Demonstrating a Digital Curation Workflow using the BitCurator Environment

Christopher A. Lee
School of Information and Library Science
University of North Carolina
216 Lenoir Drive, CB #3360
1-(919)-966-3598
callee@ils.unc.edu

ABSTRACT

This demonstration will highlight several key steps in a digital curation workflow that incorporates digital forensics tools and methods. Using the open-source BitCurator environment, I will demonstrate several discrete tasks, how they can feed into each other, and considerations related to incorporating them into a larger set of curation practices within collecting institutions. A strong emphasis will be placed on features of the software that have been added or enhanced over the past year, including mounting and exporting of files from forensically packaged disk images, identification of duplicate files, generation of PREMIS metadata and initial steps toward redaction of potentially sensitive information.

Categories and Subject Descriptors

H.3.7 [**Information Storage and Retrieval**]: Digital Libraries – *collection, dissemination, systems issues.*

General Terms

Provenance, Data Triage, Digital Forensics.

Keywords

Forensics, preservation, DFXML, metadata, privacy, collections, acquisition

1. BITCURATOR PROJECT

The BitCurator Project, a collaborative effort led by the School of Information and Library Science at the University of North Carolina at Chapel Hill and Maryland Institute for Technology in the Humanities at the University of Maryland, is addressing two fundamental needs and opportunities for collecting institutions: (1) integrating digital forensics tools and methods into the workflows and collection management environments of libraries, archives and museums and (2) supporting properly mediated public access to forensically acquired data [4].

2. BITCURATOR ENVIRONMENT

We are developing and disseminating a suite of open source tools. These tools are being developed and tested in a Linux environment; the software on which they depend can readily be compiled for Windows environments (and in most cases are currently distributed as both source code and Windows binaries). We intend the majority of the development for BitCurator to support cross-platform use of the software. We are freely disseminating the software under an open source (GPL, Version 3) license. BitCurator provides users with two primary paths to integrate digital forensics tools and techniques into archival and library workflows.

First, the BitCurator software can be run as a ready-to-run Linux environment that can be used either as a virtual machine (VM) or installed as a host operating system. This environment is customized to provide users with graphic user interface (GUI)-based scripts that provide simplified access to common functions associated with handling media, including facilities to prevent inadvertent write-enabled mounting (software write-blocking).

Second, the BitCurator software can be run as a set of individual software tools, packages, support scripts, and documentation to reproduce full or partial functionality of the ready-to-run BitCurator environment. These include a software metapackage (.deb) file that replicates the software dependency tree on which software sources built for BitCurator rely; a set of software sources and supporting environmental scripts developed by the BitCurator team and made publicly available at via our GitHub repository (links at http://wiki.bitcurator.net); and all other third-party open source digital forensics software included in the BitCurator environment.

3. DEMONSTRATED TOOLS AND FEATURES

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 [1,2,3,5].

This demonstration place significant emphasis on features of the software that have been added or enhanced over the past year, including mounting and exporting of files from forensically packaged disk images, identification of duplicate files, generation of PREMIS metadata and initial steps toward redaction of potentially sensitive information. Other supported features that will be illustrated in the demonstration include mounting media as read-only, creating disk images, using Nautilus scripts to perform batch activities, generation of Digital Forensics XML (DFXML), generation of customized reports, and identification of sensitive data within data.

4. ACKNOWLEDGMENTS

BitCurator development has been supported by the Andrew W. Mellon Foundation. Members of the BitCurator team are Alexandra Chassanoff, Matthew Kirschenbaum, Christopher (Cal) Lee, Sunitha Misra, Porter Olsen, and Kam Woods. Members of two advisory boards have made valuable contributions: the Development Advisory Group (DAG) and Professional Experts Panel (PEP).

5. REFERENCES

[1] Cohen, M., Garfinkel, S., and Schatz, B. 2009. Extending the Advanced Forensic Format to Accommodate Multiple Data Sources, Logical Evidence, Arbitrary Information and Forensic Workflow. *Digital Investigation* 6 (2009), S57-S68.

[2] Garfinkel, S. Digital Forensics XML and the DFXML Toolset. *Digital Investigation* 8 (2012), 161-174.

[3] Garfinkel, S.L. Providing Cryptographic Security and Evidentiary Chain-of-Custody with the Advanced Forensic Format, Library, and Tools. *International Journal of Digital Crime and Forensics* 1, 1 (2009), 1-28;

[4] Lee, C.A., Kirschenbaum, M.G., Chassanoff, A., Olsen, P., and Woods, K. BitCurator: Tools and Techniques for Digital Forensics in Collecting Institutions. *D-Lib Magazine* 18, 5/6 (May/June 2012).

[5] Roussev, V. An Evaluation of Forensic Similarity Hashes. *Digital Investigation* 8 (2011), S34-S41.