points out that archivists' knowledge about various aspects of the "origins of a record" are "bathed in hypothesis."[19]

Archivists must increasingly apply their professional principles to collections composed – in whole or in part – of born-digital materials. Among other activities, this includes moving records that are stored on removable media into more sustainable preservation environments. This can involve media that are already in their holdings (e.g. disks stored in boxes along with paper materials), as well as materials that they are acquiring for the first time from individual donors or other producers.

The literature on digital archives tends to place a great emphasis on the "virtual" (i.e. intangible) nature of electronic resources. Computer systems have "an illusion of immateriality by detecting error and correcting it,"[20] but it is essential to recognize that digital objects are created and perpetuated through physical things (e.g. charged magnetic particles, pulses of light, holes in disks). This materiality brings challenges, because data must be read from specific artifacts, which can become damaged or obsolete. However, the materiality of digital objects also brings unprecedented opportunities for description, interpretation and use.[21]

If records are "persistent representations of activities or other occurrents," it is important to recognize that one "can expect to find representations at many different levels."[22] These are not just levels in the

functional hierarchy of records but also levels of representation. 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.[23] There is a substantial body of information within the underlying data structures of computer systems that can often be discovered or recovered, revealing new types of records or essential metadata associated with existing record types.

Recovery of data from physical media has been a topic of discussion in the professional library and archives literature for several years. More than a decade ago, a report by Seamus Ross and Ann Gow discussed the potential relevance of advances in data recovery and digital forensics to collecting institutions.[24] More recently, there has been an active stream of literature related to the use of forensic tools and methods for acquiring and managing digital collections.[25] A project called "Computer Forensics and Born-Digital Content in Cultural Heritage Collections" hosted a symposium and generated a report,[26] which provided significant contributions to this discussion. The Born Digital Collections: An Inter-Institutional Model for Stewardship (AIMS) project developed a framework for the stewardship of born-digital materials that includes the incorporation of digital forensics methods.[27] The Digital Records Forensics project has also articulated a variety of connections between the concepts of digital forensics and archival science.[28]

Forensic Tools and Methods to Support Archival Functions

Access to data from a storage device normally involves mounting a volume and then copying or opening files through the file system. There must be hardware to detect signals on the medium, hardware and software to translate the signals into bitstreams, and hardware and software to move the bitstreams into the current working computer environment. One can then interact with data as entire files or components of files. The filesystem usually plays a mediating role between the user and the underlying data, and it is designed to facilitate interaction at the file level (e.g. file naming, viewing timestamps, access controls). The filesystem serves to "hide" complicated information from the user about "where and how it stores information."[29] For most purposes, the filesystem is a very valuable abstraction mechanism, because it does not require users to understand or directly access the underlying data.

Those who are interested in the underlying data that is hidden by the filesystem can instead generate and interact with disk images, which are low-level, sector-by-sector copies of all the data that resides on the storage medium. Inspection of the disk image can reveal a significant amount of information that users of the drive did not consciously or intentionally leave there, but can serve as traces of valuable contextual information.[30] Forensic workflows often involve creation of a disk image to serve as a baseline copy of the data from the disk, upon which many further extraction and analysis tasks can be performed. Digital

forensics professionals use hardware write blockers to ensure that no data on the disk – including essential metadata such as timestamps – are altered or overwritten during the process of copying the disk's contents.

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.[31] Using write blockers, creating full disk images and extracting data associated with files is 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.

Emerging Emphasis on Personal Archives

Much of the recent innovation in the application of digital forensics to archives has been undertaken within the context of acquiring personal archives, as opposed to institutional records.[32] Personal papers and manuscripts have long been part of the archival profession's charge, but their status has often been ambiguous. There has been much debate in the archival literature over the status of both non-institutional records and the "manuscripts tradition" more generally.[33] This tension has been addressed most explicitly in a special issue of *Archives and Manuscripts* in 1996 on "Personal Recordkeeping: Issues and Perspectives."[34]

Practical guidance and empirical findings about the archival treatment of personal papers and manuscripts has tended to be underrepresented in the published professional literature. This has been particularly true in the last several decades, as notions of evidential value, recordkeeping systems, and institutional accountability have driven much of the research. In 2006, Toby Burrows lamented, "Though the range of issues relating to personal electronic archives has been relatively well-documented, there is as yet little in the way of systematic investigation of solutions and approaches."[35] However, there has been a recent influx of publications in the archival literature related to personal archives, with much of the focus being on born-digital records created by individuals.

BitCurator

The BitCurator project 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.[36]

Digital forensics offers valuable methods that can advance the archival goals of maintaining authenticity, describing born-digital records and providing responsible access.[37] 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. Much 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.

BitCurator – and the efforts of many of the project partners – also aim to address two fundamental needs of archives that are not priorities for digital forensics industry software developers:

(1) Incorporation into the workflows of archives and libraries, e.g. supporting metadata conventions, connections to existing content management system (CMS) environments. This includes exporting forensic data in ways that can then be imported into archival descriptive systems, as well as modifying forensics triage techniques to better meet the needs of archivists.

(2) Provision of public access to the data. The typical digital forensics scenario is a criminal investigation in which the public never gets access to the evidence that was seized. By contrast, archives that are creating disk images face issues of how to provide access to the data. This includes not only access interface issues, but also how to redact or restrict access to components of the image, based on confidentiality, intellectual property or other sensitivities.

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. The core project team met with the PEP in December of 2011 and the DAG in January of 2012 to discuss the design assumptions and goals of the project. We have also received comments and suggestions from individuals in a variety of organizational settings. These various forms of input have helped us to refine the project's requirements and clarify the goals and expectations of working professionals.

The project 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.[38] 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.

Conclusion

As archivists take on the curation of born-digital materials such as floppy disks found in boxes and new acquisitions on media such as hard drives and flash drives, they are now learning and applying many methods that have been used within digital forensics for many years. Digital forensics tools and methods hold great promise for enhancing and improving the work practices of archivists who are responsible for digital records.

There are a variety of changes within the archival profession that are implied by the above trend. First, the professional vocabulary of archivists is evolving to now include terms such as disk image, hex[adecimal] viewer, cryptographic hash, and filesystem. Second, archivists are gaining access to new professional communities and sources of guidance, e.g. papers from the annual Digital Forensics Research Workshop and instructions from gaming enthusiasts about how to create, read and mount disk images of old storage media. The first and second points are closely related; having the right vocabulary can open up many new mechanisms for learning and sharing information.

A third change in the archival profession comes from the use of tools that were designed to treat data at a very low level – as raw bitstreams off media – rather than treating data at the file level. Archivists have long argued that the essential content, structure and context elements of an electronic record can reside in multiple data sources and not just in a single file.[39] Digital forensics greatly enables such thinking; for example, it allows archivists to bypass the filesystem and read data as a raw stream to be decomposed into records as appropriate.

Finally, the introduction of digital forensics into archives has the potential to shift the "center of gravity" about electronic records in the archival literature from the design of institutional recordkeeping systems toward the acquisition and management of records from a much more diverse and unpredictable set of sources.

The intersection between digital forensics and archives can be characterized as a "trading zone" that resides between different streams of activity.[40] Actors from different streams of activity can agree to use a common set of terms, concepts and methods in order to share ideas and coordinate their work, even if they still hold dramatically different worldviews, values or assumptions of their own responsibilities. It is likely that fundamental elements of digital forensics language and practice will ultimately become so embedded in the archival enterprise that archivists no longer perceive them as being borrowed from elsewhere; they will simply be part of what archivists do. As archivists develop new methods and tools that are based on forensics building blocks, hopefully they will also be to make contributions to the field of digital forensics that it can ultimately adopt as established practice. However, it is also likely that the frontiers of digital forensics archival research will continue to develop independently, based on distinct values, mandates and constraints. There is the potential for creative and well-informed translation work across the two streams for many years ahead.

Acknowledgements

The BitCurator project is supported by a grant from the Andrew W. Mellon Foundation.

[1] Maynard Brichford, "The Provenance of Provenance in Germanic Areas," *Provenance* 7, no. 2 (1989): 54– 70; Shelley Sweeney, "The Ambiguous Origins of the Archival Principle of 'Provenance,'" *Libraries and the Cultural Record* 43, no. 2 (2008): 193–213.
[2] Nancy Bartlett, "Respect Des Fonds: The Origins of the Modern Archival Principle of Provenance," *Primary Sources and Original Works* 1, no. 1/2 (1991): 107–15.
[3] Tom Nesmith, "Still Fuzzy, but More Accurate: Some Thoughts on the 'Ghosts' of Archival Theory," *Archivaria* 47 (1999): 146.
[4] *Oxford English Dictionary*, Draft Revision (June 2008), s.v. "provenance."
[5] *Oxford English Dictionary*, s.v. "provenance."
[6] Chris Hurley, "Problems with Provenance," *Archives and Manuscripts* 23, no. 2 (1995): 256–57.
[7] Peter Buneman, Sanjeev Khanna, and Wang-Chiew Tan, "Why and Where: A Characterization of Data Provenance," in *Database Theory - ICDT 2001: 8th International Conference, London, UK, January 2001. Proceedings*, ed. Jan van den Bussche and Victor Vianu (Berlin: Springer, 2001), 316–30.
[8] Maria Guercio, "Archival Theory and the Principle of Provenance for Current Records: Their Impact on Arranging and Inventorying Electronic Records," in *The Principle of Provenance: Report from the First Stockholm Conference on Archival Theory and the Principle of Provenance, 2-3 September 1993*, ed. Kerstin Abukhanfusa and Jan Sydbeck (Stockholm: Swedish National Archives, 1994), 82.
[9] Michel Duchein, "The History of European Archives and the Development of the Archival Profession in Europe," *American Archivist* 55, no. 1 (1992): 20.
[10] Ernst Posner, "Max Lehmann and the Genesis of the Principle of Provenance," in *Archives and the Public Interest: Selected Essays by Ernst Posner*, ed. Ken Munden (Chicago: Society of American Archivists, 2006), 36–44.
[11] Samuel Muller, Johan Adriaan Feith, and R. Fruin, *Manual for the Arrangement and Description of Archives: Drawn up by Direction of the Netherlands Association of Archivists*, trans. Arthur H. Leavitt (Chicago: Society of American Archivists, 2003).
[12] Pearce-Moses, *Glossary of Archival and Records Terminology*, 280–81.
[13] Hugh A. Taylor, "The Collective Memory: Archives and Libraries as Heritage," *Archivaria* 15 (1982–83): 122.
[14] Pearce-Moses, *Glossary of Archival and Records Terminology*, 281.
[15] Peter Horsman, "Dirty Hands: A New Perspective on the Original Order," *Archives and Manuscripts* 27, no. 1 (1999): 42–53.
[16] Peter Horsman, "The Last Dance of the Phoenix, or the De-Discovery of the Archival Fonds," *Archivaria* 54 (2002): 19.
[17] Pearce-Moses, *Glossary of Archival and Records Terminology*, 67.
[18] Hilary Jenkinson, *A Manual of Archive Administration: Including the Problems of War Archives and Archive Making* (Oxford: Clarendon Press, 1922), 11.
[19] Nesmith, "Still Fuzzy, but More Accurate," 141. Likewise, a forensic investigator can take great care to ensure that the bits stored on a computer that was seized at a crime scene have not changed since the time that the computer was seized, but anything that happened on the computer before that point is a matter of informed speculation. See, e.g. Michael A. Caloyannides, "Digital 'Evidence' Is Often Evidence of Nothing," in *Digital Crime and Forensic Science in Cyberspace*, ed. Panagiotis Kanellis (Hershey, PA: Idea Group, 2006), 334–39; Brian D. Carrier, "A Hypothesis-Based Approach to Digital Forensic Investigations" (doctoral dissertation, Purdue University, 2006).
[20] Matthew Kirschenbaum, *Mechanisms: New Media and the Forensic Imagination* (Cambridge, MA: MIT Press, 2008).
[21] John Lavagnino, "The Analytical Bibliography of Electronic Texts." Paper presented at the Joint Annual Conference of the Association for Literary and Linguistic Computing and the Association for Computers and the Humanities, Bergen, Norway, 1996.
[22] Geoffrey Yeo, "Concepts of Record (2): Prototypes and Boundary Objects," *American Archivist* 71, no. 1 (2008): 118-43.
[23] Christopher A. Lee, "Digital Curation as Communication Mediation," In *Handbook of Technical Communication*, edited by Alexander Mehler, Laurent Romary, and Dafydd Gibbon (Mouton De Gruyter, forthcoming).

[24] Seamus Ross and Ann Gow, "Digital Archaeology: Rescuing Neglected and Damaged Data Resources" (London: British Library, 1999).
[25] William E. Underwood and Sandra L. Laib, "PERPOS: An Electronic Records Repository and Archival Processing System." Paper presented at the International Symposium on Digital Curation (DigCCurr 2007), Chapel Hill, NC, April 18-20, 2007; Kam Woods and Geoffrey Brown, "Migration Performance for Legacy Data Access." *International Journal of Digital Curation* 3, no. 2 (2008): 74-88; Douglas Elford, Nicholas Del Pozo, Snezana Mihajlovic, David Pearson, Gerard Clifton, and Colin Webb, "Media Matters: Developing Processes for Preserving Digital Objects on Physical Carriers at the National Library of Australia," Paper presented at the 74th IFLA General Conference and Council, Québec, Canada, August 10-14, 2008; Jeremy Leighton John, "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; William Bradley Glisson, "Use of Computer Forensics in the Digital Curation of Removable Media." In *Proceedings of DigCCurr2009: Digital Curation: Practice, Promise, and Prospects*, edited by Helen R. Tibbo, Carolyn Hank, Christopher A. Lee, and Rachael Clemens (Chapel Hill, NC: University of North Carolina, School of Information and Library Science, 2009), 110-1; Simson Garfinkel 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; Kam Woods 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* (Springfield, VA: Society for Imaging Science and Technology), 213-18; Underwood, William, Marlit Hayslett, Sheila Isbell, Sandra Laib, Scott Sherrill, and Matthew Underwood. "Advanced Decision Support for Archival Processing of Presidential Electronic Records: Final Scientific and Technical Report." Technical Report ITTL/CSITD 09-05. October 2009.
[26] Matthew G. Kirschenbaum, Richard Ovenden, and Gabriela Redwine, "Digital Forensics and Born-Digital Content in Cultural Heritage Collections" (Washington, DC: Council on Library and Information Resources, 2010).
[27] AIMS Working Group. "AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship." 2012.
[28] Luciana Duranti, "From Digital Diplomatics to Digital Records Forensics," Archivaria 68 (2009): 39-66; Luciana Duranti, and Barbara Endicott-Popovsky, "Digital Records Forensics: A New Science and Academic Program for Forensic Readiness," Journal of Digital Forensics, Security and Law 5, no. 2 (2010); Sherry L Xie, "Building Foundations for Digital Records Forensics: A Comparative Study of the Concept of Reproduction in Digital Records Management and Digital Forensics," *American Archivist* 74, no. 2 (2011): 576-99.
[29] Dan Farmer and Wietse Venema, *Forensic Discovery* (Upper Saddle River, NJ: Addison-Wesley, 2005).
[30] Simson L. Garfinkel and Abhi Shelat, "Remembrance of Data Passed: A Study of Disk Sanitization Practices," *IEEE Security and Privacy* 1 (2003): 17-27.
[31] Kam Woods, 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* (New York, NY: ACM Press, 2011), 57-66.
[32] This is not a distinction that holds consistently across languages, nations or recordkeeping traditions. However, it is a division that has had significant professional implications in many countries. For a discussion of differences in terminology around this issue, see Christopher A. Lee, "Introduction," in *I, Digital: Personal Collections in the Digital Era*, edited by Christopher A. Lee (Chicago, IL: Society of American Archivists, 2011), 1-26.
[33] Curtis W. Garrison, "The Relation of Historical Manuscripts to Archival Materials," *American Archivist* 2, no. 2 (1939): 97–105; Richard C. Berner, "Manuscript Collections, Archives, and Special Collections: Their Relationships," *Archivaria* 18 (1984): 248–54; Luke J. Gilliland-Swetland, "The Provenance of a Profession: The Permanence of the Public Archives and Historical Manuscripts Traditions in American History," *American Archivist* 54 (1991): 160–75.
[34] Adrian Cunningham, "Editorial: Beyond Corporate Accountability," *Archives and Manuscripts* 24, no. 1 (1996): 6–11.
[35] Toby Burrows, "Personal Electronic Archives: Collecting the Digital Me," *OCLC Systems & Services* 22, no. 2 (2006): 87.
[36] Christopher A. Lee, 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).
[37] 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.

[38] Tools that BitCurator is incorporating include Guymager, a program for capturing disk images; bulk extractor, for extracting features of interest from disk images (including private and individually identifying information); fiwalk, for generating Digital Forensics XML (DFXML) output describing filesystem hierarchies contained on disk images; The Sleuth Kit (TSK), for viewing, identifying and extraction information from disk images; Nautilus scripts to automate the actions of command-line forensics utilities through the Ubuntu desktop browser; and sdhash, a fuzzing hashing application that can find partial matches between similar files. For further information about several of these tools, see Michael Cohen, Simson Garfinkel, and Bradley Schatz, "Extending the Advanced Forensic Format to Accommodate Multiple Data Sources, Logical Evidence, Arbitrary Information and Forensic Workflow," *Digital Investigation* 6 (2009): S57-S68; Simson Garfinkel, "Digital Forensics XML and the DFXML Toolset," *Digital Investigation* 8 (2012): 161-174; Simson L. Garfinkel, "Providing Cryptographic Security and Evidentiary Chain-of-Custody with the Advanced Forensic Format, Library, and Tools," *International Journal of Digital Crime and Forensics* 1, no. 1 (2009): 1-28; Vassil Roussev, "An Evaluation of Forensic Similarity Hashes," Digital Investigation 8 (2011): S34-S41.

[39] See e.g., David Bearman, "Record-Keeping Systems," *Archivaria* 36 (1993): 16-36; John McDonald, "Towards Automated Record Keeping, Interfaces for the Capture of Records of Business Processes," *Archives and Museum Informatics* 11 (1997): 277-85.

[40] Peter Louis Galison, *Image and Logic: A Material Culture of Microphysics* (Chicago: University of Chicago Press, 1997).