

Archival Application of Digital Forensics Methods for Authenticity, Description and Access Provision

Cal Lee

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

International Council on Archives Congress

August 20-24, 2012

Brisbane, Australia

BitCurator 



What is an archivist to do with things like this?



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 Acquiring Analog Materials

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

Same Fundamental Archival Principles Apply

- | | |
|------------------|--|
| Provenance | <ul style="list-style-type: none">• 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 | <ul style="list-style-type: none">• “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 digital is different...

“No computation without representation”

Smith, Brian Cantwell. "Limits of Correctness in Computers." In *Computerization and Controversy: Value Conflicts and Social Choices*, edited by Rob Kling, 810-25. San Diego, CA: Academic Press, 1996. 815.

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

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

ContextMiner Alpha 3.0

[\[Home\]](#)[\[Publications\]](#)[\[Reports\]](#)[\[Add\]](#)[\[View\]](#)[\[Search\]](#)[\[Profile\]](#)[\[Visualize\]](#)[\[Monitor\]](#)[\[Tools\]](#)[\[Developer\]](#)

This page lists all the seed queries that are used for monitoring videos related to elections on YouTube. Clicking on a query will show all the results collected over several crawls. Total number of these results are also listed here for each query. The last column in the following table shows how many total results YouTube had for a given query during our latest crawl. Clicking on 'Setup' associated with a query will bring up an interface where the curator can specify what constitutes as a "significant" change for a video of that query.

#	Query	Setup	Total results so far	Max results on last crawl
1	election 2008	Setup	574	6150
2	US election 2008	Setup	349	795
3	United States election 2008	Setup	216	257
4	presidential election 2008	Setup	206	1820
5	campaign 2008	Setup	273	2530
6	decision 2008	Setup	168	142
7	Joe Biden	Setup	209	1080
8	Hillary Rodham Clinton	Setup	193	353
9	Christopher Dodd	Setup	267	815
10	John Edwards	Setup	902	7540
11	Mike Gravel	Setup	301	1210
12	Dennis Kucinich	Setup	229	1600
13	Barack Obama	Setup	861	9140
14	Bill Richardson	Setup	287	1100
15	Wesley Clark	Setup	191	375
16	Al Gore	Setup	613	4910
17	Tom Vilsack	Setup	89	68
18	Sam Brownback	Setup	254	404
19	John McCain	Setup	22	16

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

ContextMiner Alpha 3.0

[\[Home\]](#) [\[Publications\]](#) [\[Reports\]](#) [\[Add\]](#) [\[View\]](#) [\[Search\]](#) [\[Profile\]](#) [\[Visualize\]](#) [\[Monitor\]](#) [\[Tools\]](#) [\[Developer\]](#)

This page presents contextual information for a video captured over a number of days. Contextual information is defined as the information about a video that may change with time. Usually this information is contributed by the visitors of the video page. [See](#) the metadata information for this video. Description of various attributes displayed is given [here](#).



Query: *Rudy Giuliani*

[I Got A Crush On.... Giuliani](#)

Collaboration with the very talented JackDanyells, who came up with the concept for this video. Check out his channel at: <http://www.youtube.com/jackdanyells> -Lyrics by JackDanyells -Vocal melody composed and sung by me -Royalty free background music from sounddogs.com

Comedy

Crawling since 2007-07-19

Color coding for % changes

< 0.05 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 5.0 >

Crawl #	Crawl date	Rank	Views	Ratings	Avg Rating	Comments	Links	Favorited	Honors	Change
1	2007-07-31	5	27357	301	3.74	288	5	44	0	--
2	2007-08-01	5	27452	303	3.73	290	5	44	0	--
3	2007-08-02	5	27780	307	3.72	291	5	45	0	--
4	2007-08-03	5	28048	309	3.71	291	5	45	0	--
5	2007-08-04	2	28398	310	3.71	291	5	45	0	--
6	2007-08-05	2	28443	314	3.69	294	5	45	0	--
7	2007-08-06	3	28980	314	3.69	296	5	45	0	--
8	2007-08-07	3	29265	318	3.65	298	5	45	0	--
9	2007-08-08	3	29551	319	3.65	299	5	46	0	--
10	2007-08-09	3	30094	320	3.64	300	5	47	0	--
11	2007-08-10	3	30384	323	3.61	302	5	47	0	--
12	2007-08-10	5	30419	324	3.62	303	5	48	0	--
13	2007-08-11	3	30540	324	3.62	305	5	49	0	--
14	2007-08-12	3	30697	326	3.61	306	5	49	0	--
15	2007-08-13	3	30848	326	3.61	306	5	49	0	--
16	2007-08-14	3	31036	326	3.61	306	5	49	0	--
17	2007-08-15	2	31181	326	3.61	306	5	49	0	--
18	2007-08-16	2	31321	326	3.61	307	5	51	0	--
19	2007-08-17	2	31459	327	3.61	307	5	51	0	--
20	2007-08-18	2	31662	331	3.59	308	5	51	0	--
21	2007-08-19	2	31792	332	3.58	308	5	51	0	--
22	2007-08-20	2	31937	335	3.57	310	5	51	0	--
23	2007-08-21	2	32135	335	3.57	311	5	52	0	--
24	2007-08-22	2	32404	335	3.57	311	5	54	0	--

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



The screenshot shows the YouTube interface for a video titled "Vote Different". The video features a woman rowing a boat, wearing a white tank top with a blue and red Obama campaign logo. The video player includes a progress bar at 0:16 / 1:14, a volume icon, and a download icon. Below the video, the rating is shown as four stars (Rate: ★★★★★) with 12,058 ratings, and the view count is 5,268,816.

YouTube
Broadcast Yourself™

Sign Up | QuickList (0) | Help | Sign In | Site: 

Home Videos Channels Community

Videos Search advanced Upload


Vote Different


From: **ParkRidge47**
Joined: 1 year ago
Videos: 3 [Subscribe](#)


Added: **March 05, 2007** ([More info](#))
Make up your own mind. Decide for yourself who ...
Embed: [Customize](#)
<object width="425" height="344"><param name="movie" value="http:

► More From: **ParkRidge47**

▼ Related Videos

 **Barack Obama Hillary Clinton - Umbrella**
01:56 From: wolf084
Views: 11,179,757

 **The Shocking Video Hillary Does NOT Want You To See! (1of2)**
10:28 From: NufftRespect
Views: 3,401,587

 **Obama Girl Returns for Iowa! (Why Obama Won)**
02:19 From: barelypolitical
Views: 2,451,439

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

```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

G:\>dir /a
Volume in drive G is KINGSTON
Volume Serial Number is 17E9-242F

Directory of G:\

03/12/2009  08:54 AM                4,096  ._.Trashes
03/12/2009  08:54 AM          <DIR>      .Trashes
03/12/2009  08:54 AM          <DIR>      .Spotlight-V100
03/11/2009  07:07 PM      1,023,213  nc-busmodels-jpw2009.pptx
03/12/2009  08:55 AM                4,096  .nc-busmodels-jpw2009.pptx
03/31/2009  01:23 PM      6,442,496  EMSS Meeting.ppt
               4 File(s)          7,473,901 bytes
               2 Dir(s)        120,145,920 bytes free

G:\>
```



Interaction Examples

Level

Aggregation of objects

Object or package

In-application rendering

File through filesystem

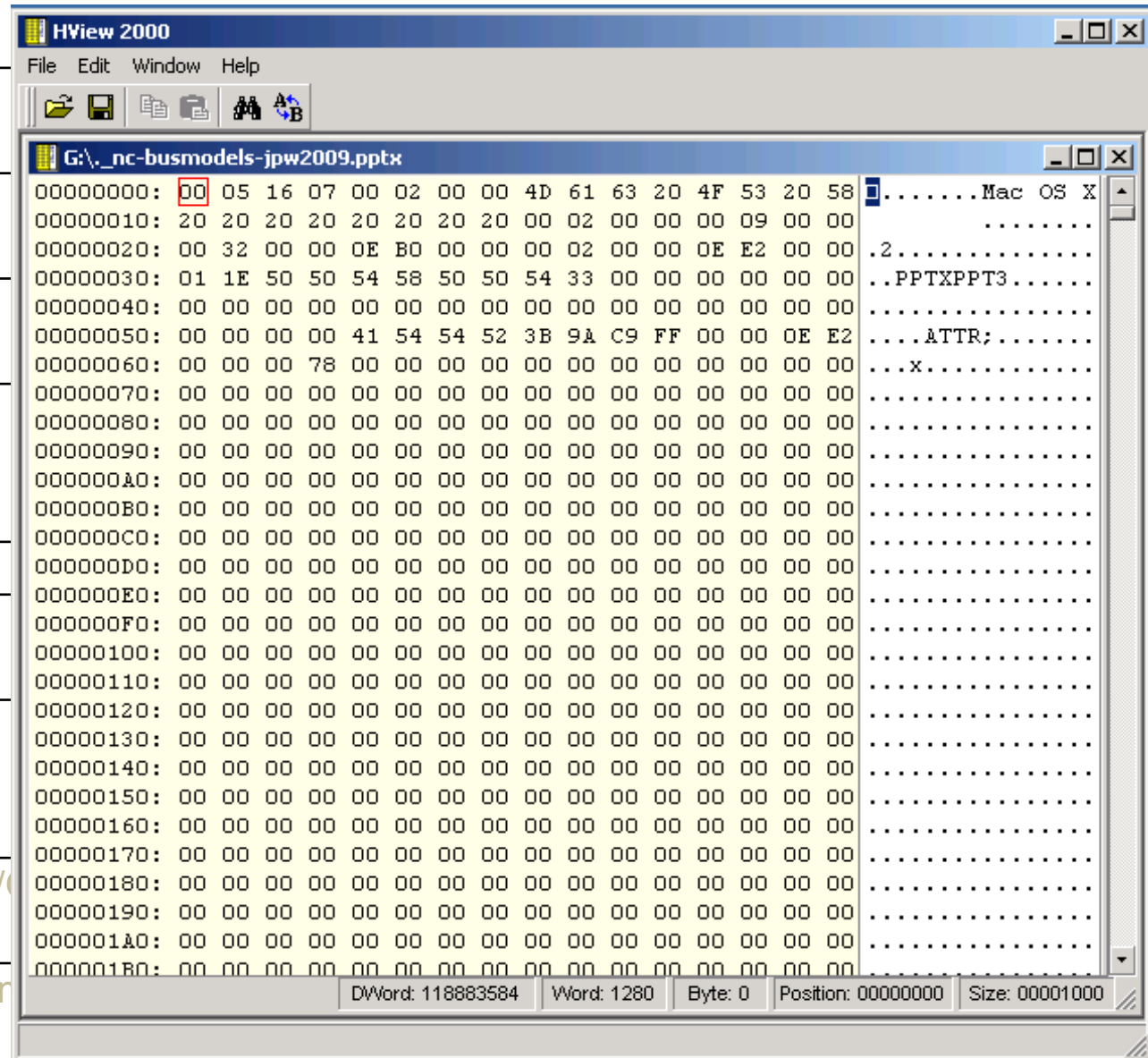
File as “raw” bitstream

Sub-file data structure

Bitstream through I/O
equipment

Raw signal stream through I/O
equipment

Bitstream on physical medium



Interaction Examples

Level

Aggregation of objects

Object or package

In-application rendering

File through filesystem

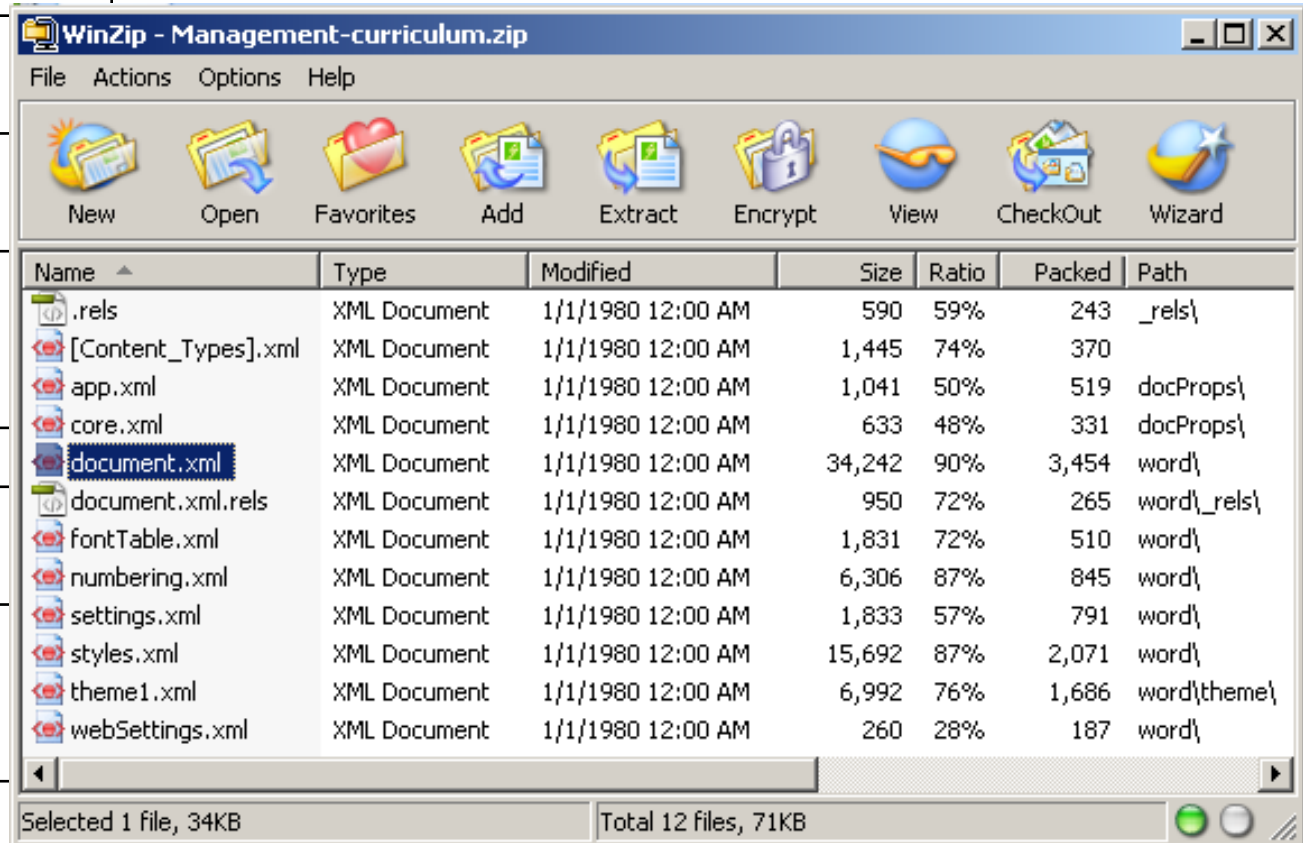
File as “raw” bitstream

Sub-file data structure

Bitstream through I/O
equipment

Raw signal stream through
equipment

Bitstream on physical medium



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

Segment =====	arg =====	length =====	data =====
afflib_version	0	7	"3.3.3"
aff_file_type	0	3	AFF
acquisition_commandline	0	36	aimage /dev/sda /mnt/charlie-002
acquisition_device	0	8	/dev/sda
sectorsize	1024	0	
pagesize	16777216	0	
devicesectors	2	8	= 9999864 (64-bit value)
acquisition_macaddr	0	18	00:0b:db:4f:6b:10.
acquisition_dmesg	0	27298	[0.000000] Initializing cgro
image_gid	0	16	7256 F895 DE4F E304 233E 21C0 2347 CCC5
acquisition_date	0	20	2009-11-12 19:12:18.
md5	0	16	0609 2DFE AA4F B183 946F 95D8 AD84 519E
acquisition_seconds	1570	0	= 00:26:10 (hh:mm:ss)
imagesize	2	8	= 10239860736 (64-bit value)

Level

Aggregat

Object or

In-applica

File throu

File as "r

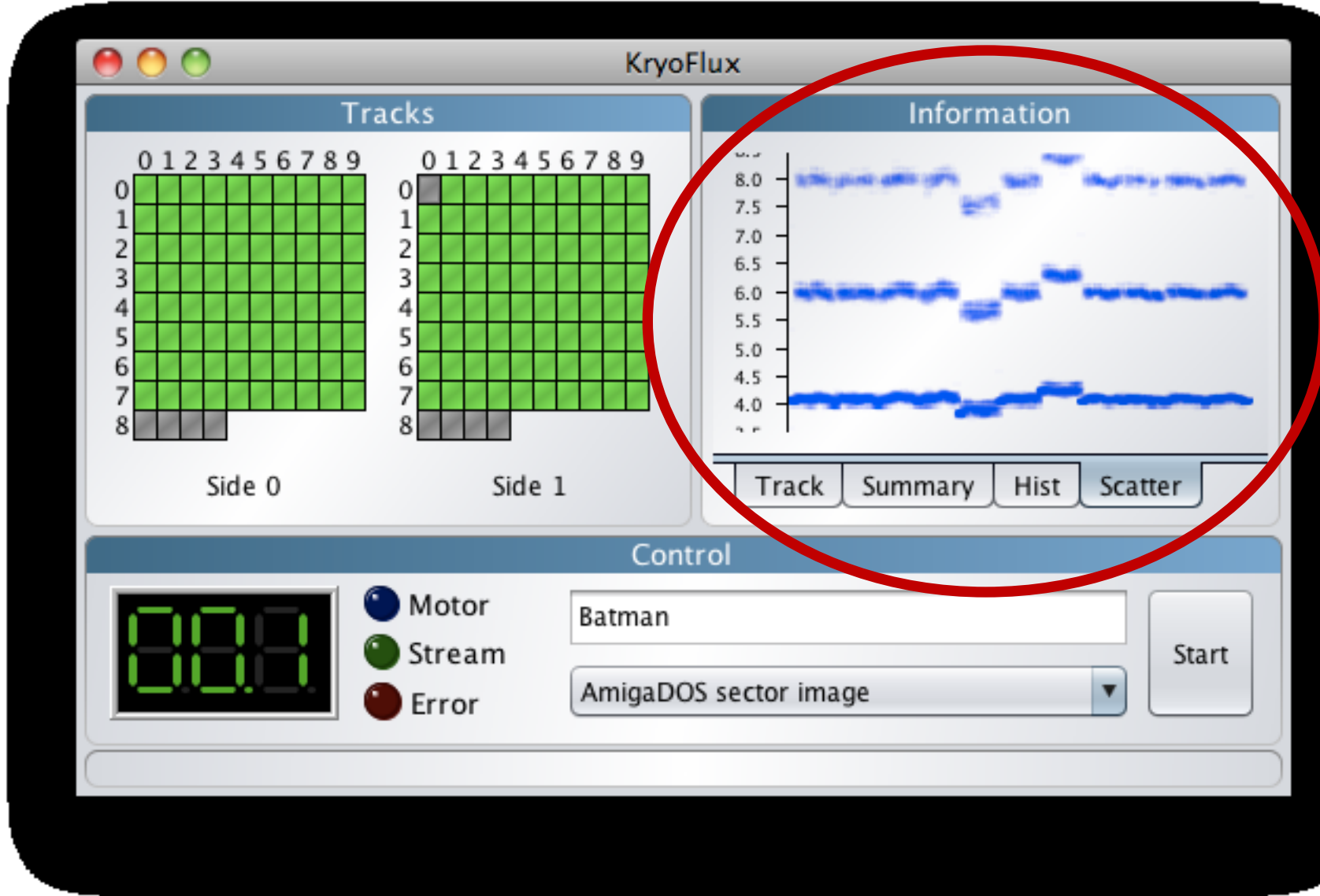
Sub-file c

Bitstream
equipment

Raw sign

I/O equipment

Bitstream on physical medium



Interaction Examples

Level

Aggregation of objects

Object or package

In-application rendering

File through filesystem

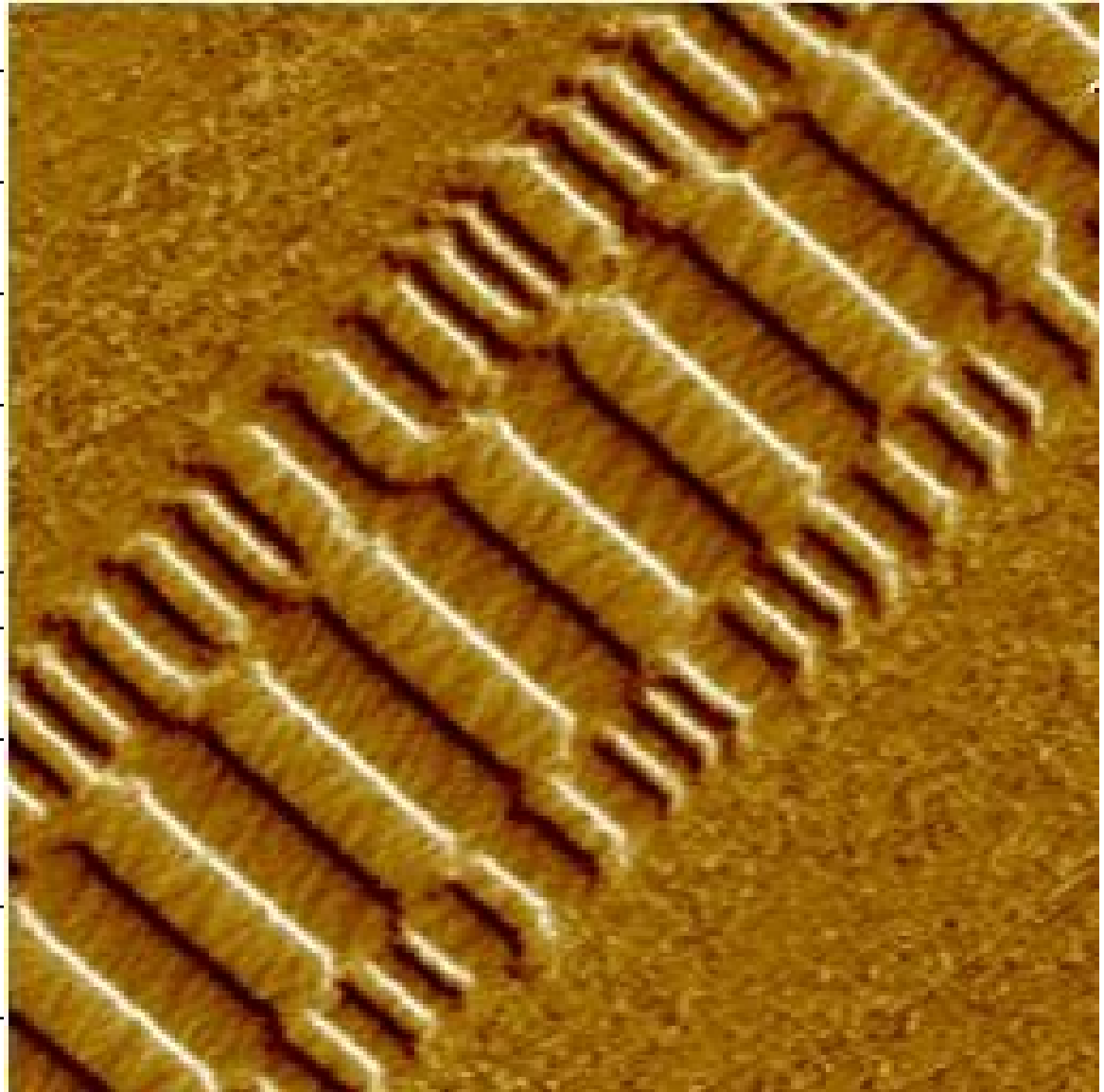
File as “raw” bitstream

Sub-file data structure

Bitstream through I/O
equipment

Raw signal stream through I/O
equipment

**Bitstream on physical
medium**



Digital Forensics Principles Can Help Archivists to 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 | <ul style="list-style-type: none">• 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 | <ul style="list-style-type: none">• Identify personally identifying information, regardless of where it appears• Flag for removal, redaction, closure or restriction |

Previous Work*

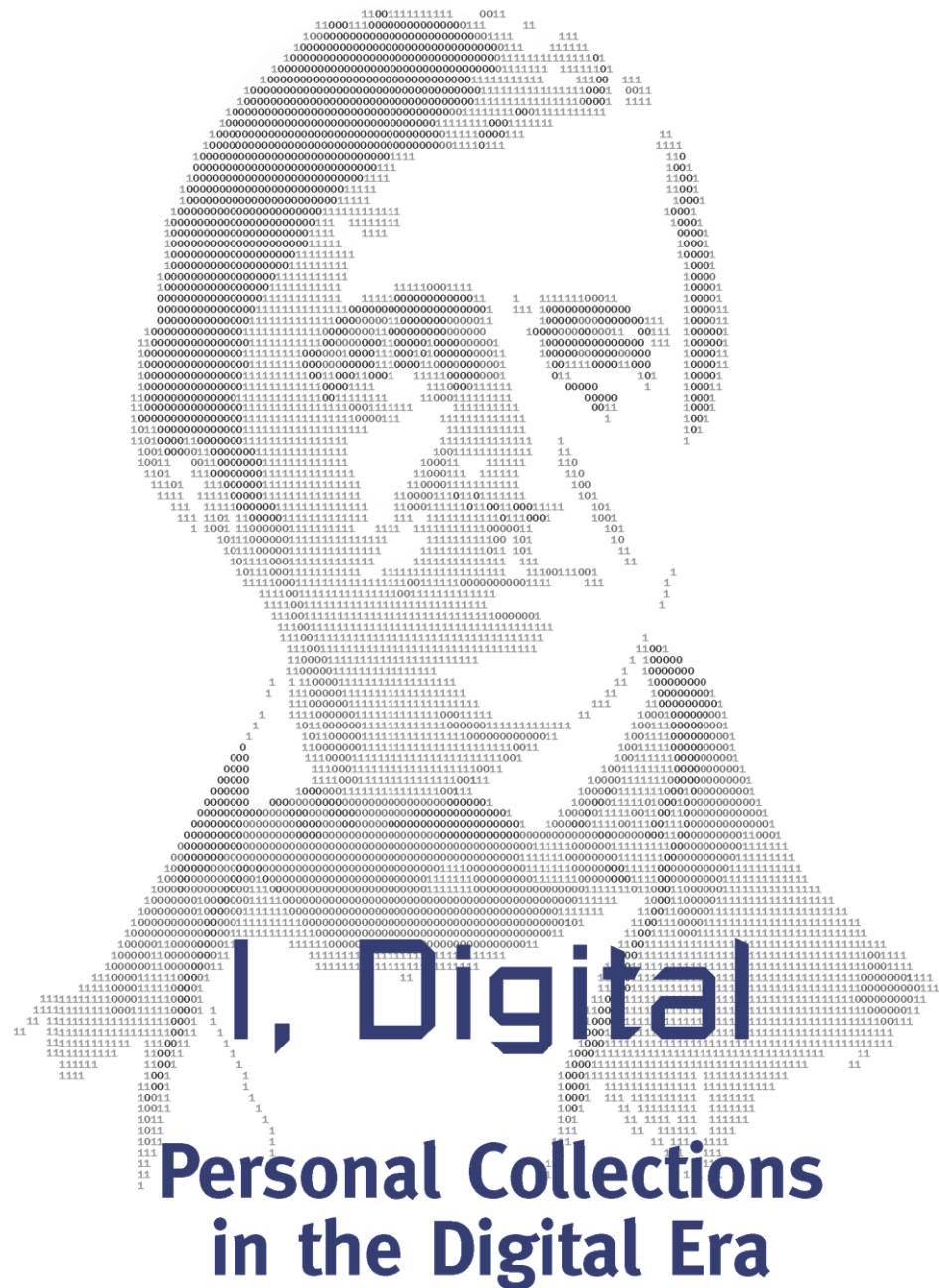
- Some library/archives literature on recovering data from media
- Report by Ross and Gow (1999) on relevance of advances in data recovery and digital forensics to collecting institutions
- More recently, stream of literature related to use of forensic tools and methods for acquiring and managing digital collections
- Related projects:
 - Computer Forensics and Born-Digital Content in Cultural Heritage Collections
 - Born Digital Collections: An Inter-Institutional Model for Stewardship (AIMS)
 - Digital Records Forensics project

*See citations in full paper:

<http://www.ica2012.com/files/data/Full%20papers%20upload/ica12Final00290.pdf>

Renewed Energy and Attention around Personal Archives

- Arguably under-represented in first few decades of literature about electronic records
- Technical challenges of unruly acquisitions from individuals are channeling new energy into personal archives issues



Edited by Christopher A. Lee

What is Digital Forensics (aka Forensic Computing)?

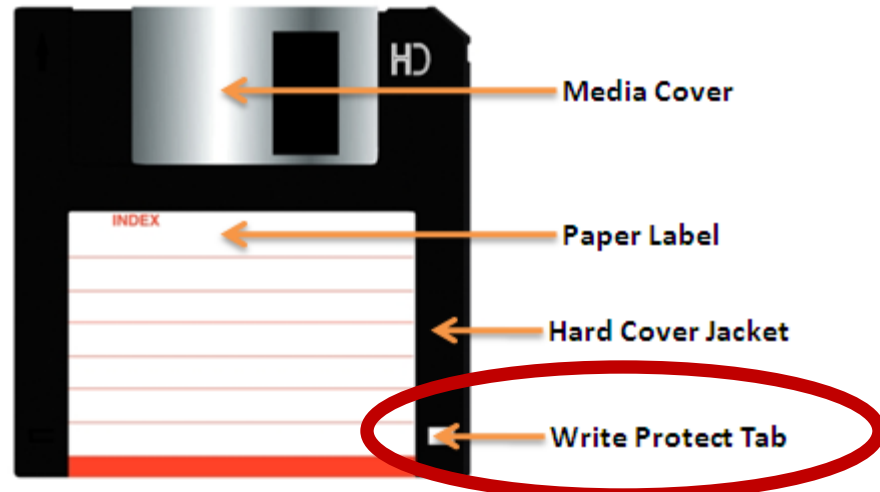
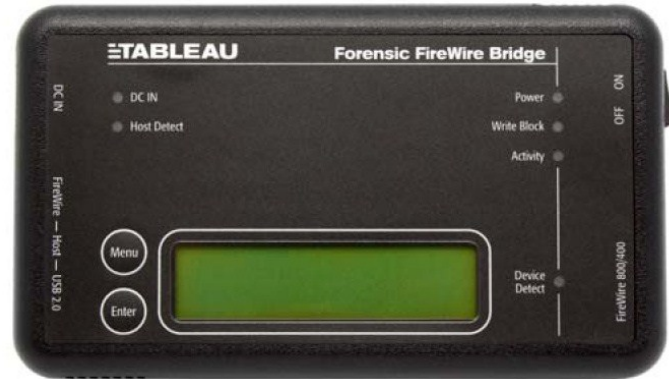
- “The process of identifying, preserving, analysing and presenting digital evidence in a manner that is legally acceptable.”¹
- “Involves multiple methods of
 - **Discovering digital data (computer system, mobiles)**
 - **Recovering deleted, encrypted, or damaged file information**
 - Monitoring live activity
 - Detecting violations of corporate policy”²

1. McKemmish, R. “What is Forensic Computing?” *Trends and Issues in Crime and Criminal Justice* 118 (1999).
2. Brad Glisson, Introduction to Computer Forensics & E-discovery, University of Glasgow, Week 1 Lecture, September 2008. (emphasis mine)

Guidelines for Evidence Collection & Archiving (RFC 3227) – Main Lessons

- “Such collection represents a considerable efforts on the part of the System Administrator.”
- “Keep detailed notes.”
- “Minimise changes to the data as you are collecting it.”
- “Do collection first and analysis later.”
- “Proceed from the volatile to the less volatile.”
- Computer evidence should be: admissible, authentic, complete, reliable, believable

Write Blocking – One-Way Streets for Data



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

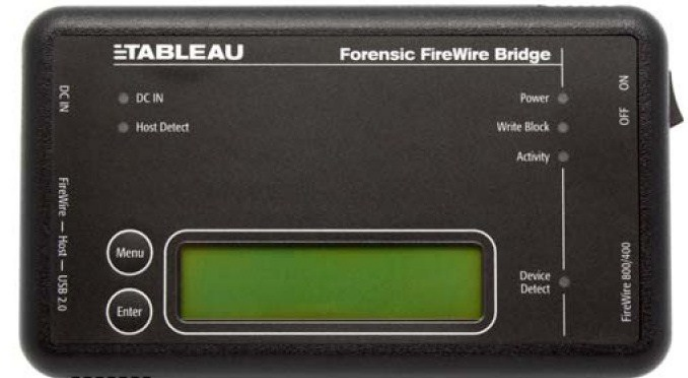
File System – An Essential Layer for Metadata

- 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)

Getting below the File System – Low-Level Copying

- Getting an “image” of a storage medium involves working at a level below the file system
- Image is a copy of all of the storage sectors from the drive, rather than just copying the files
- Can get at file attributes and deleted files not visible through higher-level copy operations

Digital Forensics Tools – Hardware and Software



AFFLIB

Open Source Computer Forensics Software

Digital forensics tools are designed primarily to be used
in places like this:



El Paso County Sheriff's Office (Colorado)

<http://shr.elpasoco.com/Law+Enforcement+Bureau/Investigations+Division/Computer+Crime+Lab.htm>

But they're also be used in places
like this:

Stanford University Libraries and Academic Information Resources (SULAIR)



British Library, London



BitCurator Project

- 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), Porter Olsen, GA (MITH)

Professional Experts Panel

- 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, 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

- 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 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

BitCurator Environment*

- Bundles, integrates and extends functionality 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. Virtual Box 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/>



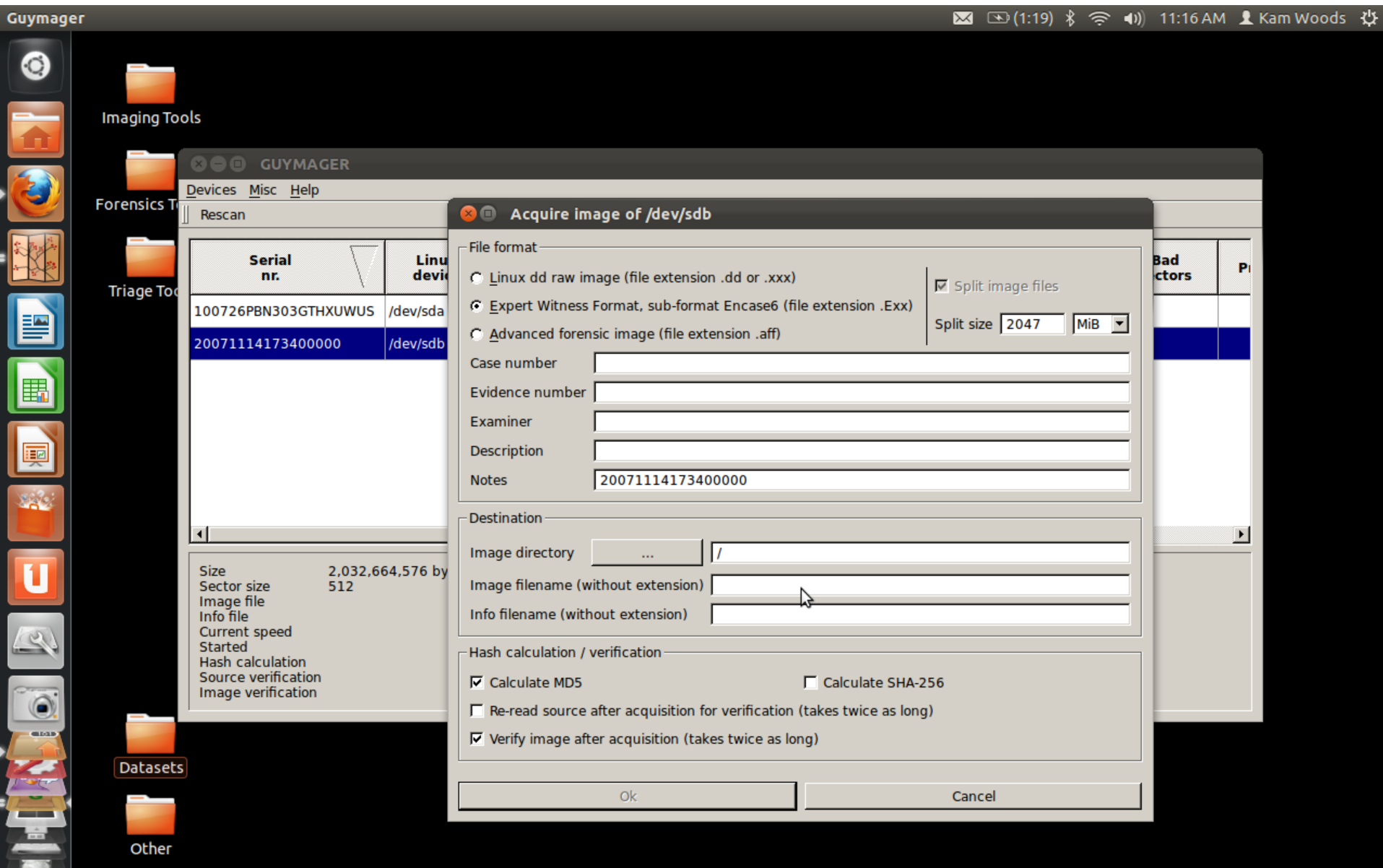
- Imaging Tools
- Forensics Tools
- Triage Tools

- Datasets
- Other

BitCurator



Main Acquisition Interface for Guymager



Guymager Showing Technical Metadata about an SD Card (Right Click)

Guymager

Imaging Tools

Forensics Tools

Triage Tools

Size
Sector size
Image file
Info file
Current
Started
Hash calc
Source v
Image v

Datasets

Other

GUYMAGER

Devices Misc Help

Rescan

Serial nr.	Linux device	Model	State	Size	Hidden Areas	Bad sectors	P
100726PBN303GTHXUWUS	/dev/sda	ATA HITACHI HTS545032B9A300	Idle	320.1GB	unknown		
20071114173400000	/dev/sdb	Generic- Multi-Card	Idle	2.0GB	unknown		

Device info

Media Manufacturer: 8E1E999E999E999E
Transport: 0xea00

Standards:
Used: unknown (minor revision code 0x03ee)
Supported: 13 12 8 7
Likely used: 13

Configuration:
CHS addressing not supported
LBA user addressable sectors: 65941624
Logical/Physical Sector size: 512 bytes
device size with M = 1024*1024: 32198 MBytes
device size with M = 1000*1000: 33762 MBytes (33 GB)
cache/buffer size = unknown

Capabilities:
LBA, IORDY(may be)(cannot be disabled)
Queue depth: 32
Standby timer values: spec'd by Vendor
R/W multiple sector transfer: Max = 255 Current = 255
Recommended acoustic management value: 234, current value: 0
DMA: *mdma0 *mdma1 *mdma2 *mdma3 *mdma4 *mdma5 *mdma6 *mdma7 *udma1 *udma4 *udma5 (?)
Cycle time: min=1006ns recommended=59904ns
PIO: pio0 pio1 pio2 pio7 pio8
Cycle time: no flow control=65535ns IORDY flow control=12520ns
* reserved 69[1]
* reserved 69[2]
* reserved 69[3]
* reserved 69[6]

Exporting Filesystem Metadata - Output from fiwalk (XML)*

```
<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>
  <ctime>1257729636</ctime>
  <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

Bulk Extractor
(Created by Simson Garfinkel)

Bulk Extractor Scanning Options

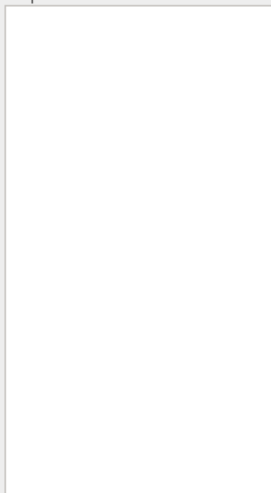
Run bulk_extractor

File Edit View Tools Help



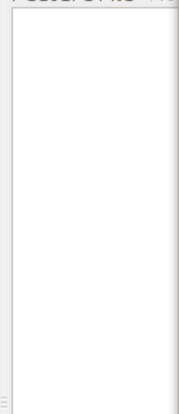
Highlight:

Reports



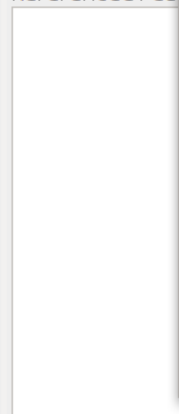
Feature Filter

Feature File No



Referenced Fea

Referenced Fea



Run bulk_extractor

Scan Options Processing Options

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 Context Stop List File

☐ Use Find Regex Text File

☐ Use Find Regex Text

Tuning Parameters

☐ Use Context Window Size

☐ Use Margin Size

☐ Use Min Word Size

☐ Use Max Word Size

☐ Use Number of Threads

Plugins

☐ Use Plugin Directory

Scanners

☐ net

☐ wordlist

☒ accts

☒ base64

☒ kml

☒ email

☒ gps

☒ aes

☒ json

☒ exif

☒ zip

☒ gzip

☒ pdf

☒ hiber

☒ winprefetch

Restore Defaults

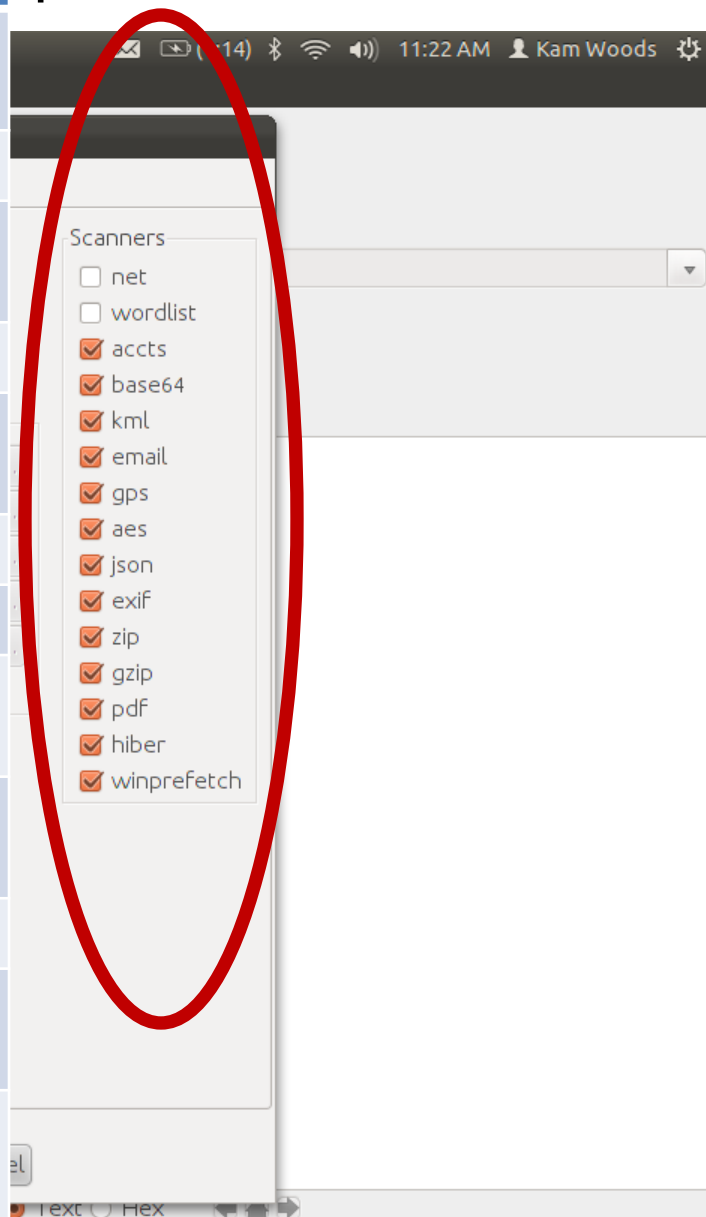
Start bulk_extractor

Cancel

11:22 AM Kam Woods

Options

Scanner	Description
scan-accts	Looks for phone numbers, credit card numbers, etc
scan_base64	Decodes BASE64 text
scan_kml	Detects KML (Keyhole Markup Language) files – used to identify geographic locations
scan_gps	Detects XML from Garmin GPS devices
scan_aes	Detects in-memory AES (Advanced Encryption Standard) keys from the key schedules
scan_json	Detects JavaScript Object Notation files
scan_exif	Detects EXIF structures from JPEG files
scan_zip	Detects and decompresses ZIP files and zlib streams
scan_gzip	Detects and decompresses GZIP files and gzip streams
scan_pdf	Extracts text from some kinds of PDF files
scan_hiber	Detects and decompresses Windows hibernation file fragments
scan_winprefetch	Detects and extracts fields from windows prefetch fields from Windows prefetch files and file fragments



Email Addresses Identified - With Repeats Highlighted

Bulk Extractor Viewer File Edit View Tools Help

Highlight: ☒ Match case

Reports

- May-2012-SD-Image-Output
 - domain.txt
 - domain_histogram.txt
 - email.txt**
 - email_histogram.txt
 - ether.txt
 - ether_histogram.txt
 - exif.txt
 - gps.txt
 - json.txt
 - rfc822.txt
 - telephone.txt
 - telephone_histogram.txt
 - url.txt
 - url_histogram.txt
 - url_services.txt
 - zip.txt

Feature Filter ☒ Match case

Feature File	email.txt
257769	modperl-cvs@perl.apache.org
287442	modperl-subscribe@perl.apache.org
307958	dougm@pobox.com
1513(46518312-ZIP)	archaeology@yorkat.cc
40285(2277376-GZIP)	simsong@ncr.nps.edu
41649(2277376-GZIP)	simsong@ncr.nps.edu
43311(2277376-GZIP)	simsong@arlington-8-3
43464(2277376-GZIP)	simsong@t.nitroba.org
46127(2277376-GZIP)	root@arlington-8-30-77
46384(2277376-GZIP)	simsong@arlington-8-3
47184(2277376-GZIP)	simsong@arlington-8-3
47400(2277376-GZIP)	simsong@ncr.nps.edu
48407(2277376-GZIP)	simsong@dhcp184-49-
49359(2277376-GZIP)	simsong@Alphonse-Mu
50141(2277376-GZIP)	simsong@t.nitroba.org
50907(2277376-GZIP)	simsong@163.sub-75-1
51057(2277376-GZIP)	simsong@ps14412.dre

Referenced Feature File None

Referenced Feature None

Navigation

image.E01, 2277376-GZIP-41949, simsong@ncr.nps.edu

Image File image.E01

Feature File email.txt

Feature Path 2277376-GZIP-41949

Feature simsong@ncr.nps.edu

Image

```
40960 re_recorder::write): added support for opt_offset_add to allow o
41024 utput to be shifted (for parallelizing across multiple systems.)
41088 ...* src/sbuf.h (class pos0_t): removed snprintf; now uses strin
41152 gstream...(operator +): changed most functions to take const & r
41216 ather than a new object...* src/feature_recorder.cpp (feature_r
41280 e_recorder::write): now always writes out the second \t for the con
41344 text, even if there is no context...2011-11-21 Simson Garfinkel
41408 <simsong@Alphonse-Mucha.local>...* configure.ac: advanced to b
41472 eta9..added AC_PROG_CC AC_PROG_CXX and AC_PROG_INSTALL...* src/M
41536 akefile.am (.flex.o): FlexLexer.h moved to MyFlexLexer.h to supp
41600 ort CentOS where an out-of-date flex is installed...2011-11-16
41664 Simson Garfinkel <simsong@FC15>...* src/bulk_extractor.cpp (pro
41728 cess_path): fixed handling of /h and /r with -p option..2011-11-
41792 12 Simson Garfinkel <simsong@imac3.home>...* configure.ac: rem
41856 oved pcap.h tests becuase its not needed..increased to beta4..20
41920 11-11-05 Simson Garfinkel <simsong@ncr.nps.edu>...* src/scan_e
41984 mail.flex (Host): now only writes domains>0...* src/scan_zip.cp
42048 p (scan_zip): zip components with no name are now given <NONAME>
42112 ...* src/scan_winprefetch.cpp (scan_winprefetch): modified to on
42176 ly write out prefect files with non-zero exec name...* src/scan_
42240 net.cpp (scan_net): significant update --- I don't need libpcap
42304 to do packet carving!..2011-11-09 Simson Garfinkel <simsong@Al
42368 phonse-Mucha.local>...* configure.ac: updated to beta3..2011-11-
42432 08 Simson Garfinkel <simsong@Alphonse-Mucha.local>...* src/ima
42496 ge_process.cpp (sbuf_alloc): added a new iterator method it->pos
42560 0() returns the pos0 of the sbuf to be allocated by it->sbuf_all
42624 oc()..(sbuf_alloc): changed calloc to malloc for performance..(p
42688 rocess_aff::sbuf_alloc): now thorws bad_alloc if an exception is
42752 encountered..(process_ewf::sbuf_alloc): now thorws bad_alloc..(
42816 process_raw::sbuf_alloc): now thorws bad_alloc..2011-11-07 Sims
42880 on Garfinkel <simsong@alphonse-mucha>...* src/bulk_extractor.cp
42944 p: removed scanner_enabled()...* src/Makefile.am (bulk_extracto
43008 r_SOURCES): removed checkpoint.h...* src/bulk_extractor.cpp (mai
43072 n) checkpoint removed, protecting now done through <feature_
```

Text Hex

Histogram of Email Addresses*

Bulk Extractor Viewer

File Edit View Tools Help



Highlight: ☒ Match case

Reports

- May-2012-SD-Image-Output
 - domain.txt
 - domain_histogram.txt
 - email.txt
 - email_histogram.txt**
 - ether.txt
 - ether_histogram.txt
 - exif.txt
 - gps.txt
 - json.txt
 - rfc822.txt
 - telephone.txt
 - telephone_histogram.txt
 - url.txt
 - url_histogram.txt
 - url_services.txt
 - zip.txt

Feature Filter ☒ Match case

Feature File email_histogram.txt

```
n=808 forensics@hoffmannbv.nl
n=647 ahuggel@gmx.net
n=490 jbmertz@users.sourceforge.net
n=314 rpj@callisto.canberra.edu.au
n=77 bugs@afflib.org
n=76 autoconf@gnu.org
n=74 simsong@m.ern.nps.edu
n=70 config-patches@gnu.org
n=54 brad@robotbattle.com
n=50 cholet@logilune.com
n=49 simsong@domex.nps.edu
n=48 bug-libtool@gnu.org
n=44 simsong@t.eecs.harvard.edu
n=42 randy@theoryx5.uwinnipeg.ca
n=40 domexuser1@gmail.com
n=33 jrb3@best.com
n=32 bug-automake@gnu.org
n=32 rpj@special.ise.canberra.edu.au
```

Referenced Feature File email.txt

Referenced Feature None

```
257769 modperl-cvs@perl.apache.org
287442 modperl-subscribe@perl.apache.org
307958 dougm@pobox.com
1513(46518312-ZIP) archaeology@yorkat.co.uk
40285(2277376-GZIP) simsong@ncr.nps.edu
41949(2277376-GZIP) simsong@ncr.nps.edu
43311(2277376-GZIP) simsong@arlington-8-30-77-13
43464(2277376-GZIP) simsong@t.nitroba.org
46127(2277376-GZIP) root@arlington-8-30-77-137.nc
46384(2277376-GZIP) simsong@arlington-8-30-77-13
47184(2277376-GZIP) simsong@arlington-8-30-77-13
47400(2277376-GZIP) simsong@ncr.nps.edu
48407(2277376-GZIP) simsong@dhcpl84-49-148-156
49359(2277376-GZIP) simsong@Alphonse-Mucha.cus
50141(2277376-GZIP) simsong@t.nitroba.org
50907(2277376-GZIP) simsong@163.sub-75-195-180
51057(2277376-GZIP) simsong@psl4412.dreamhost
```

Navigation

image.E01, 13103104-GZIP-4848128-GZIP-758149, callisto.canberra.edu.au

Image File image.E01

Feature File domain.txt

Feature Path 13103104-GZIP-4848128-GZIP-758149

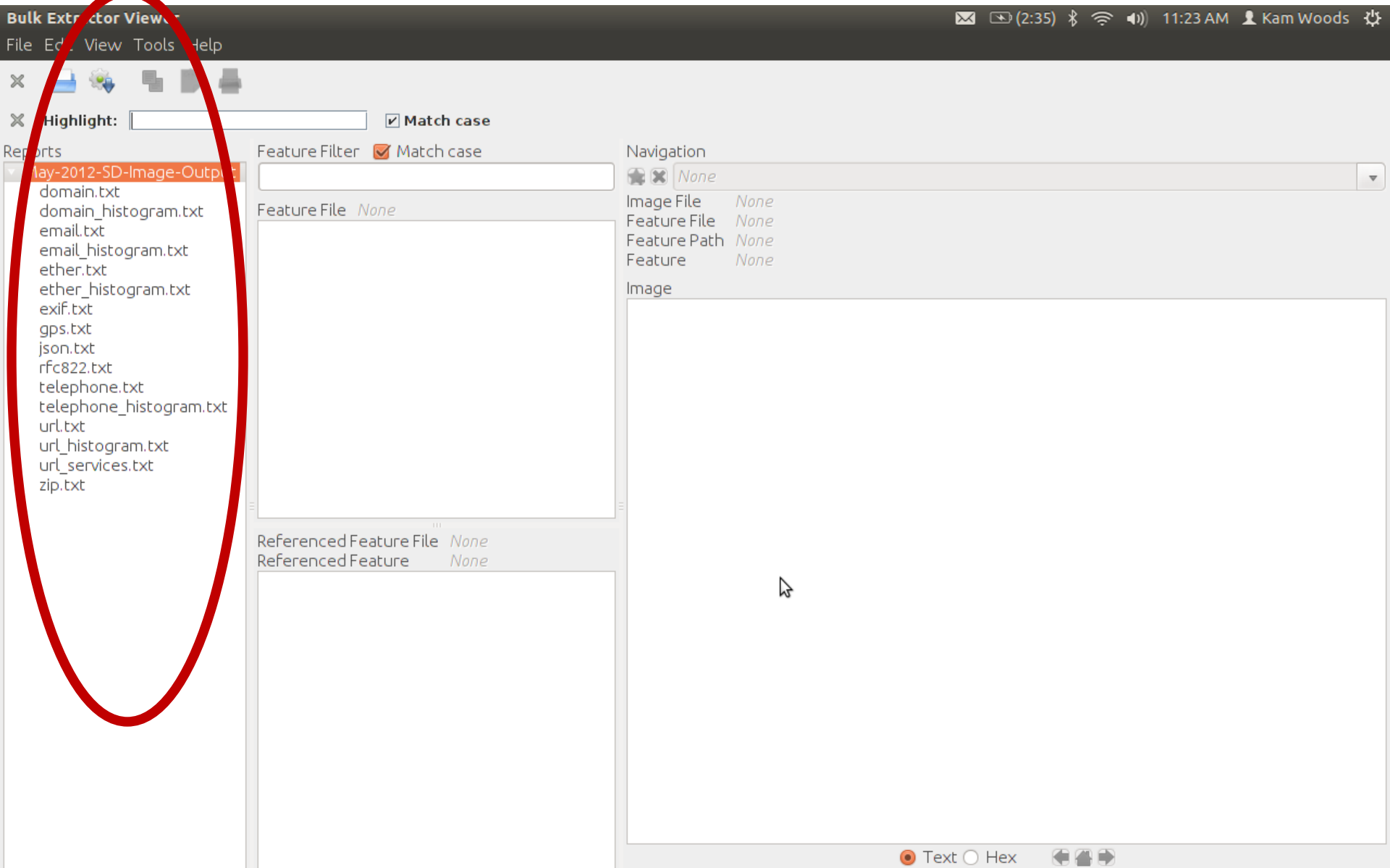
Feature callisto.canberra.edu.au

Image

```
757760 le for the list of contributors...As much as possible, the Chang
757824 eLog file attributes.contributions and patches that have been in
757888 corporated.in the library to the individuals responsible...Final
757952 ly, thanks to all those who work on and contribute to the.POSIX
758016 and Single Unix Specification standards. The maturity of an.indu
758080 stry can be measured by it's open standards.....Ross Johnson.
758144 <rpj@callisto.canberra.edu.au>.....
758208 .....
758272 pthreads-w32-2-8-0-release/README.Borland.....
758336 .....0000664.0000764.0000764.0000
758400 0004752.10233325107.016374. 0.....
758464 .....
758528 .ustar .ross.....ross.....
758592 .....
758656 .....
758720 .....
758784 In ptw32_InterlockedCompareExchange.c, I've added a section for.
758848 Borland's compiler; it's identical to that for the MS compiler e
758912 xcept.that it uses /* ... */ comments instead of ; comments...[R
758976 PJ: need to define HAVE_TASM32 in config.h to use the above.]...
759040 The other file is a makefile suitable for use with Borland's com
759104 piler.(run "make -fBmakefile" in the directory). It builds a si
759168 ngle version.of the library, pthreadBC.dll and the corresponding
759232 pthreadBC.lib.import library, which is comparable to the pthrea
759296 dVC version; I can't.personally see any demand for the versions
759360 that include structured or.C++ exception cancellation handling s
759424 o I haven't attempted to build.those versions of the library. {
759488 I imagine a static version might be.of use to some, but we can't
759552 legally use that on my commercial.projects so I can't try that
759616 out, unfortunately...[RPJ: Added tests\Bmakefile as well.]..Bor
759680 land C++ doesn't define the ENOSYS constant used by pthreads-win
759744 32;.rather than make more extensive patches to the pthreads-win3
759808 2 source I.have a mostly-arbitrary constant for it in the makefi
759872 1. However, this doesn't make it suitable to the application and
```

☒ Text ☐ Hex

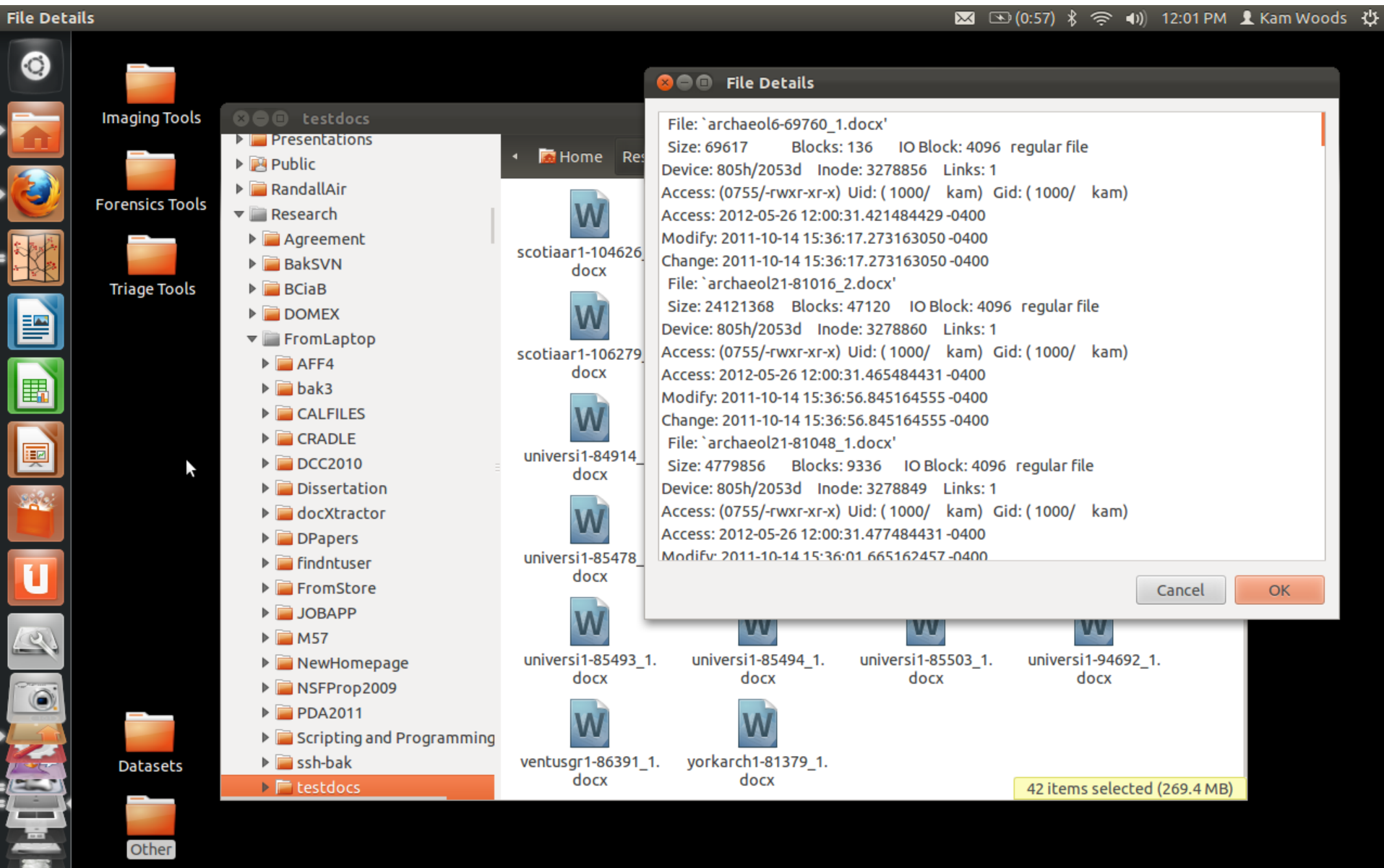
Bulk Extractor Report Options



Nautilus Scripts

- In addition to the specialized forensics tools in the BitCurator environment, there are a variety of scripts that can be run using the GNOME file manager called Nautilus (Linux analog to Windows Explorer or Mac OS X Finder)
- Can be used in the BitCurator environment or your own Linux environment

File Details for Word Documents in a Directory (Nautilus Script)



MD5 Hashes of Files (Nautilus Script)

Text Edit File Edit View Search Tools Documents Help (0:53) 12:05 PM Kam Woods

md5Results.md5 (~ /Research/FromLaptop/testdocs) - gedit

Open Save Undo

md5Results.md5

```
k210f06f75bfc5ed09ef853280010bbd /home/kam/Research/FromLaptop/testdocs/archaeol6-69760_1.docx
bacd0311caaadfd0de8d7c38bf15d64 /home/kam/Research/FromLaptop/testdocs/archaeol21-81016_2.docx
77a39ca3ac7124c7374ad146be17c1b8 /home/kam/Research/FromLaptop/testdocs/archaeol21-81048_1.docx
a3d9c64b823c95a59c9309de6d3842b0 /home/kam/Research/FromLaptop/testdocs/archaeol21-81740_1.docx
7f76bebb6ad6c1bb30c14a6a35befad3 /home/kam/Research/FromLaptop/testdocs/archaeol21-83991_1.docx
b6279b34c082c6abfa9686e90fb91798 /home/kam/Research/FromLaptop/testdocs/archaeol21-85873_2.docx
b393d82ca450dfe85dbbd82fbfaa4e82 /home/kam/Research/FromLaptop/testdocs/archaeol21-105745_1.docx
7e8933260fea7f3ade57f255f8c102a4 /home/kam/Research/FromLaptop/testdocs/cameraona1-88343_1.docx
9ec5781e7d6ca810da7ae3da43c3e1e7 /home/kam/Research/FromLaptop/testdocs/contexto1-66217_1.docx
d08bf83b14c92221b8f7147bec7973c5 /home/kam/Research/FromLaptop/testdocs/contexto1-69471_1.docx
3c85ecd7e128bf4b3ccbba8da3993010 /home/kam/Research/FromLaptop/testdocs/contexto1-79006_1.docx
61ed85e3d39fa22fc43b74ae30836f96 /home/kam/Research/FromLaptop/testdocs/contexto1-80765_1.docx
b19b44cd870acd5d546d829a7b7f2e0d /home/kam/Research/FromLaptop/testdocs/contexto1-85055_1.docx
b5aea481497bf36f936e545653a6a3c2 /home/kam/Research/FromLaptop/testdocs/dennisp1-58597_2.docx
51bf7f5bc623c5ff9db88a154541a42f /home/kam/Research/FromLaptop/testdocs/dennisp1-70846_3.docx
adf11d9c854486c70db3746e252e8e18 /home/kam/Research/FromLaptop/testdocs/dennisp1-76116_2.docx
c28d100f1f064de7fc365a660e470789 /home/kam/Research/FromLaptop/testdocs/dennisp1-83743_1.docx
b9345190327e0debb113e287c258ea7a /home/kam/Research/FromLaptop/testdocs/dennisp1-91925_1.docx
b224123f788d7c0fd7a73a2dfb34a219 /home/kam/Research/FromLaptop/testdocs/national3-99879_1.docx
b4bce311d0c97b4009b3cdc6d50529ae /home/kam/Research/FromLaptop/testdocs/quaterna1-62216_1.docx
e245d76ce027cd79309024ab9e83c6d1 /home/kam/Research/FromLaptop/testdocs/scotiaar1-104626_2.docx
bb0eccd64d98db2660c909c276f30398 /home/kam/Research/FromLaptop/testdocs/scotiaar1-104630_1.docx
5b7bb2476374212e849e1513923c670d /home/kam/Research/FromLaptop/testdocs/scotiaar1-104632_2.docx
638cc2ed4562380fdad3bb87d570250a /home/kam/Research/FromLaptop/testdocs/scotiaar1-104662_10.docx
e440d9f4f6eaa9dc188f9b5fb8ad23c7 /home/kam/Research/FromLaptop/testdocs/scotiaar1-106279_2.docx
4ed53bc15578a68f1ec8cc14ac2a9503 /home/kam/Research/FromLaptop/testdocs/scotiaar1-106334_1.docx
bb1e322e94e13bc91a57be59f7a905ab /home/kam/Research/FromLaptop/testdocs/universi1-84908_1.docx
bb1e322e94e13bc91a57be59f7a905ab /home/kam/Research/FromLaptop/testdocs/universi1-84908_2.docx
5f8ef69582e68b68d2ad9ecf8e37e564 /home/kam/Research/FromLaptop/testdocs/universi1-84914_1.docx
```

Plain Text Tab Width: 8 Ln 1, Col 1 INS

quaterna1-62216_1.docx scotiaar1-104626_2.docx scotiaar1-104630_1.docx scotiaar1-104632_2.docx

"md5Results.md5" selected (12.1 kB)

testdocs

- Presentations
- Public
- RandallAir
- Research
 - Agreement
 - BakSVN
 - BCiaB
 - DOMEX
 - FromLaptop
 - AFF4
 - bak3
 - CALFILES
 - CRADLE
 - DCC2010
 - Dissertation
 - docXtractor
 - DPapers
 - findntuser
 - FromStore
 - JOBAPP
 - M57
 - NewHomepage
 - NSFProp2009
 - PDA2011
 - Scripting and Programming
 - ssh-bak
 - testdocs

Conclusions: Implied Changes with the Archival Profession

- Professional vocabulary evolving to include terms such as disk image, hex viewer, cryptographic hash, and filesystem
- Gaining access to new professional communities and guidance
- Use of tools designed to treat data at a low level – as raw bitstreams off media – rather than at the file level
- Potential to shift “center of gravity” about electronic records from design of institutional recordkeeping systems toward acquisition and management of records from a more diverse and unpredictable set of sources

Thank you!

<http://bitcurator.net>

<http://wiki.bitcurator.net>

Twitter: @bitcurator