

BitCurator

Tools for Digital Forensics Methods and Workflows
in Real-World Collecting Institutions

Kam Woods

Research Associate / BitCurator Technical Lead

University of North Carolina School of Information and Library Science

A.R.T. & LLF Program Presentation

November 19, 2013



MARYLAND INSTITUTE FOR
TECHNOLOGY IN THE HUMANITIES

The Andrew W. Mellon Foundation

Many archivists know how to process this stuff...



Source: The Processing Table: Reflections on a manuscripts internship at the Lilly Library.
<https://processingtable.wordpress.com/tag/archival-processing/>

How about processing 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/>

Same goals as when processing analog materials...

- Ensure integrity of materials
- Allow users to make sense of materials and understand their context
- Prevent inadvertent disclosure of sensitive data

...and the same archival principles apply:

Provenance

- Reflect “life history” of records
- Records from a common origin or source should be managed together as an aggregate unit

Original Order

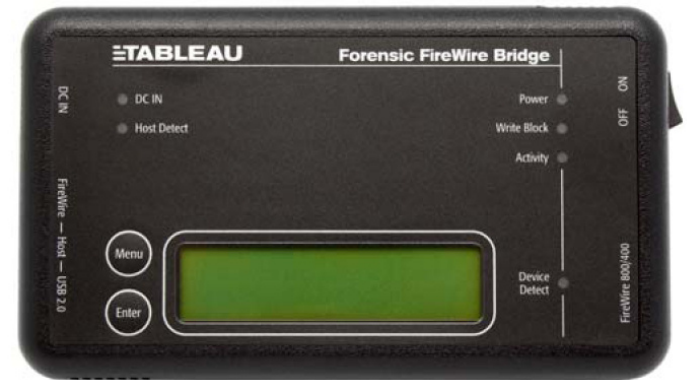
Organize and manage records in ways that reflect their arrangement within the creation/use environment

Chain of Custody

- “Succession of offices or persons who have held materials from the moment they were created”¹
- Ideal recordkeeping system would provide “an unblemished line of responsible custody”²

1. Pearce-Moses, Richard. *A Glossary of Archival and Records Terminology*. Chicago, IL: Society of American Archivists, 2005.
2. Hilary Jenkinson, *A Manual of Archive Administration: Including the Problems of War Archives and Archive Making* (Oxford: Clarendon Press, 1922), 11.

But you might need some of this stuff:



AFFLIB

Open Source Computer Forensics Software

**Here's what it looks like in some institutions
right now...**

Stanford University Libraries and Academic Information Resource (SULAIR)



The British Library (London)



UNC School of Information and Library Science



Motivation

- Archivists are increasingly responsible for acquiring and providing access to materials on digital media
- Information is often not packaged nor described as we might hope
- Information professionals must extract whatever useful information resides on the medium, while avoiding the accidental alteration of data or metadata

Digital forensics can help archivists fulfill their principles

Provenance

- Identify, extract and save essential information about context of creation

Original Order

- Reflect original folder structures, files associations, related applications and user accounts

Chain of Custody

- Documentation of how records were acquired and any transformations to them
- Use well-established hardware and software mechanisms to ensure that data haven't been changed inadvertently

Identifying Sensitive Information

- Identify personally identifying information, regardless of where it appears
- Flag for removal, redaction, closure or restriction

**Digital forensics technologies are often
expensive and difficult to use...**

How about an alternative?

The BitCurator Project brings open source digital forensics tools and techniques to archives

BitCurator

- Partners:
 - School of Information and Library Science (SILS) at UNC
 - Maryland Institute for Technology in the Humanities (MITH)
- Funded by Andrew W. Mellon Foundation

Goals of BitCurator

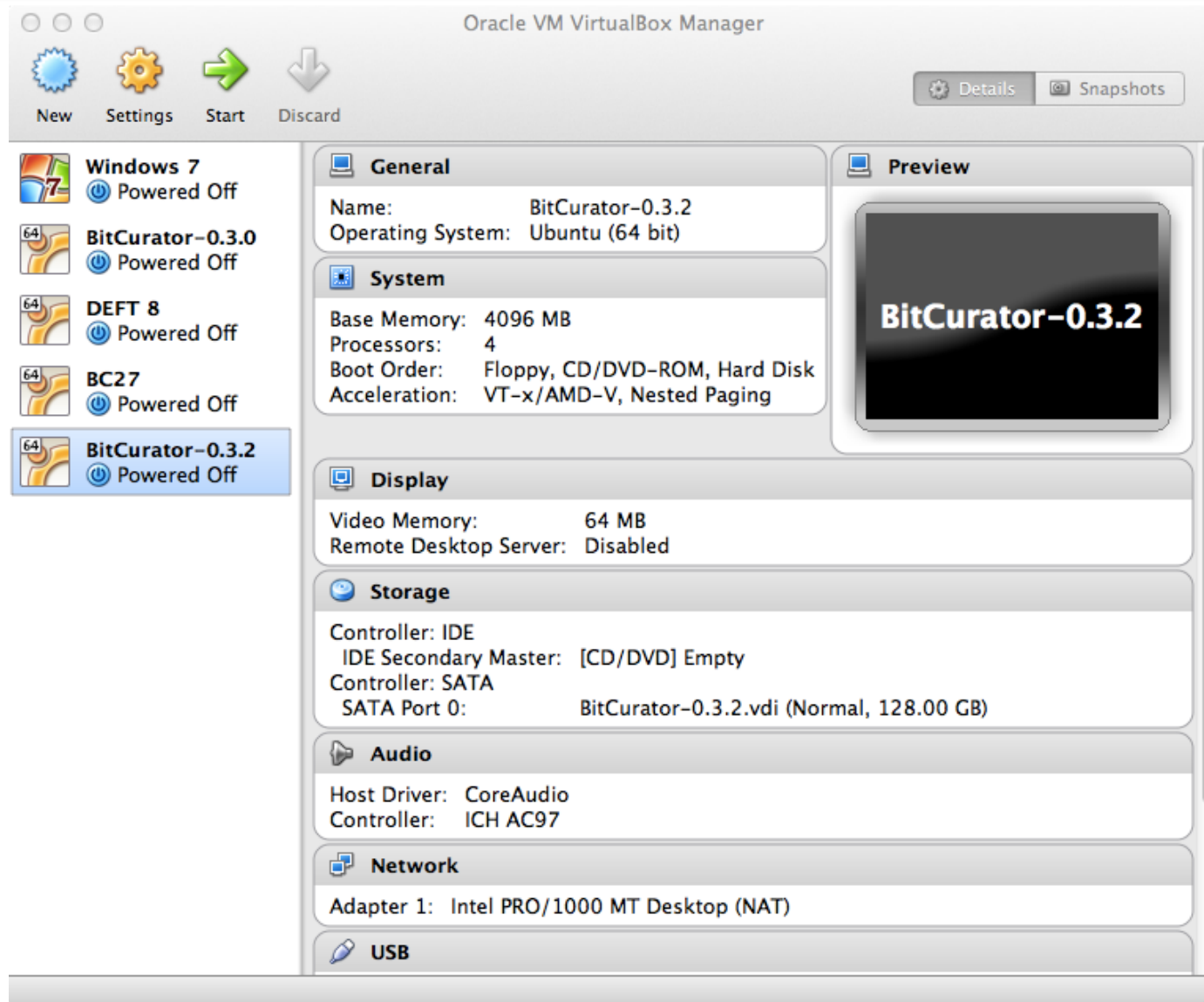
- Develop a **system for collecting professionals** that incorporates the functionality of **open-source digital forensics tools**
- Address two fundamental needs not usually addressed by the digital forensics industry:
 - Incorporation into the **workflow of archives/library ingest and collection management** environments
 - Provision of **public access** to the data

The BitCurator Environment

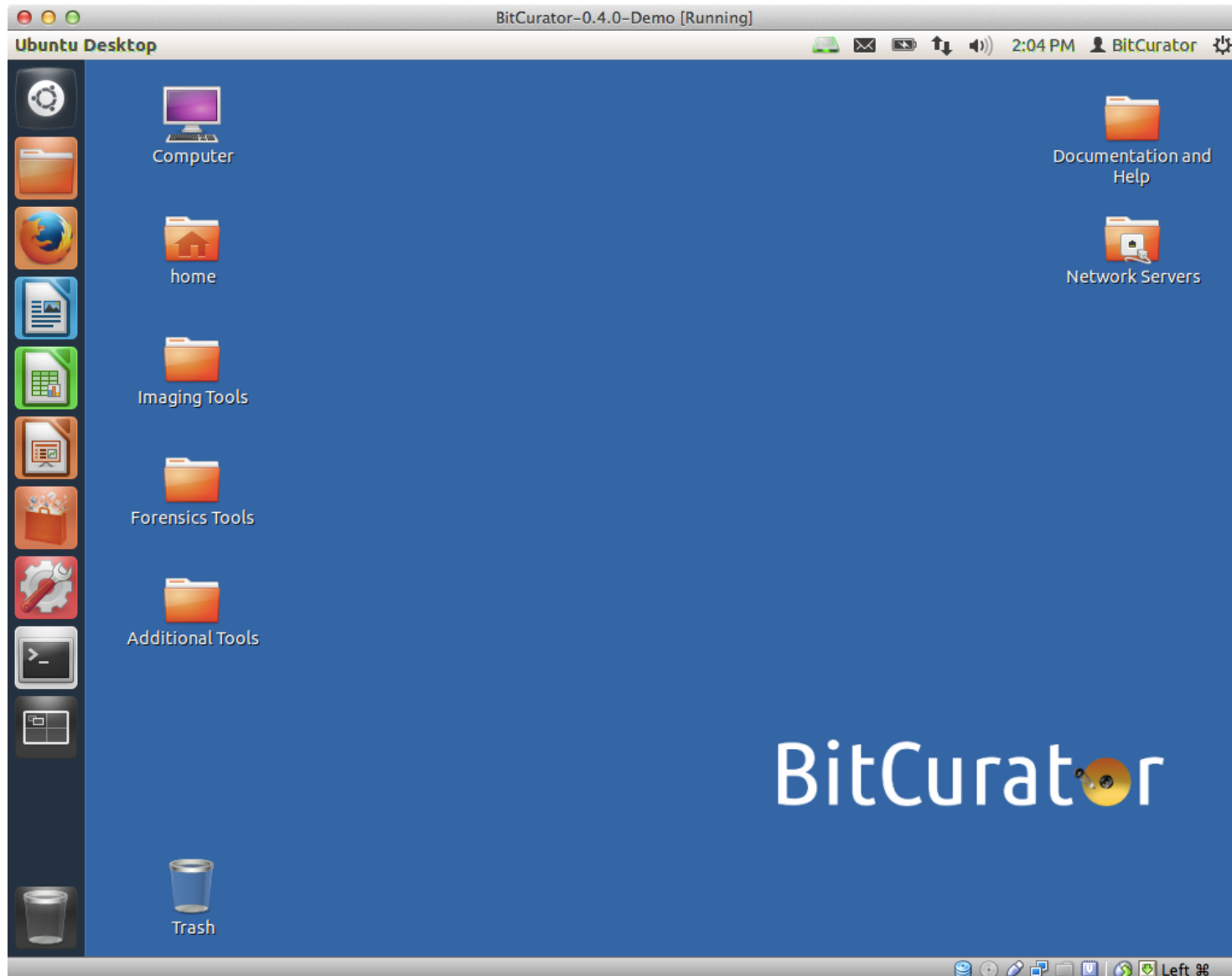
- BitCurator uses such as Guymager to generate bit-identical copies of source media (CD-ROMs, floppy disks, hard drives). Specialized software (and optional hardware) prevents writing back to the disk.
- BitCurator incorporates open source forensics tools including fiwalk, bulk_extractor, and The Sleuth Kit to perform fast file system analysis and help you analyze large quantities of data
- The BitCurator team has written custom reporting tools to help manage the results.
- BitCurator can be run as a self-contained virtual environment (built using Ubuntu Linux). You can run BitCurator on any modern computer using freely available virtualization software such as VirtualBox, or install it on a dedicated machine.

Here's what it looks like...

Start up the virtual machine in VirtualBox.



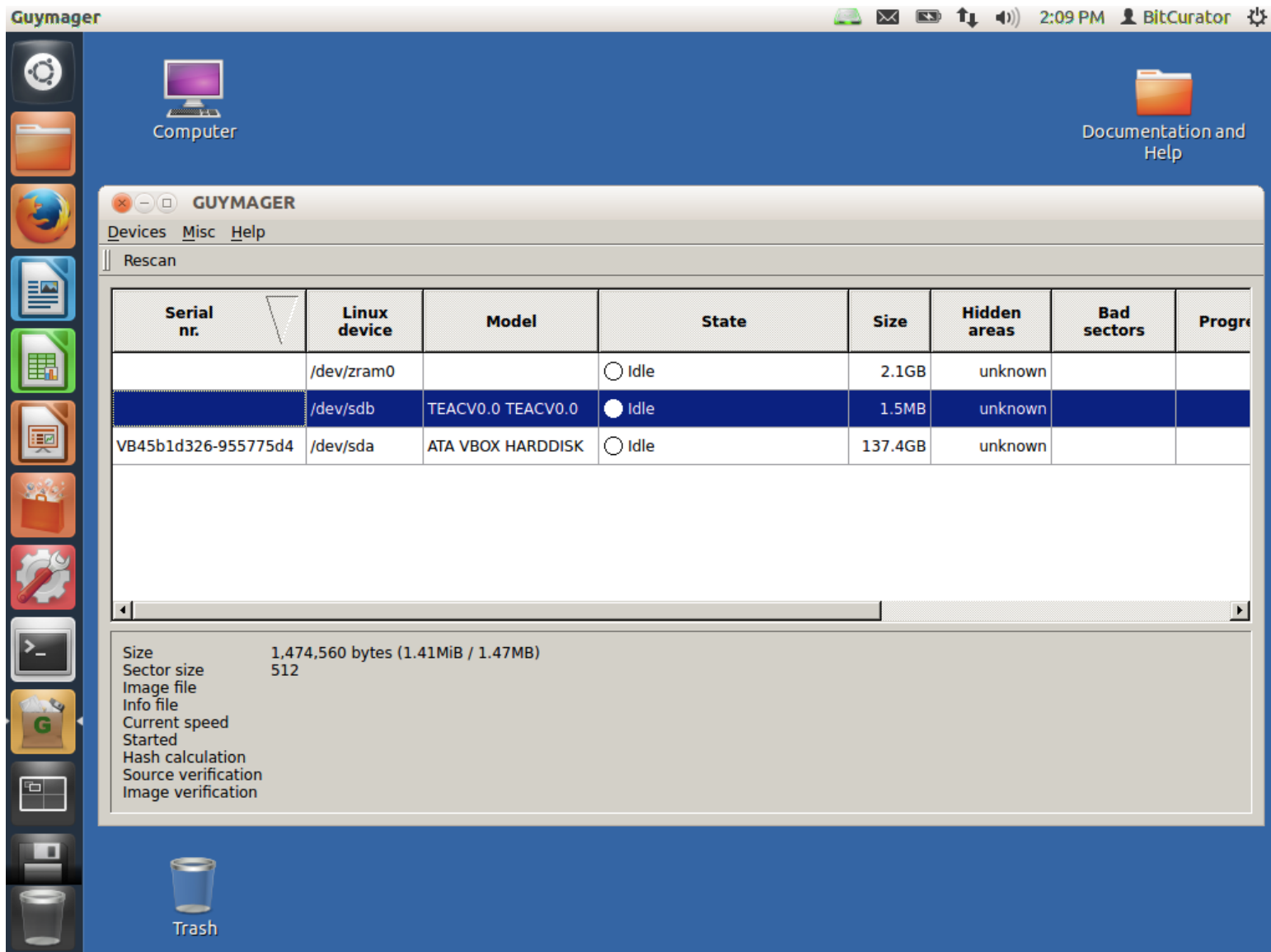
You'll see a desktop that looks like this



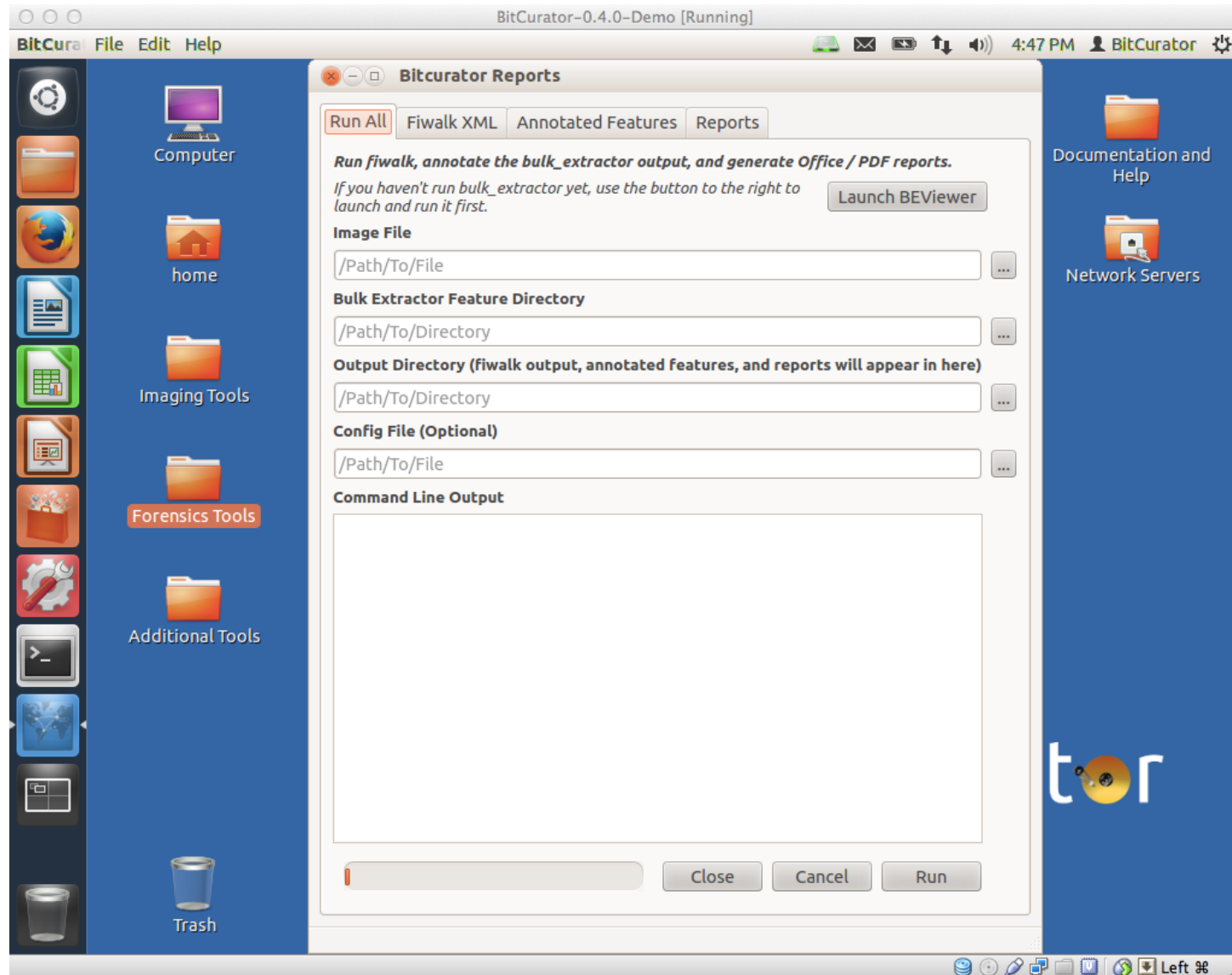
Attached devices will NOT be mounted; writing to the device is prevented in-software



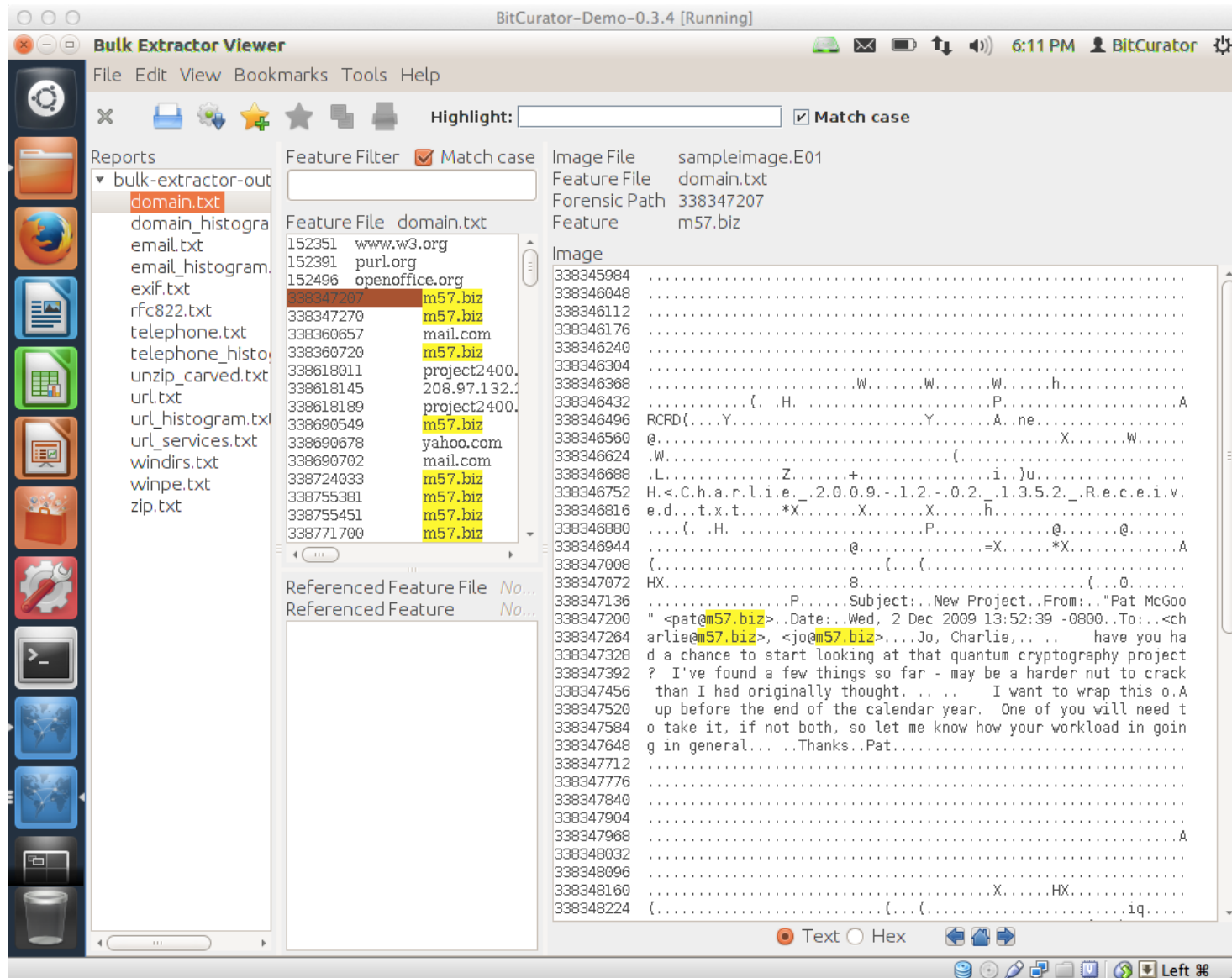
Forensic imaging software allows you to capture the entire disk, and package it securely



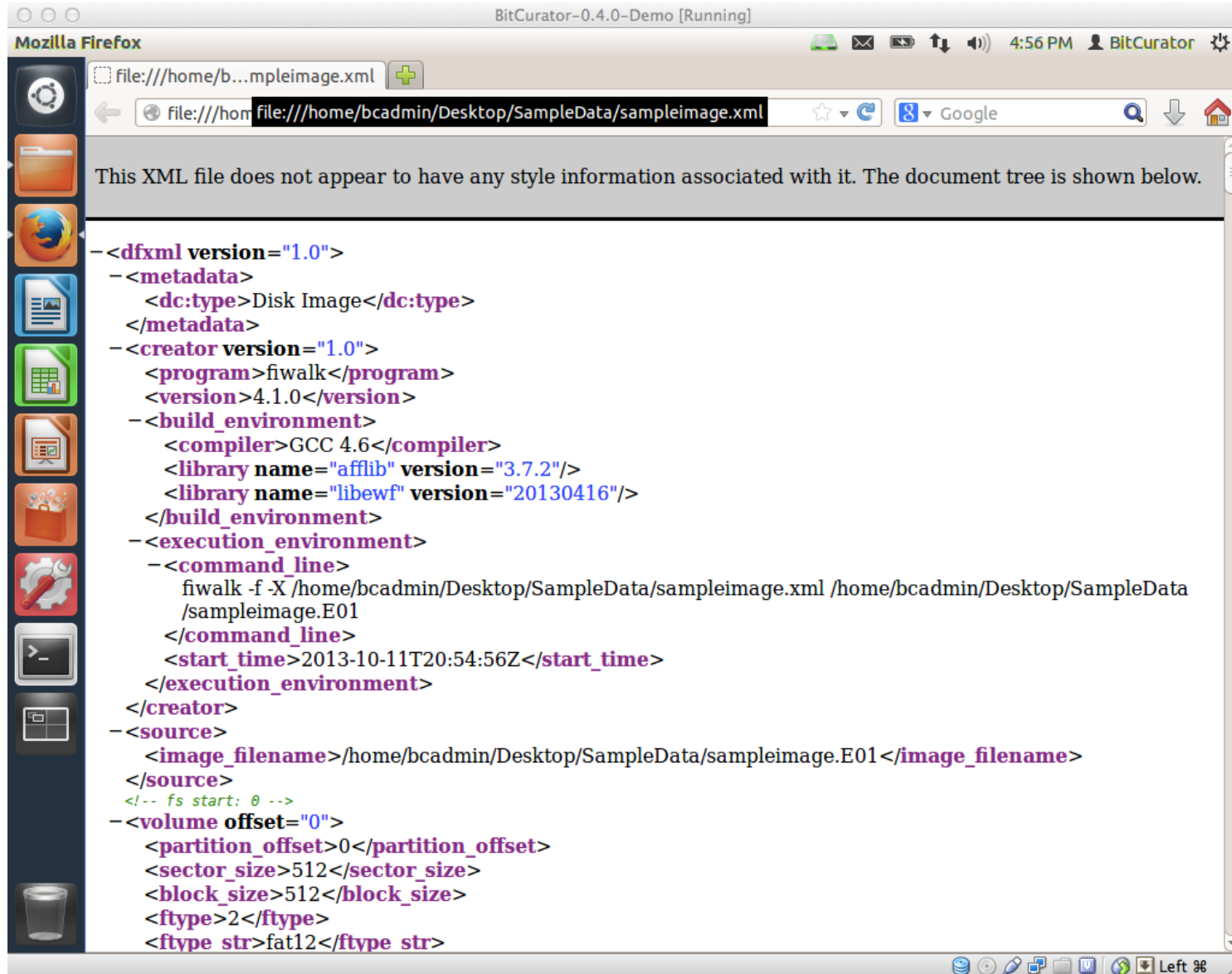
A front-end developed by BitCurator provides simplified access to sophisticated forensics tools...



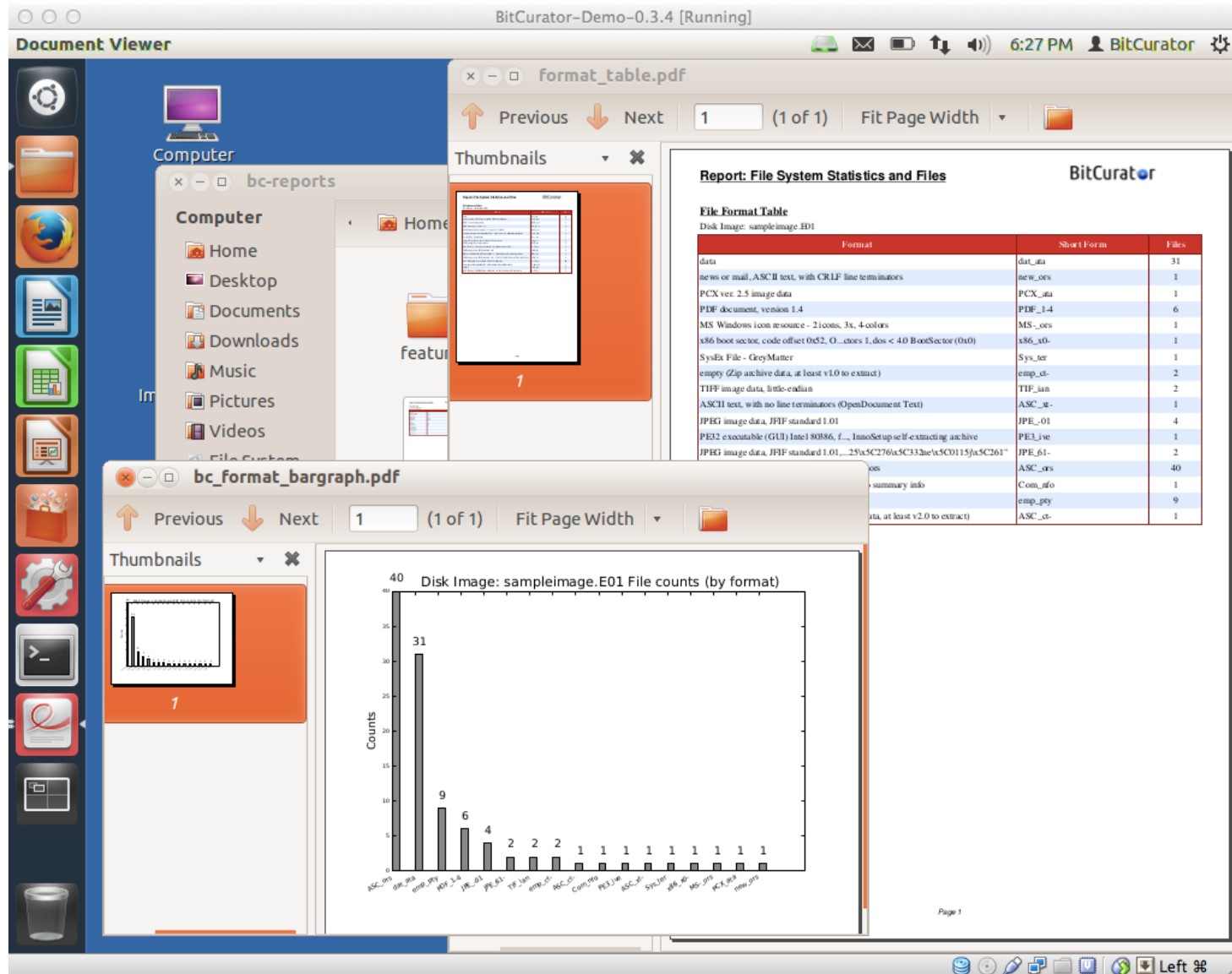
...including bulk_extractor, which can be used to identify potentially sensitive information and other features of interest...



...and fiwalk, which produces XML output on file system metadata, including timestamps and hashes

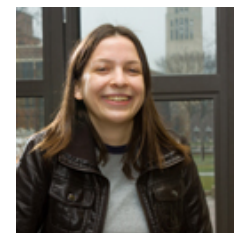
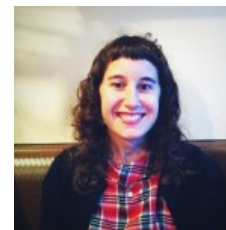


BitCurator produces easy-to-read reports from this data which may be used for analysis or stored in AIPs



The BitCurator Team

- Christopher (Cal) Lee – PI
- Matt Kirschenbaum - Co-PI
- Kam Woods - Technical Lead
- Porter Olsen - Community Lead
- Alex Chassanoff - Project Manager
- Sunitha Misra - GA (UNC)
- Amanda Visconti - GA (MITH)



Two groups of advisors

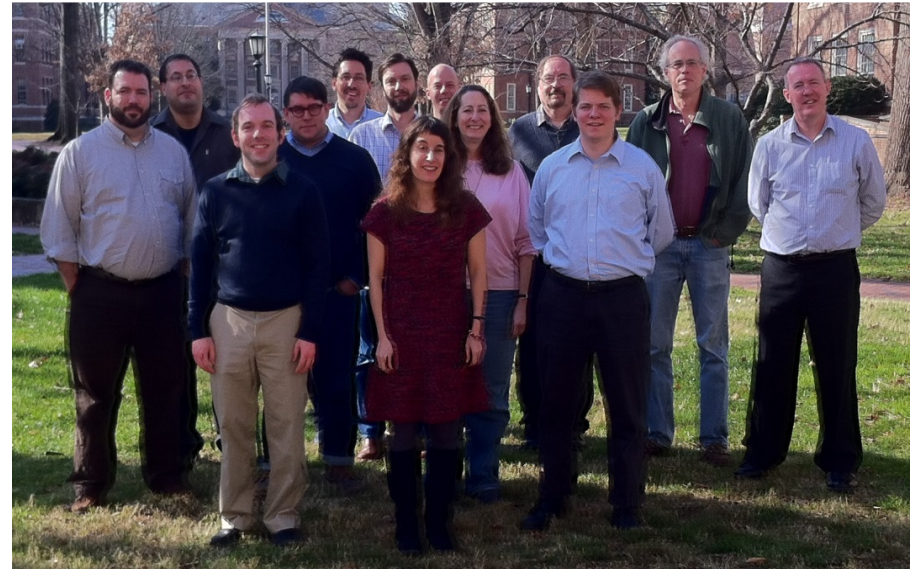
Professional Experts Panel

- Bradley Daigle, University of Virginia Library
- Erika Farr, Emory University
- Jennie Levine Knies, University of Maryland
- Jeremy Leighton John, British Library
- Leslie Johnston, Library of Congress
- Naomi Nelson, Duke University
- Erin O'Meara, Gates Archive
- Michael Olson, Stanford University Libraries
- Gabriela Redwine, Harry Ransom Center, University of Texas
- Susan Thomas, Bodleian Library, University of Oxford



Development Advisory Group

- Barbara Guttman, National Institute of Standards and Technology
- Jerome McDonough, University of Illinois
- Mark Matienzo, Yale University
- Courtney Mumma, Artefactual Systems
- David Pearson, National Library of Australia
- Doug Reside, New York Public Library
- Seth Shaw, University Archives, Duke University
- William Underwood, Georgia Tech



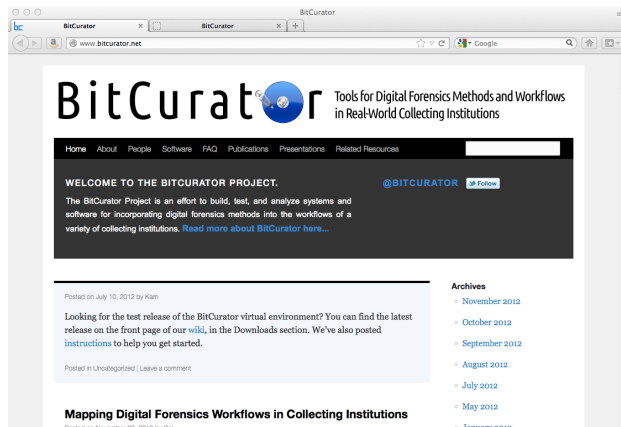
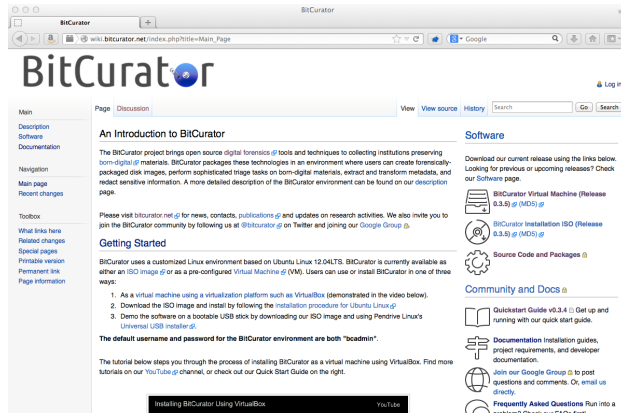
Find Updated BitCurator Information and Documentation Online

Get the software
Documentation and technical
specifications
Google Group

<http://wiki.bitcurator.net/>

People
Project overview
News

<http://www.bitcurator.net/>



...and on Twitter, @bitcurator and @kamwoods