Bitstreams Beyond Borders: The Value of Digital Forensics to Archivists

Cal Lee

School of Information and Library Science

University of North Carolina, Chapel Hill

Society of American Archivists Annual Meeting

August 6-11, 2012

San Diego, CA

BitCurator

UNC SCHOOL OF INFORMATION AND LIBRARY SCIENCE

Archivists now have to deal with this stuff

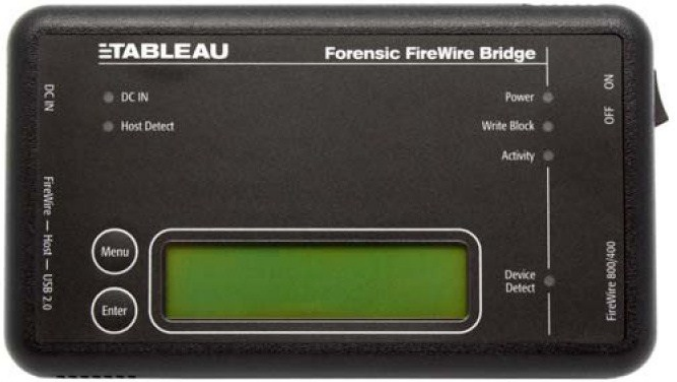
Source: Simson Garfinkel

Source: “Digital Forensics and creation of a narrative.” *Da Blog: ULCC Digital Archives Blog*. http://dablog.ulcc.ac.uk/2011/07/04/forensics/

Born Digital Materials Acquired/Held on Storage Media

- Acquiring and processing information from raw digital sources (e.g. hard disks, floppies) is a common task for archives and other collecting institutions
- Media often contain significant contextual information and potentially private and sensitive information
- Identification and management of this information can be critical to ensure compliance with donor or submission agreements, establish provenance, and enable future access

Luckily, there are many tools and methods



AFFLIB

Open Source Computer Forensics Software

Multiple Paths to Digital Forensics

- Speakers have many different disciplinary backgrounds:
  - Courtney Mumma – English and archival science
  - Kam Woods – computer science
  - Matt Kirshenbaum – English and humanities computing
  - Jeremy Leighton John – zoology
- All have found important connections based on application of digital forensics to archival collections

My own path

- Inspired to take on the challenge of digital preservation while a masters student
- Electronic records archivists for state of Kansas
- Digital preservation research while a doctoral student:
  - CAMiLEON project
  - Case study of development of OAIS Reference Model (dissertation)
- Teaching and research as faculty member at University of North Carolina

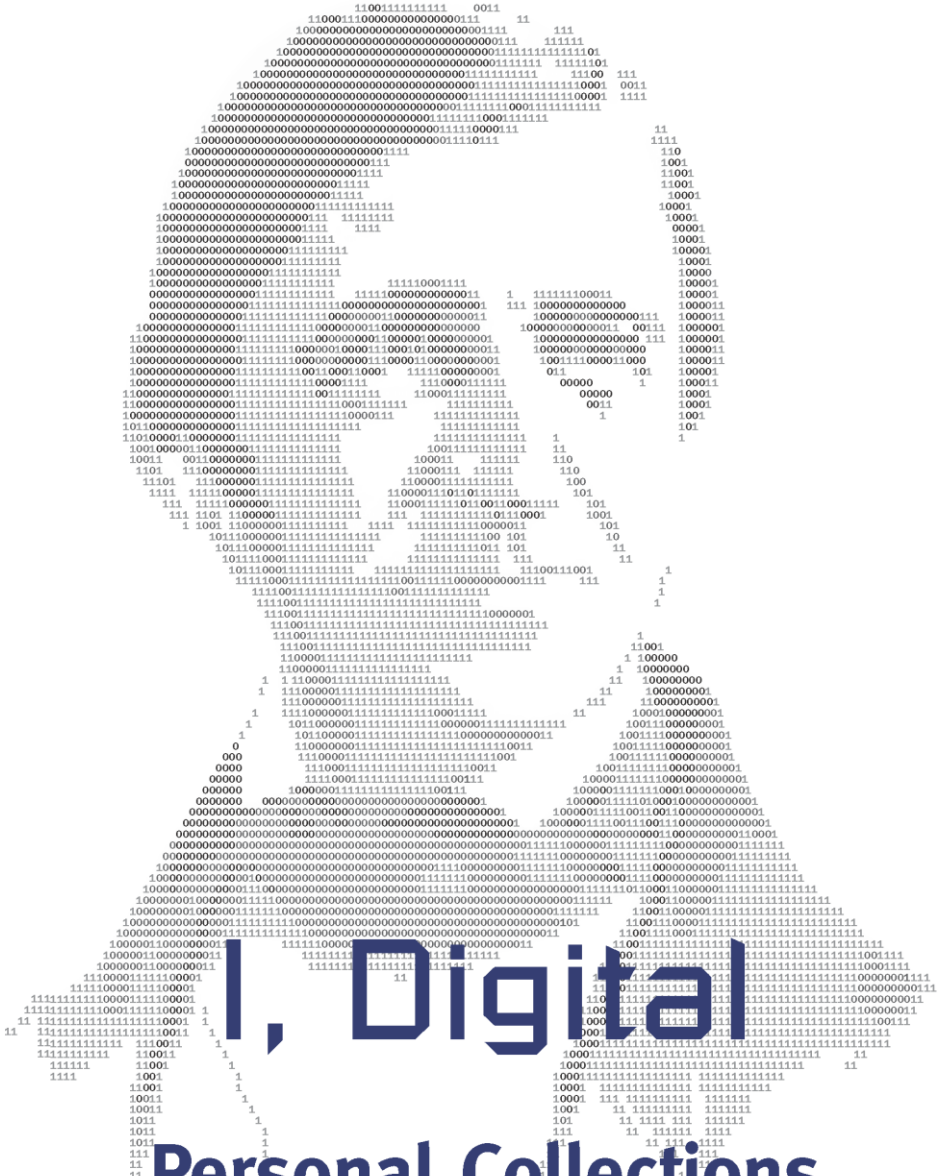
Why I Got Here

- Goal: enabling archivists to take on practical challenges that they face
- Belief that archivists cannot act professionally or responsibly unless they understand and can engage with the information technology that enacts recordkeeping
- Two persistent motivations:
  - Diffusion of innovations[1]
  - System of professions[2]

1. Rogers, Everett M. *Diffusion of Innovations. 5th ed. New York: Free Press, 2003.*
2. Abbott, Andrew Delano. *The System of Professions: An Essay on the Division of Expert Labor. Chicago: University of Chicago Press, 1988.*

Personal Digital Archives

- Arguably neglected by archival literature until quite recently
- Rich and (often at-risk) materials that have historically served as essential part of the archival endeavor
- Growing importance as existing types of records are now born-digital and new types of personal traces emerge



I, Digital

Personal Collections in the Digital Era

Edited by Christopher A. Lee

Digital Forensics - Some Basics

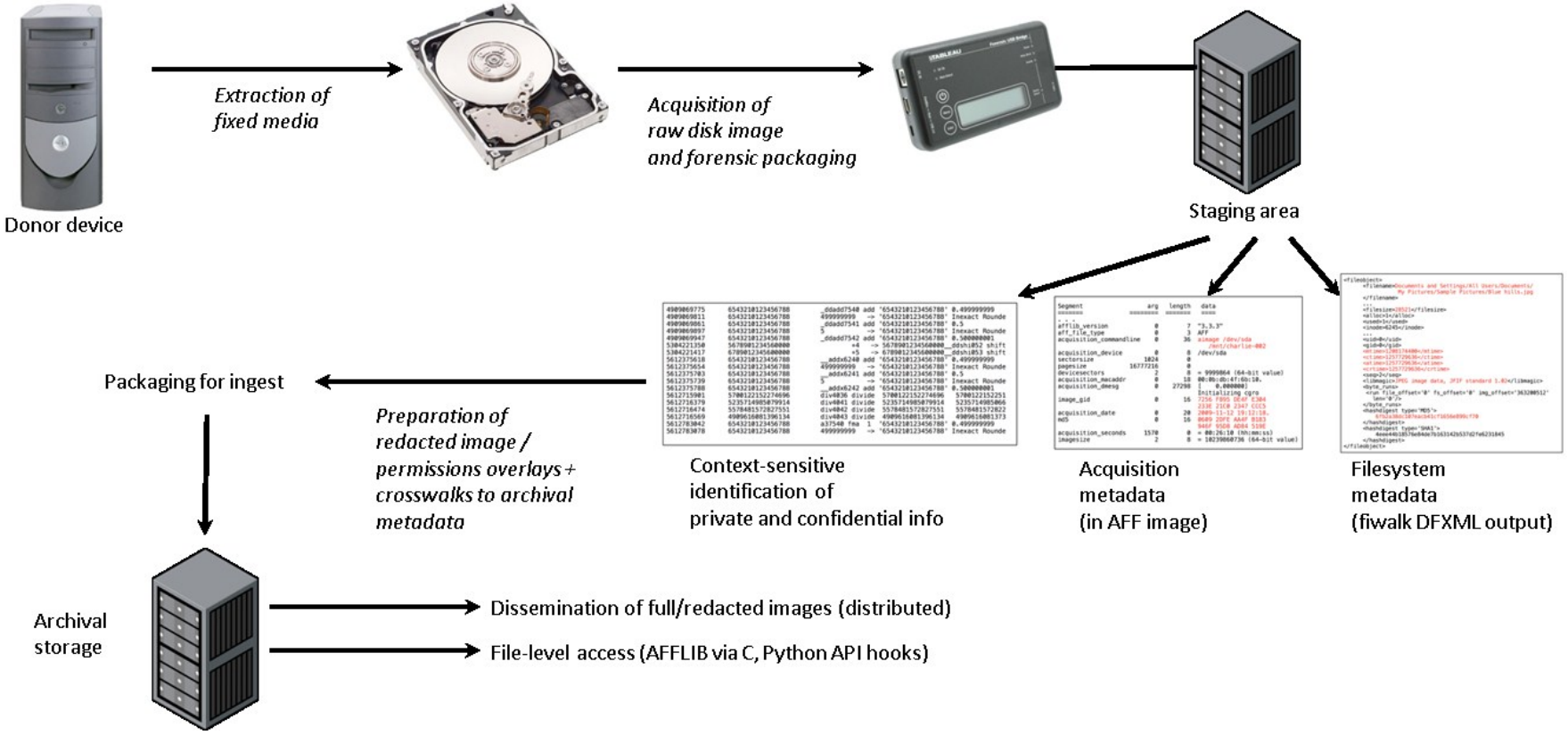
Low-Level Copying of Data Off Media – Disk Imaging

- Getting an “image” of a medium involves working at a level below the file system
- Can get at file attributes and data not visible through higher-level copy operations
- Tools for creating forensic images:
  - aimage
  - FTK Imager
  - Guymager
  - Imaging utilities in commercial applications

Disk Imaging Benefits

- Data processing and information extraction benefits:
  - More efficient automation
  - Increased accuracy in data triage
  - Assurance of data integrity
  - Identifying personally identifying and sensitive information
  - Establishing environmental and technical context
- Information located in disk images can assist in linking digital objects to other data sources and activities:
  - Versioning information
  - Backups
  - Related local and network user activity
  - System logs
- Open source tools to create, manage and process disk images

High-Level Workflow Based on Acquisition of Disk Images



BitCurator

[http://bitcurator.net]

- Funded by Andrew W. Mellon Foundation - October 1, 2011 – September 30, 2013
- Partners: SILS and Maryland Institute for Technology in the Humanities (MITH)
- Core Team:
  - Cal Lee, PI
  - **Matt Kirschenbaum, Co-PI**
  - **Kam Woods, Technical Led**
  - Alex Chassonoff, Project Manager (UNC), Sunitha Misra, GA (UNC), and Porter Olsen, GA (MITH)

Professional Experts Panel	Development Advisory Group
• Bradley Daigle, University of Virginia Library • Erika Farr, Emory University • **Jeremy Leighton John, British Library** • Leslie Johnston, Library of Congress • **Courtney Mumma, Artefactual Systems** • Naomi Nelson, Duke University • Erin O'Meara, University of North Carolina • Michael Olson, Stanford University Libraries • Gabriela Redwine, Harry Ransom Center, University of Texas, Austin • Susan Thomas, Digital Archivist, Bodleian Library, University of Oxford	• Geoffrey Brown, Indiana University • Barbara Guttman, National Institute of Standards and Technology • Jerome McDonough, University of Illinois • Mark Matienzo, Yale University • David Pearson, National Library of Australia • Doug Reside, New York Public Library • Seth Shaw, University Archives, Duke University • William Underwood, Georgia Tech • Peter Van Garderen, Artefactual Systems

BitCurator Goals

- Develop a system for librarians and archivists that incorporates the functionality of many digital forensics tools
- Address two fundamental needs:
  - incorporation into the workflow of archives/library ingest and collection management environments
  - provision of public access to the data

BitCurator Environment

- Bundles, integrates and extends some functionality of open source software, including Guymager, fiwalk, bulk extractor, The Sleuth Kit and others
- Can be run as a self-contained environment (based on Ubuntu Linux) in a virtual machine using e.g. Virtual Box or VMWare
- Individual components also available to run directly in your own Linux environment or (whenever possible) Windows environment

Further information:
http://bitcurator.net
http://wiki.bitcurator.net

Software and installation instructions:
http://wiki.bitcurator.net/index.php?title=Software