

```
$> Preserve, redact, discard? [P/r/d]  
$> Human-Readable Reporting on Complex Data Sources  
to Support Data Triage Decisions
```

Kam Woods, BitCurator Technical Lead
University of North Carolina at Chapel Hill SILS
CurateGear 2013 - January 9



The Andrew W. Mellon Foundation

BitCurator

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

The BitCurator Project is an effort to build, test, and analyze systems and software for incorporating digital forensics methods into the workflows of a variety of collecting institutions.

BitCurator uses open source and public domain software drawn from the digital forensics community, along with internally developed GPLv3 licensed software to assist in analyzing born-digital materials.

Intended to work alongside or augment existing preservation and archiving environments.

Two ways to use the software: as part of a pre-configured virtual machine, or as independent software modules and scripts.

BitCurator

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

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

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

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

Professional Experts Panel	Development Advisory Group
<ul style="list-style-type: none">• Bradley Daigle, University of Virginia Library• Erika Farr, Emory University• Jeremy Leighton John, British Library• Leslie Johnston, Library of Congress• Courtney Mumma, City of Vancouver Archives• 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	<ul style="list-style-type: none">• 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



Source: Simson Garfinkel

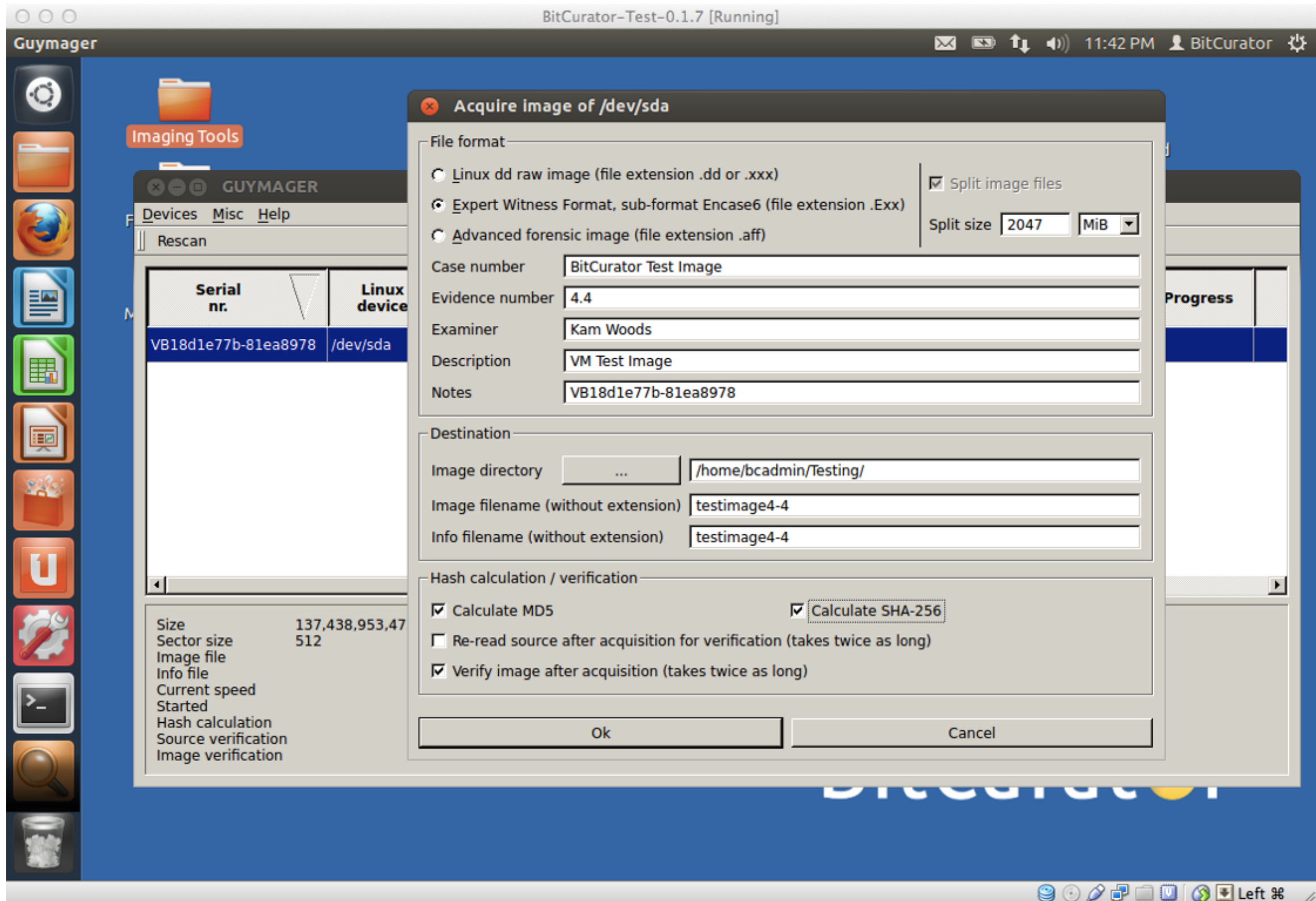


Source: BitCurator

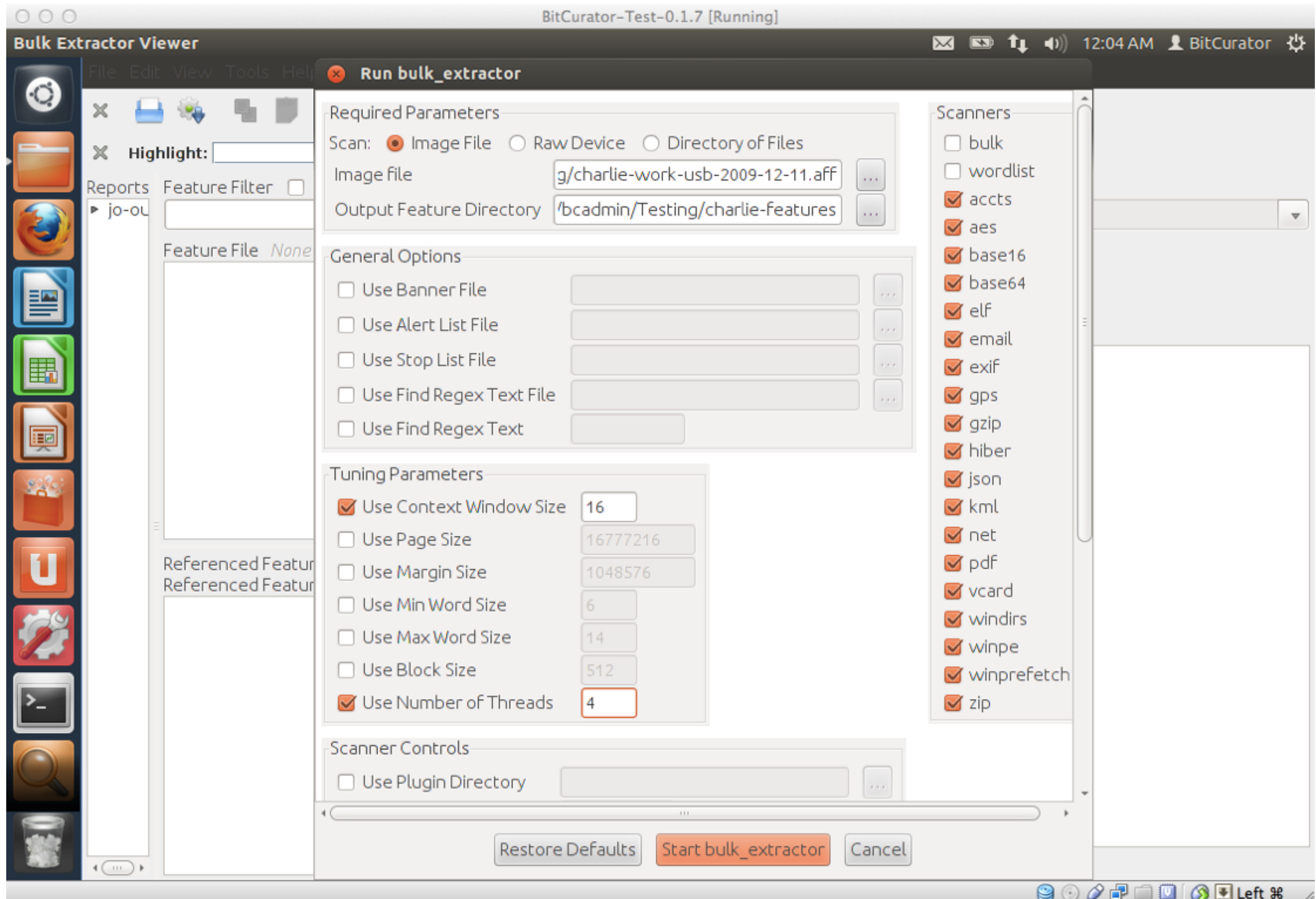


Source: "Digital Forensics and creation of a narrative." *Da Blog: ULCC Digital Archives Blog*. <http://dablog.ulcc.ac.uk/2011/07/04/forensics/>

How do I know I'm not changing anything on the disk?



How do I know I've found everything there is to find?



How do I figure out where the important things are?

BitCurator-Test-0.1.7 [Running]

Bulk Extractor Viewer

File Edit View Tools Help

Highlight: ☒ Match case

Reports

- jo-output
 - domain.txt
 - domain_histogram
 - email.txt
 - email_histogram.txt**
 - json.txt
 - url.txt
 - url_histogram.txt
 - url_services.txt
 - windirs.txt

Feature Filter ☒ Match case

Histogram File email_histogram.txt

n=10	gross.joshua.b@gmail.com
n=8	hous-daccq-1369054661@craigslist.org
n=6	3c8ab9a1f3055711468ef4a7185fba9f13
n=6	amsuich@nps.edu
n=4	3c4a527e0e.4000600@pitt.edu
n=4	cnbc-all@cnbc.cmu.edu
n=4	leonardochiesi@gmail.com
n=4	mathbio@math.pitt.edu
n=2	3c2acb011c0907060808k4ece07eal4334
n=2	bard@math.pitt.edu
n=2	buy.com_offers@enews.buy.com
n=2	bw3maggars@gmail.com
n=2	cherylseekingforoom1@gmail.com
n=2	daughtry@psu.edu
n=2	glenn.gunzelmann@mesa.afmc.af.mil

Referenced Feature File email.txt

Referenced Feature leonardochiesi@gm...

727925	leonardochiesi@gmail.com
733303	leonardochiesi@gmail.com
727925	leonardochiesi@gmail.com
733303	leonardochiesi@gmail.com

Navigation

jo-work-usb-2009-12-11.aff, 727925, leonardochiesi@gmail.com

Image File jo-work-usb-2009-12-11.aff

Feature File email.txt

Feature Path 727925

Feature leonardochiesi@gmail.com

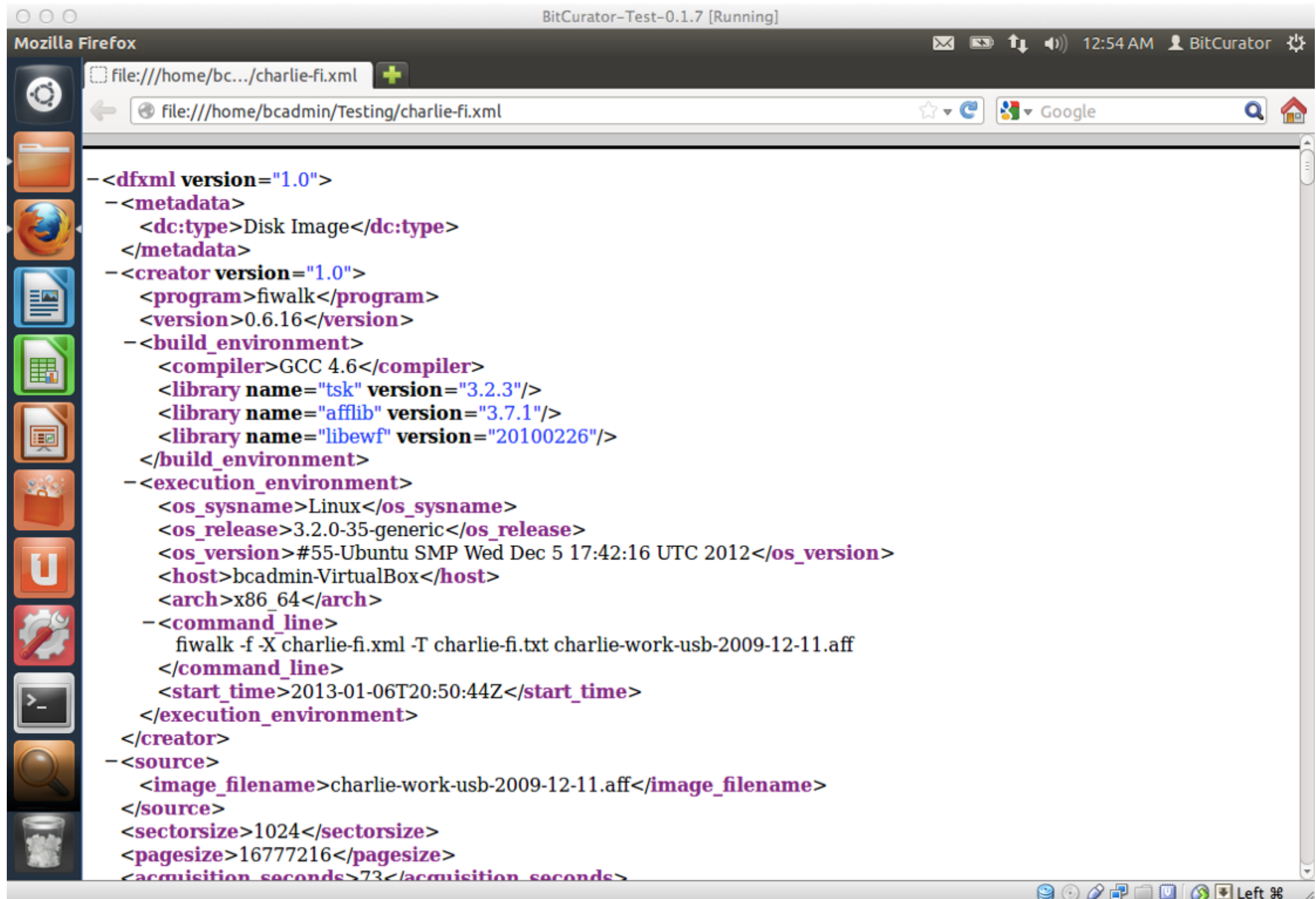
Image

```
726208 ...R.I.....?.....ig.oyment Site.....Q.....
726272 .n...A...7.....hræku.edu...0.....V...P.....
726336 .9.....gross.joshua.b+job@gmail.com...9.....p...A.....
726400 .....A...0.....y.....0...&...o...'..w.....h/...
726464 ...@.....2...L...+.....-...../...0...1.....2
726528 ...3.....4...5.....6.....8.....Joshua Gross....
726592 ...../...Gunzelmann, Glenn F Civ USAF AFMC 711 HPW/R
726656 HAC...../.....46.emlx.....?.....RE: update a
726720 nd CV.....R.....N...A...3.....RE: update and CV.....
726784 .....Q.....|...A...7..."...Glenn.Gunzelmann@mesa.afmc.
726848 af.mil...0.....V...P.....9.....gross.joshua.
726912 b@gmail.com...9.....p{.A.....N...A...0.....
726976 ..y.....0...}.....(.w.....y/.....@.....2...L...+..
727040 .....-...../...0...1.....2...3.....4...5.....6..
727104 .....8.....mathbio@math.pitt.edu.cnbc-all@cnbc.cmu.edu
727168 .....G. Bard Ermentrout...../.....
727232 .....23.emlx.....?.....Special seminar.....R.....
727296 A...3.....Special seminar.....Q.....?...A..
727360 .7.....bard@math.pitt.edu...0.....V...P.....
727424 .9.....mathbio@math.pitt.edu.cnbc-all@cnbc.cmu.edu...9.....
727488 .....A.....A...0.....y.....0.....n~.w
727552 .....7.....@.....2...L...+.....-...../..
727616 ..0...1.....2...3.....4...5.....6.....8.....Jo
727680 shua Gross.....Leonardo Chiesi...../.....
727744 .....40.partial.emlx.....?.....Re: hello and congra
727808 tulations.....R.....K..A...3.....Re: hello and congratu
727872 lations.....Q.....Ih.A...7.....leonardochi
727936 esi@gmail.com...0.....V...P.....9.....gross.
728000 joshua.b@gmail.com...9.....p...A.....K..A...0
```

Text Hex

Left %

What if I just want the metadata?



```
<?xml version="1.0"?>
<dfxml version="1.0">
  <metadata>
    <dc:type>Disk Image</dc:type>
  </metadata>
  <creator version="1.0">
    <program>fiwalk</program>
    <version>0.6.16</version>
    <build_environment>
      <compiler>GCC 4.6</compiler>
      <library name="tsk" version="3.2.3"/>
      <library name="afflib" version="3.7.1"/>
      <library name="libewf" version="20100226"/>
    </build_environment>
    <execution_environment>
      <os_sysname>Linux</os_sysname>
      <os_release>3.2.0-35-generic</os_release>
      <os_version>#55-Ubuntu SMP Wed Dec 5 17:42:16 UTC 2012</os_version>
      <host>bcadmin-VirtualBox</host>
      <arch>x86_64</arch>
    </execution_environment>
    <command_line>
      fiwalk -f -X charlie-fi.xml -T charlie-fi.txt charlie-work-usb-2009-12-11.aff
    </command_line>
    <start_time>2013-01-06T20:50:44Z</start_time>
  </creator>
  <source>
    <image_filename>charlie-work-usb-2009-12-11.aff</image_filename>
  </source>
  <sectorsize>1024</sectorsize>
  <pagesize>16777216</pagesize>
  <acquisition_seconds>73</acquisition_seconds>
</dfxml>
```


What if I'm not familiar with this kind of metadata?

BitCurator-Test-0.1.7 [Running]

LibreOffice Calc

Documentation and Help

DFXML tag library v3.xlsx - LibreOffice Calc

File Edit View Insert Format Tools Data Window Help

A62 f(x) Σ = <compiler>

	A	B	C	D
1	Tag name	Element name	Description	May contain
2	<dfxml>	DFXML	Root element, marks the beginning and end of the DFXML metadata file. The <dfxml> element contains the primary elements reported in fiwalk's xml structure: <metadata>, <creator>, <source>, <volume>, and <runstats>.	<metadata>, <creator>, <source>, <volume>, <runstats>, <sectorsize>, <pagesize>, <acquisition seconds>
3	<metadata>	Metadata	The <metadata> tag provides header information that defines the metadata in the DFXML document. Includes namespace declaration, namespace schema location, and other information that is used to define the elements used in the XML file.	<dc:type>, <dc:creator>, <dc:title>, <dc:description>; for more information on Dublin Core element set, see (21).
4			These declarations provide information on the types of standardization schemes used to convey information in the DFXML document. The <metadata> tag may also contain high level descriptive information about the DFXML document rendered in Dublin Core (dc), in order to increase interoperability.	
	<creator>	Creator	The Creator element provides documentation about the program and computing environment in which the disk analysis (or capture) take place. <Creator> includes tags documenting the program that initiated the capture creating the DFXML file, and other contextual information about the system on which	<program>, <version>, <build_environment>, <execution_environment>

Sheet 1 / 5 PageStyle_fiwalk STD Sum=0 100%

What if I need information on the file system(s)?

BitCurator-Test-0.1.7 [Running]

Document Viewer

FiwalkReport.pdf — Bitcurator Report

Previous Next 1 (1 of 6) 125%

Thumbnails

1

2

3

Report: File System Statistics and Files

BitCurator

Technical Metadata

Disk Image: image_filename: charlie-work-usb-2009-12-11.aff

Feature	Value
SECTORSIZE	1024
FTYPE STR	ntfs
PARTITION OFFSET	512
BLOCK SIZE	4096
ACQUISITION SECONDS	73
FIRST BLOCK	0
BLOCK COUNT	258559
LAST BLOCK	258558
PAGESIZE	16777216
FTYPE	1
IMAGE FILENAME	charlie-work-usb-2009-12-11.aff
Number of Files	128
Total Directories	23
Total Deleted Files	0
Total Unused Files	0
Files with Nlinks > 1	0

How about just the private and sensitive information?

BitCurator-Test-0.1.7 [Running]

Document Viewer

BeReport.pdf

Previous Next 1 (1 of 1) Fit Page Width

Thumbnails

1

Report: Bulk Extractor Features

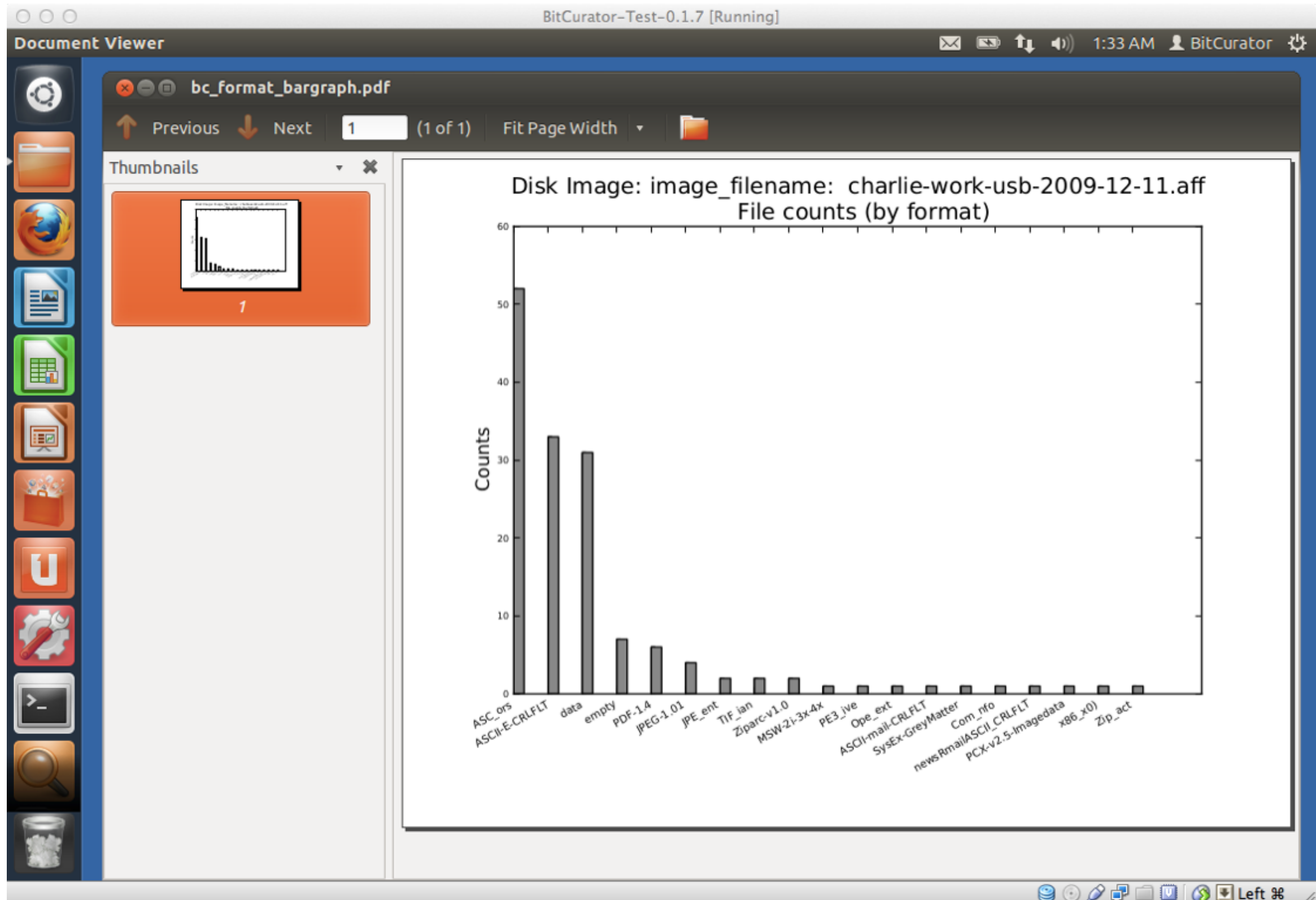
BitCurator

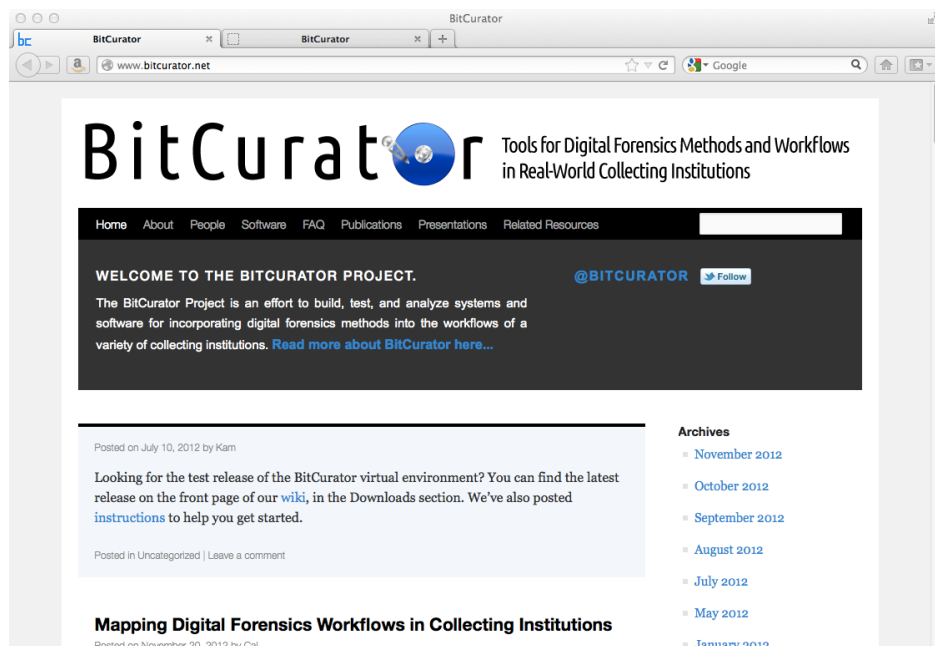
Note:
FIUF: Total features unallocated to files
FIUF: Total features unallocated to files
FICR: Total features in compressed regions

Bulk Extractor Report Files	Feature Instances	FLTF	FIUF	FICR
annotated_telephone.txt	5	4	1	2
annotated_rfc822.txt	258	39	219	110
annotated_zip.txt	127	8	119	3
annotated_windirs.txt	466	13	453	180
annotated_domain.txt	653	48	605	317
annotated_exif.txt	2	2	0	0
annotated_winpe.txt	1	1	0	0
annotated_email.txt	500	42	458	224

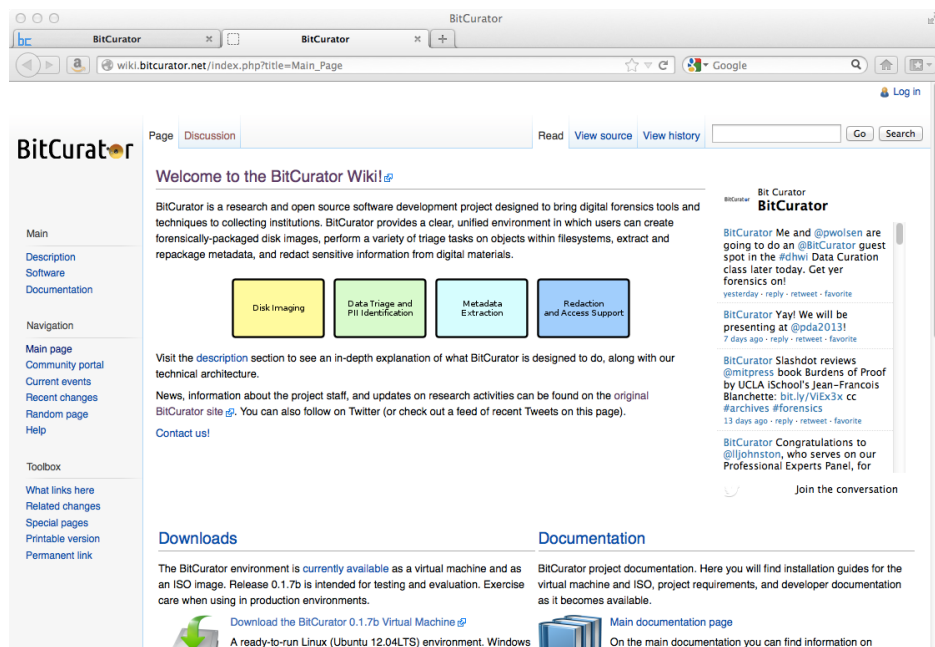
Left

What if I just need a visual cue for what's on the media?





People
Project overview
News
<http://www.bitcurator.net/>



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

BitCurator

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

Thank you!



The Andrew W. Mellon Foundation