Up Close and Personal: Individual Digital Traces as Cultural Heritage and Discovery through Forensics Tools

Cal Lee
School of Information and Library Science
University of North Carolina, Chapel Hill

24 February 2014
Personalized Access to Cultural Heritage (PATCH)
Haifa, Israel

BitCurator

UNC
SCHOOL OF INFORMATION AND LIBRARY SCIENCE

Personal Traces

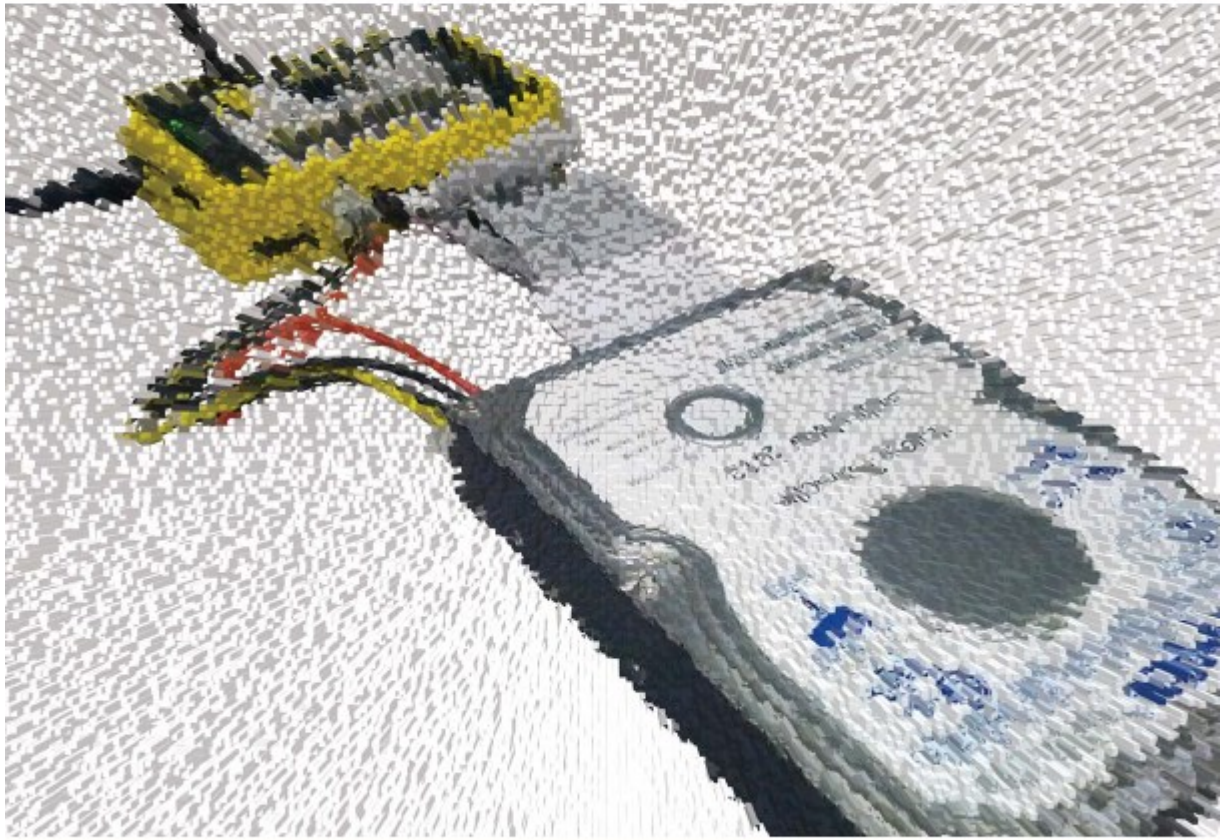
- Documentary traces of individuals (personal traces) have long been recognized and preserved as fundamental components of cultural heritage
- The nature of personal documentary traces has undergone dramatic evolution in recent years, including various aspects of one's "digital footprint."

Applying Forensics to Personal Traces

- Cultural institutions (libraries, archives, museums) have begun applying digital forensics to:
  - create authentic copies of data on disks
  - reflect the original order of materials
  - establish more trustworthy chains of custody
  - discover and expose associated contextual information
  - identify sensitive information that should be filtered, redacted or masked in appropriate ways.
- Many of the same approaches can be adapted and applied by individuals and families who are managing their own collections of personal traces.

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

BitCurator

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

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 archives/library ingest and collection management environments
  - provision of public access to the data

Core BitCurator Team

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

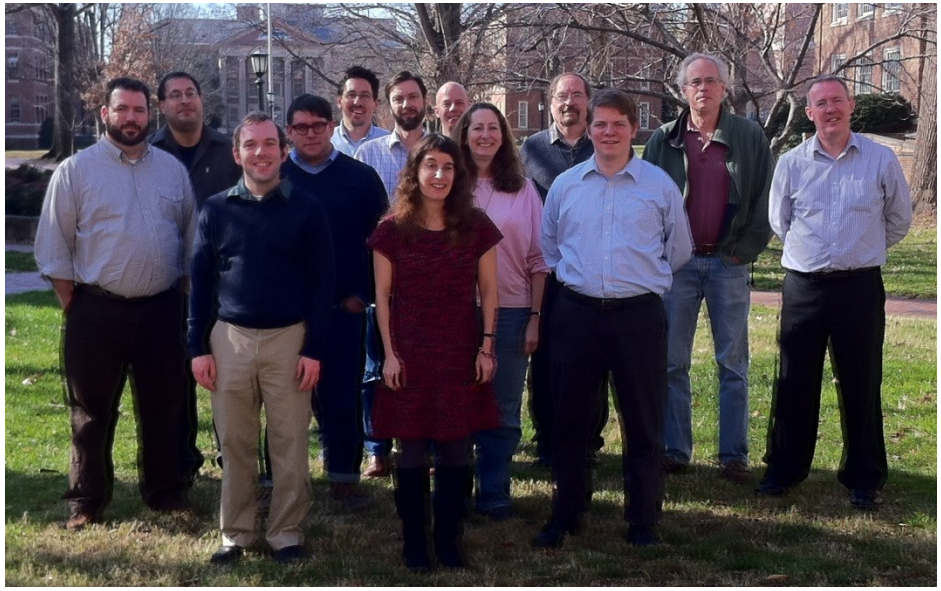






Two Groups of Advisors

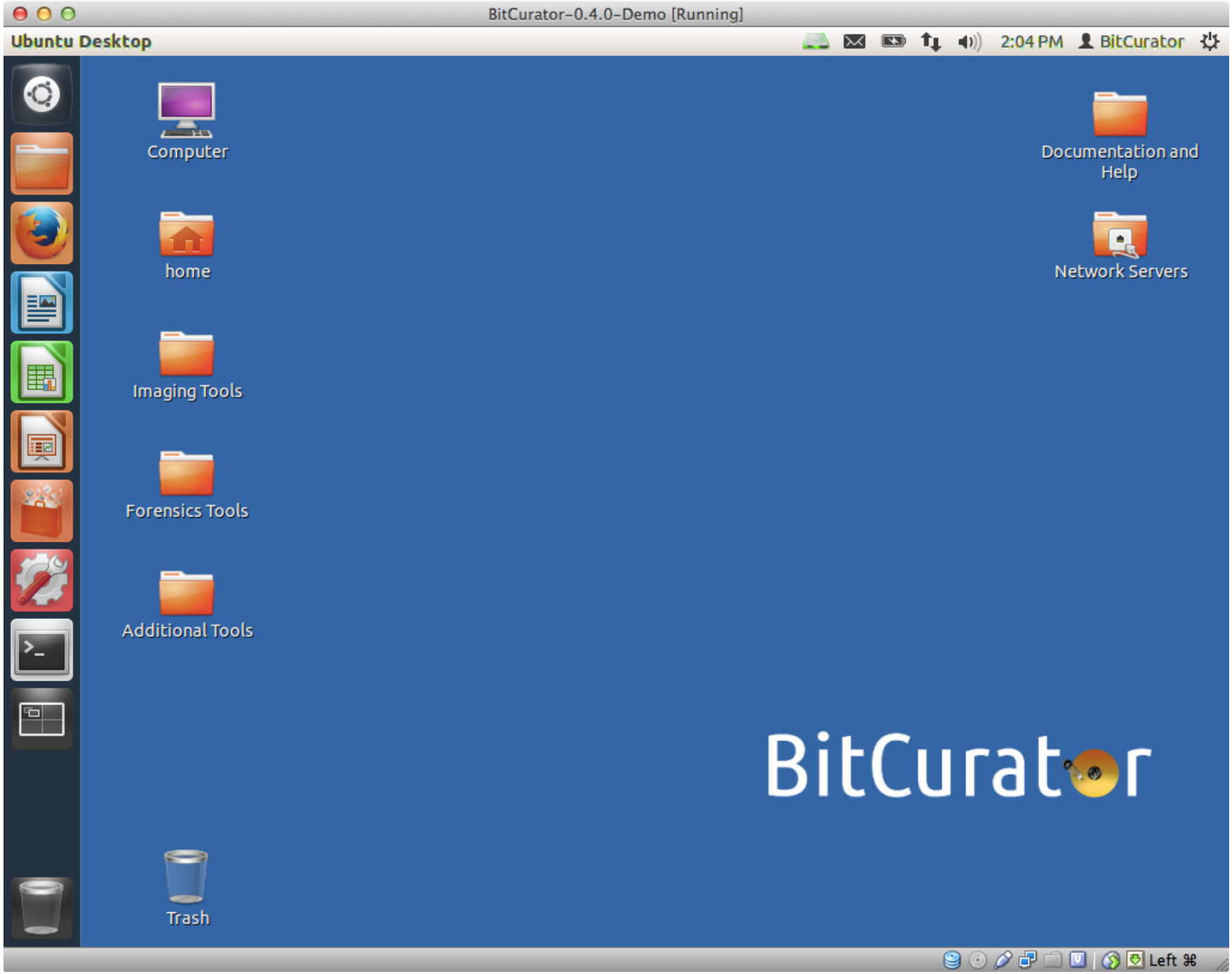
Professional Experts Panel	Development Advisory Group
• Bradley Daigle, University of Virginia Library • Erika Farr, Emory University • Jennie Levine Knies, University of Maryland • Jeremy Leighton John, British Library • Leslie Johnston, Library of Congress • Naomi Nelson, Duke University • Erin O'Meara,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



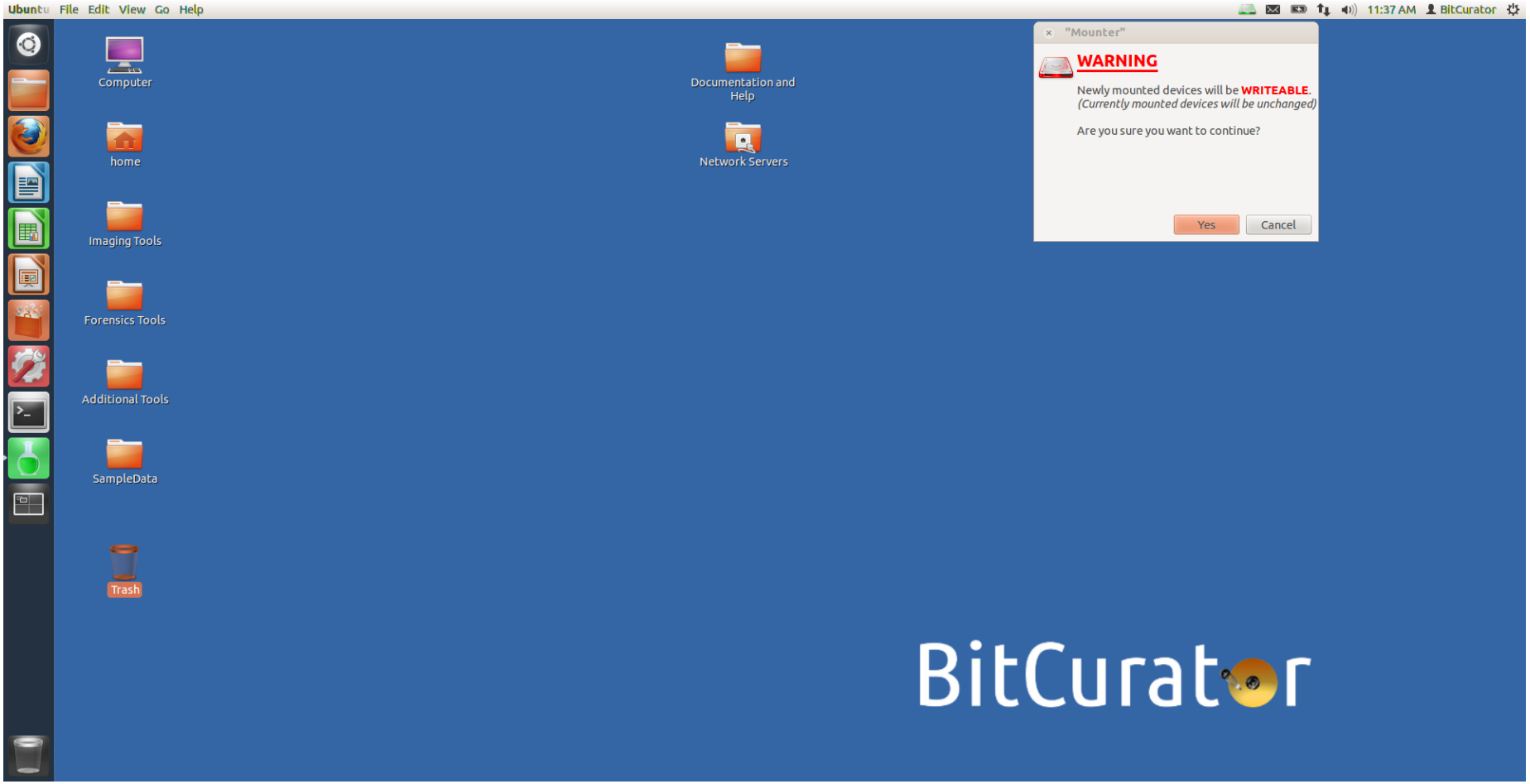
BitCurator Environment*

- Bundles, integrates and extends functionality (primarily data capture and reporting) of open source software: fiwalk, bulk extractor, Guymager, The Sleuth Kit, sdhash and others
- 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. VirtualBox or VMWare
  - As individual components run directly in your own Linux environment or (whenever possible) Windows environment

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


BitCurator-0.4.0-Demo [Running]
Ubuntu Desktop
2:04 PM
BitCurator
Computer
home
Imaging Tools
Forensics Tools
Additional Tools
Trash
Documentation and Help
Network Servers
BitCurator
Left ⌘

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



*Not intended to replace use of hardware-based write blockers

Acquiring Disk Images with Guymager

BitCurator-0.3.0 [Running]

Guymager

1:15 AM BitCurator

GUYMAG

Devices Misc Help

Rescan

Serial nr.	Progress
VB20bfff24-3f8cb5fa	
	2%

Size
Sector size
Image file
Info file
Current speed
Started
Hash calculation
Source verification
Image verification

Acquire image of /dev/sdb

File format

- Linux dd raw image (file extension .dd or .xxx)
- Expert Witness Format, sub-format Guymager (file extension .Exx)
- Advanced forensic image (file extension .aff)

Split image files

Split size 2047 MiB

Case number 1

Evidence number 1

Examiner BitCurator test user

Description A sample floppy disk image

Notes Additional notes go here

Destination

Image directory ... /home/bcadmin/Desktop/SampleData/

Image filename (without extension) sampleimage

Info filename (without extension) sampleimage

Hash calculation / verification

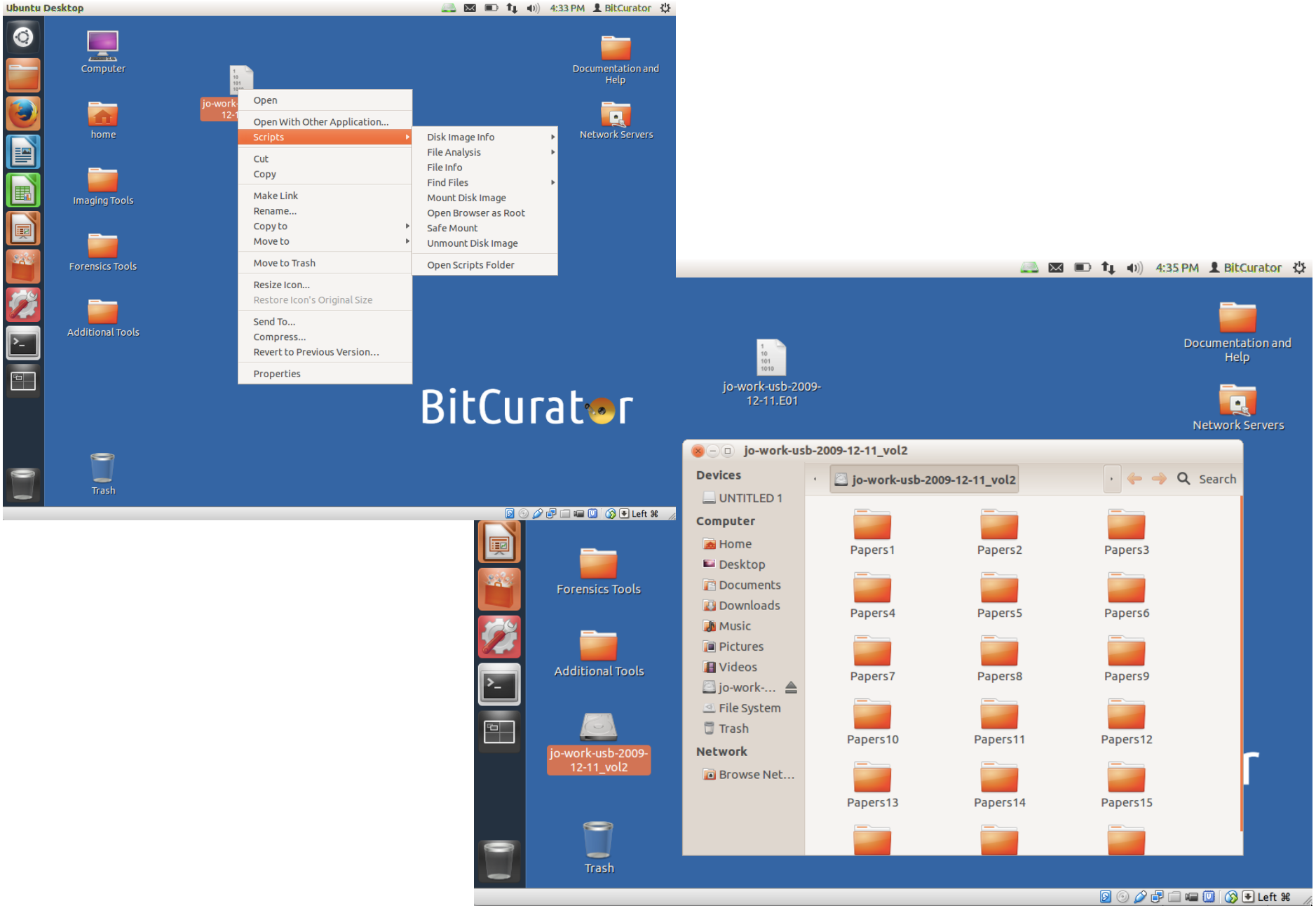
- [x] Calculate MD5
- [ ] Calculate SHA-1
- [ ] Calculate SHA-256
- [ ] Re-read source after acquisition for verification (takes twice as long)
- [x] Verify image after acquisition (takes twice as long)

Cancel | Duplicate image... | Start

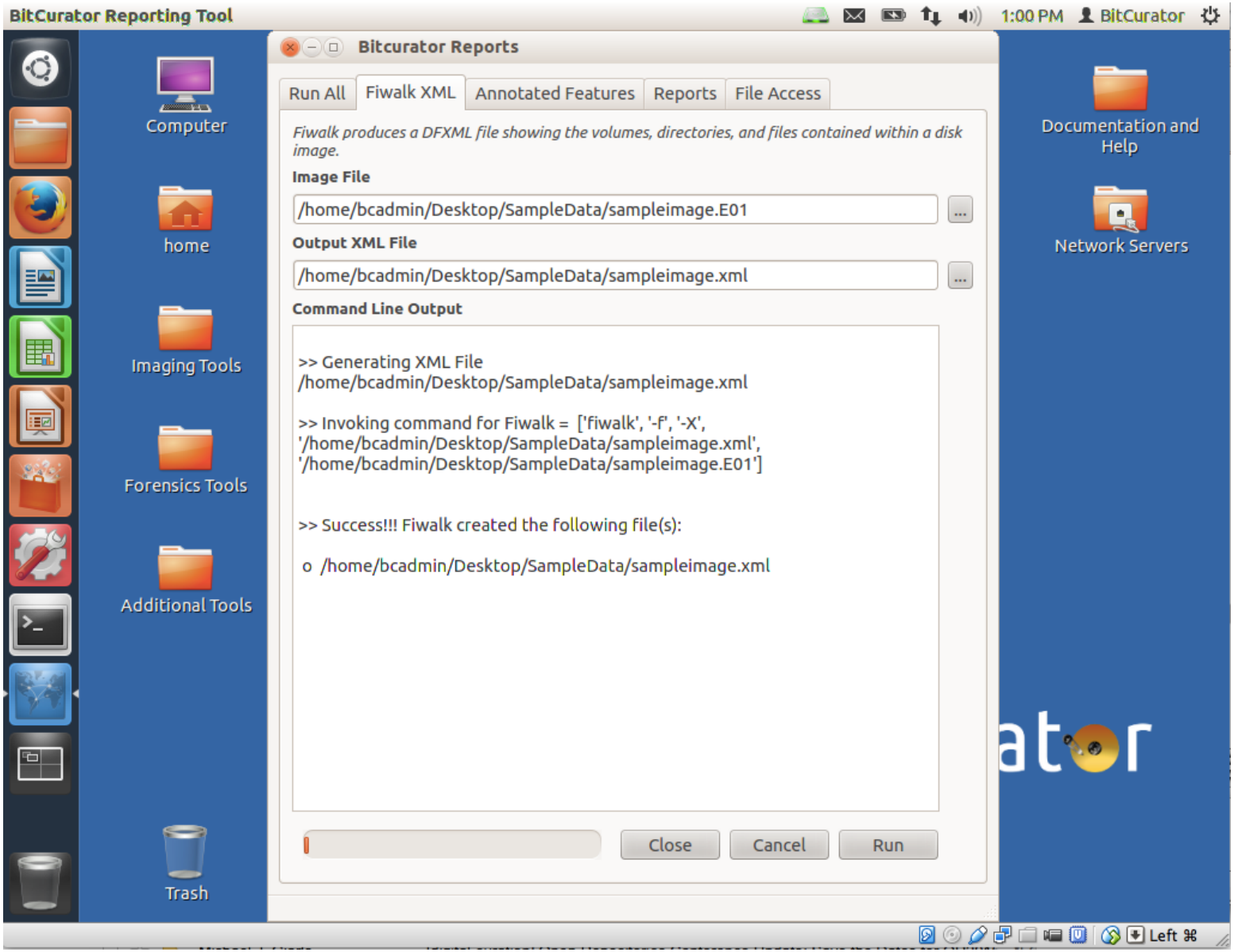
Trash

Left ⌘

Mounting a Forensically Packaged Disk Image



Exporting Filesystem Content Using fiwalk

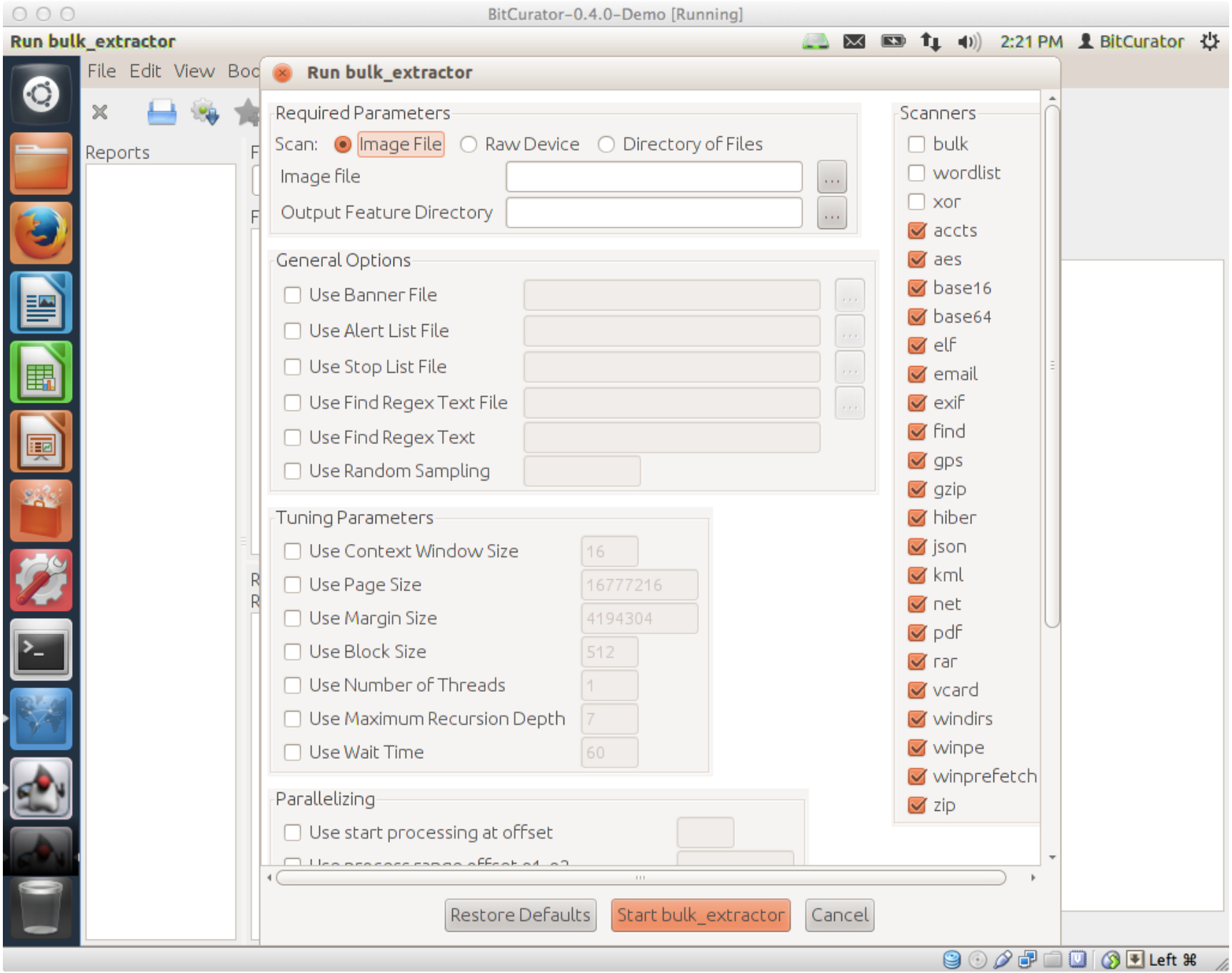


Fiwalk Output for a Specific File

```
<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>
```

Identifying “Features” of Interest in Disk Images or Directories

Bulk Extractor


BitCurator-0.4.0-Demo [Running]
Run bulk_extractor
2:21 PM
BitCurator
File Edit View Boo
Reports
Run bulk_extractor
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
1
Use Maximum Recursion Depth
7
Use Wait Time
60
Parallelizing
Use start processing at offset
Scanners
bulk
wordlist
xor
accts
aes
base16
base64
elf
email
exif
find
gps
gzip
hiber
json
kml
net
pdf
rar
vcard
windirs
winpe
winprefetch
zip
Restore Defaults
Start bulk_extractor
Cancel
Left ⌘

BitCurator-0.4.0-Demo [Running]

Run bulk_extractor

2:21 PM BitCurator

File Edit View Boo

Reports

Run bulk_extractor

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 1

Use Maximum Recursion Depth 7

Use Wait Time 60

Parallelizing

Use start processing at offset

Scanners

- bulk
- wordlist
- xor
- accts
- aes
- base16
- base64
- elf
- email
- exif
- find
- gps
- gzip
- hiber
- json
- kml
- net
- pdf
- rar
- vcard
- windirs
- winpe
- winprefetch
- zip

Left ⌘

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

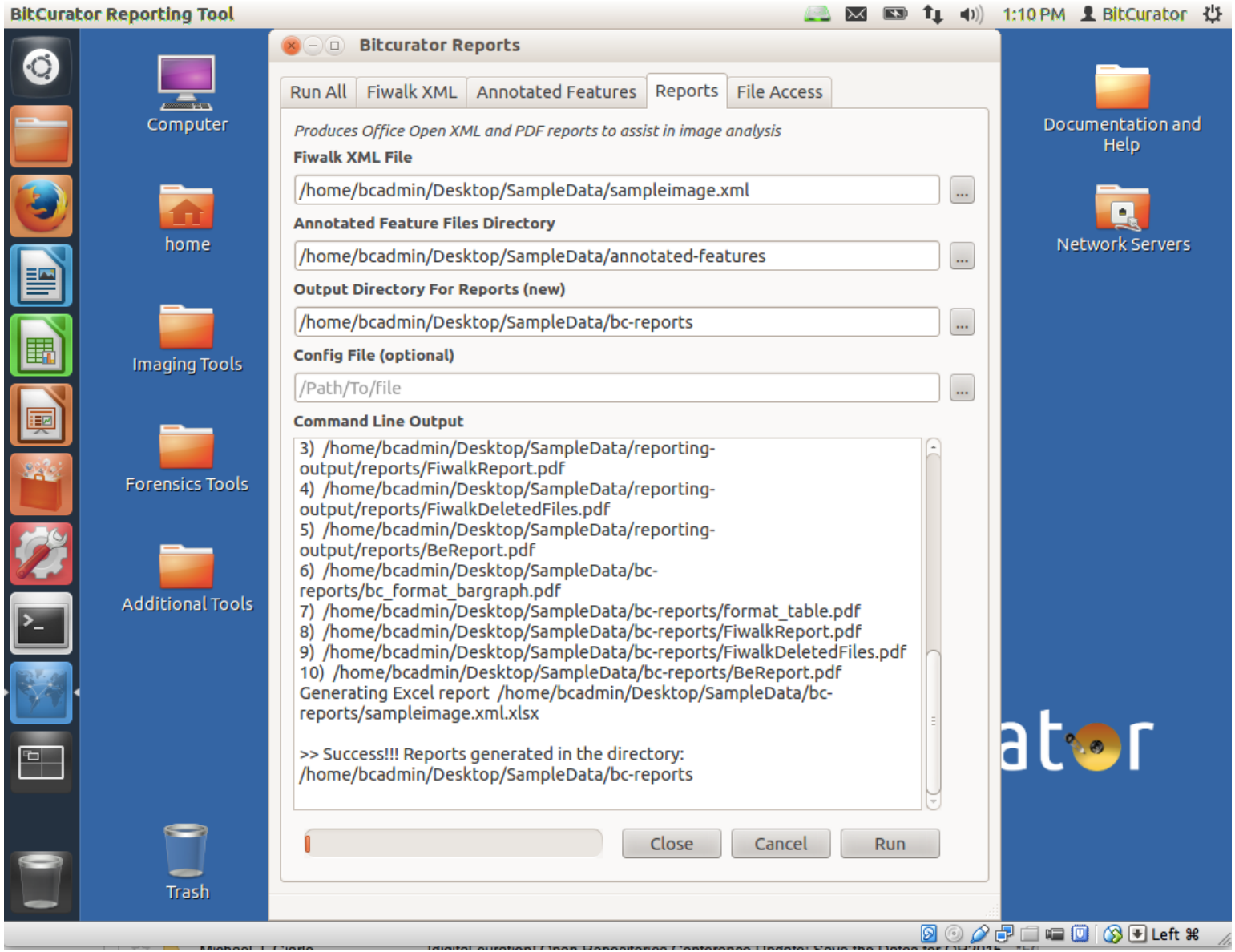
Image

```
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 ⌘

Generating BitCurator Reports



BitCurator-Demo-0.3.4 [Running]

Document Viewer

6:27 PM BitCurator

Computer

bc-reports

Computer
Home
Desktop
Documents
Downloads
Music
Pictures
Videos

format_table.pdf

Previous Next 1 (1 of 1) Fit Page Width

Thumbnails

Report: File System Statistics and Files

BitCurator

File Format Table

Disk Image: sampleimage.E01

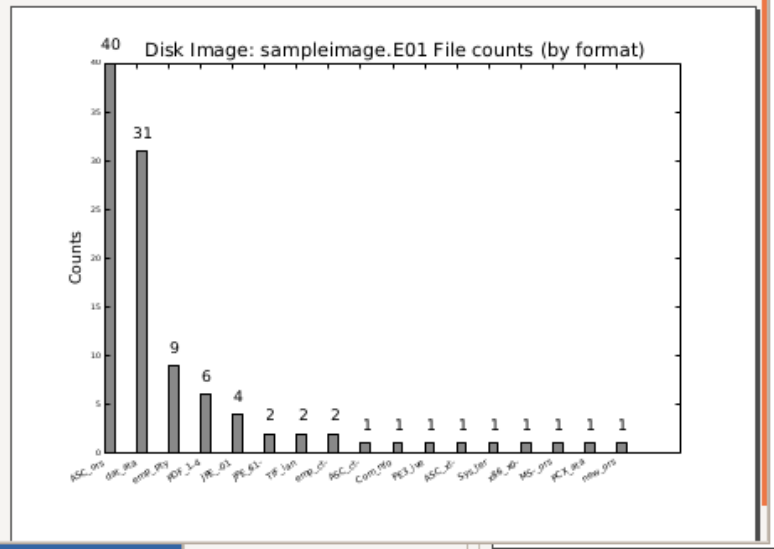
Format	Short Form	Files
data	dat_ata	31
news or mail, ASCII text, with CRLF line terminators	new_ors	1
PCX ver. 2.5 image data	PCX_ata	1
PDF document, version 1.4	PDF_1-4	6
MS Windows icon resource - 2 icons, 3x, 4-colors	MS-_ors	1
x86 boot sector, code offset 0x52, O...ctors 1, dos < 4.0 BootSector (0x0)	x86_x0-	1
SysEx File - GreyMatter	Sys_ter	1
empty (Zip archive data, at least v1.0 to extract)	emp_ct-	2
TIFF image data, little-endian	TIF_ian	2
ASCII text, with no line terminators (OpenDocument Text)	ASC_xt-	1
JPEG image data, JFIF standard 1.01	JPE_-01	4
PE32 executable (GUI) Intel 80386, f..., InnoSetup self-extracting archive	PE3_ive	1
JPEG image data, JFIF standard 1.01,...25\x5C276\x5C332ne\x5C0115j\x5C261"	JPE_61-	2
ors	ASC_ors	40
summary info	Com_nfo	1
	emp_pty	9
ata, at least v2.0 to extract)	ASC_ct-	1

Page 1

bc_format_bargraph.pdf

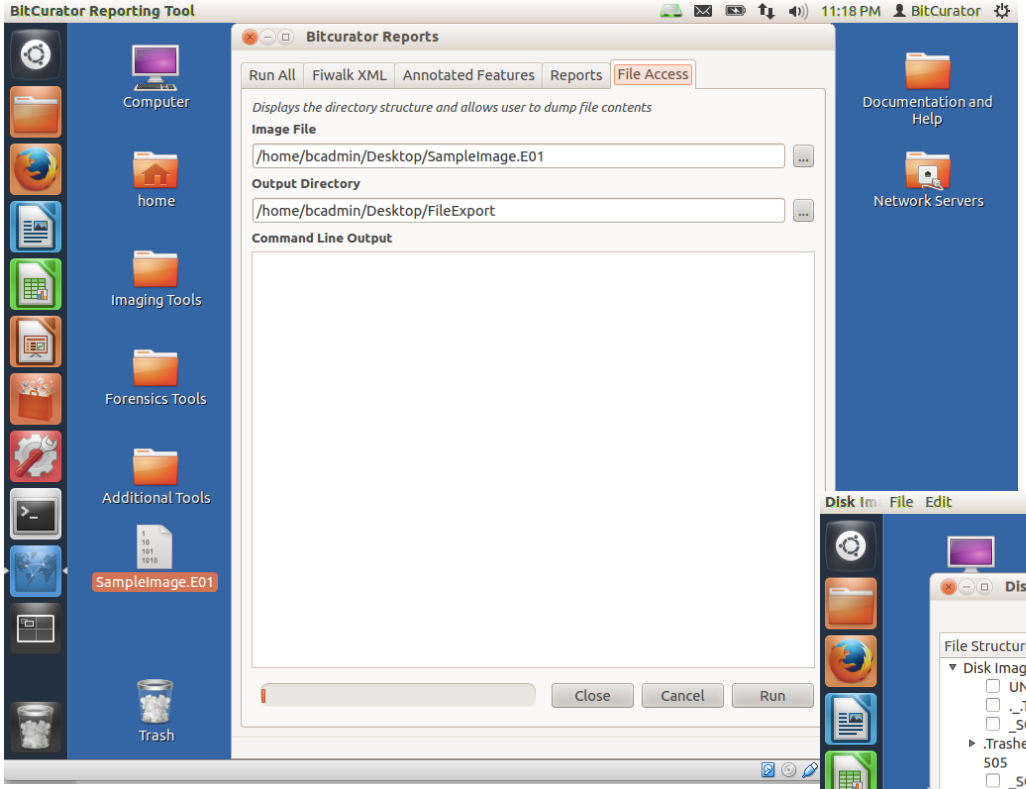
Previous Next 1 (1 of 1) Fit Page Width

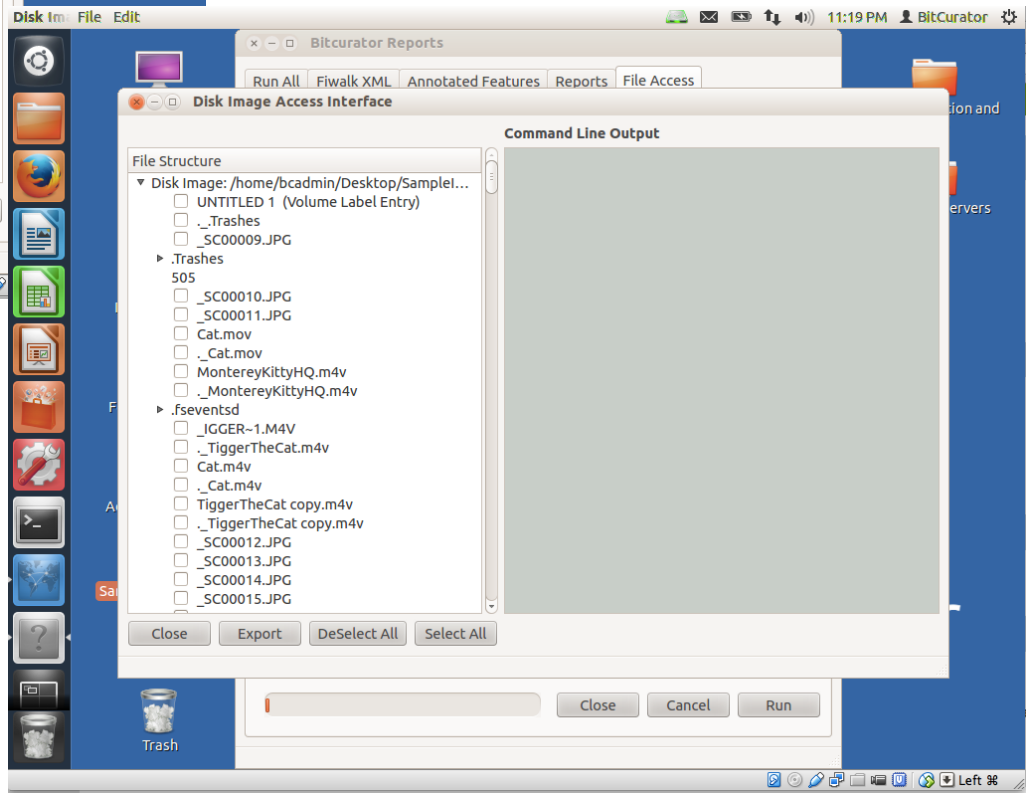
Thumbnails



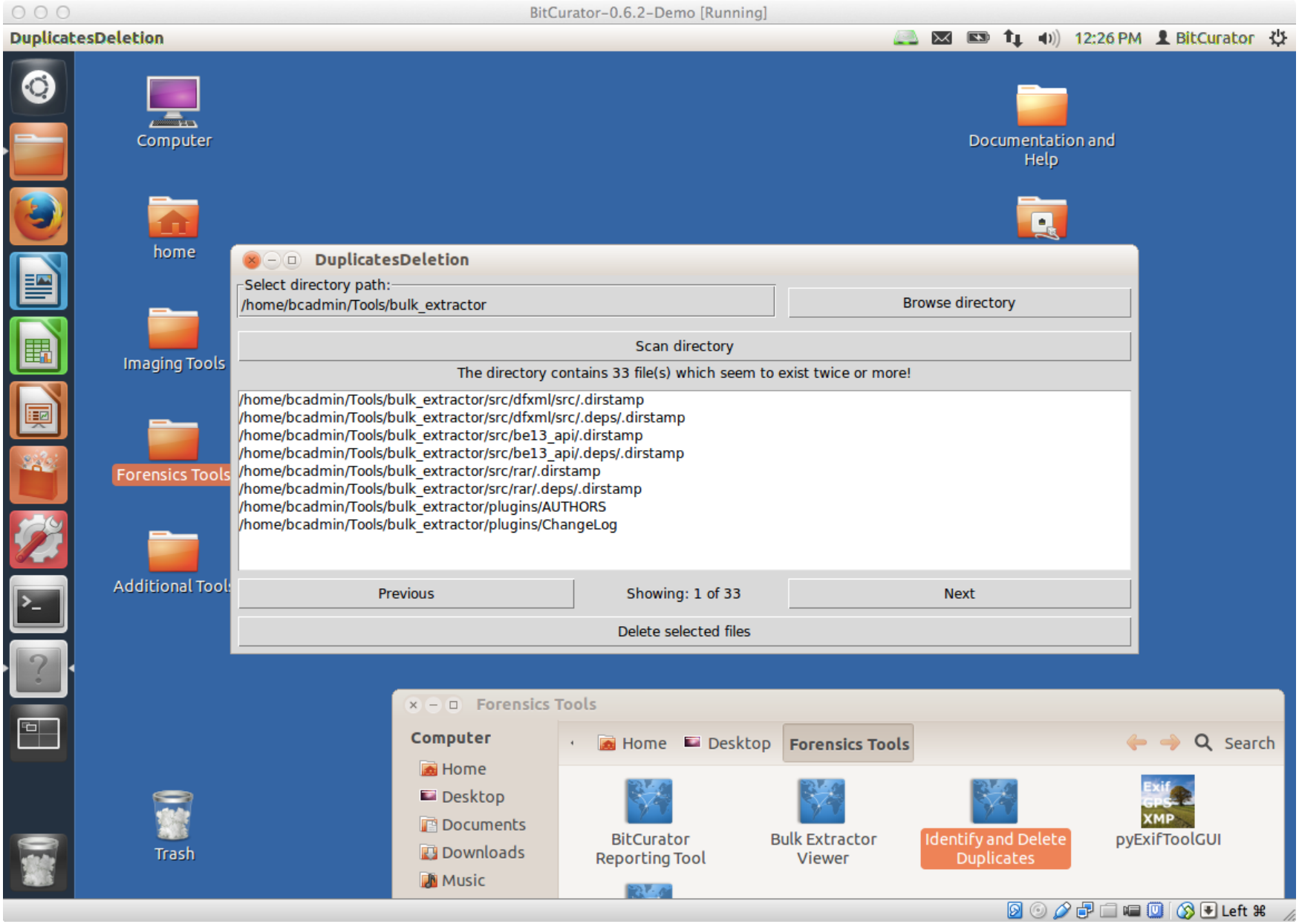
Left ⌘

Exporting Files from a Disk Image





Identifying (and Possibly Deleting) Duplicate Files



Viewing EXIF Metadata with pyExifToolGUI

BitCurator-0.6.2-Demo [Running]

pyExifToolGUI 12:17 PM BitCurator

pyExifToolGui 0.4.0.2

File Extra Help

	Thumb	file name
1		media.tiff
2		PANO_20130221_170748.tiff

Load Images Display Image

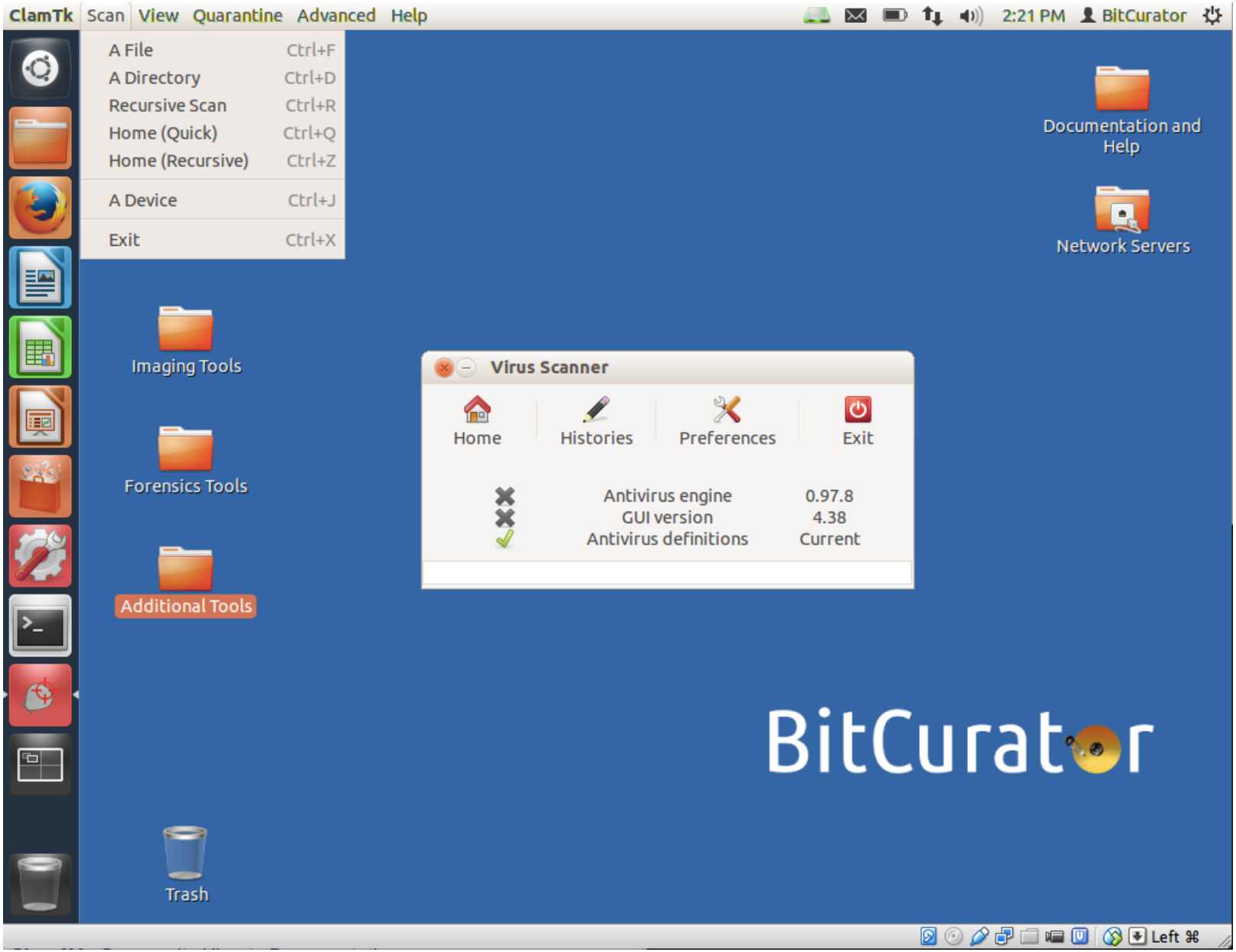
View Data | Edit Data | Your Commands | Preferences

All Exif xmp IPTC GPS/Location GPano ICC_profile Makernotes

	Descriptor	Description
1	Subfile Type	Full-resolution Image
2	Image Width	821
3	Image Height	650
4	Bits Per Sample	8 8 8
5	Compression	LZW
6	Photometric Interpretation	RGB
7	Document Name	/Users/kamwoods/Downloads/media.tiff
8	Strip Offsets	(Binary data 55 bytes, use -b option to extract)
9	Orientation	Horizontal (normal)
10	Samples Per Pixel	3
11	Rows Per Strip	64
12	Strip Byte Counts	(Binary data 51 bytes, use -b option to extract)
13	X Resolution	72
14	Y Resolution	72
15	Planar Configuration	Chunky
16	Resolution Unit	inches
17	Predictor	Horizontal differencing

Left ⌘

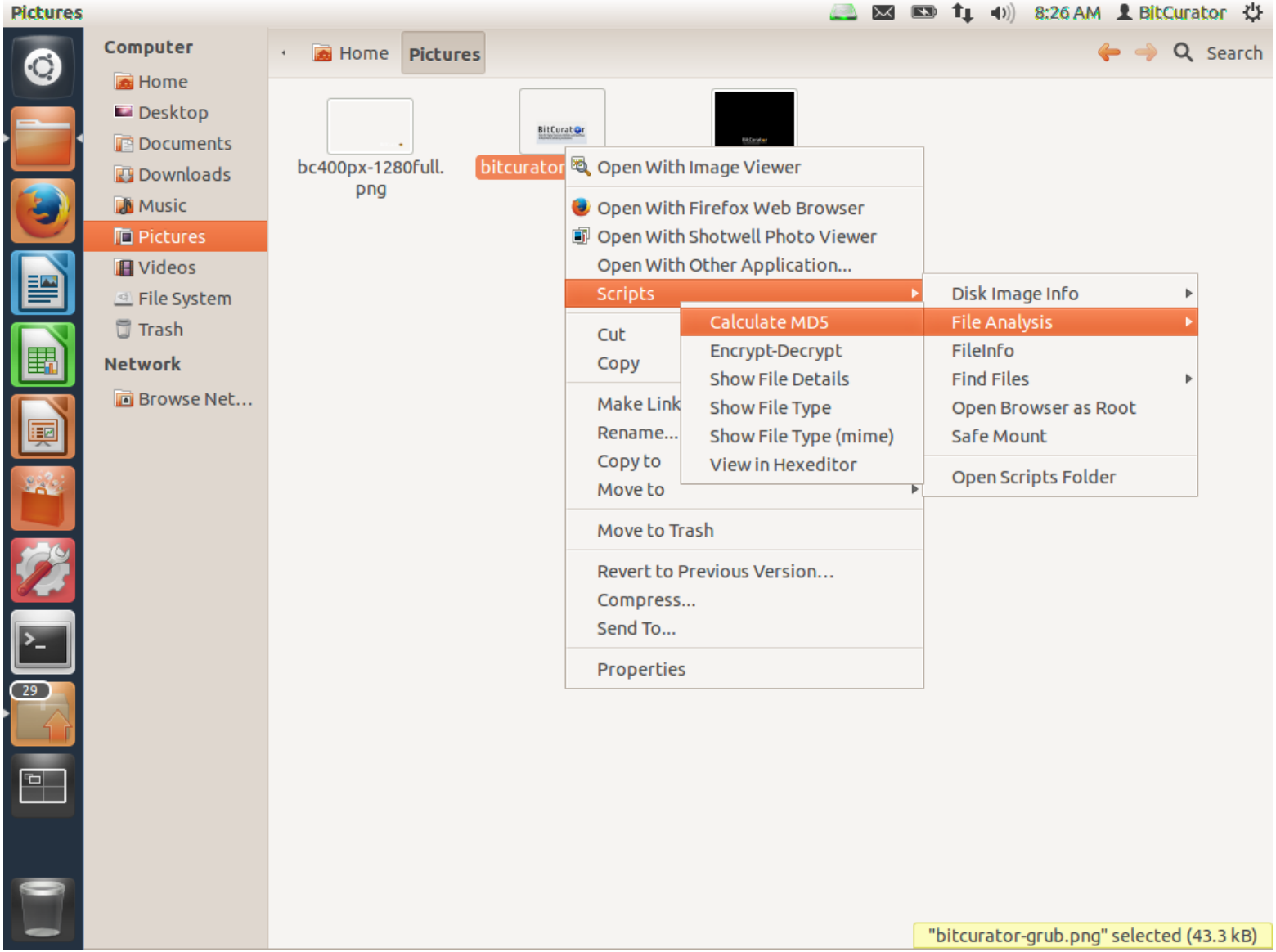
Scanning for Viruses (ClamTK)

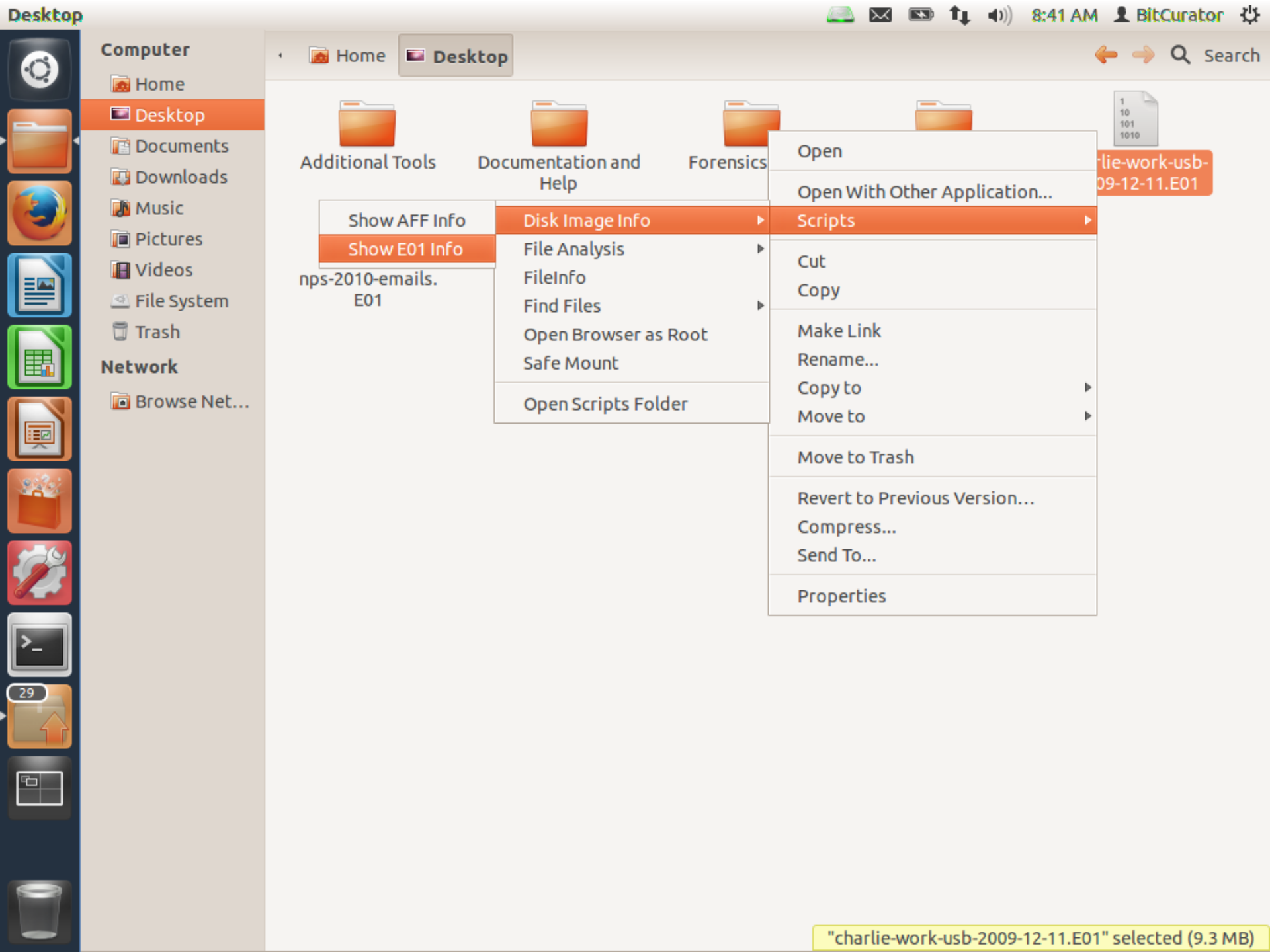


Nautilus Scripts

- Scripts that can be run using the GNOME's file manager, Nautilus
- Can be used in the BitCurator environment or your own Linux environment
- You've already seen one of these (viewing EXIF metadata), but here are some others…

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




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)

Quick Start Guide
Most recent version always available at:
http://wiki.bitcurator.net/

BitCurator

Quick Start Guide v0.7.6

Last updated: February 21, 2014

UNC SCHOOL OF INFORMATION AND LIBRARY SCIENCE

MITH MARYLAND INSTITUTE FOR TECHNOLOGY IN THE HUMANITIES

Related Development: DIMAC (Disk Image Access for the Web)

- Developed by Sunitha Misra and Kam Woods
- Allows the user to dynamically navigate and download contents of a disk image, without having to download the image or mount it
- See: https://github.com/kamwoods/dimac
- Demo at:

https://www.youtube.com/watch?v=W2Gd7eY8XOI

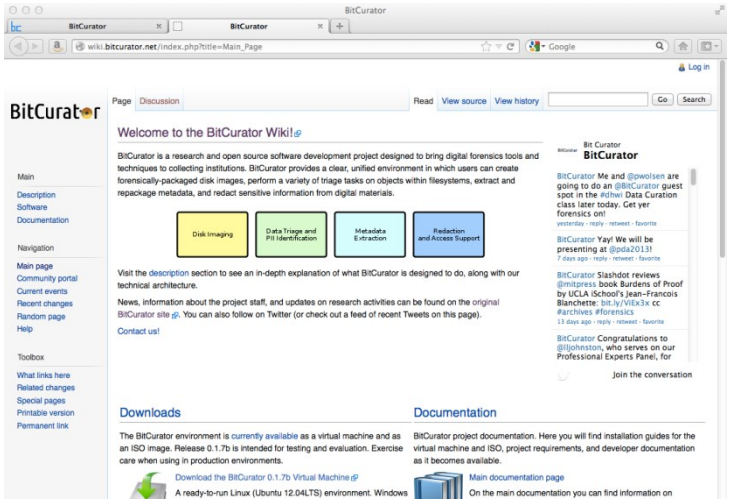
Telling Meaningful Stories

- Any given digital trace may have limited value in isolation, but analysis and comparison of multiple traces can support many valuable inferences.
- For example, an email address that appears on a disk can be combined with various other forms of contextual information, such as:
  - text that surrounded the string
  - histograms showing how often the address appeared elsewhere on the disk
  - timestamps of when files were created and when specific applications were run
  - browser history files that reflect an entire session of use rather than just a discrete transaction
  - user account information
- This can include forensic investigation of your own activities.

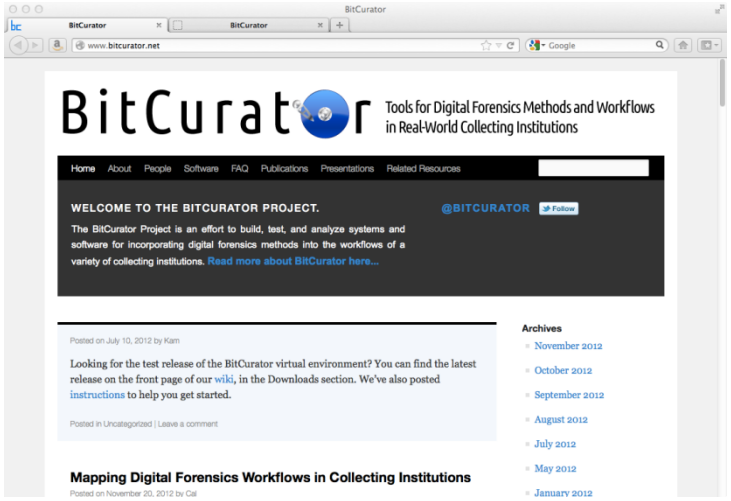
Protecting Privacy

- Forensic investigation can reflect many aspects of one's life that he or she might not want to share with others.
- Forensics tools can be used to identify (e.g. using bulk_extractor) a given pattern and then overwrite those patterns so that they cannot be found by others.
- A major objective of the BitCurator environment is to support such privacy-sensitive processing of materials.

Sources for BitCurator Information:



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



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

Twitter: @bitcurator