Bringing Bits to the User: BitCurator and BitCurator Access

Christopher (Cal) Lee
UNC School of Information and Library Science

Coalition for Networked Information (CNI) Membership Meeting
December 14-15, 2015
Washington, DC

BitCurator Access

UNC SCHOOL OF INFORMATION AND LIBRARY SCIENCE

The Andrew W. Mellon Foundation

What are we to do with this stuff?

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

Goals When Acquiring Materials

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

Fundamental Archival Principles

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.

Bit digital is different.

See:

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.

Digital Resources - Levels of Representation

Level	Label	Explanation
8	Aggregation of objects	Set of objects that form an aggregation that is meaningful encountered as an entity
7	Object or package	Object composed of multiple files, each of which could also be encountered as individual files
6	In-application rendering	As rendered and encountered within a specific application
5	File through filesystem	Files encountered as discrete set of items with associate paths and file names
4	File as “raw” bitstream	Bitstream encountered as a continuous series of binary values
3	Sub-file data structure	Discrete “chunk” of data that is part of a larger file
2	Bitstream through I/O equipment	Series of 1s and 0s as accessed from the storage media using input/output hardware and software (e.g. controllers, drivers, ports, connectors)
1	Raw signal stream through I/O equipment	Stream of magnetic flux transitions or other analog electronic output read from the drive without yet interpreting the signal stream as a set of discrete values (i.e. not treated as a digital bitstream that can be directly read by the host computer)
0	Bitstream on physical medium	Physical properties of the storage medium that are interpreted as bitstreams at Level 1

Interaction Examples

Level

Aggregation of objects

gazette.govt.nz

HOME | ABOUT US | HOW TO SUBMIT | DEADLINES | HELP | CONTACT US | FEEDBACK

Search all notices: Advanced search

Enter a keyword

Find a specific notice:

Year | Page number | Or | Notice number | Find

Browse Gazette issues

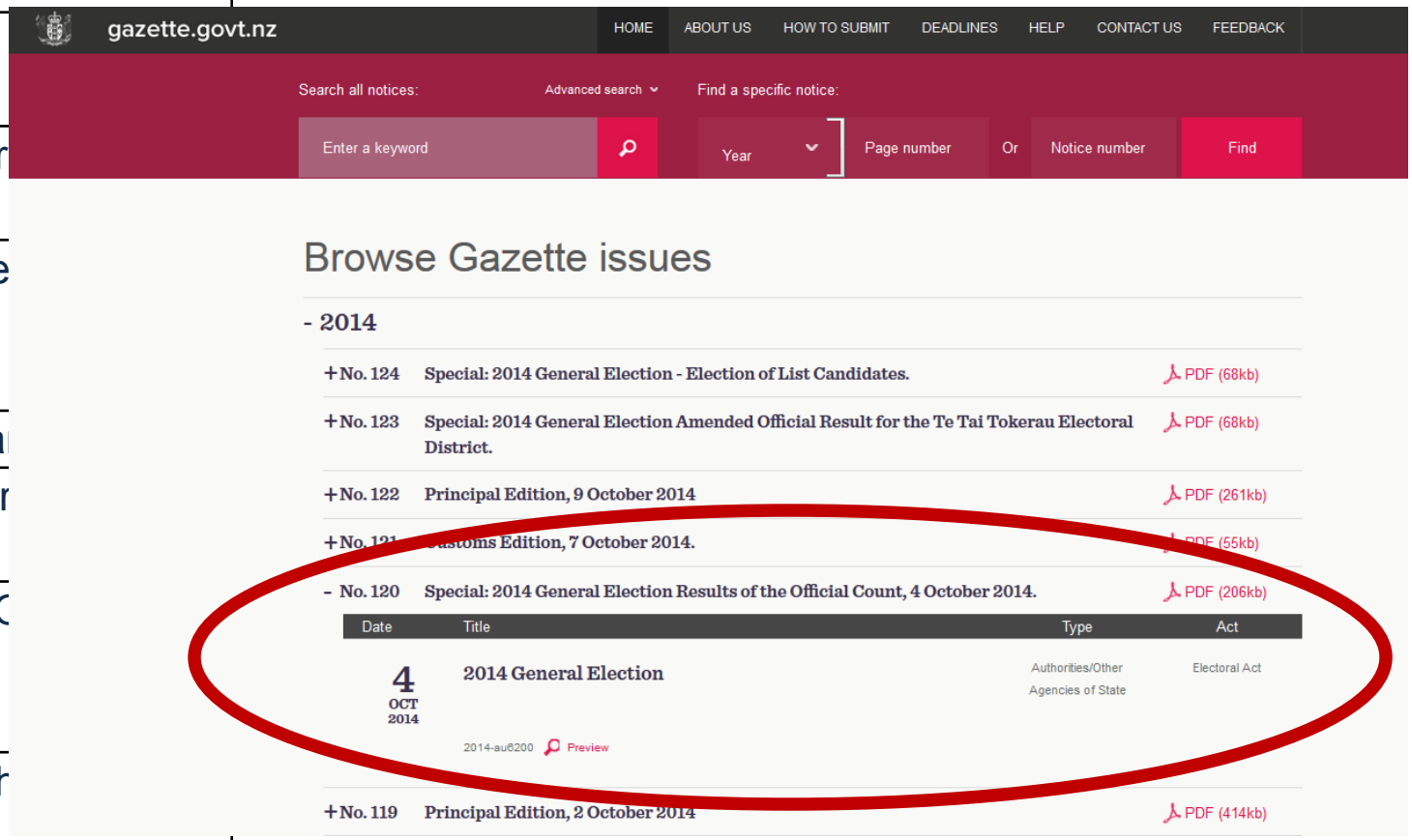
- 2014

+No. 124	Special: 2014 General Election - Election of List Candidates.	PDF (68kb)
+No. 123	Special: 2014 General Election Amended Official Result for the Te Tai Tokerau Electoral District.	PDF (68kb)
+No. 122	Principal Edition, 9 October 2014	PDF (261kb)
+No. 121	Customs Edition, 7 October 2014.	PDF (55kb)
+No. 120	Special: 2014 General Election Results of the Official Count, 4 October 2014.	PDF (206kb)
+No. 119	Principal Edition, 2 October 2014	PDF (414kb)
+No. 118	Supplement: Financial Markets Authority: Authorised Futures Dealers Notices, 30 September 2014	PDF (69kb)

Interaction Examples

Level

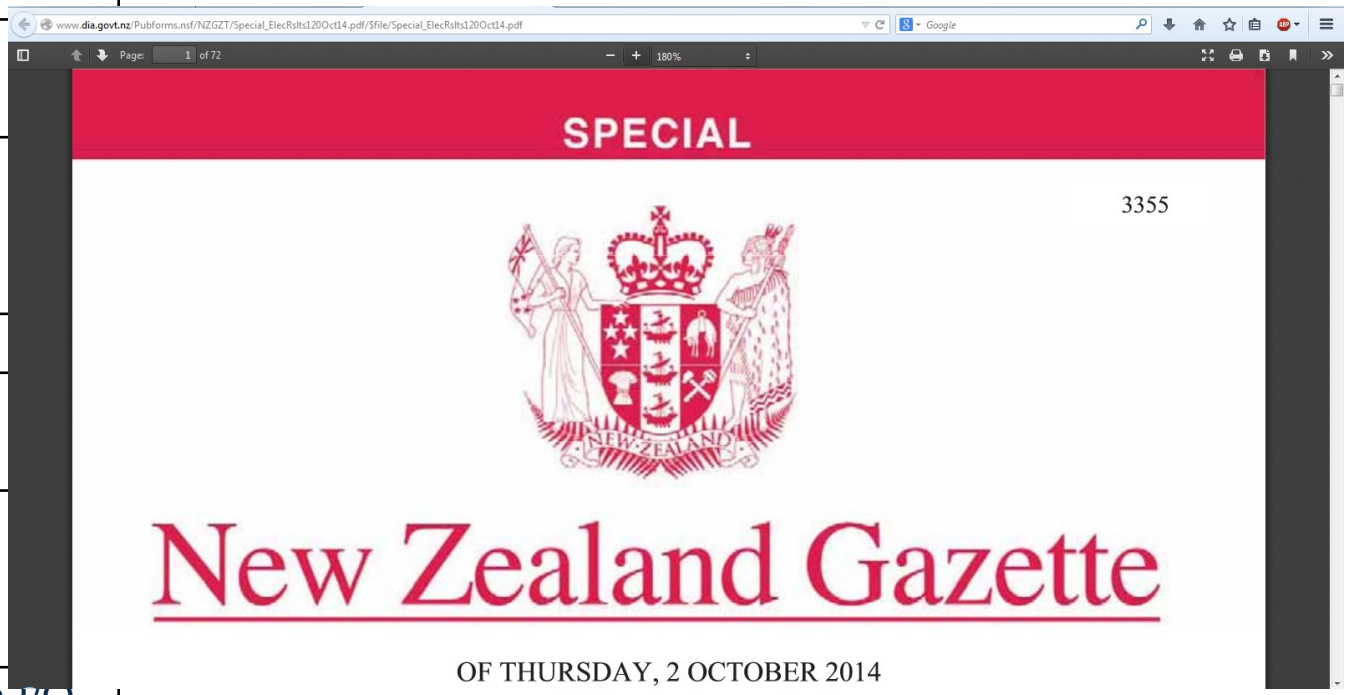
Aggregation of objects
Object or package
In-application render
File through filesyste
File as “raw” bitstrea
Sub-file data structur
Bitstream through I/C equipment
Raw signal stream th equipment
Bitstream on physical medium



Interaction Examples

Level

Aggregation of objects
Object or package
In-application rendering
File through filesystem
File as “raw” bitstream
Sub-file data structure
Bitstream through I/O equipment
Raw signal stream through I/O equipment
Bitstream on physical medium



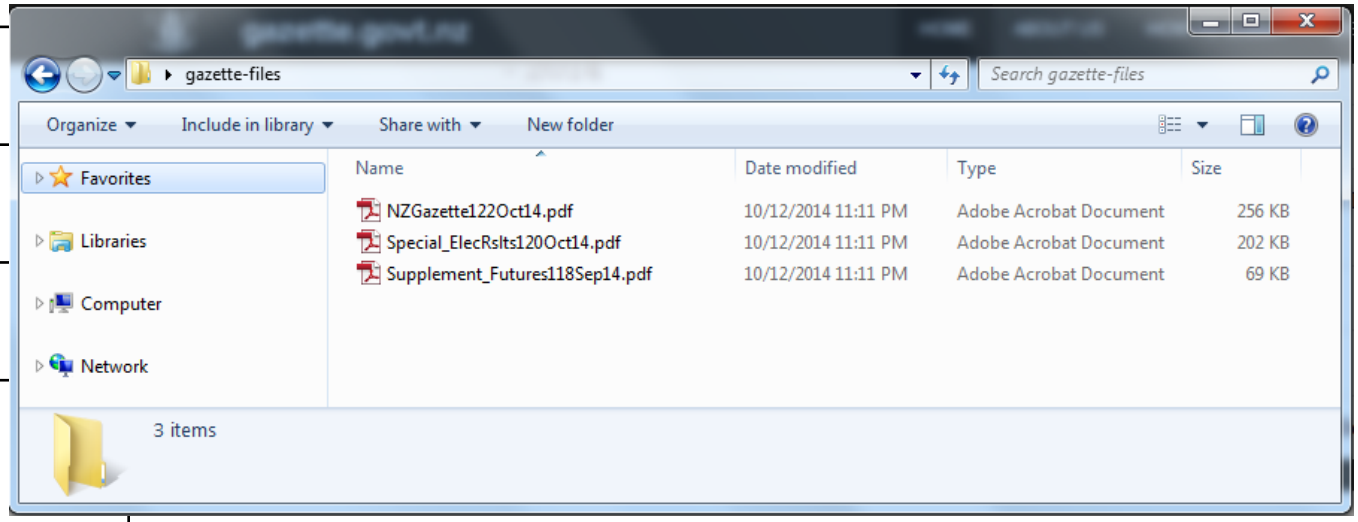
File System

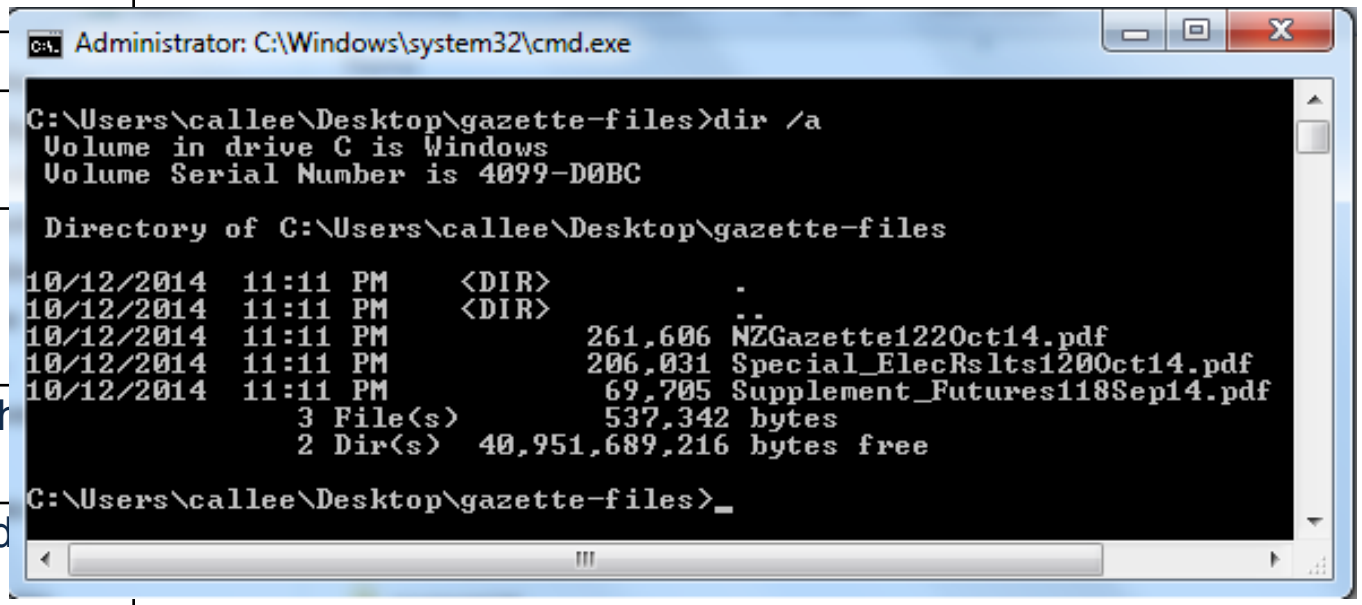
- Access controls
- File names & identifiers
- File size (length)
- Where to find files in storage (sectors and clusters)
- MAC times
  - Modified – when the content was last changed
  - Accessed – time file was last accessed (by person or software)
  - Changed – last time metadata changed
  - Created – (implemented inconsistently, if at all, across different file systems)

Interaction Examples

Level

Aggregation of objects
Object or package
In-application rendering
File through filesystem
File as “raw” bitstream
Sub-file data structure
Bitstream through I/O equipment
Raw signal stream through equipment
Bitstream on physical med

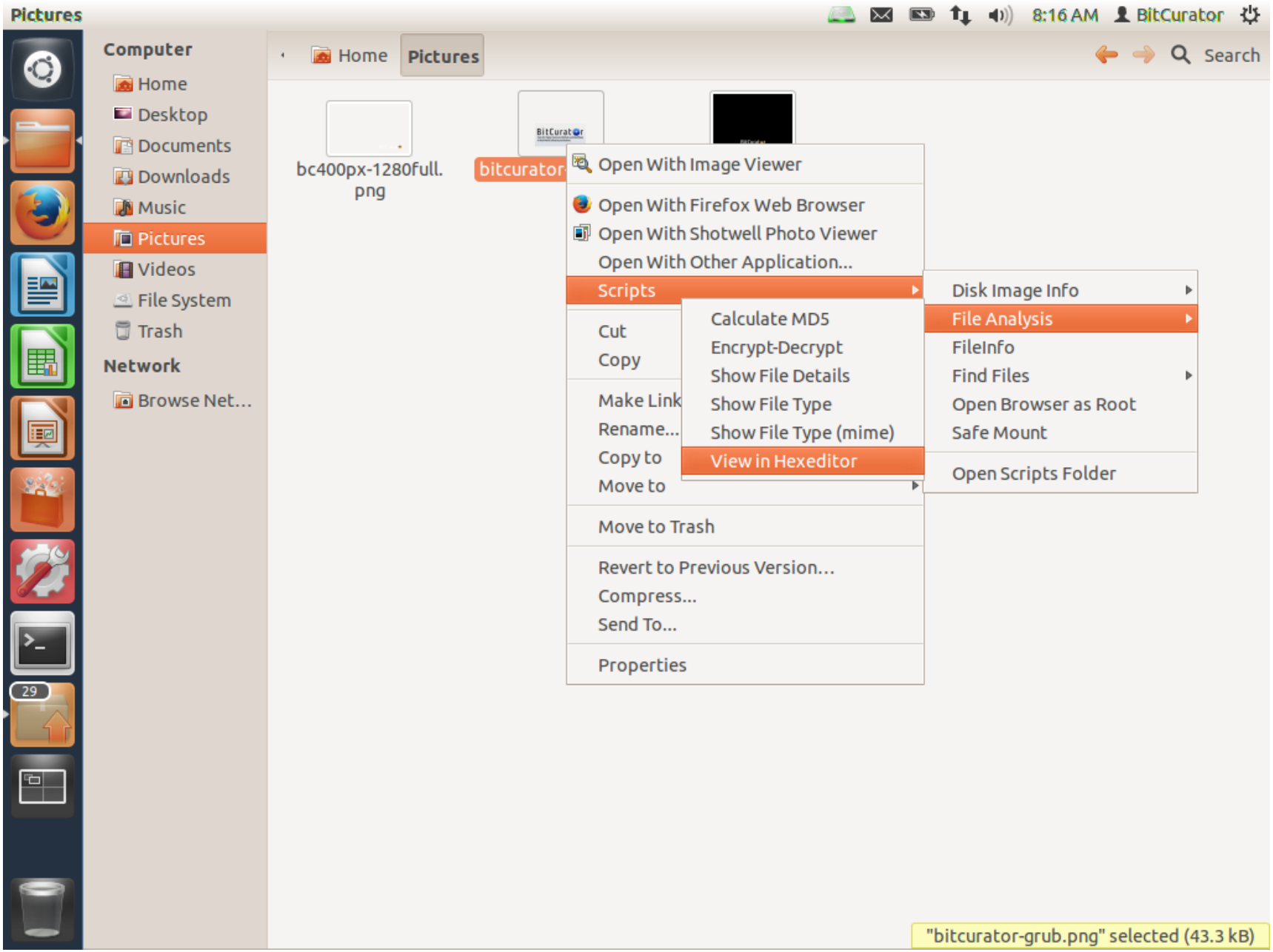




Hex Dump

- A more compact and more humanly readable way of conveying a stream of bits
- Uses hexadecimal notation
  - Each character represents one of 16 possible values (0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F)
  - Conveniently, a series of two characters represented in hexadecimal can represent exactly one byte (2^8 = 256 possible values) of data, because 16^2 = 256
- Hex dumps from computer's memory often used for debugging or reverse engineering software and for data recovery

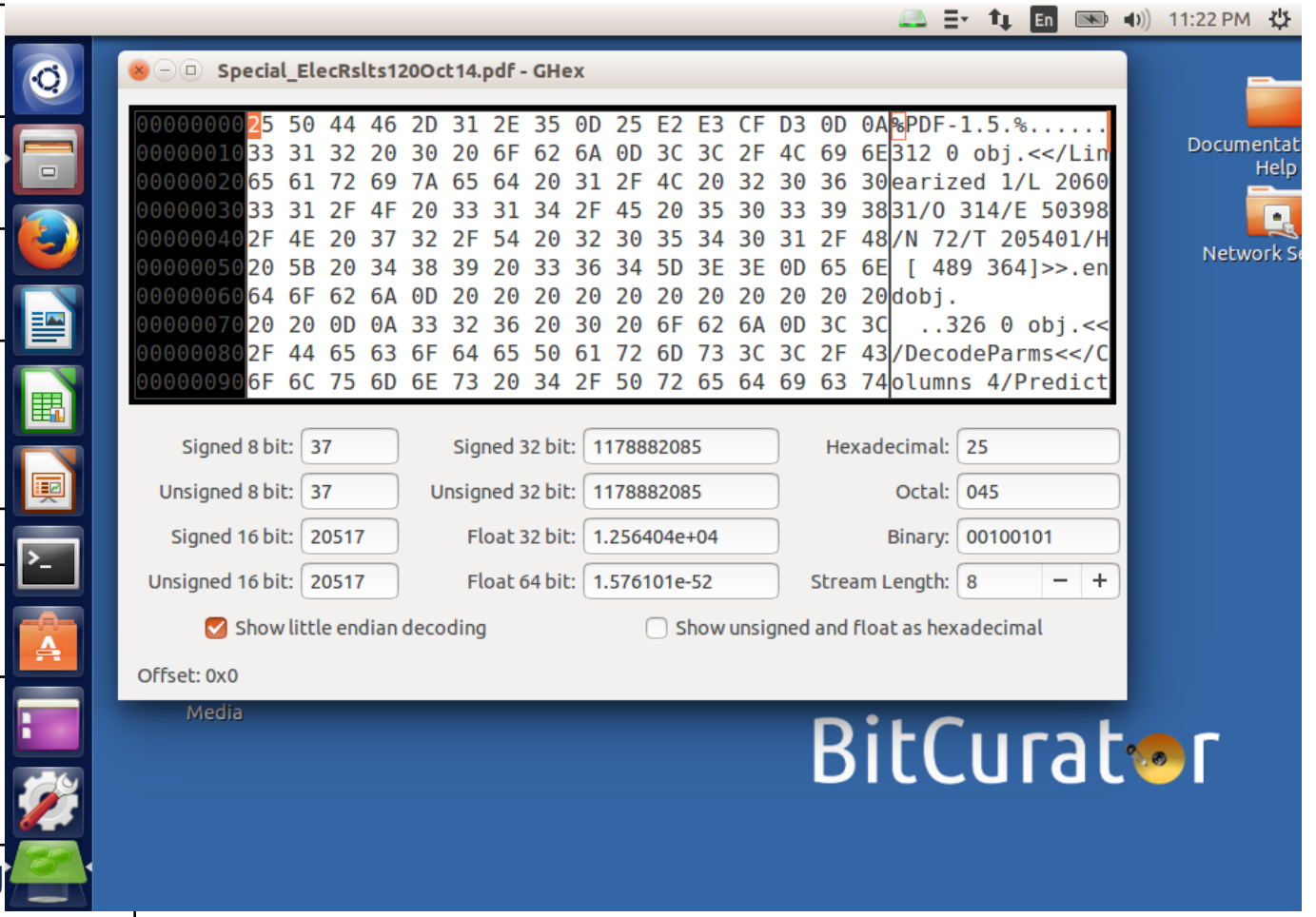
In the BitCurator environment:



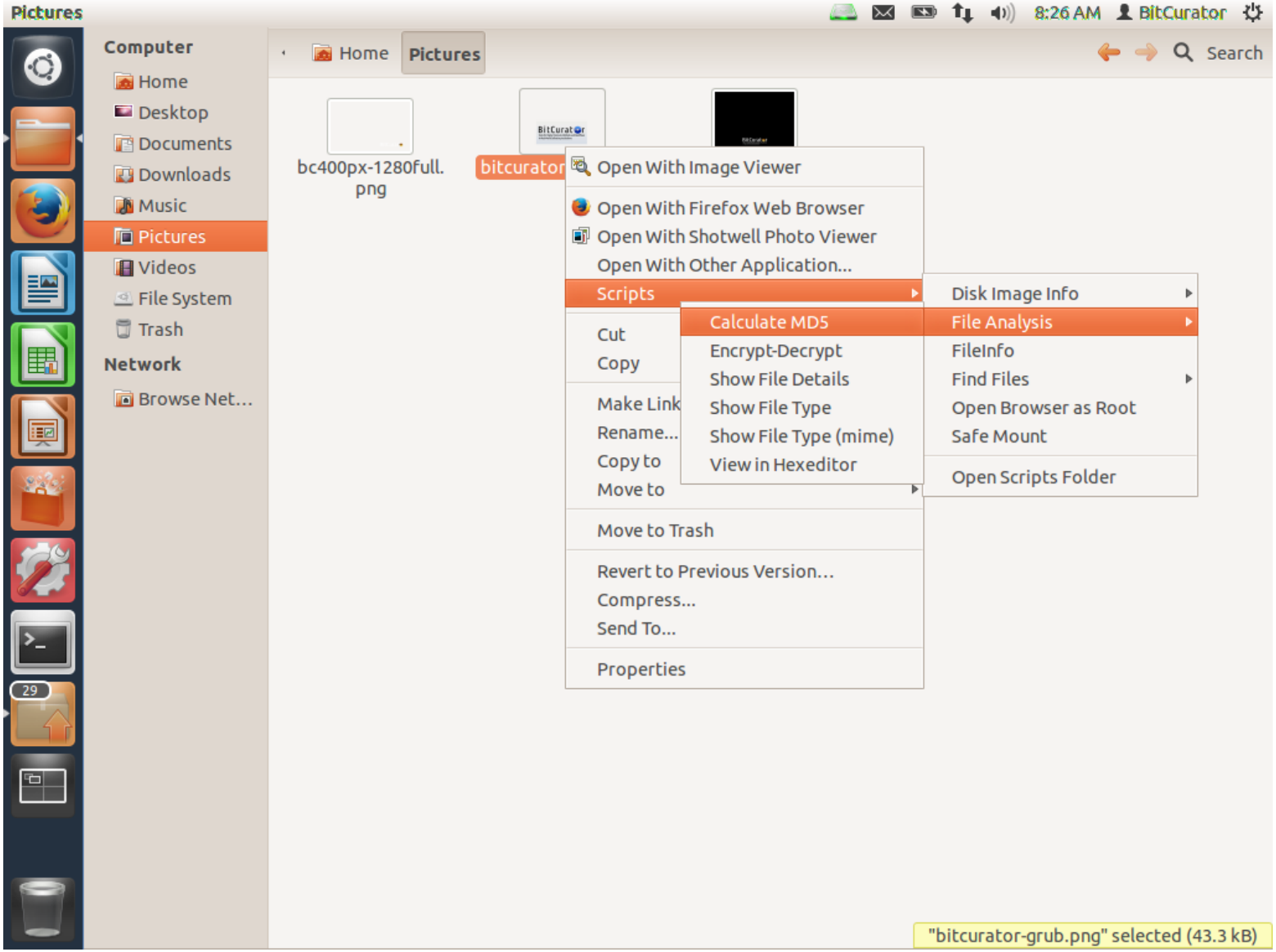
Interaction Examples

Level

Aggregation of objects
Object or package
In-application rendering
File through filesystem
File as “raw” bitstream
Sub-file data structure
Bitstream through I/O equipment
Raw signal stream throug equipment
Bitstream on physical medium



In BitCurator environment: Right Click on File or Directory and Calculate MD5



Calculate MD5 (Files and Directories)

8:31 AM BitCurator

Computer
- Home
- Desktop
- Documents
- Downloads
- Music
- Pictures
- Videos
- File System
- Trash

Network
- Browse Net...

Home Pictures Search

bc400px-1280full.png

bitcurator-grub.png

bitcurator-grub-new.png

Calculate MD5 (Files and Directories)

Please choose the way you want the MD5 hash to be presented:
(1 file(s) selected)

Handling
(•) Display on screen
( ) Save to file (the selected filename + .md5 extenion)

Cancel OK

"bitcurator-grub.png" selected (43.3 kB)

Calculate MD5 (Files and Directories)

8:30 AM BitCurator

Computer
- Home
- Desktop
- Documents
- Downloads
- Music
- Pictures
- Videos
- File System
- Trash

Network
- Browse Net...

Home Pictures Search

bc400px-1280full.png

bitcurator-grub.png

bitcurator-grub-new.png

Calculate MD5 (Files and Directories)

The MD5 hash of the selected file:

```
ceb2622125be1231b0fc9babee27942d  /home/bcadmin/Pictures/bitcurator-grub.png
```

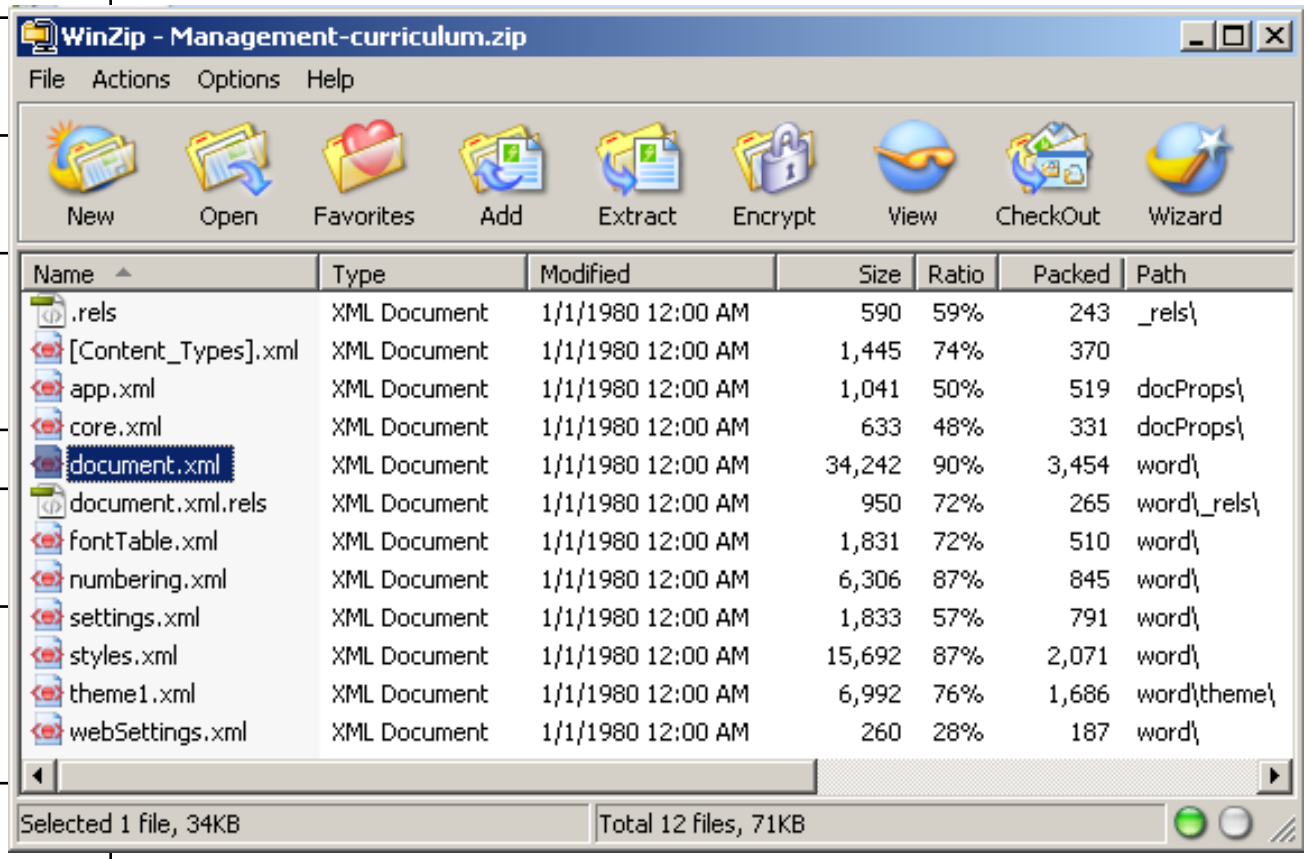
Cancel OK

"bitcurator-grub.png" selected (43.3 kB)

Interaction Examples

Level

Aggregation of objects
Object or package
In-application rendering
File through filesystem
File as “raw” bitstream
Sub-file data structure
Bitstream through I/O equipment
Raw signal stream through equipment
Bitstream on physical medium



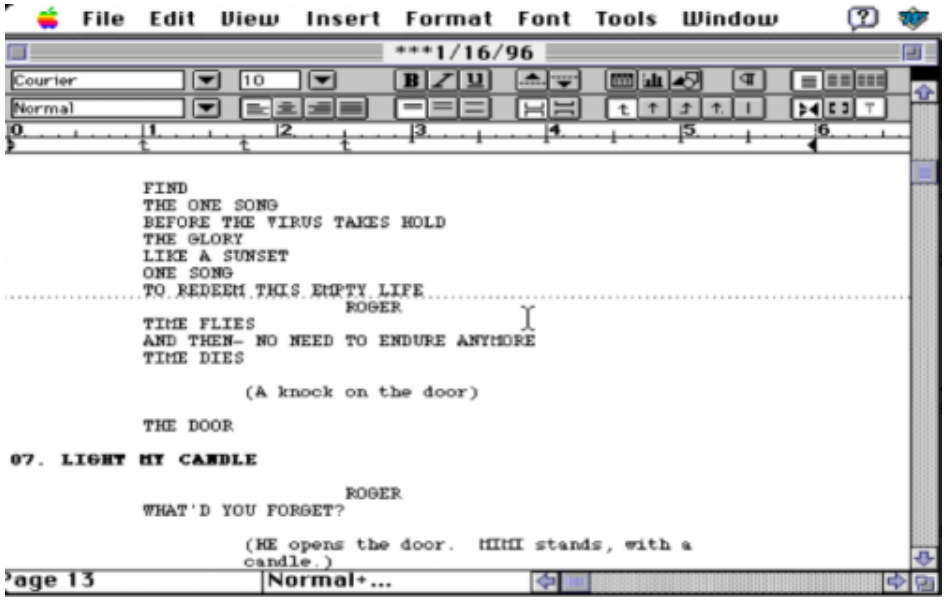
Name	Type	Modified	Size	Ratio	Packed	Path
.rels	XML Document	1/1/1980 12:00 AM	590	59%	243	_rels\
[Content_Types].xml	XML Document	1/1/1980 12:00 AM	1,445	74%	370	
app.xml	XML Document	1/1/1980 12:00 AM	1,041	50%	519	docProps\
core.xml	XML Document	1/1/1980 12:00 AM	633	48%	331	docProps\
document.xml	XML Document	1/1/1980 12:00 AM	34,242	90%	3,454	word\
document.xml.rels	XML Document	1/1/1980 12:00 AM	950	72%	265	word_rels\
fontTable.xml	XML Document	1/1/1980 12:00 AM	1,831	72%	510	word\
numbering.xml	XML Document	1/1/1980 12:00 AM	6,306	87%	845	word\
settings.xml	XML Document	1/1/1980 12:00 AM	1,833	57%	791	word\
styles.xml	XML Document	1/1/1980 12:00 AM	15,692	87%	2,071	word\
theme1.xml	XML Document	1/1/1980 12:00 AM	6,992	76%	1,686	word\theme\
webSettings.xml	XML Document	1/1/1980 12:00 AM	260	28%	187	word\

Example of EXIF Metadata from a JPEG File (Generated Using exiftool*)

```
---- ExifTool ----
ExifTool Version Number         : 9.38
---- System ----
File Name                       : IMG_20130823_151811.jpg
Directory                       : C:/Users/callee/Documents/images/digital-forensics-lab
File Size                       : 1785 kB
File Modification Date/Time     : 2013:08:23 16:36:44-04:00
File Access Date/Time           : 2013:10:14 17:13:02-04:00
File Creation Date/Time         : 2013:08:23 16:36:44-04:00
File Permissions                : rw-rw-rw-
---- File ----
File Type                       : JPEG
MIME Type                       : image/jpeg
Exif Byte Order                 : Big-endian (Motorola, MM)
Image Width                     : 2592
Image Height                    : 1944
Encoding Process                : Baseline DCT, Huffman coding
Bits Per Sample                 : 8
Color Components                : 3
Y Cb Cr Sub Sampling            : YCbCr4:2:0 (2 2)
---- GPS ----
GPS Img Direction               : 83
GPS Img Direction Ref           : Magnetic North
GPS Latitude Ref                : North
GPS Latitude                    : 35 deg 55' 2.24"
GPS Longitude Ref               : West
GPS Longitude                   : 79 deg 2' 57.55"
GPS Altitude Ref                : Above Sea Level
GPS Altitude                    : 0 m
GPS Time Stamp                  : 19:18:06
GPS Processing Method           : NETWORK
GPS Date Stamp                  : 2013:08:23
---- IFD0 ----
Orientation                     : Unknown (0)
Camera Model Name               : Galaxy Nexus
Modify Date                     : 2013:08:23 15:18:11
Y Cb Cr Positioning             : Centered
Y Resolution                    : 72
Resolution Unit                 : inches
X Resolution                    : 72
Make                            : Samsung
---- ExifIFD ----
Create Date                     : 2013:08:23 15:18:11
Date/Time Original              : 2013:08:23 15:18:11
Exif Version                    : 0220
Flash Energy                    : 0
Image Unique ID                 : OAEL01
Exposure Time                   : 1/17
ISO                             : 125, 0, 0
Scene Type                      : Directly photographed
Exposure Index                  : undef
Components Configuration        : Y, Cb, Cr, -
F Number                        : 2.8
Compressed Bits Per Pixel       : 0
Sensing Method                  : One-chip color area
Exposure Program                : Aperture-priority AE
Aperture Value                  : 2.6
Brightness Value                : 0
Subject Distance Range          : Unknown
Shutter Speed Value             : 1/15
Subject Distance                : 0 m
Saturation                      : Normal
Color Space                     : sRGB
Contrast                        : Normal
Metering Mode                   : Multi-spot
Flashpix Version                :
Exposure Compensation           : 0
Exif Image Height               : 1944
Max Aperture Value              : 2.6
Sharpness                       : Normal
Exif Image Width                : 2592
Focal Length                    : 3.4 mm
Digital Zoom Ratio              : 1
Light Source                    : Fluorescent
Scene Capture Type              : Standard
Flash                           : Off, Did not fire
Custom Rendered                 : Custom
White Balance                   : Auto
Exposure Mode                   : Auto
---- IFD1 ----
Compression                     : JPEG (old-style)
Image Width                     : 160
Image Height                    : 120
Thumbnail Offset                : 1239
Thumbnail Length                : 7164
---- Composite ----
Aperture                        : 2.8
GPS Altitude                    : 0 m Above Sea Level
GPS Date/Time                   : 2013:08:23 19:18:06Z
GPS Latitude                    : 35 deg 55' 2.24" N
GPS Longitude                   : 79 deg 2' 57.55" W
GPS Position                    : 35 deg 55' 2.24" N, 79 deg 2' 57.55" W
Image Size                      : 2592x1944
Shutter Speed                   : 1/17
Thumbnail Image                 : (Binary data 7164 bytes, use -b option to extract)
Focal Length                    : 3.4 mm
Light Value                     : 6.7
```

*http://www.sno.phy.queensu.ca/~phil/exiftool/ (Also available through the BitCurator environment)

Jonathan Larson Fast Save Example




```
    FIND
    THE ONE SONG
    BEFORE YOU ENTER THE LIGHT
    THE GLORY
    LIKE A SUNSET
    ONE SONG
    TO REDEEM THIS EMPTY LIFE

    TIME FLIES
    AND THEN- NO NEED TO ENDURE ANYMORE
    TIME DIES
        (A knock on the door)

    THE DOOR
08. LIGHT MY CANDLE

            ROGER
    WHAT'D YOU FORGET?

        (HE opens the door.  MIMI stands, with a            candle.)
```

```
00028b60  09 09 09 2a 2a 2a 31 2f  31 36 2f 39 36 4f 55 52  |...***1/16/96OUR|
00028b70  20 57 45 44 44 49 4e 47  4f 4e 20 54 48 45 20 53  | WEDDINGON THE S|
00028b80  4f 46 41 53 4f 46 41 54  48 45 20 56 49 52 55 53  |OFASOFATHE VIRUS|
00028b90  20 54 41 4b 45 53 20 48  4f 4c 44 4d 45 45 54 20  | TAKES HOLDMEET |
00028ba0  59 4f 55 20 41 54 20 54  48 45 20 53 48 4f 57 49  |YOU AT THE SHOWI|
00028bb0  27 4c 4c 20 54 52 59 20  41 4e 44 20 43 4f 4e 56  |'LL TRY AND CONV|
00028bc0  49 4e 43 45 20 52 4f 47  45 52 20 54 4f 20 47 4f  |INCE ROGER TO GO|
00028bd0  43 4c 4f 53 45 20 4f 4e  43 41 4e 20 49 20 48 45  |CLOSE ONCAN I HE|
00028be0  4c 50 4d 69 73 73 20 50  6f 72 74 65 72 27 73 46  |LPMiss Porter'sF|
00028bf0  4f 52 47 45 54 20 49 54  50 41 55 4c 2a 2a 2a 2a  |ORGET ITPAUL****|
```

http://www.nypl.org/blog/2011/04/22/no-day-today-look-jonathan-larsons-word-files

Interaction Examples

Level

Aggregation of objec

Object or package

In-application render

File through filesyste

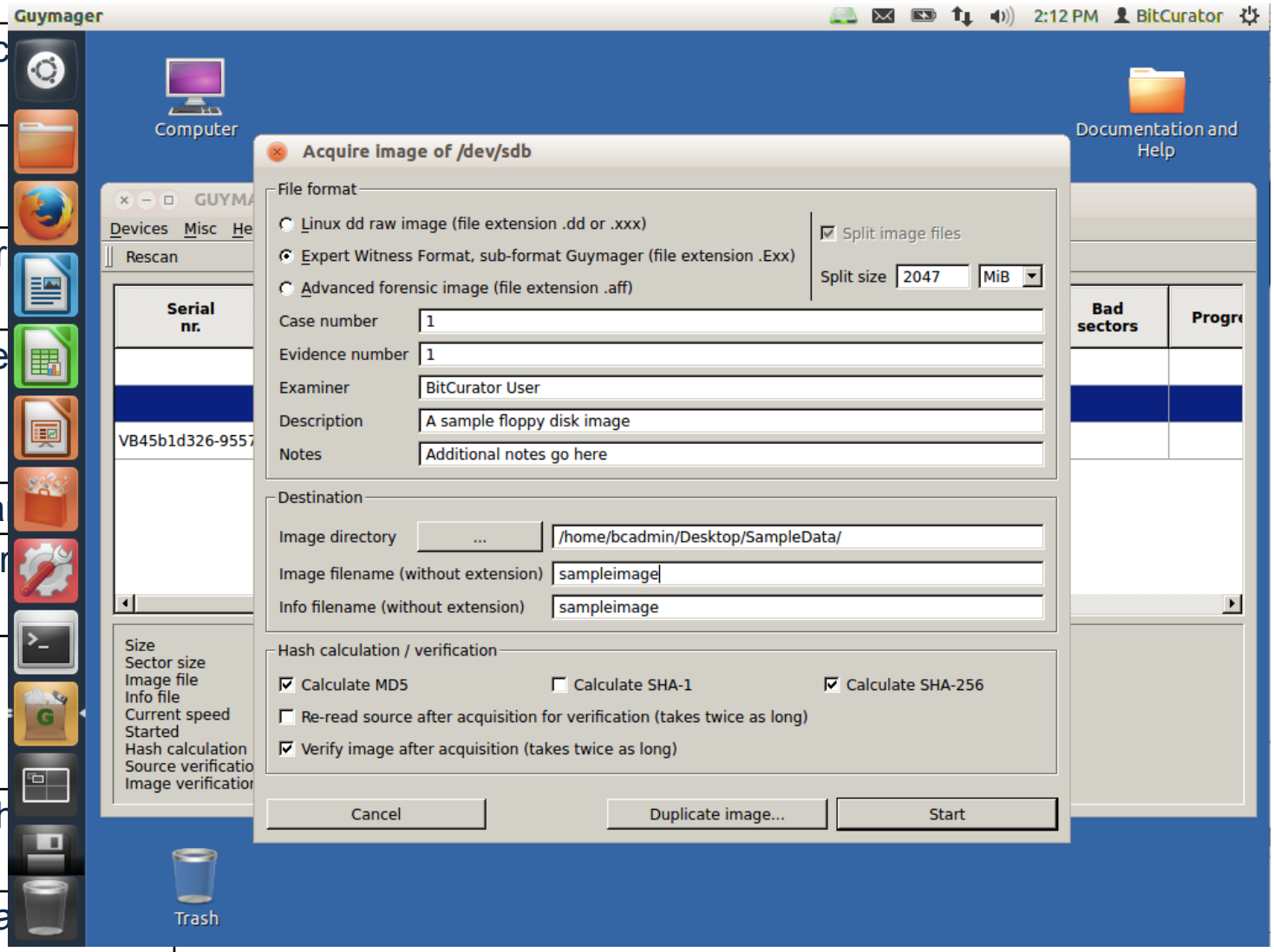
File as “raw” bitstrea

Sub-file data structur

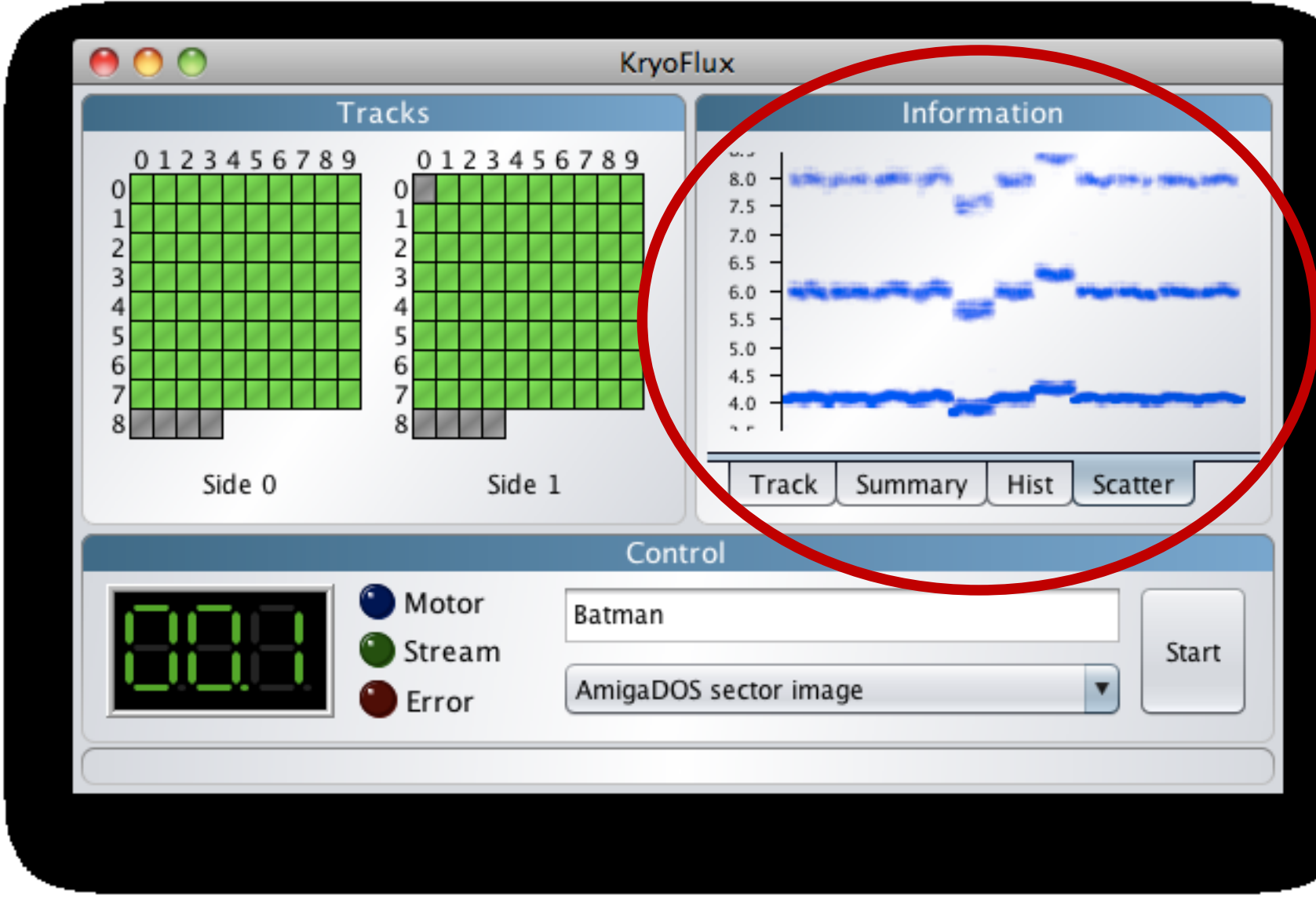
Bitstream through I/O equipment

Raw signal stream th I/O equipment

Bitstream on physica

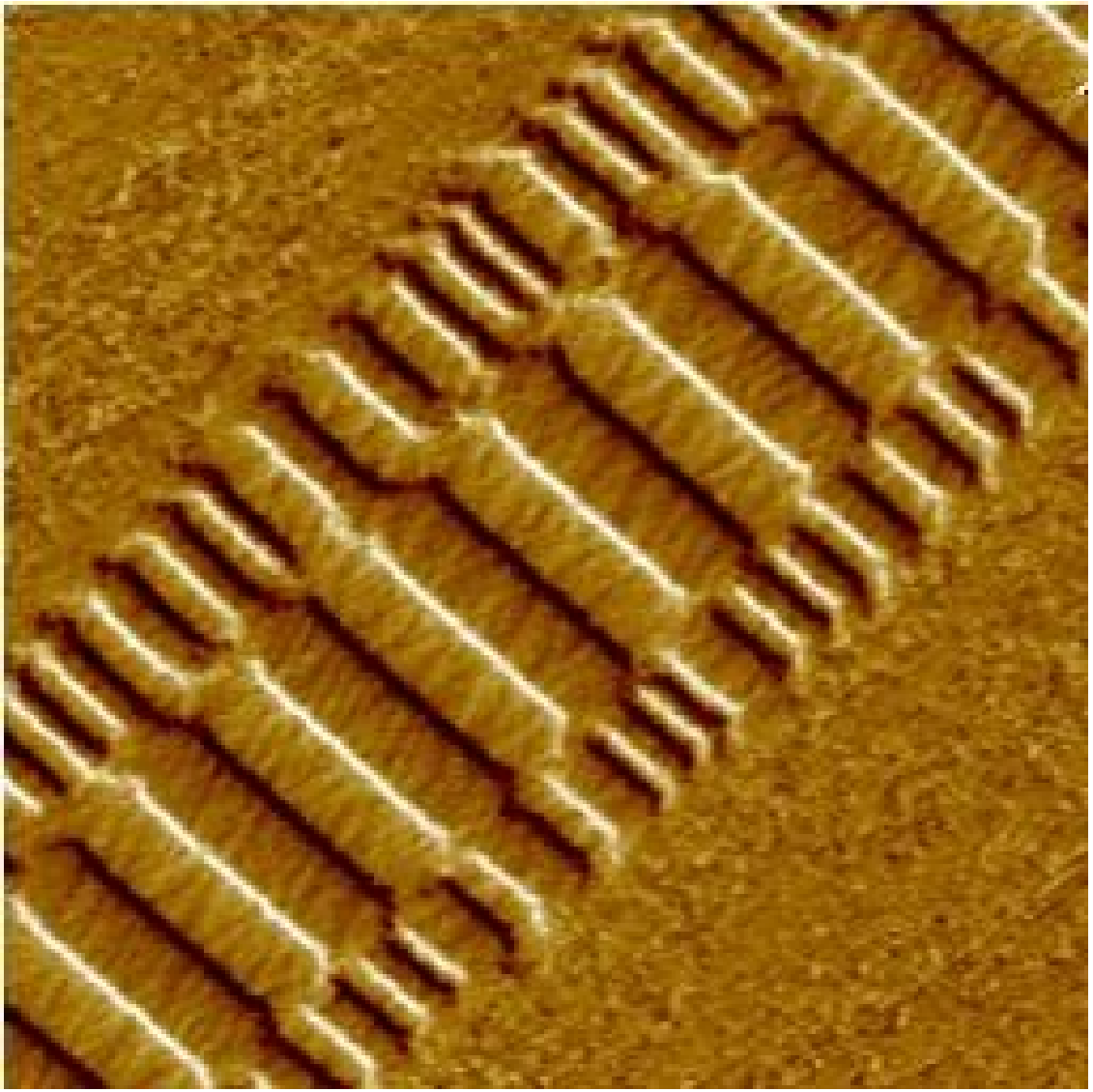


Level
Aggregat
Object or
In-applica
File throu
File as “r
Sub-file c
Bitstream equipmen
Raw sig I/O equipment
Bitstream on physical medium



Interaction Examples

Level
Aggregation of objects
Object or package
In-application rendering
File through filesystem
File as “raw” bitstream
Sub-file data structure
Bitstream through I/O equipment
Raw signal stream through I/O equipment
Bitstream on physical medium



Veeco Instruments. http://www.veeco.com/library/nanotheater_detail.php?type=application&id=78&app_id=34

Three Complicating Factors:

1. Medium Failure / Bit Rot
2. Obsolescence
3. Volatility

Bit Rot

- Preventing measures can help (proper storage and handling), but bits on a given medium will eventual flip or become unreadable
- Bit rot and advantages of newer media both call for periodic refreshing and reformatting
- Ensuring the **integrity of the bit stream** when transferring from one medium to another is extremely important

Obsolescence

“Obsolete power corrupts obsoletely.”

- Ted Nelson

The technology associated with interpreting the representation at each of the layers can change or become less available

Order of Volatility

- Some types of data change much more quickly and often than others
- Important to recognize in order to recover data from a computer system or media, while ensuring that actions don’t make irreversible changes to their record characteristics
- Example: If the contents of the browser cache are important to you, capture the cache before using the browser

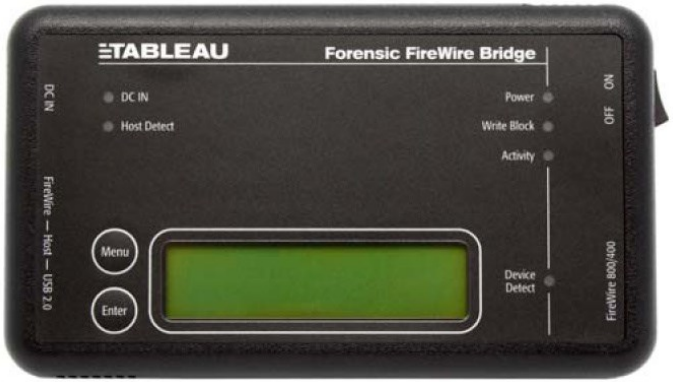
Digital Forensics in Libraries, Archives and Museums (LAMs)

- In recent years, LAMs have been applying various digital forensics methods, for example:
  - use of write blockers
  - generation of disk images
  - applying cryptographic hashes to files
  - capture of Digital Forensics XML (DFXML)
  - scanning bitstreams for personally identifying information

Need for Adaptation of Digital Forensics Tools and Tasks for LAMs

- Existing digital forensics tools provide valuable functionality, but they don’t always fit well into primary workflows of LAMs.
- For example, LAMs are particularly concerned with:
  - structure and persistence of metadata
  - provisions for providing public access to data
  - support for older technologies (e.g. floppy disks, HFS)

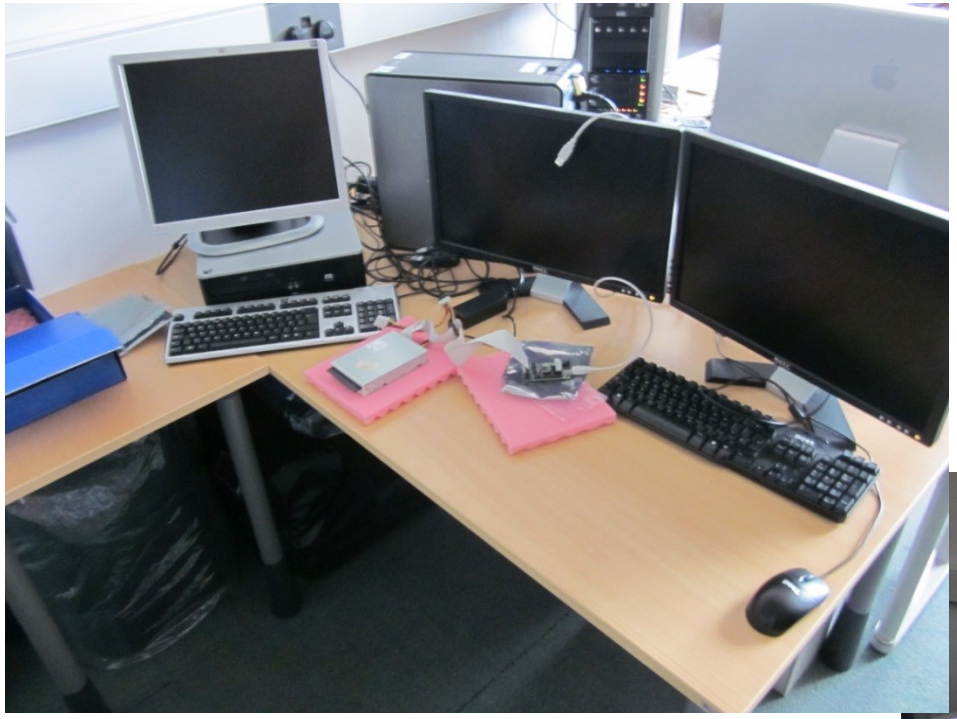
They're using a lot of this stuff:



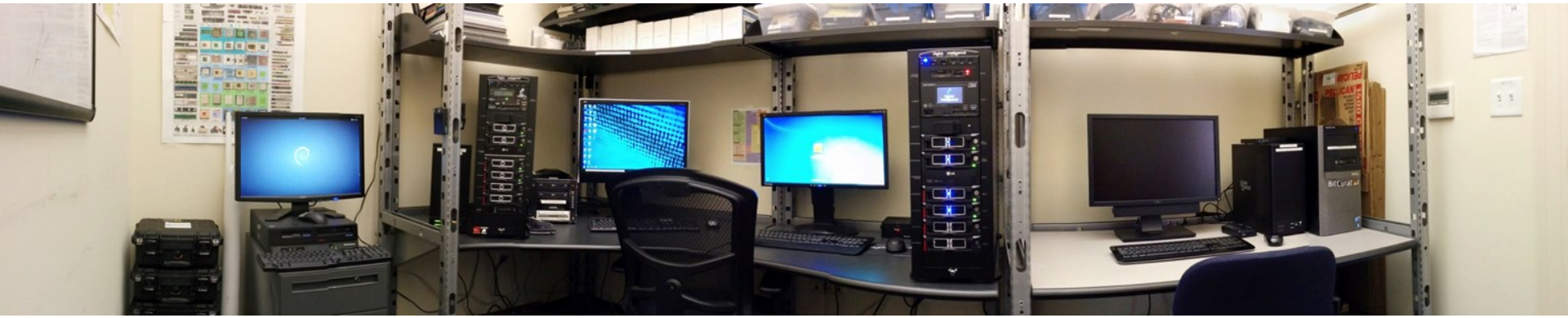
Here’s what it looks like in libraries and archives:

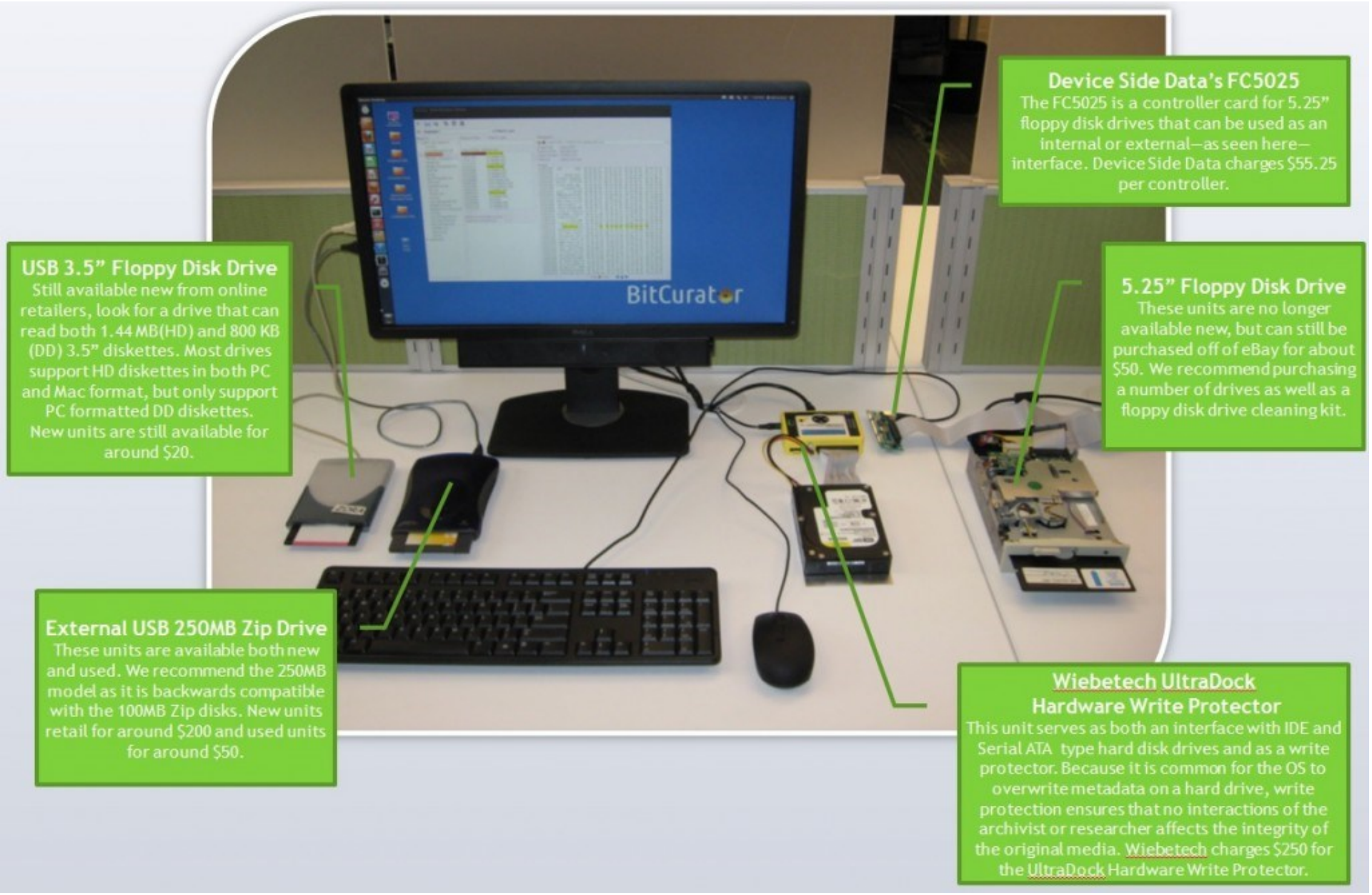
Stanford University Libraries and Academic Information Resources (SULAIR)

British Library, London



UNC School of Information and Library Science (Manning 213)

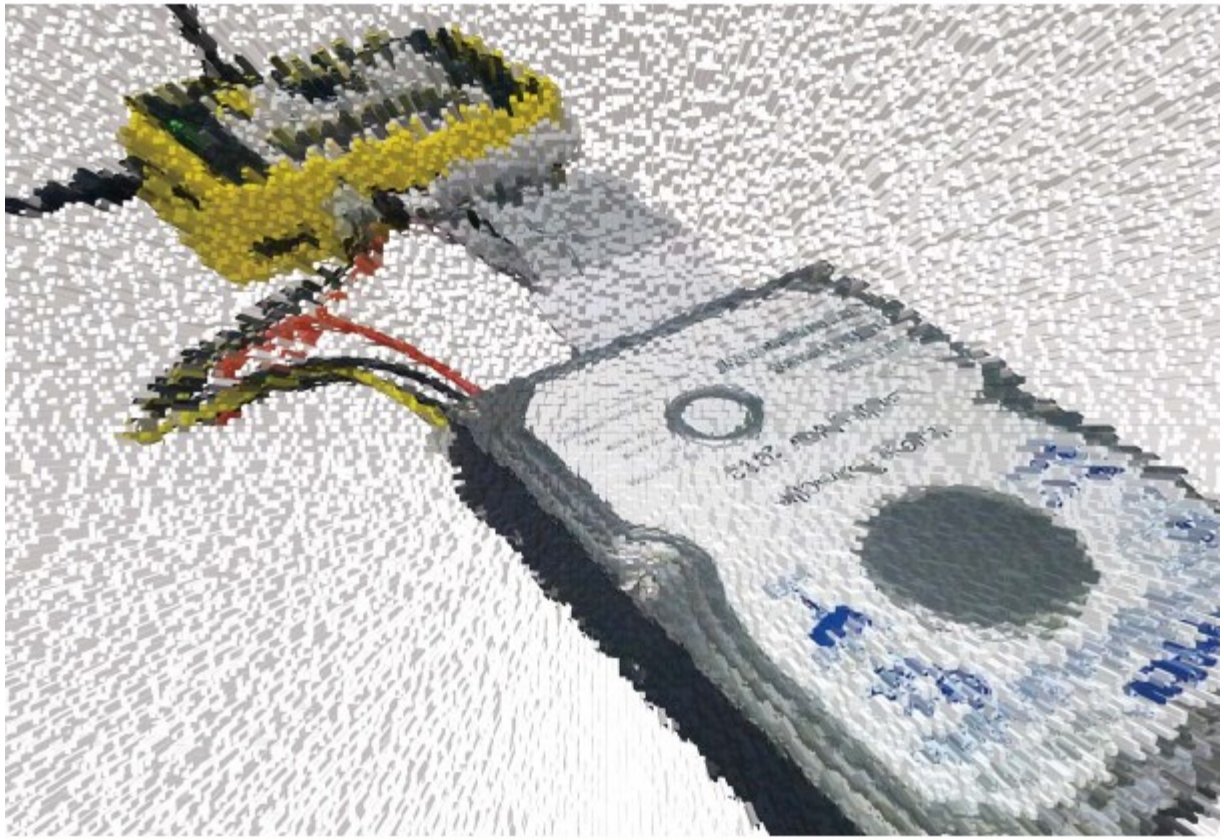




Porter Olsen, Building a Digital Curation Workstation with BitCurator (update)
http://www.bitcurator.net/building-a-digital-curation-workstation-with-bitcurator-update/

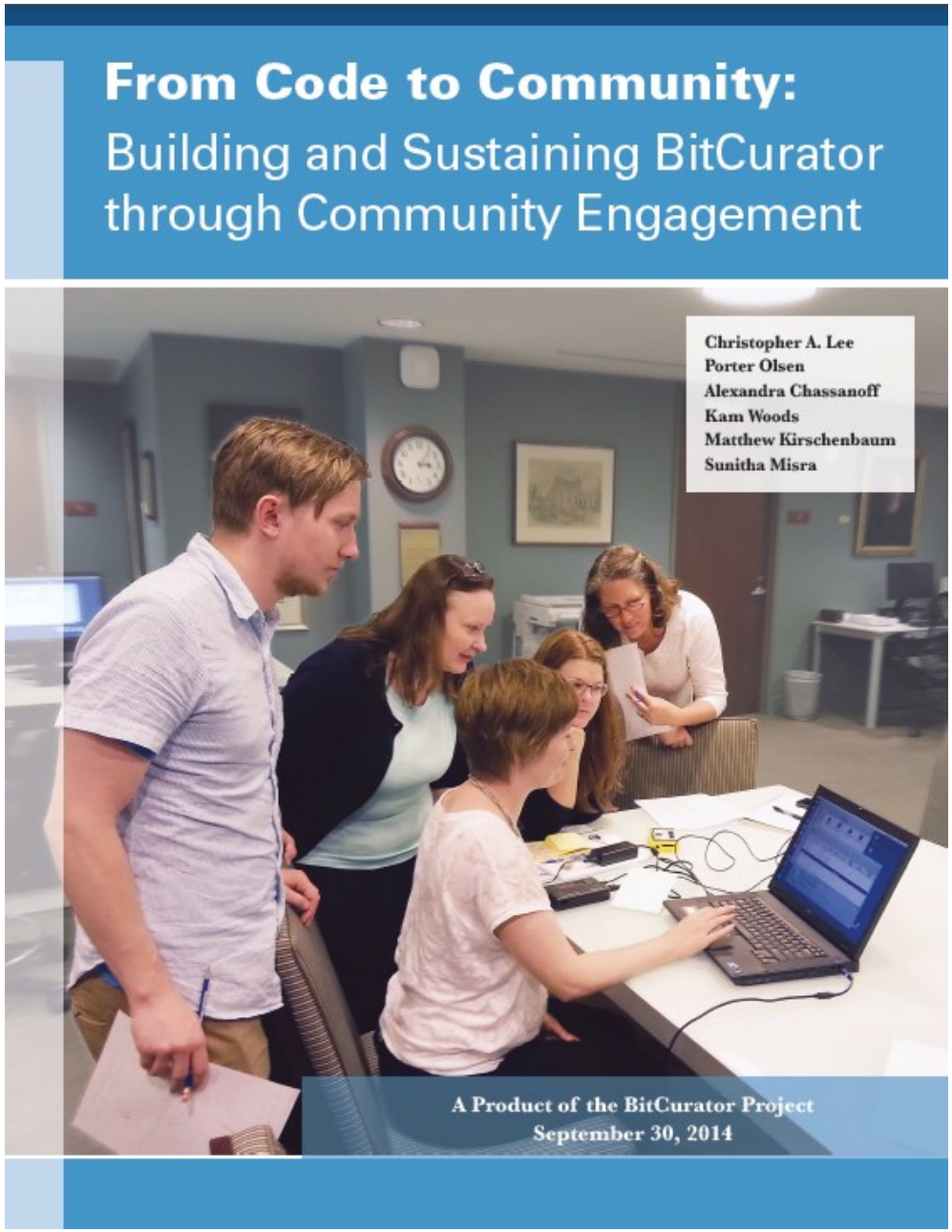
From Bitstreams to Heritage:

Putting Digital Forensics into Practice in Collecting Institutions



Christopher A. Lee, Kam Woods, Matthew Kirschenbaum, and Alexandra Chassanoff

http://www.bitcurator.net/docs/bitstreams-to-heritage.pdf



http://www.bitcurator.net/wp-content/uploads/2014/11/code-to-community.pdf

BitCurator

- Funded by Andrew W. Mellon Foundation
  - Phase 1: October 1, 2011 – September 30, 2013
  - Phase 2 – October 1, 2013 – September 30, 2014
- Partners: School of Information and Library Science (SILS) at UNC and Maryland Institute for Technology in the Humanities (MITH)

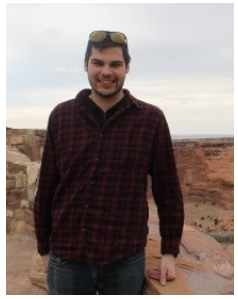
Core BitCurator Team

- Cal Lee, PI
- Matt Kirschenbaum, Co-PI
- Kam Woods, Technical Lead
- Porter Olsen, Community Lead
- Alex Chassanoff, Project Manager
- Sunitha Misra, Software Developer (UNC)
- Kyle Bickoff, GA (MITH)
- Amanda Visconti, GA (MITH)









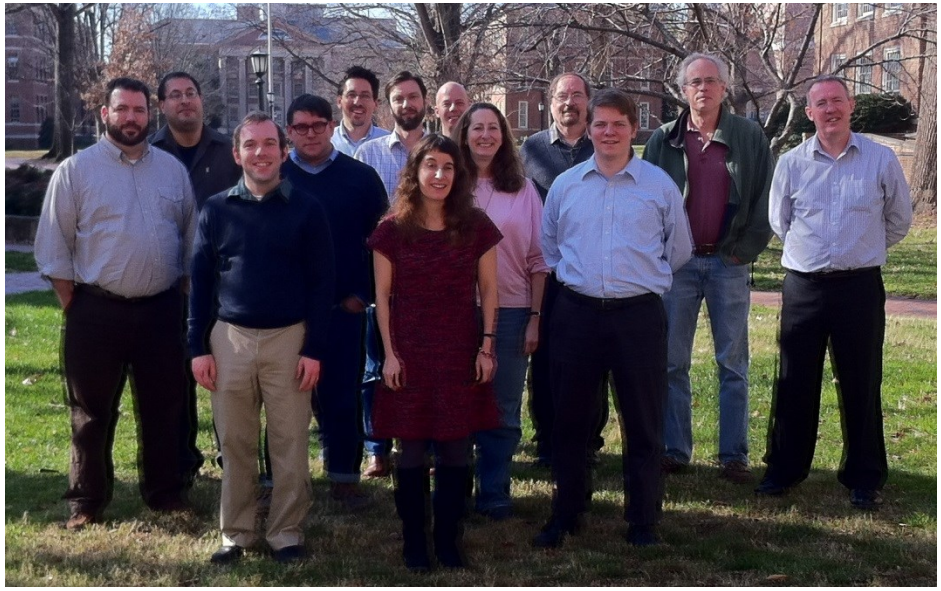

Two Groups of Advisors

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, US National Archives and Records Administration
- Naomi Nelson, Duke University
- Erin O'Meara, Gates Archive
- Michael Olson, Stanford University Libraries
- Gabriela Redwine, Beinecke, Yale University
- 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, Digital Public Library of America
- 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



BitCurator Goals

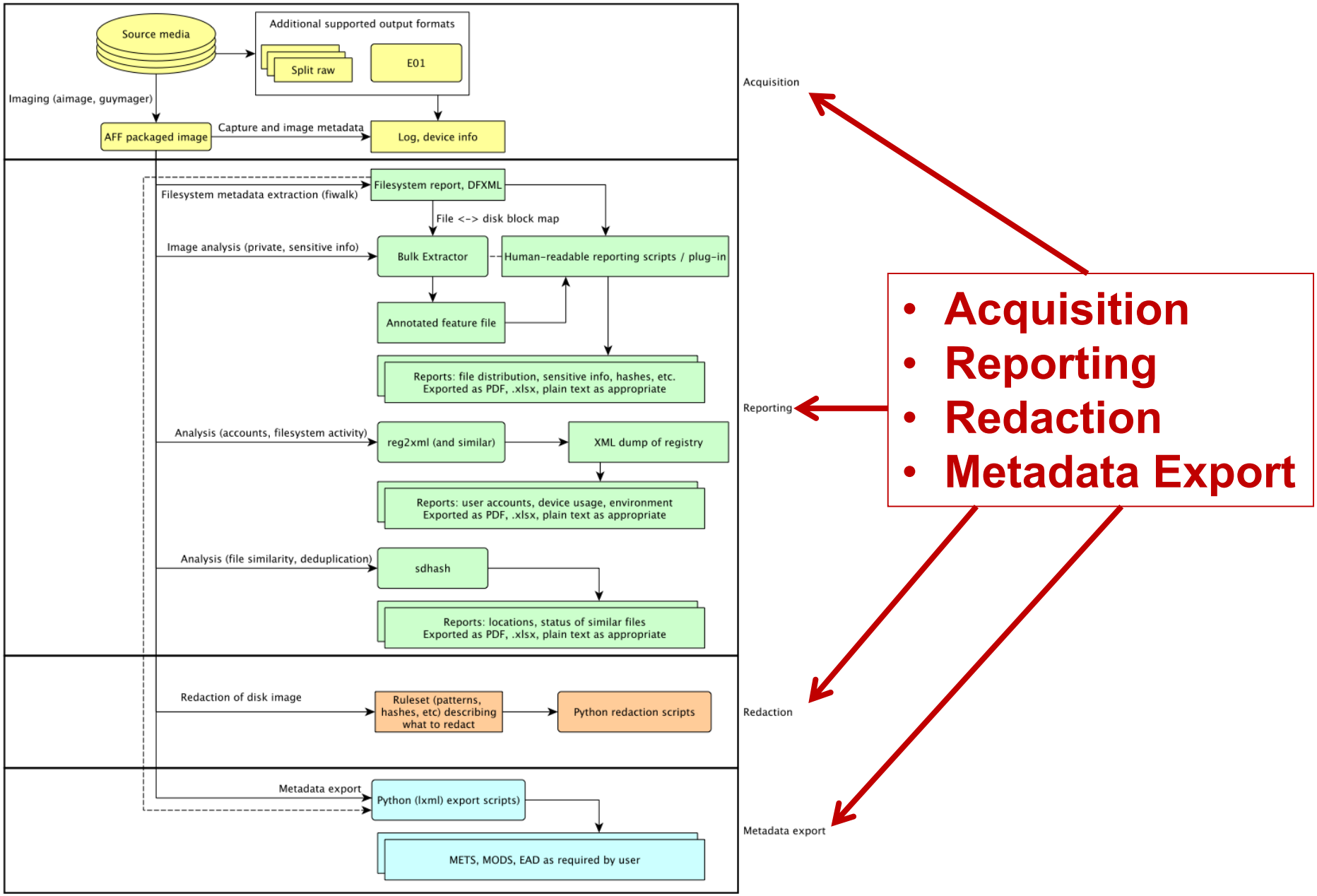
- Develop a system for collecting professionals that incorporates the functionality of open-source digital forensics tools
- Address two fundamental needs not usually addressed by the digital forensics industry:
  - Incorporation into the workflow of LAM ingest and collection management environments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

BitCurator Environment*

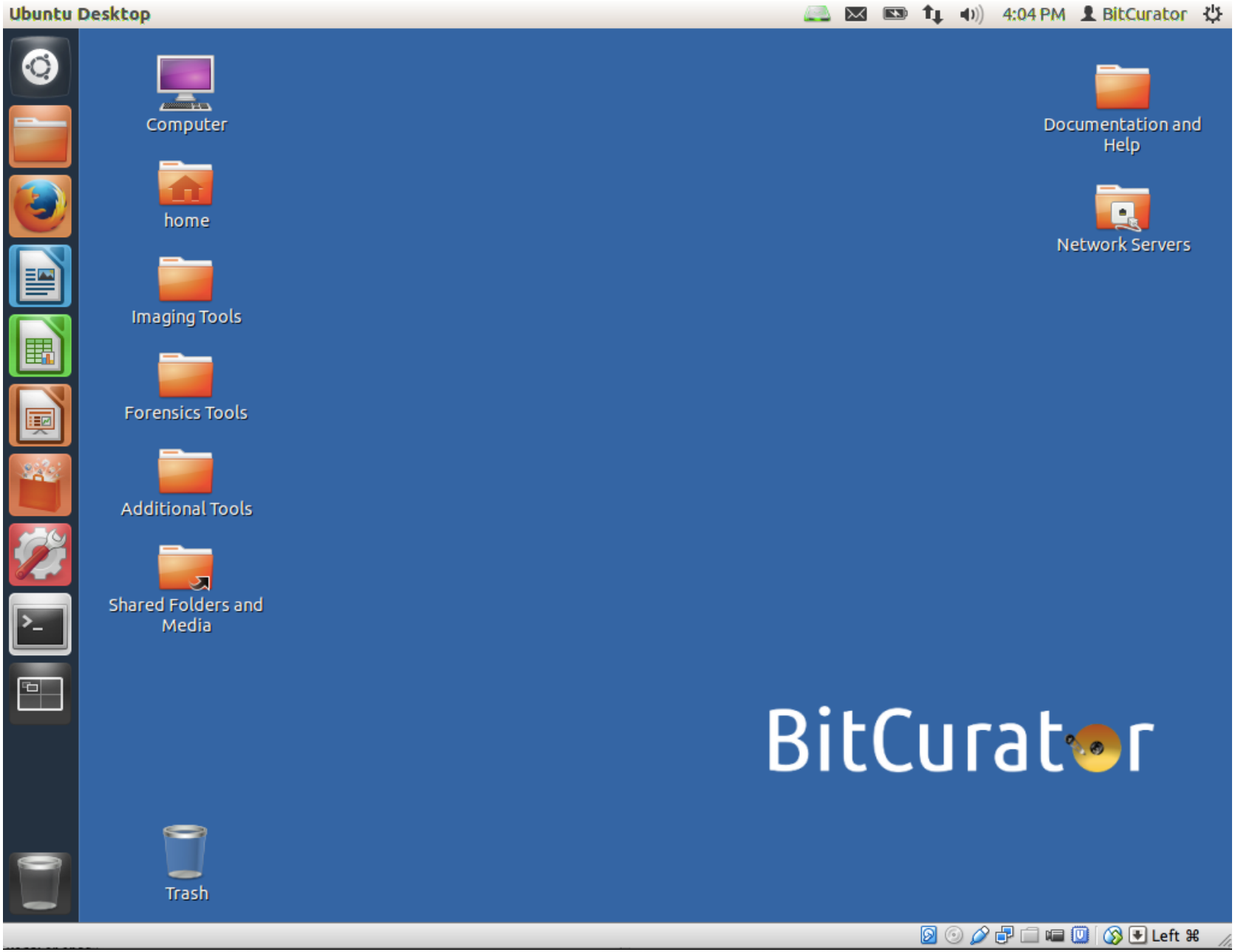
- Bundles, integrates and extends functionality of open source software
- Can be run as:
  - Self-contained environment (based on Ubuntu Linux) running directly on a computer (download installation ISO)
  - Self-contained Linux environment in a virtual machine using e.g. Virtual Box or VMWare
  - As individual components run directly in your own Linux environment or (whenever possible) Windows environment
- We'll be looking at specific BitCurator tools and functions later

*To read about and download the environment, see: http://wiki.bitcurator.net/

BitCurator-Supported Workflow



See: http://bitcurator.net


Ubuntu Desktop
4:04 PM
BitCurator
Computer
home
Imaging Tools
Forensics Tools
Additional Tools
Shared Folders and Media
Trash
Documentation and Help
Network Servers
BitCurator
Left ⌘

BitCurator Consortium

- Continuing home for hosting, stewardship and support of BitCurator (and BitCurator Access) tools and associated user engagement
- Administrative home: Educopia Institute
- Funding based on membership dues
- Institutions as members, with two categories of membership: Charter and General
- The most important member benefit is assurance that the BitCurator software will persist in future years

http://www.bitcurator.net/bitcurator-consortium/

A Growing Community

The BitCurator Consortium provides spaces for members to share documentation, develop their skills, and improve the BitCurator environment.

Membership is open >

Membership is open to libraries, archives, museums, and other institutions worldwide that seek a collaborative community within which they may explore and apply forensics approaches and solutions to their digital collections.

Become a member now >

How to Use BitCurator

- Acquire and process digital collections.
- Maintain the original order of digital materials.
- Survey the extent and composition of digital collections.
- Redact personally identifiable information.
- Extract technical and preservation metadata.
- Package digital materials for archival storage.

Learn more about getting started.

Member Benefits

- Use of the members-only BCC mailing list and help desk
- Access to the members-only videos and documentation
- Prioritized requests for BitCurator feature development
- Opportunities to serve on the BCC committees
- Voting rights for community governance
- Professional development opportunities
- Discounts for events including the BitCurator User Forum

How our members are using BitCurator

Members

McMaster University

Penn State University

Massachusetts Institute of Technology

Duke University

The University of Maryland, MITH

Stanford University

Yale University

The University of Manchester Library

University of

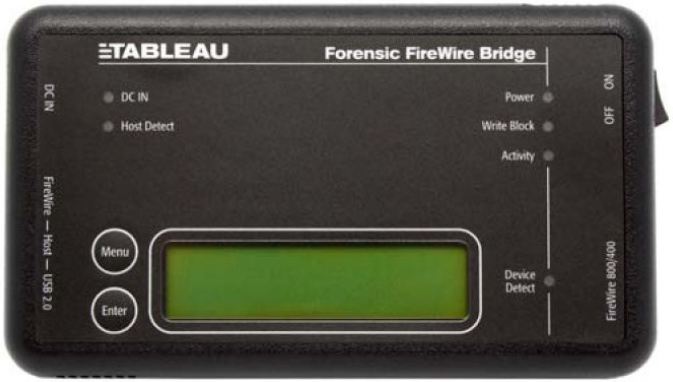
So let’s now look at specific curation actions and processes…

Write Blocking – One-Way Streets for Data

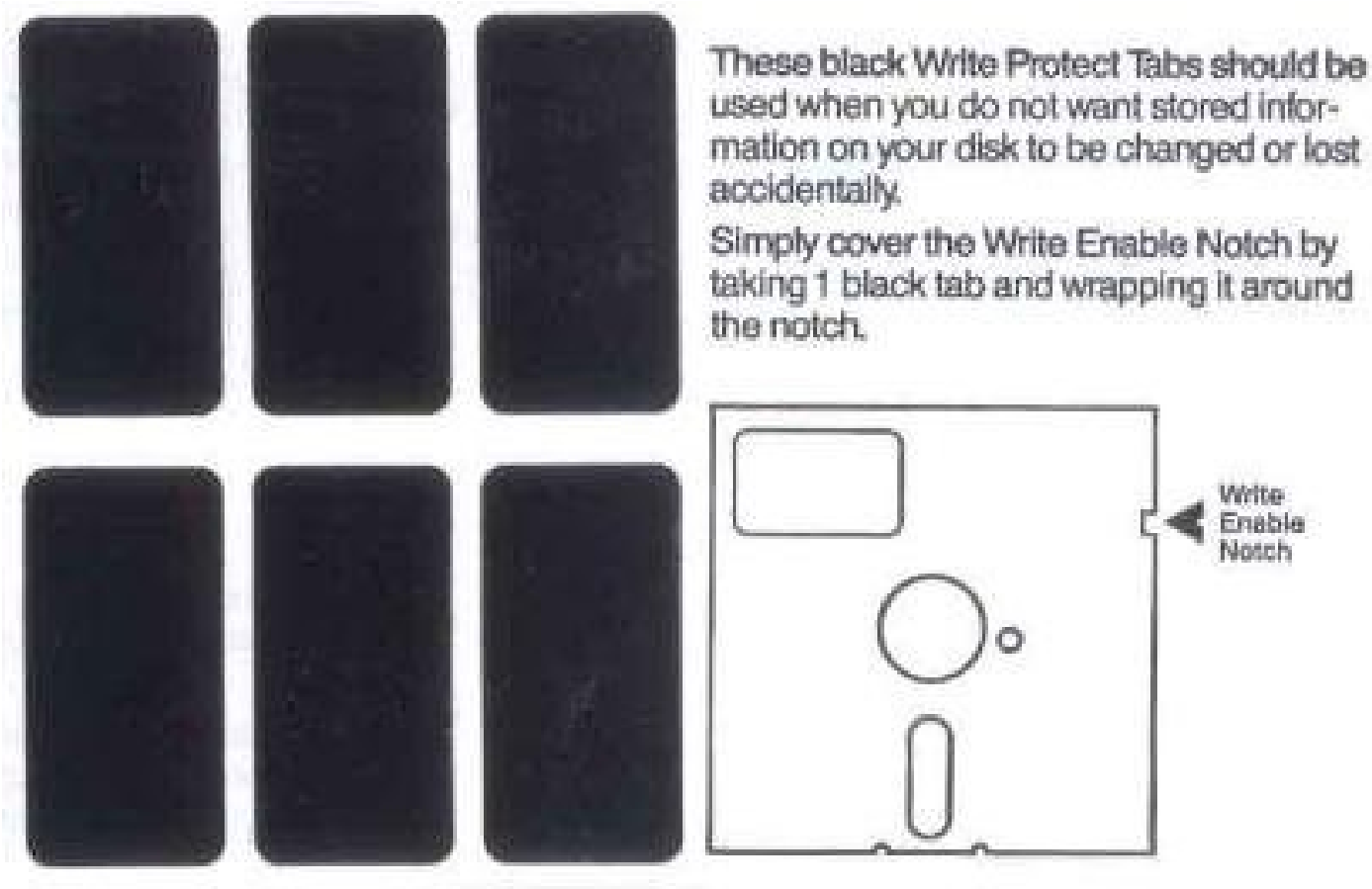
- Ensures that data can be read from the device, but no bits can be changed
- Doesn't just prevent changes conscious made by user but also changes made by the system
- Options for write blocking (in order of most to least certain to prevent writes to the drive):
  - Dedicated write blockers
  - Writing blocking tabs or settings on the device itself
  - Software-based write blocking

Image source: http://thinng.com/1555-one-way-sign-seat

Dedicated Hardware Write Blockers

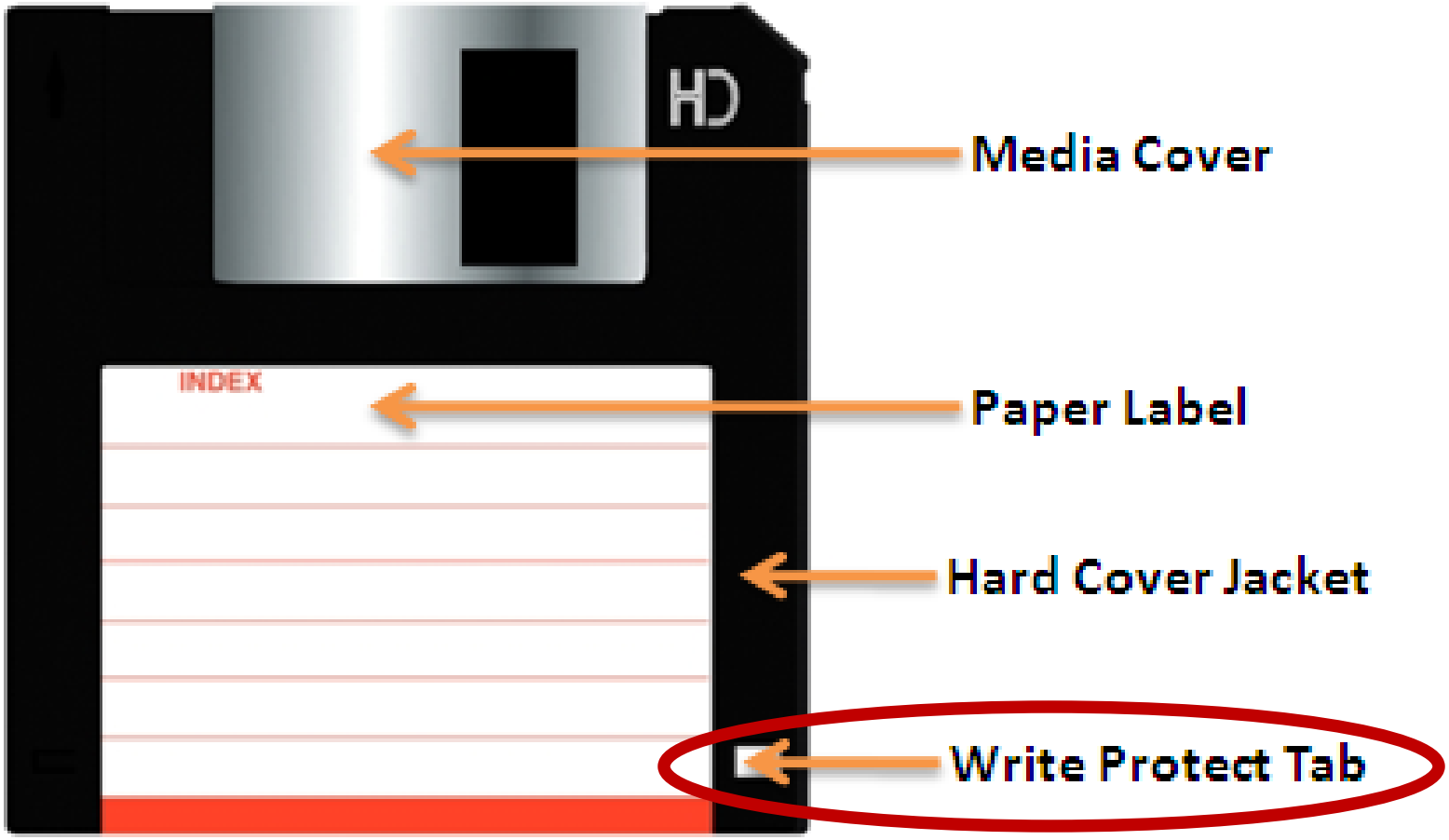


5.25 Inch Floppy – If light can get through, it's **not** write protected



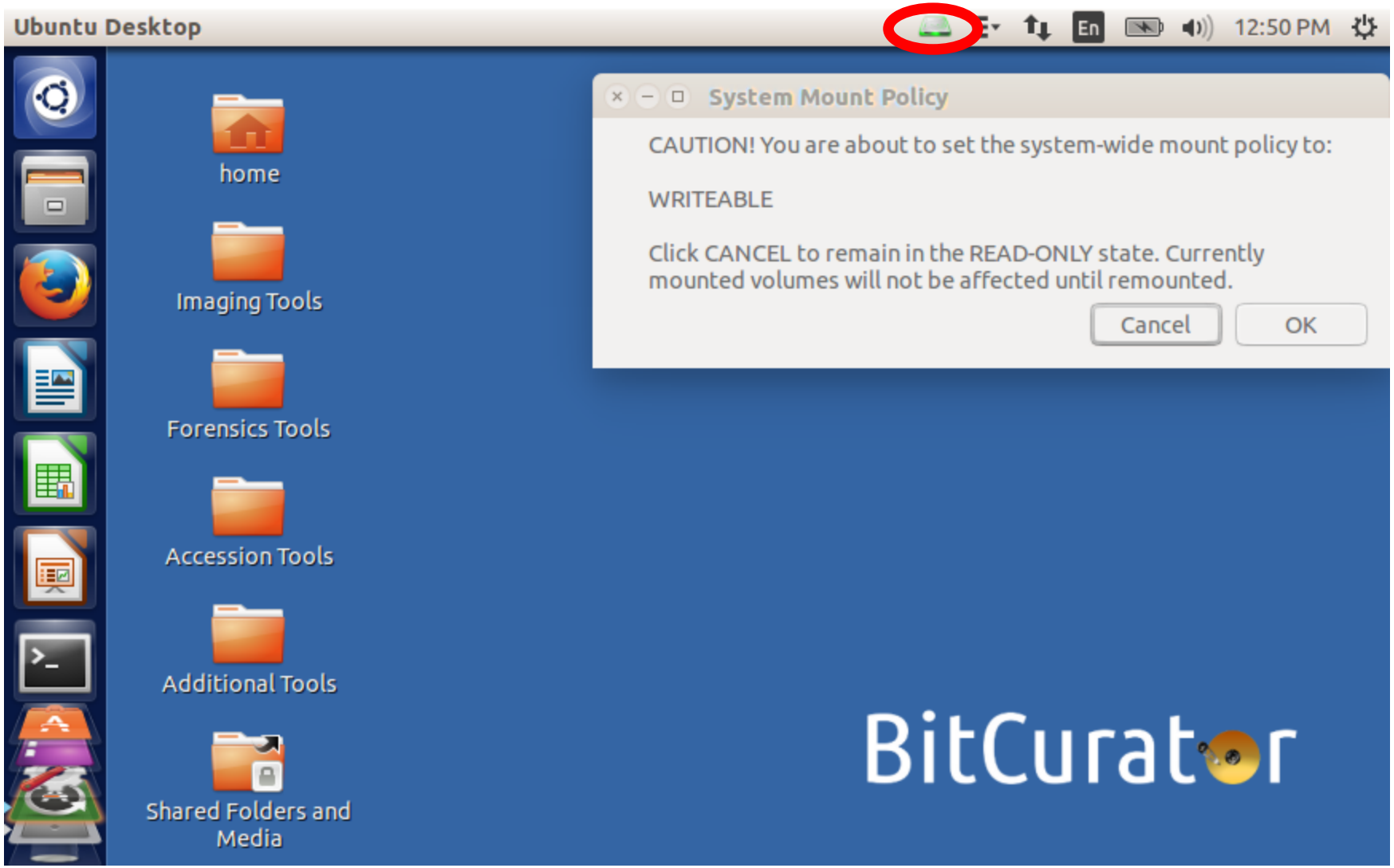
http://en.wikipedia.org/wiki/File:Floppy_tabs_3x2.jpg

3.5 Inch Floppy – If light can get through, it **is** write protected



http://www.techmint.info/2009/09/security-write-protecting-floppy-disks.html

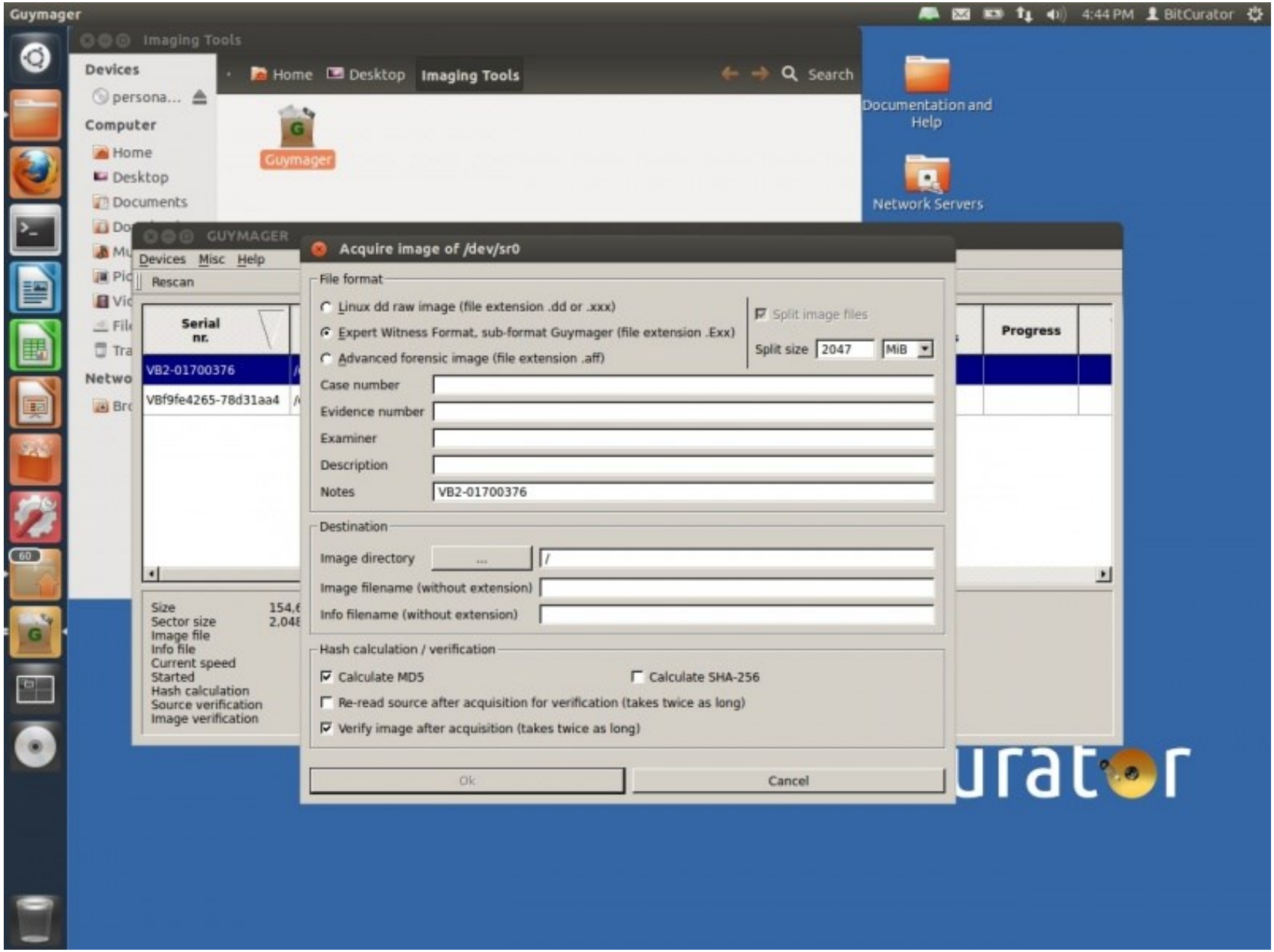
Example of Software Write Blocking – Mounted Devices set to Read-Only by Default

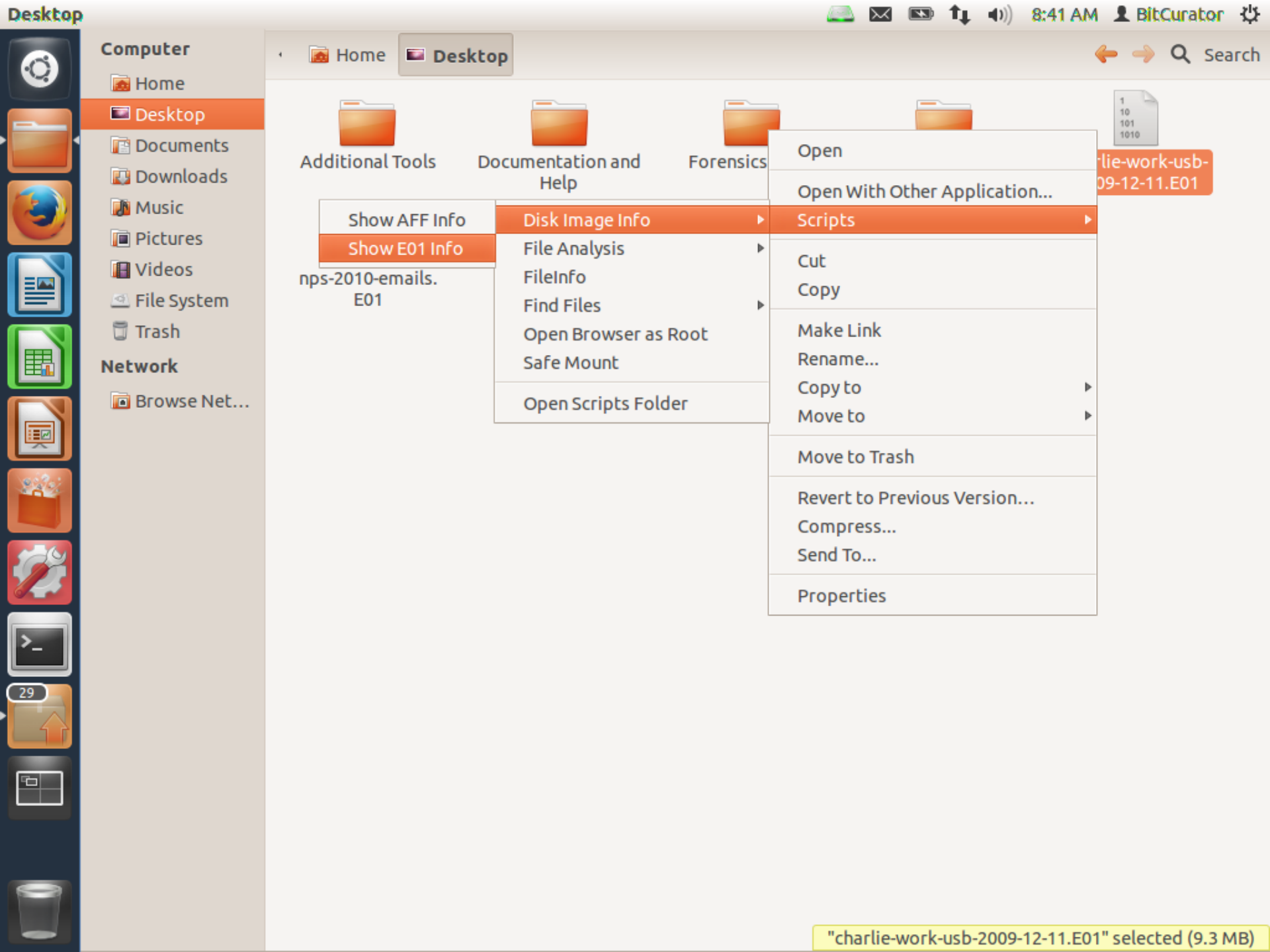


Getting below the File System – Low-Level Copying

- Getting an “image” of a storage medium involves working at a level below the file system
  - Can get at file attributes and deleted files not visible through higher-level copy operations
- Most commonly used tool is dd (or variant) - UNIX program for low-level copying and conversion of data from a storage device
- More specialized tools for creating forensic images include:
  - FTK Imager
  - Guymager
  - Imaging utilities in commercial applications (e.g. EnCase)

Main Acquisition Interface for Guymager




Desktop
8:41 AM
BitCurator
Computer
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
File System
Trash
Network
Browse Net...
Home
Desktop
Search
Additional Tools
Documentation and Help
Forensics
nps-2010-emails.E01
lie-work-usb-09-12-11.E01
Show AFF Info
Show E01 Info
Disk Image Info
File Analysis
FileInfo
Find Files
Open Browser as Root
Safe Mount
Open Scripts Folder
Open
Open With Other Application...
Scripts
Cut
Copy
Make Link
Rename...
Copy to
Move to
Move to Trash
Revert to Previous Version...
Compress...
Send To...
Properties
"charlie-work-usb-2009-12-11.E01" selected (9.3 MB)

EnCase Disk Image Info

8:43 AM BitCurator

Computer
- Home
- Desktop
- Documents
- Downloads
- Music
- Pictures
- Videos
- File System
- Trash

Network
- Browse Net...

Home | Desktop | Search

charlie-work-usb-2009-12-11.E01

EnCase Disk Image Info

```
ewfinfo 20130416

Acquiry information
	Acquisition date:	Wed Jan 19 12:09:18 2011
	System date:		Wed Jan 19 12:09:18 2011
	Operating system used:	Linux
	Software version used:	20100226
	Password:		N/A

EWF information
	File format:		EnCase 6
	Sectors per chunk:	64
	Error granularity:	64
	Compression method:	deflate
	Compression level:	best compression
	Set identifier:		4eb6701d-6cf0-2f4a-a0c6-0cb5d5e20959

Media information
	Media type:		fixed disk
	Is physical:		yes
	Bytes per sector:	512
	Number of sectors:	2068480
	Media size:		1010 MiB (1059061760 bytes)

Digest hash information
	MD5:		9c0de6c8532d7a66ddcf01861dfb6535
```

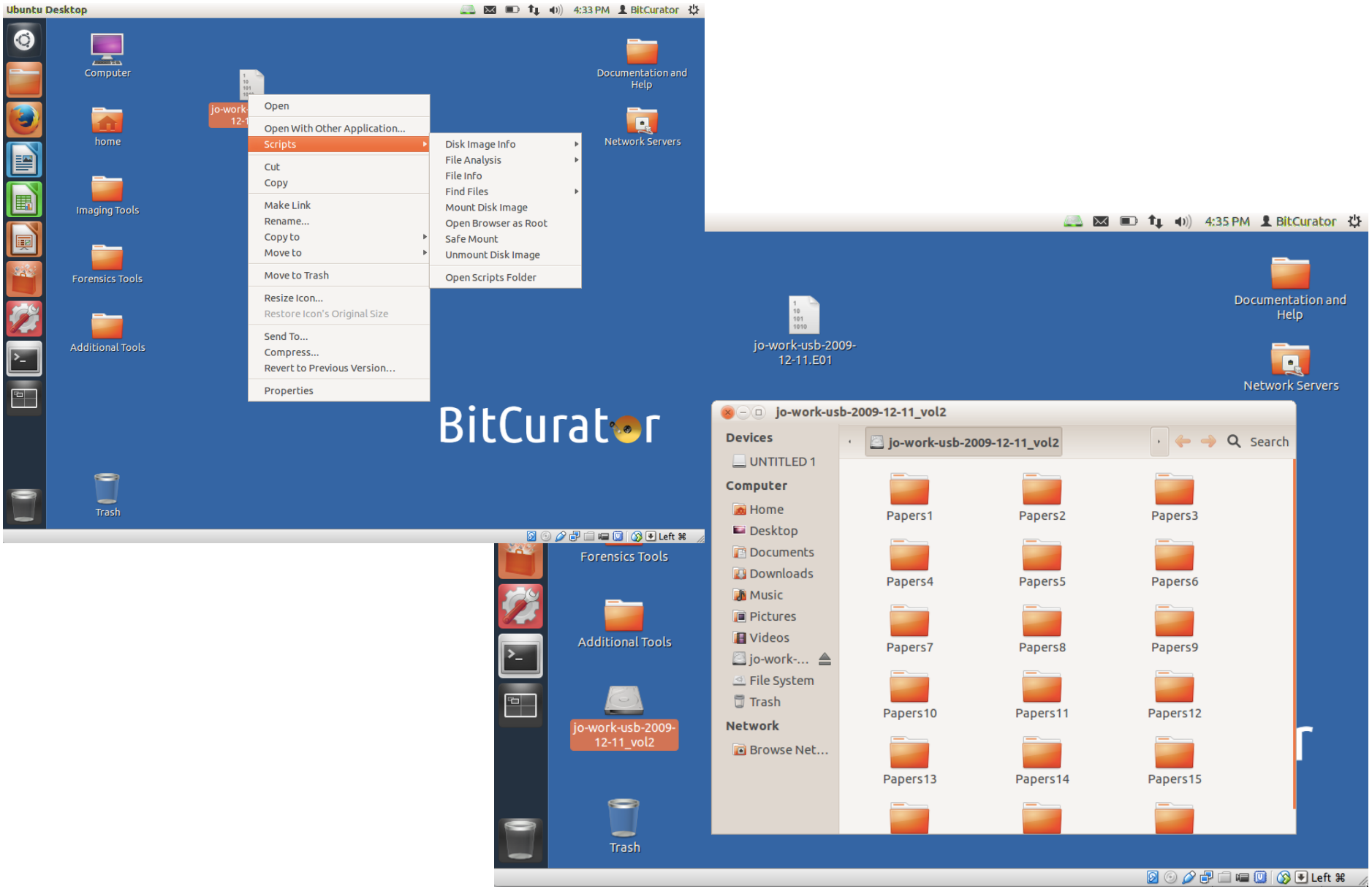
Cancel OK

"charlie-work-usb-2009-12-11.E01" selected (9.3 MB)

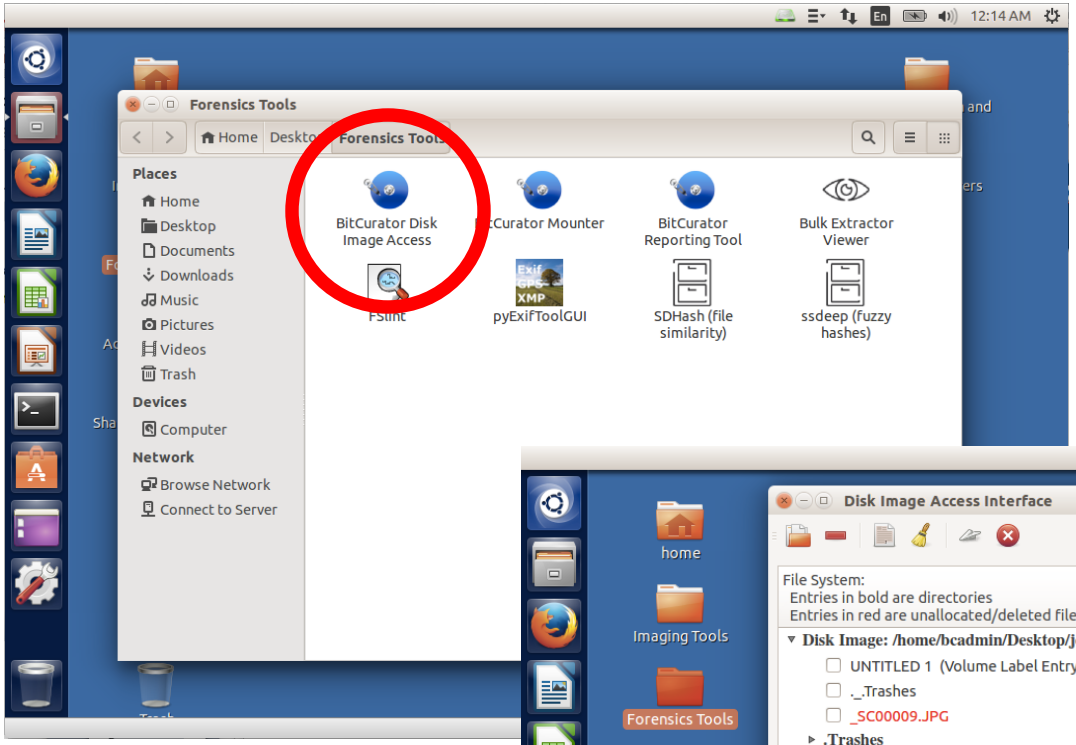
Two Ways to Interact with Disk Images

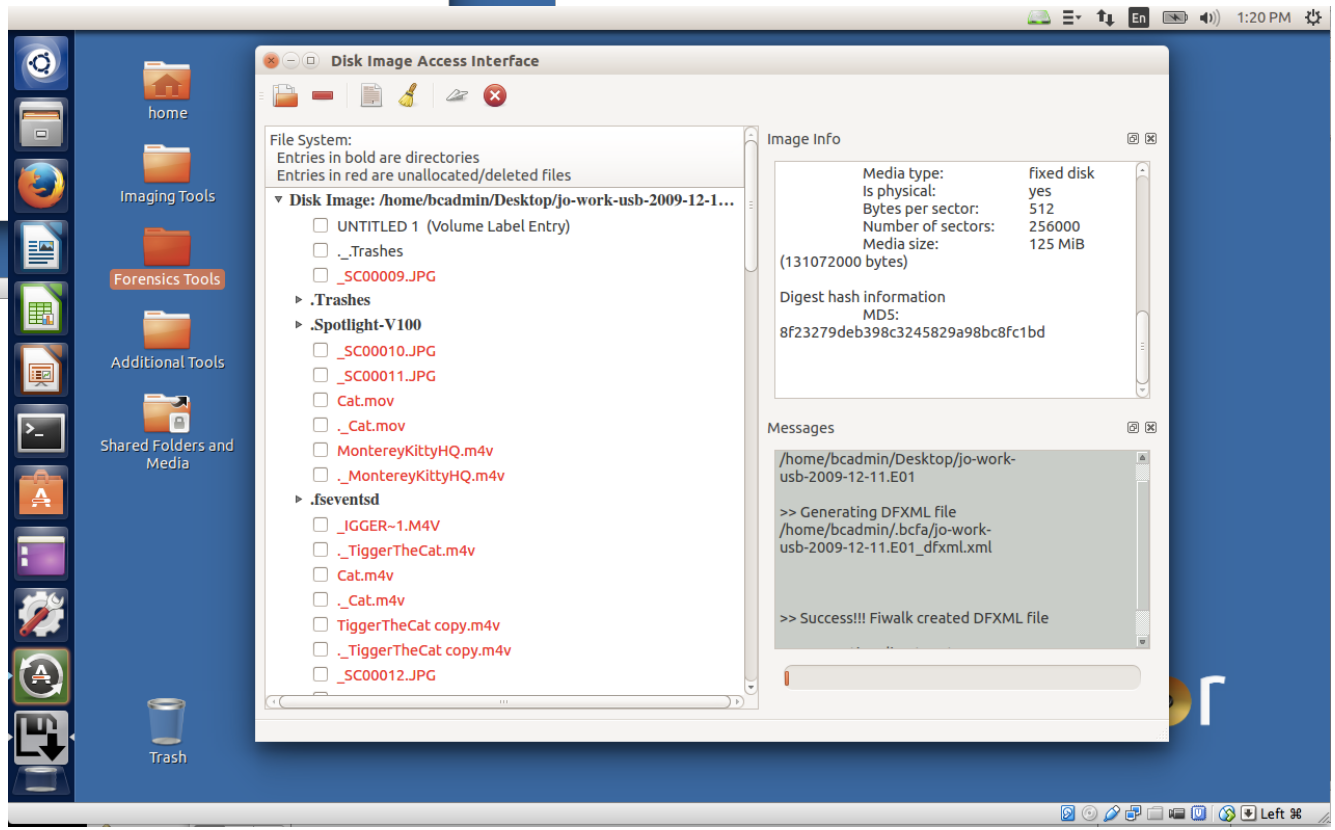
- Mount them like regular drives:
  - Disk Utility in Mac OS X (for ISO images)
  - ewfmount
  - MagicDisc (for ISO images)
  - OSFMount
  - BitCurator (mounting scripts built into the environment)
- Inspect them as forensic objects
  - FTK Imager
  - The Sleuth Kit (TSK)
  - BitCurator (Disk Image Access tool)

Mounting a Forensically Packaged Disk Image in the BitCurator Environment



Exporting Files from a Disk Image





Identifying Potentially Sensitive Data using Bulk Extractor - Scanning Options

File Edit View Bookmarks Tools Help

Reports

Feature Filter

Feature File

Referenced Fea
Referenced Fea

Required Parameters

Scan: Image File Raw Device Directory of Files

Image File

Output Feature Directory

General Options

- Use Banner File
- Use Alert List File
- Use Stop List File
- Use Find Regex Text File
- Use Find Regex Text
- Use Random Sampling

Tuning Parameters

- Use Context Window Size 16
- Use Page Size 16777216
- Use Margin Size 4194304
- Use Block Size 512
- Use Number of Threads 2
- Use Maximum Recursion Depth 7
- Use Wait Time 60

Parallelizing

- Use start processing at offset
- Use process range offset o1-o2
- Use add offset to reported feature offsets

Debugging Options

- Start on Page Number 0
- Use Debug Mode Number 1

Scanners

- base16
- facebook
- outlook
- sceadan
- wordlist
- xor
- accts
- aes
- base64
- elf
- email
- exif
- find
- gps
- gzip
- hiberfile
- httplogs
- json
- kml
- msxml
- net
- pdf
- rar
- sqlite
- vcard
- windirs
- winlnk
- winpe
- winprefetch
- zip

See: http://www.forensicswiki.org/wiki/Bulk_extractor

Histogram of Email Addresses (Specific Instances in Context on Right)

BitCurator-0.2.0 [Running]

Bulk Extractor Viewer

8:44 PM BitCurator

File Edit View Tools Help

Highlight: Match case

Reports

- beoutput
  - domain.txt
  - domain_histogram
  - email.txt
  - email_histogram.tx
  - ether.txt
  - ether_histogram.t
  - json.txt
  - packets.pcap
  - rfc822.txt
  - tcp.txt
  - tcp_histogram.txt
  - url.txt
  - url_histogram.txt
  - url_services.txt
  - windirs.txt
  - winpe.txt

Feature Filter Match case

Histogram File email_histog...

n=12	privacy@motorola.cor
n=3	0mj5nj@0itgx.ib.dj
n=3	73t@fo.pa
n=3	john@humaniz.com
n=3	newton@planetb.fr
n=3	sales@integrationnew
n=1	5kda_c@kqahw.sl
n=1	dqf@40mt.ro
n=1	fodfv@nwa4.ck
n=1	imki@73yjt.lr
n=1	jqnmq@17.pn
n=1	kjph@sj.gr
n=1	nq9@5c7k.sg
n=1	pdcnfb@tft.ao
n=1	qyf@j65.de
n=1	tw+4vsa@xf.ms

Referenced Feature File e...
Referenced Feature pri...

34804080	privacy@Motor
34807246	privacy@Motor
34808676	privacy@Motor
42271602	privacy@Motor
42273785	privacy@Motor
42274743	privacy@Motor
42347307	privacy@Motor
42349490	privacy@Motor
42350448	privacy@Motor
74735841	privacy@Motor
74738019	privacy@Motor
74738989	privacy@Motor

Navigation

sampleimage.E01, 42273785, privacy@Motorola.com

Image File	sampleimage.E01
Feature File	email.txt
Feature Path	42273785
Feature	privacy@Motorola.com

Image

```
42271936  your credit card number, so this information can only be viewed
42272000  by Motorola. .Motorola uses Secure Sockets Layer (SSL) encrypti
42272064 on technology, the highest level of security on the Internet. Th
42272128 e SSL protocol provides server authentication, data integrity, a
42272192 nd privacy on the Web. This security measure helps ensure that n
42272256 o impostors, eavesdroppers, or vandals get your personal informa
42272320 tion. SSL not only encrypts your personal and financial informat
42272384 ion transmitted, including credit card information, but also ver
42272448 ifies the identity of the server and that the original message a
42272512 rrives safely at its destination. .However, no data transmission
42272576  over the Internet can be guaranteed to be 100% secure. As a res
42272640 ult, while we strive to protect your personal information, Motor
42272704 ola cannot ensure or warrant the security of any information you
42272768  transmit to us or from our Web site, and therefore you use our
42272832 site at your own risk. Once we receive your transmission, we use
42272896  our best effort to ensure its security on our systems. ..000200
42272960 0007AE000038B6.7A8,As a global company Motorola has internationa
42273024 l sites and users all over the world. When you give Motorola per
42273088 sonal information, that information may be sent electronically t
42273152 o servers outside of the country where you originally entered th
42273216 e information. In addition, that information may be used, stored
42273280  and processed outside of the country where you entered that inf
42273344 ormation. Whenever Motorola handles personal information, regard
42273408 less of where this occurs, it takes steps to ensure that your in
42273472 formation is treated securely and in accordance with the relevan
42273536 t Terms of Use and this Privacy Policy. .How can I correct or ch
42273600 ange my personal information? .If you would like to review, corr
42273664 ect or change any personal information you have provided, or rem
42273728 ove your name from our mailing list, please e-mail us at privacy
42273792 @Motorola.com. If you have established a "user profile" on a Mot
42273856 orola website, you may change the information you provided at an
```

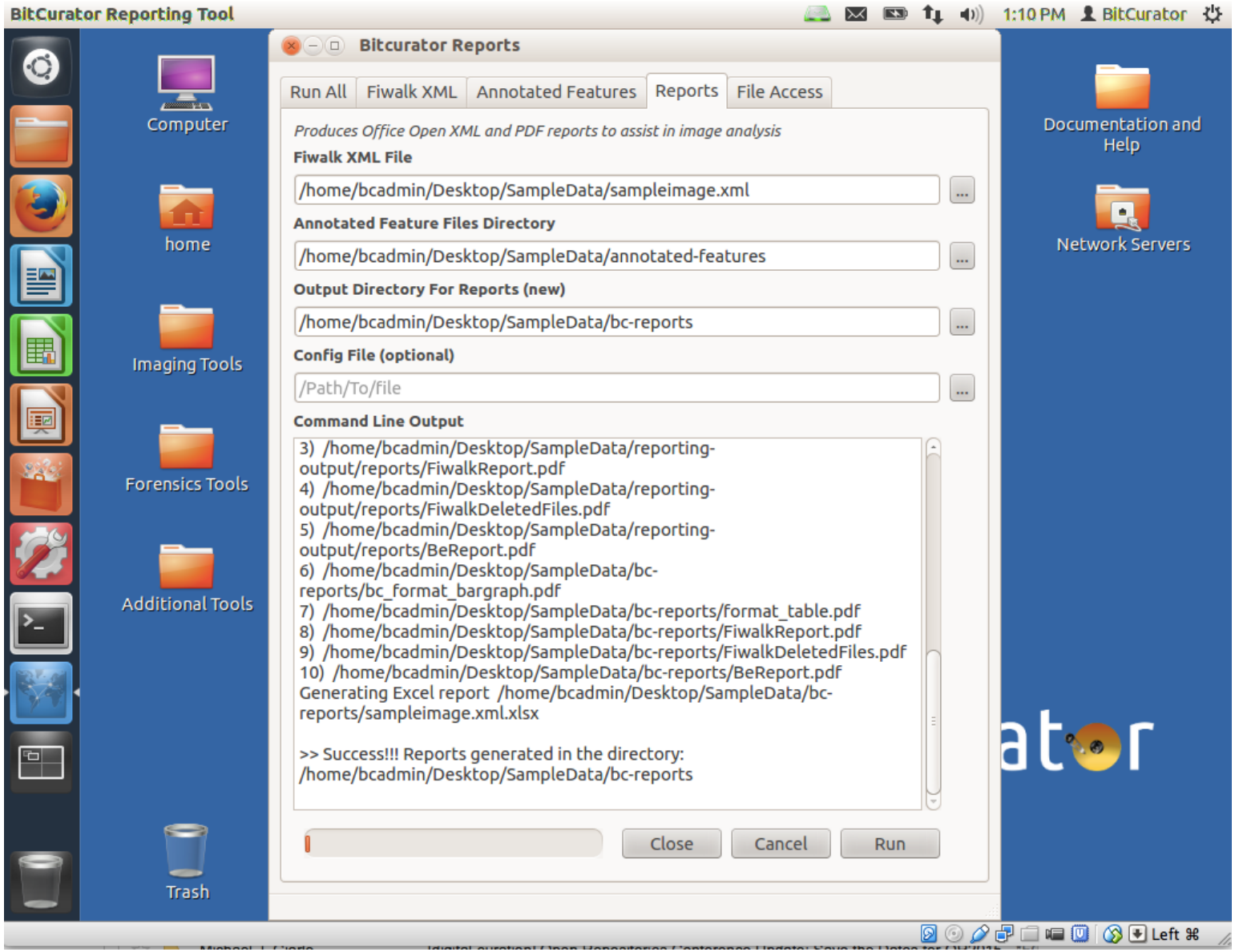
Text Hex

Left ⌘

```
OUTLOOK-36[redacted]2  DOB:  [redacted]
OUTLOOK-36[redacted]8  SSN:  [redacted]
OUTLOOK-36[redacted]8  DOB:  [redacted]
OUTLOOK-36[redacted]9  SSN:  [redacted]
OUTLOOK-36[redacted]9  DOB:  [redacted]
OUTLOOK-36[redacted]7  SSN:  [redacted]
OUTLOOK-36[redacted]7  DOB:  [redacted]
OUTLOOK-46[redacted]   SSN:  [redacted]
OUTLOOK-46[redacted]   SSN:  [redacted]
OUTLOOK-46[redacted]   SSN:  [redacted]
OUTLOOK-46[redacted]   SSN:  [redacted]
OUTLOOK-46[redacted]   SSN:  [redacted]
OUTLOOK-46[redacted]   SSN:  [redacted]
OUTLOOK-46[redacted]   SSN:  [redacted]
OUTLOOK-46[redacted]   SSN:  [redacted]
OUTLOOK-46[redacted]   SSN:  [redacted]
OUTLOOK-46[redacted]   SSN:  [redacted]
OUTLOOK-46[redacted]   SSN:  [redacted]
OUTLOOK-55[redacted]   SSN:  [redacted]
OUTLOOK-55[redacted]   SSN:  [redacted]
OUTLOOK-55[redacted]   SSN:  [redacted]
OUTLOOK-55[redacted]   SSN:  [redacted]
OUTLOOK-55[redacted]   SSN:  [redacted]
OUTLOOK-55[redacted]   SSN:  [redacted]
OUTLOOK-55[redacted]   SSN:  [redacted]
OUTLOOK-55[redacted]   SSN:  [redacted]
OUTLOOK-55[redacted]   SSN:  [redacted]
OUTLOOK-55[redacted]   SSN:  [redacted]
OUTLOOK-56[redacted]   SSN:  [redacted]
```

SSNs and DOBs identified in large PST collection using bulk_extractor

Generating BitCurator Reports



- Provenance metadata - about the disk capture process
- Technical metadata - about the specific storage partition(s) on the disk

BitCurator-0.2.0 [Running]

Mozilla Firefox 8:08 PM BitCurator

file:///home/b...mpleimage.xml

file:///home/bcadmin/Desktop/SampleData/sampleimage.xml

This XML file does not appear to have any style information associated with it. The document tree is shown below.


```
-<dfxml version="1.0">
  -<metadata>
     <dc:type>Disk Image</dc:type>
   </metadata>
  -<creator version="1.0">
     <program>fiwalk</program>
     <version>4.0.2</version>
   -<build_environment>
       <compiler>GCC 4.6</compiler>
       <library name="afflib" version="3.7.1"/>
       <library name="libewf" version="20130303"/>
     </build_environment>
   -<execution_environment>
     -<command_line>
         fiwalk -f -X /home/bcadmin/Desktop/SampleData/sampleimage.xml /home/bcadmin/Desktop/SampleData
         /sampleimage.E01
       </command_line>
       <start_time>2013-03-12T00:08:28Z</start_time>
     </execution_environment>
   </creator>
  -<source>
     <image_filename>/home/bcadmin/Desktop/SampleData/sampleimage.E01</image_filename>
   </source>
   <!-- fs start: 0 -->
  -<volume offset="0">
     <partition_offset>0</partition_offset>
     <block_size>2048</block_size>
     <ftype>2048</ftype>
     <ftype_str>iso9660</ftype_str>
     <block_count>36839</block_count>
```

Operationalizing Original Order - Filesystem Metadata Output from fiwalk*


```
<fileobject>
      <filename>Documents and Settings/All Users/Documents/
                  My Pictures/Sample Pictures/Blue hills.jpg
      </filename>
      ...
      <filesize>28521</filesize>
      <alloc>1</alloc>
      <used>1</used>
      <inode>6245</inode>
      ...
      <uid>0</uid>
      <gid>0</gid>
      <mtime>1208174400</mtime>
      <ctime>1257729636</ctime>
      <atime>1257729636</atime>
      <crtime>1257729636</crtime>
      <seq>2</seq>
      <libmagic>JPEG image data, JFIF standard 1.02</libmagic>
      <byte_runs>
       <run file_offset='0' fs_offset='0' img_offset='363200512'
         len='0'/>
      </byte_runs>
      <hashdigest type='MD5'>
          6fb2a38dc107eacb41cf1656e899cf70
      </hashdigest>
      <hashdigest type='SHA1'>
          4eee44b18576e84de7b163142b537d2fe6231845
      </hashdigest>
</fileobject>
```


*Developed by Simson Garfinkel

XML Schema for Digital Forensics XML

43 commits | 1 branch | 9 releases | 1 contributor

branch: master | dfxml_schema / +

Document an XML validation step ...

ajnelson authored on Dec 4, 2014 latest commit 4c8aab566e

ref	Allow offline validation with local XSD cache	2 years ago
LICENSE.txt	Add public domain license text	2 years ago
README.md	Document an XML validation step	6 months ago
dfxml.xsd	Document an XML validation step	6 months ago

Code
Issues 8
Pull requests 0
Pulse
Graphs

HTTPS clone URL
https://github.com/

You can clone with HTTPS or Subversion.

Clone in Desktop
Download ZIP

README.md

This is the schema repository for Digital Forensics XML, version 1.1.1.

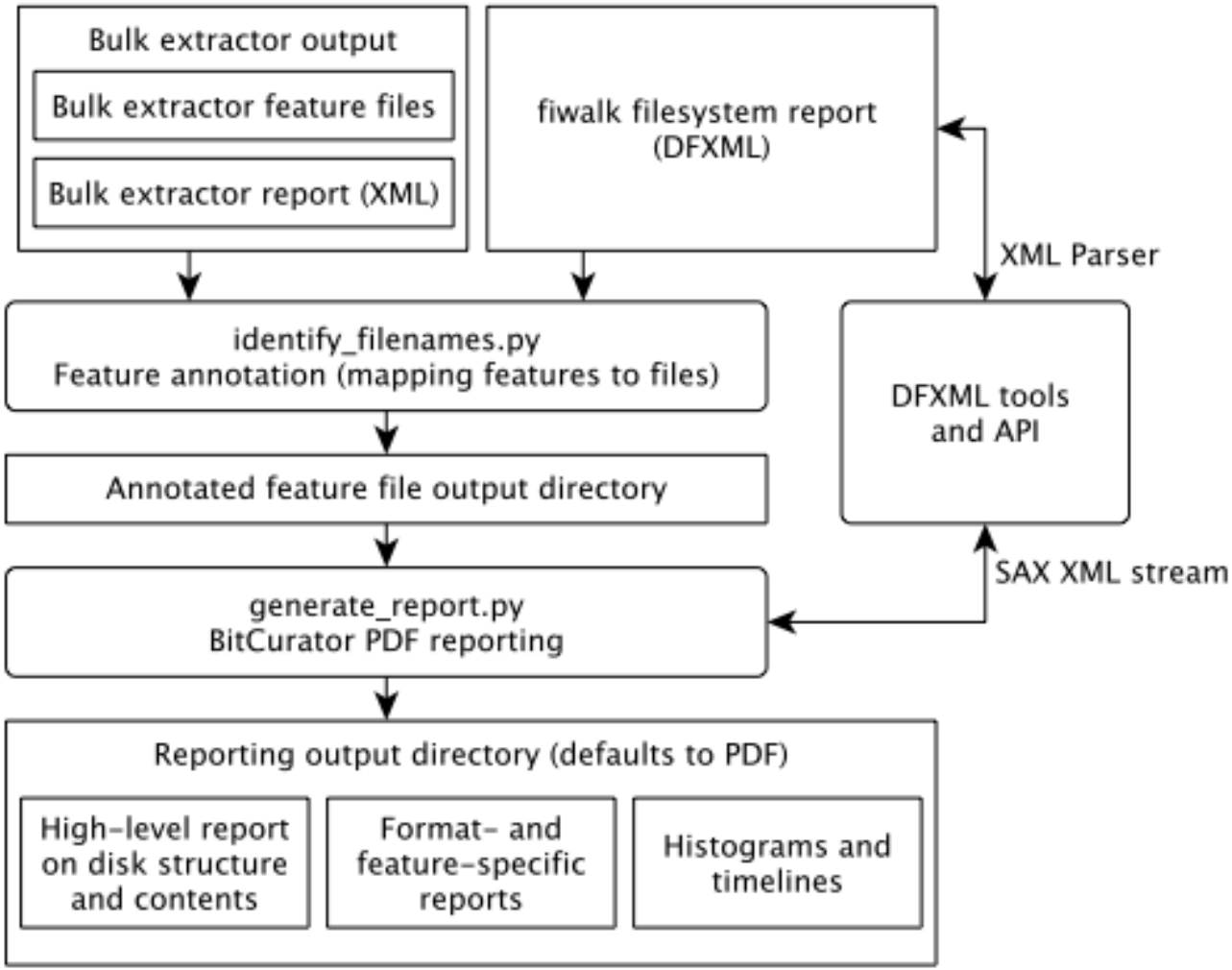
If you intend to use the dfxml.xsd file as a DFXML document validator, note that you will also need to download two accompanying .xsd files under the "ref" directory. The easiest way to do this is by downloading the repository as a Git clone, or by downloading the zip archive from the Github page.

To report issues, questions, or feature requests, please either:

- File a Github issue, seeing first if it is already filed, here.
- Email the dfxml@nist.gov mailing list. If you wish to join the mailing list, send an email to dfxml-subscribe@nist.gov (no subject or message body is necessary), and a moderator will grant access.

https://github.com/dfxml-working-group/dfxml_schema

High-Level view of Metadata Generation and Reporting



See: Woods, Kam, Christopher Lee, and Sunitha Misra. “Automated Analysis and Visualization of Disk Images and File Systems for Preservation.” In *Proceedings of Archiving 2013* (Springfield, VA: Society for Imaging Science and Technology, 2013), 239-244.

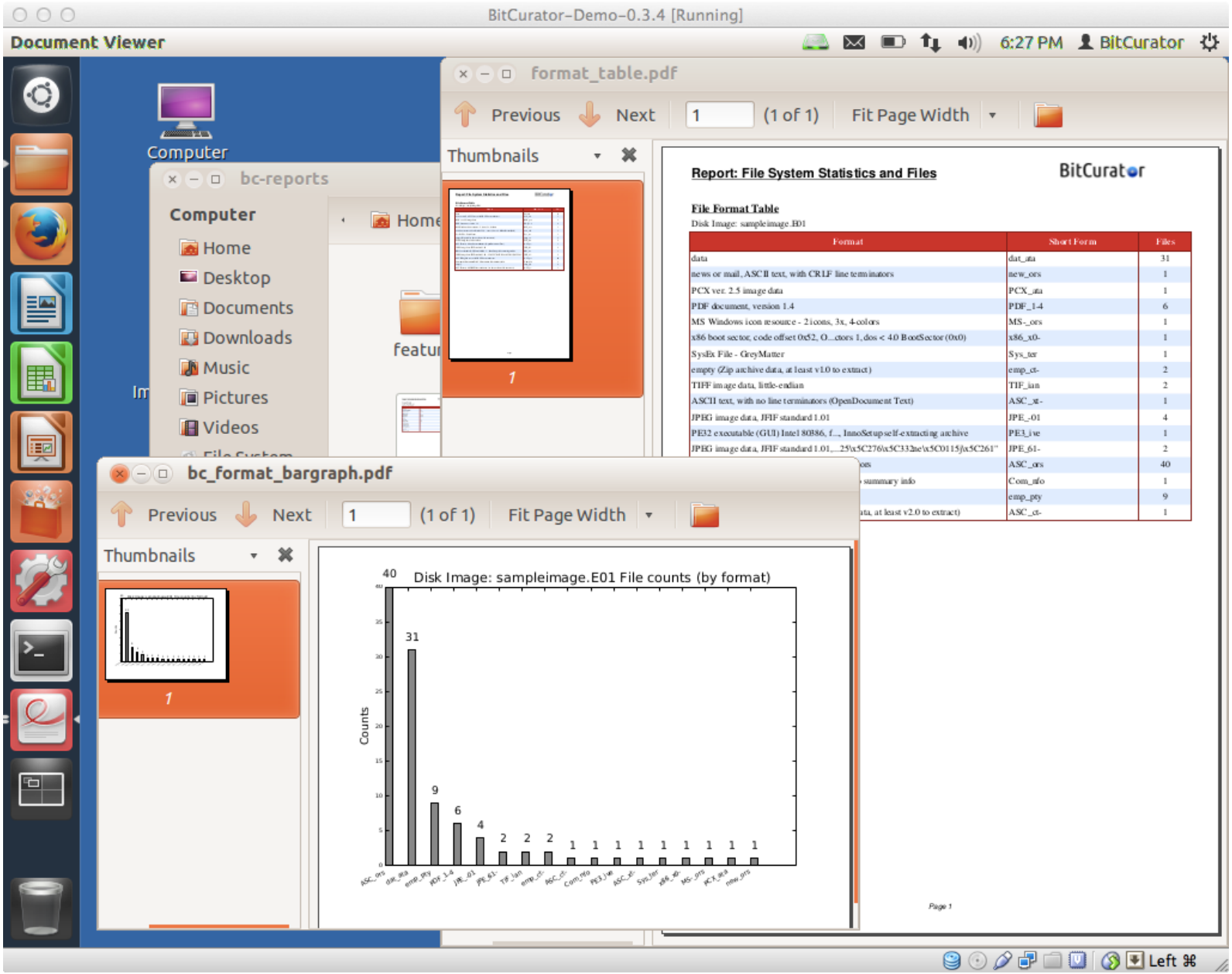
PREMIS (Preservation) Metadata Generated from Running BitCurator Tools – Recorded as PREMIS Events

premis.xml (~/Desktop/demo1/demo1reports/reports) - gedit


```xml
<?xml version="1.0" encoding="UTF-8"?>
<premis xmlns="info:lc/xmlns/premis-v2" version="2.0" xsi="http://www.w3c.org/2001/XMLSchema-instance">
  <object>
    <objectIdentifier>
      <objectIdentifierType>0d4e30d6-b8dc-11e3-a80f-080027f8dfea</objectIdentifierType>
      <objectIdentifierValue>/home/bcadmin/Desktop/terry-work-usb-2009-12-11.E01</objectIdentifierValue>
    </objectIdentifier>
  </object>
  <event>
    <eventIdentifier>
      <eventIdentifierType>0d4ea1ce-b8dc-11e3-a80f-080027f8dfea</eventIdentifierType>
      <eventIdentifierValue>E01/home/bcadmin/Desktop/terry-work-usb-2009-12-11.E01</
eventIdentifierValue>
    </eventIdentifier>
    <eventType>Capture</eventType>
    <eventDateTime>     Wed Jan 19 12</eventDateTime>
    <eventOutcomeInformation>
      <eventOutcome>E01</eventOutcome>
      <eventOutcomeDetail>Version:      20100226
, Image size: 512</eventOutcomeDetail>
    </eventOutcomeInformation>
  </event>
  <event>
    <eventIdentifier>
      <eventIdentifierType>19882604-b8dc-11e3-93f0-080027f8dfea</eventIdentifierType>
      <eventIdentifierValue>bulk_extractor -o /home/bcadmin/Desktop/demo1 /home/bcadmin/Desktop/terry-
work-usb-2009-12-11.E01</eventIdentifierValue>
    </eventIdentifier>
    <eventType>Feature Stream Analysis</eventType>
    <eventDateTime>2014-03-31T13:49:59Z</eventDateTime>
    <eventOutcomeInformation>
      <eventOutcome>Bulk Extractor Output</eventOutcome>
      <eventOutcomeDetail>version: 1.4.4</eventOutcomeDetail>
    </eventOutcomeInformation>
  </event>
</premis>
```


XML ▾ Tab Width: 8 ▾ Ln 1, Col 1 INS

Various Specialized BitCurator Reports



Other Functionality to Meet Identified User Needs:

Function	Tool(s)
Identify duplicate files	FSLint
Characterize files	FITS, FIDO
Scan for viruses	ClamTK
Examine, copy and extract information from old Mac disks	HFS Utilities (including HFS Explorer)
Capture AV file metadata	MediaInfo, FFProbe
Extract text from older binary (.doc) Word files	antiword
Read contents of Mircosoft Outlook PST files	readpst
Examine embedded header information in images	pyExifToolGUI
Generate images of problematic disks or particular disk types	dd, dcfldd, ddrescue, cdrdao (in addition to Guymager)
Extract and analyze data from Windows Registry files	regripper
Identify files that are partially similar but not identical	sdhash, ssdeep
Package files for storage and/or transfer	BagIt (Java) library, Bagger
File preview (left-click on file then hit space bar)	gnome-sushi
Play and examine metadata from AV media files	VLC media player
Damaged/lost partition recovery	TestDisk
Damaged/lost file recovery	PhotoRec
Identify the filesystem on a disk	disktype
Index and search for keywords in documents	recoll
Find blacklist data by using hashes calculated from hash blocks	hashdb
Generate hashes of files and blocks	md5deep (more features than md5sum)

BitCurator Access

- Two-year project (October 1, 2014 – September 30, 2016) at School of Information and Library Science, University of North Carolina at Chapel Hill
- Funded by Andrew W. Mellon Foundation
- Developing open-source software to support access to disk images. Core areas of focus:
  - Tools and reusable libraries to support web access services for disk images
  - Analyzing contents of file systems and associated metadata
  - Redacting complex born-digital objects (disk images)
  - Emulated access to data from disk images

BitCurator Access Team

Cal Lee – Principal investigator

Kam Woods - Technical Lead and Co-PI

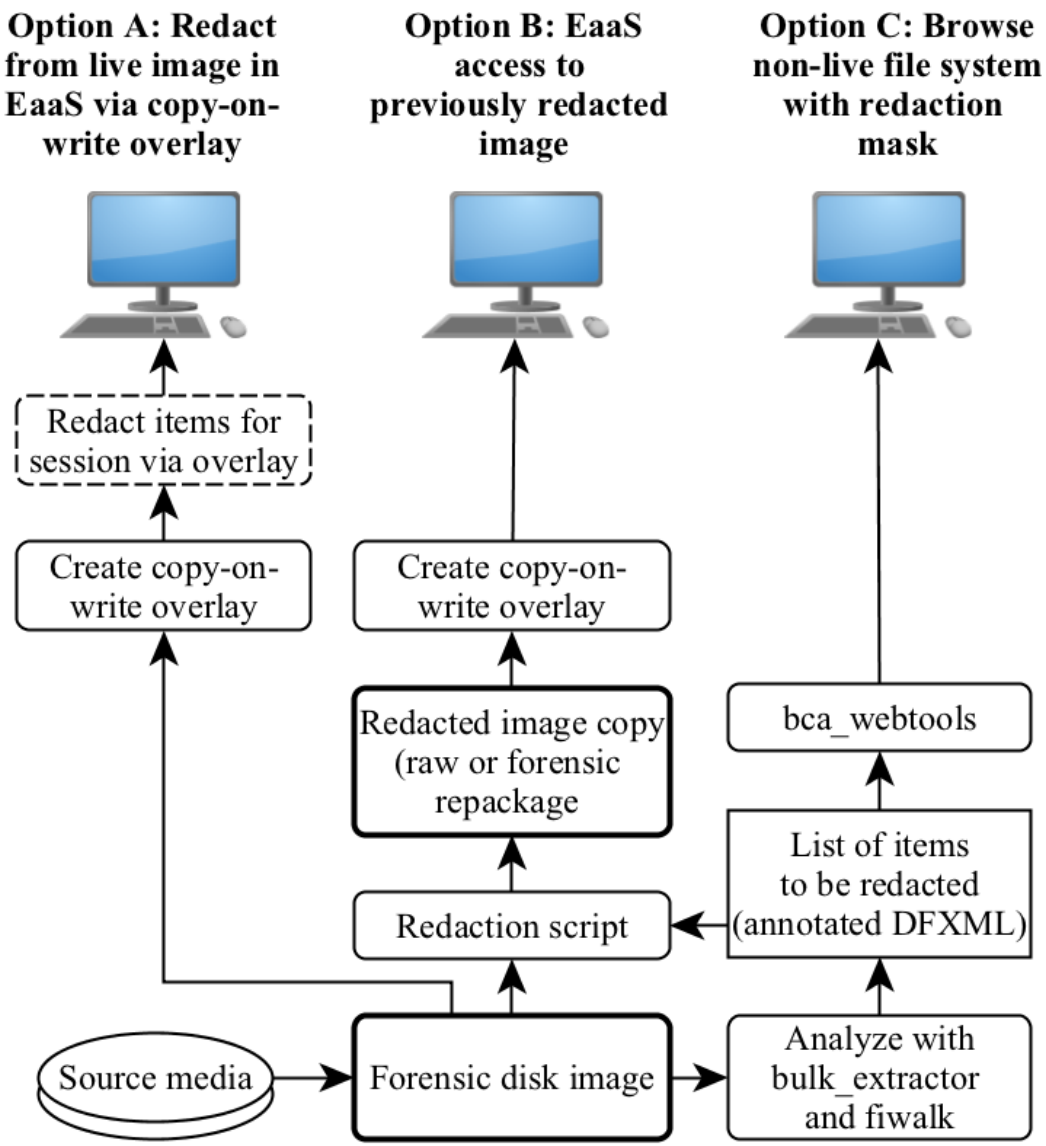
Alex Chassanoff - Project Manager

Sunitha Misra - Software Developer

BitCurator Access Advisory Board

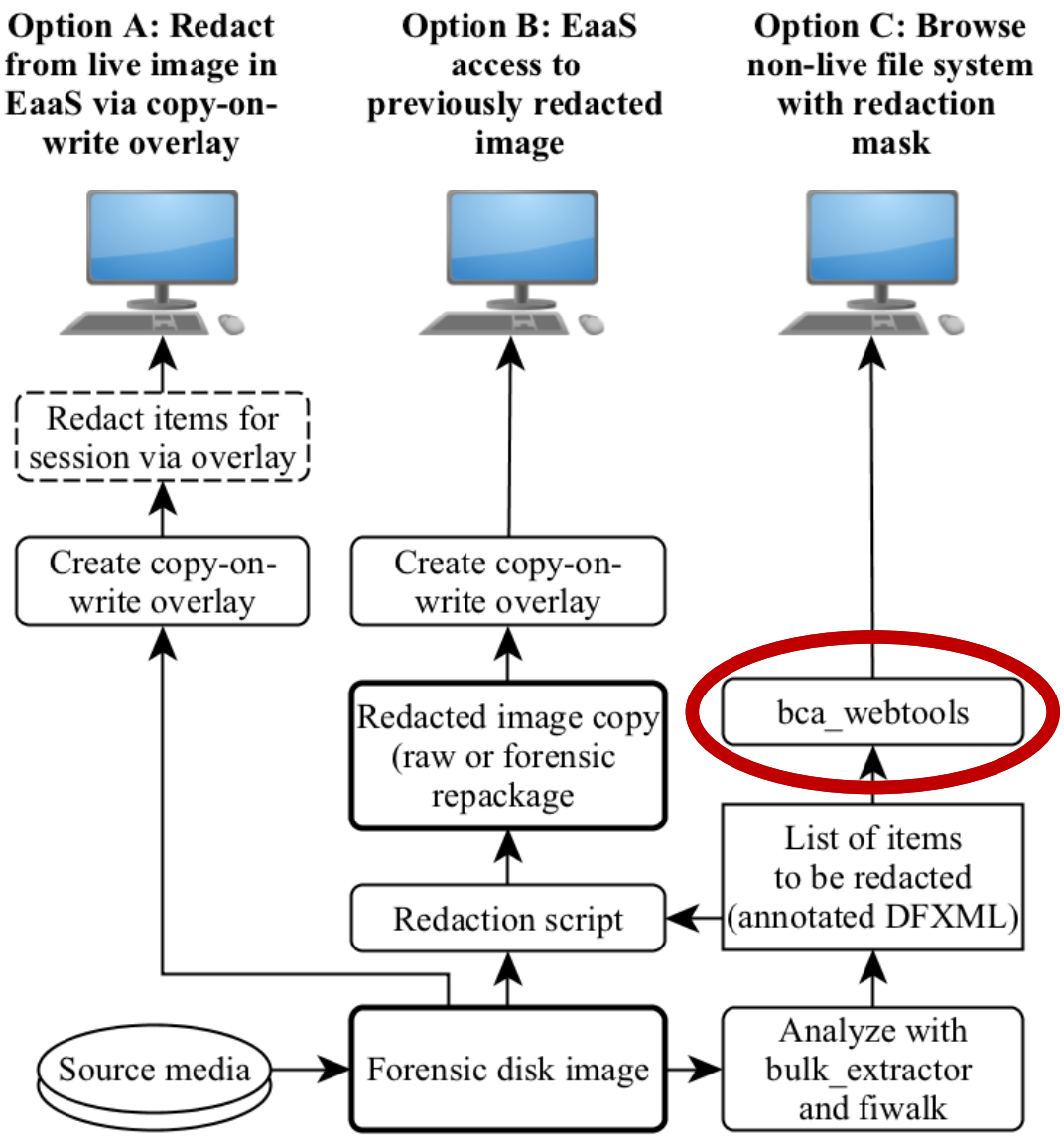
- Geoffrey Brown, Indiana University
- Mark Evans, History Associates
- Erika Farr, Emory University
- Matthew Farrell, Duke University
- Brad Glisson, University of South Alabama
- Matthew Kirschenbaum, Maryland Institute for Technology in the Humanities
- Susan Malsbury, New York Public Library
- Don Mennerich, New York University
- Klaus Rechert, University of Freiburg
- Kari Smith, Massachusetts Institute of Technology
- Bradley Westbrook, ArchivesSpace
- Doug White, National Institute of Standards and Technology
- Carl Wilson, Open Planets Foundation

Automated Redaction and Access Options



EaaS = Emulation-as-a-Service. http://bw-fla.uni-freiburg.de/

Automated Redaction and Access Options



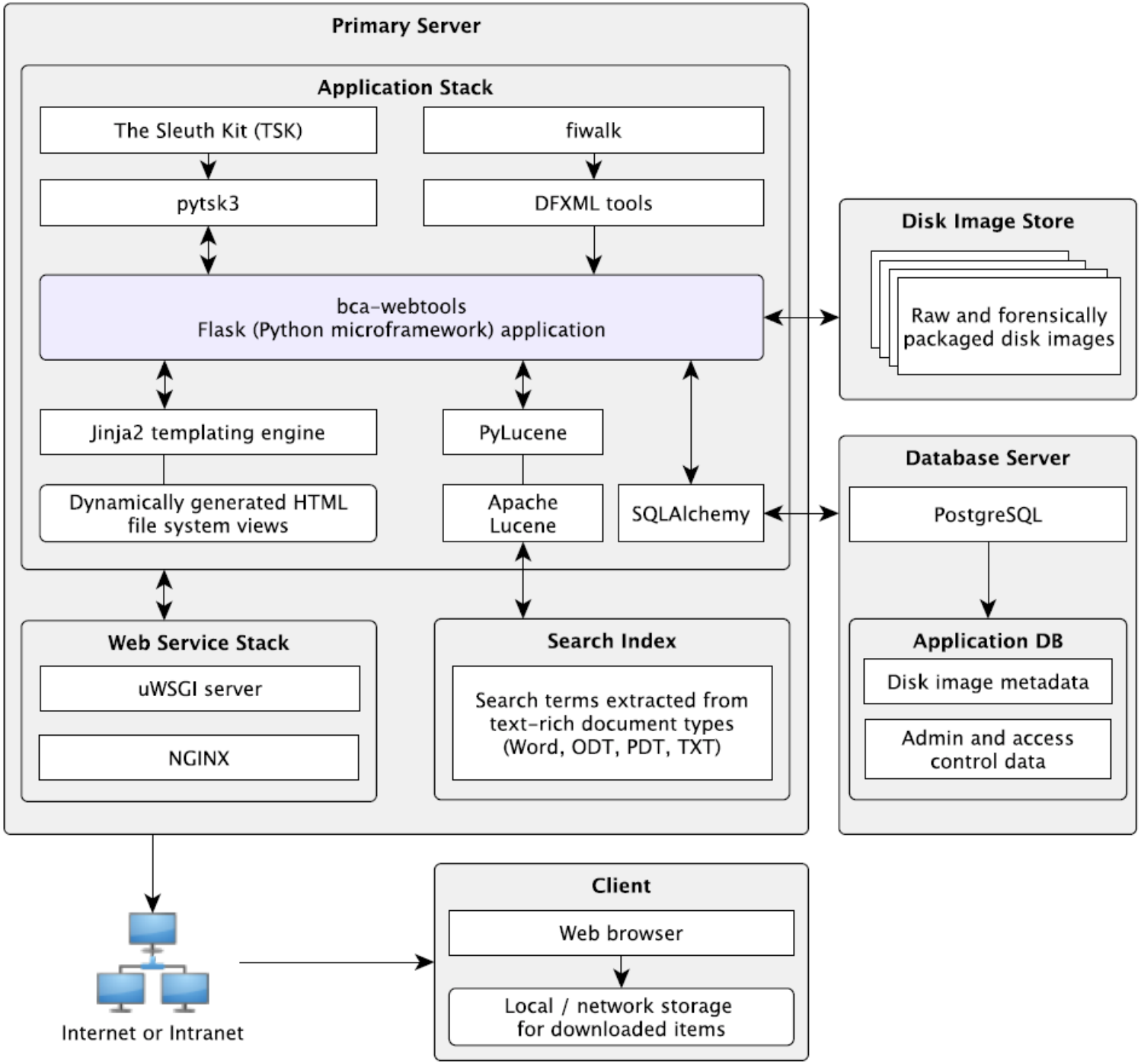
EaaS = Emulation-as-a-Service. http://bw-fla.uni-freiburg.de/

BCA (BitCurator Access) Web Tools

- Integrates digital forensics software libraries and lightweight web-services tools
- Drop disk images in a local or network-accessible location, start up the service, and start browsing
- Most analysis runs server-side (via Sleuthkit and DFXML Python bindings, among others)
- Service is database-agnostic (we're using postgres)
- Automatic metadata production – Digital Forensics XML (DFXML), PREMIS, others)

https://github.com/kamwoods/bca-webtools

Sunitha Misra, Christopher A. Lee, and Kam Woods, "A Web Service for File-Level Access to Disk Images," *Code4Lib Journal 25* (2014), http://journal.code4lib.org/articles/9773


Primary Server
Application Stack
The Sleuth Kit (TSK)
fiwalk
pytsk3
DFXML tools
bca-webtools
Flask (Python microframework) application
Jinja2 templating engine
PyLucene
Dynamically generated HTML file system views
Apache Lucene
SQLAlchemy
Web Service Stack
uWSGI server
NGINX
Search Index
Search terms extracted from text-rich document types (Word, ODT, PDT, TXT)
Disk Image Store
Raw and forensically packaged disk images
Database Server
PostgreSQL
Application DB
Disk image metadata
Admin and access control data
Internet or Intranet
Client
Web browser
Local / network storage for downloaded items

Disk Image Access for the

127.0.0.1:8080

BCA Webtools

Home

The bca-webtools application provides access to forensically-packaged and raw disk images. Supported file systems include FAT16, FAT32, NTFS, HFS+, and EXT2/3/4. Click on 'Browse' to navigate through the file system(s) within the disk image, or 'Download' to download the complete disk image.

Image Name	Info	Browse	Download
charlie-work-usb-2009-12-11.E01			
terry-work-usb-2009-12-11.E01			

Select an option below to search available disk images by filename or file contents. (Currently indexing all filenames, contents of .doc, .odt, .pdf, and .txt)

- Search by filename
- Search by content

Search

- Admin

Disk Image Access for the

127.0.0.1:8080/admin

BCA Webtools

Home

bca-webtools - Admin Tools

- Build Image Table
- Build DFXML Table
- Build All Tables
- Drop Image Table
- Drop DFXML Table
- Drop All Tables
- Generate Index
- Clear Index
- Show Image Matrix

Submit

Image Matrix

Index	Image name	Image DB?	DFXML DB?	Indexed?	Add Table	Delete Table
0	charlie-work-usb-2009-12-11.E01	True	False	True	Add	Delete
1	terry-work-usb-2009-12-11.E01	True	False	True	Add	Delete

- Admin

Disk Image Access for the

127.0.0.1:8080/image/charlie-work-usb-2009-12-11.E01/1

BCA Webtools

Home

Browse directories and download files. Items marked "r" in the first column are regular files. Items marked "d" are directories.

d/r	Filename	Size	Last Modified	Deleted?
r	$AttrDef	2560	2009-11-20T17:38:09Z	No
r	$BadClus	0	2009-11-20T17:38:09Z	No
r	$Bitmap	32320	2009-11-20T17:38:09Z	No
r	$Boot	8192	2009-11-20T17:38:09Z	No
d	$Extend	552	2009-11-20T17:38:09Z	No
r	$LogFile	7405568	2009-11-20T17:38:09Z	No
r	$MFT	262144	2009-11-20T17:38:09Z	No
r	$MFTMirr	4096	2009-11-20T17:38:09Z	No
r	$Secure	0	2009-11-20T17:38:09Z	No
r	$UpCase	131072	2009-11-20T17:38:09Z	No
r	$Volume	0	2009-11-20T17:38:09Z	No
d	.	56	2009-12-03T21:17:01Z	No
r	01.zip	108438	2009-11-24T21:21:16Z	No
r	astronaut.jpg	713418	2009-11-24T21:33:33Z	No
r	astronaut1.jpg	722717	2009-11-24T21:43:42Z	No
d	Email	56	2009-12-10T22:27:55Z	No
d	Immortality	56	2009-11-24T21:55:45Z	No
r	invsecr2.exe	1291720	2009-11-19T18:42:25Z	No
r	microscope.jpg	136274	2009-11-24T21:27:51Z	No

Upcoming Events

CurateGear: Enabling the Curation of Digital Collections

January 14, 2016 – Friday Center, Chapel Hill, North Carolina

An interactive day-long event focused on digital curation tools and methods. See demonstrations, hear about the latest developments, and discuss application in professional contexts.

http://ils.unc.edu/digccurr/curategear2016.html

BUF 2016

Event Information

Registration

Not a member?

Much of the content on BitCuratorConsortium.org is accessible to members only. Learn more about the benefits of joining the BCC.

BitCurator User Forum 2016

Join BitCurator users from around the globe as we discuss how we are using the BitCurator software environment. Hosted by the BitCurator Consortium (BCC), this event will be grounded in the practical, real-world experiences of digital archivists and digital curation experts. Come prepared to discuss your current challenges, share emerging BitCurator integrations and workflows, and address the "now what" of handling your digital forensics outputs.

Date: 15 January 2016
Location: Pleasants Family Assembly Room, Wilson Library, University of North Carolina, Chapel Hill, North Carolina

Registration
General Registration - $30
Student Registration - $15
BitCurator Consortium Member Registration - Free

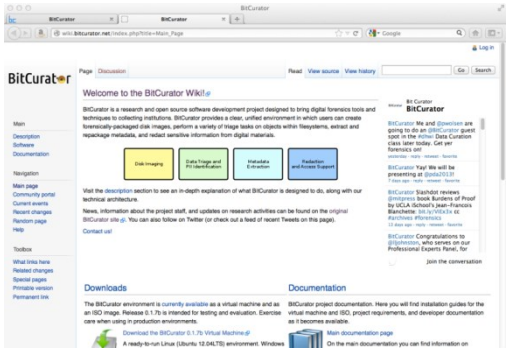
Register Now

Program
(details about speakers coming soon)

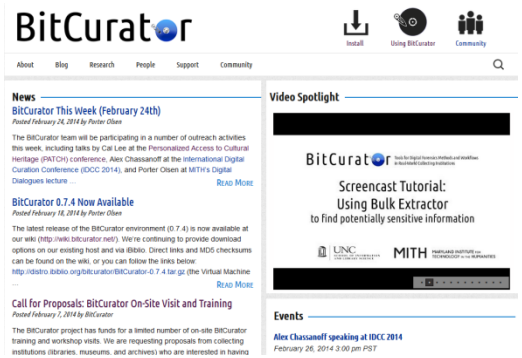
8:00 - 8:30am	**Registration and Coffee**
8:30 - 9:00am	**Welcome and Introductions**
9:00 - 10:15am	**Panel: Beyond Floppy Disks** Cultural heritage institutions are collecting a larger variety of media than ever before including external hard drives, computers,

https://bitcuratorconsortium.org/events/buf2016

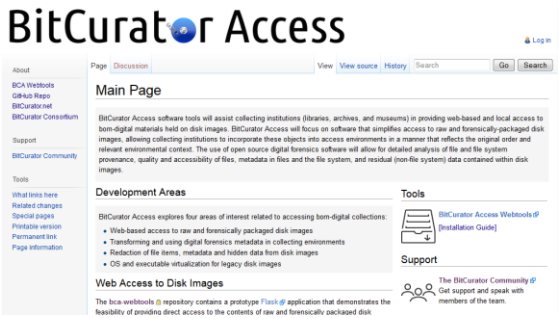
BitCurator, BitCurator Consortium and BitCurator Access Resources



Get the software
Documentation and technical specifications
Screencasts
Google Group
http://wiki.bitcurator.net/



People
Project overview
Publications
News
http://www.bitcurator.net/



BitCurator Access Project and Products
http://access.bitcurator.net/

Twitter: @bitcurator