

BitCurator: Digital Forensics for Collecting Institutions

Kam Woods

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

BitCurator



UNC
SCHOOL OF INFORMATION
AND LIBRARY SCIENCE

MITH MARYLAND INSTITUTE FOR
TECHNOLOGY IN THE HUMANITIES

The Andrew W. Mellon Foundation

CurateGEAR 2014 - January 8, 2014 - University of North Carolina at Chapel Hill

What it is

- BitCurator packages forensics software (in a Linux virtual machine) to augment digital curation workflows in real-world collecting institutions:
 - Reliable **acquisition of bitstreams** from digital media
 - **Scalable forensic analysis** of disk images / heterogeneous collections of documents
 - **Metadata acquisition** from common file systems
 - **Preservation metadata** and finding aid metadata export
 - **Data visualization** for complex data sources
 - **Redaction** of private and sensitive information
 - Identification and **removal of duplicate files**

Why it's different

Commercial forensics tools are typically **expensive**, and not always as **reliable** as you'd expect.

Lots of things important to **collecting institutions** that aren't a **priority** for forensics developers: **access, redaction, preservation metadata**, etc...

BitCurator packages open source forensics tech into a simple-to-use virtual environment to help you avoid getting bogged down.



Image credit: <http://www.quickmeme.com/>

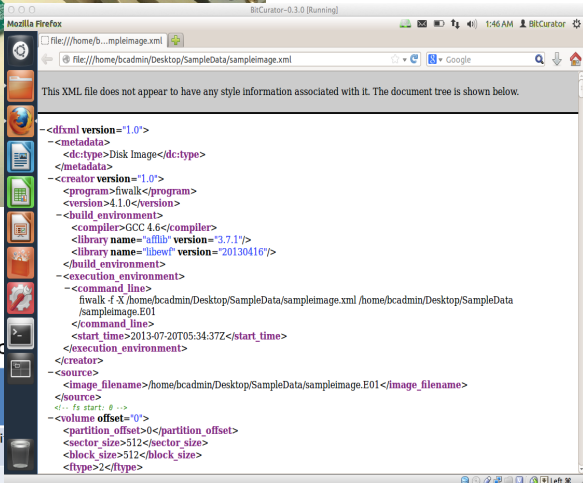
The cruising altitude view

Write-blocked physical media acquisition

Creating Digital Forensics XML to describe disk images

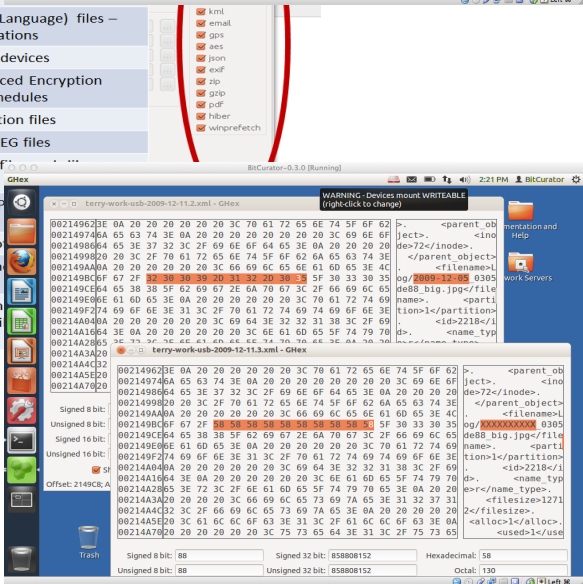
Identifying potentially private information

Selective data redaction



Bulk Extract

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 file streams
scan_gzip	Detects and decompresses GZIP file 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 file fragments

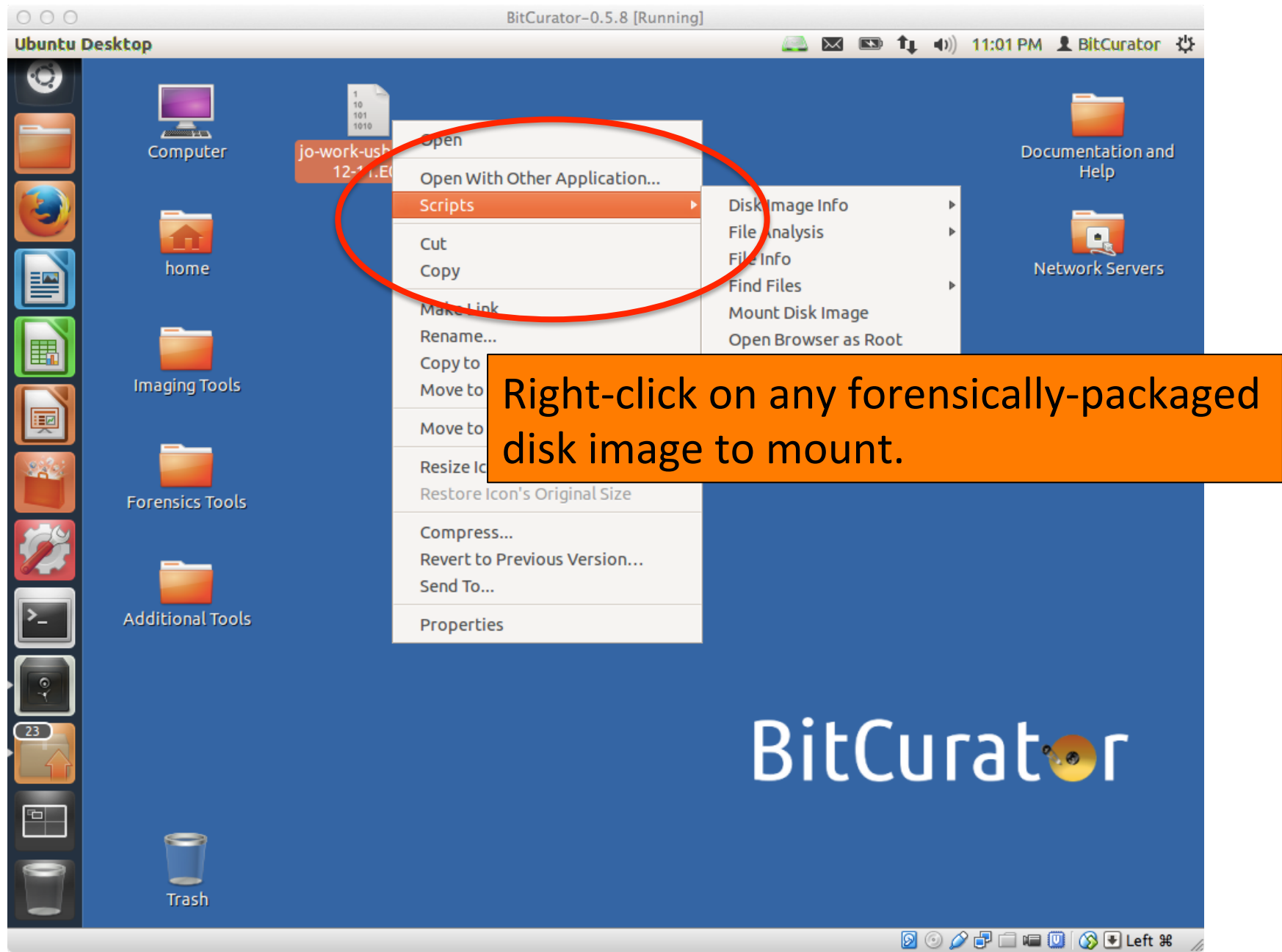


What's new in the past year

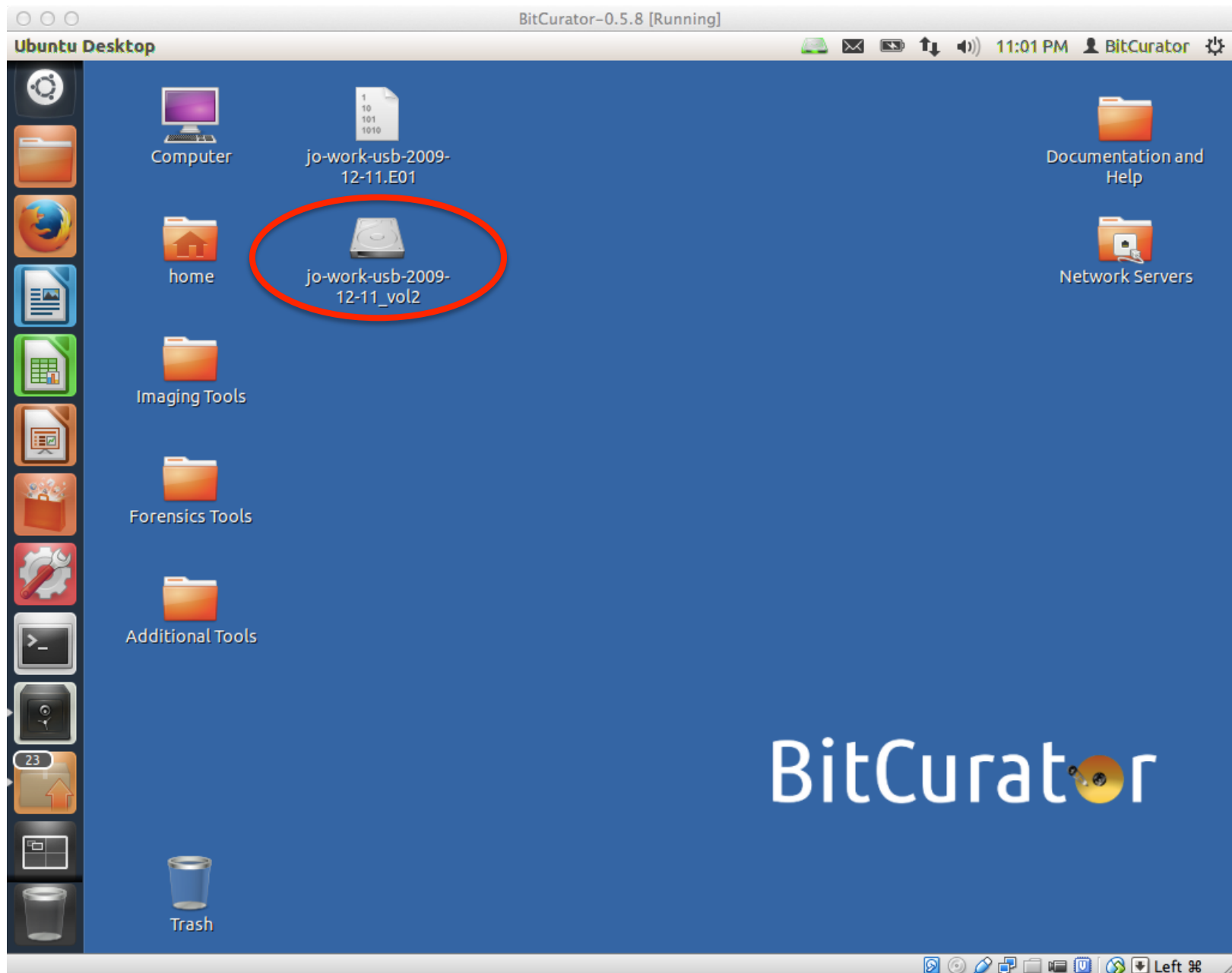
BitCurator has had 15 public releases (4 major releases) in the past year. Some highlights (as of 0.5.8):

- BitCurator-specific GUI to assist collections professionals running complex forensic tools.
- One-click mounting of forensically-packaged disk images
- Improved metadata export facilities (PDF, .xlsx, and visualizations)
- PREMIS metadata generation for forensic events and objects
- DFXML improvements and participation in the DFXML working group
- Duplicate file detection and removal
- Byte-level redaction of pattern-matched features (e.g. potentially sensitive information)

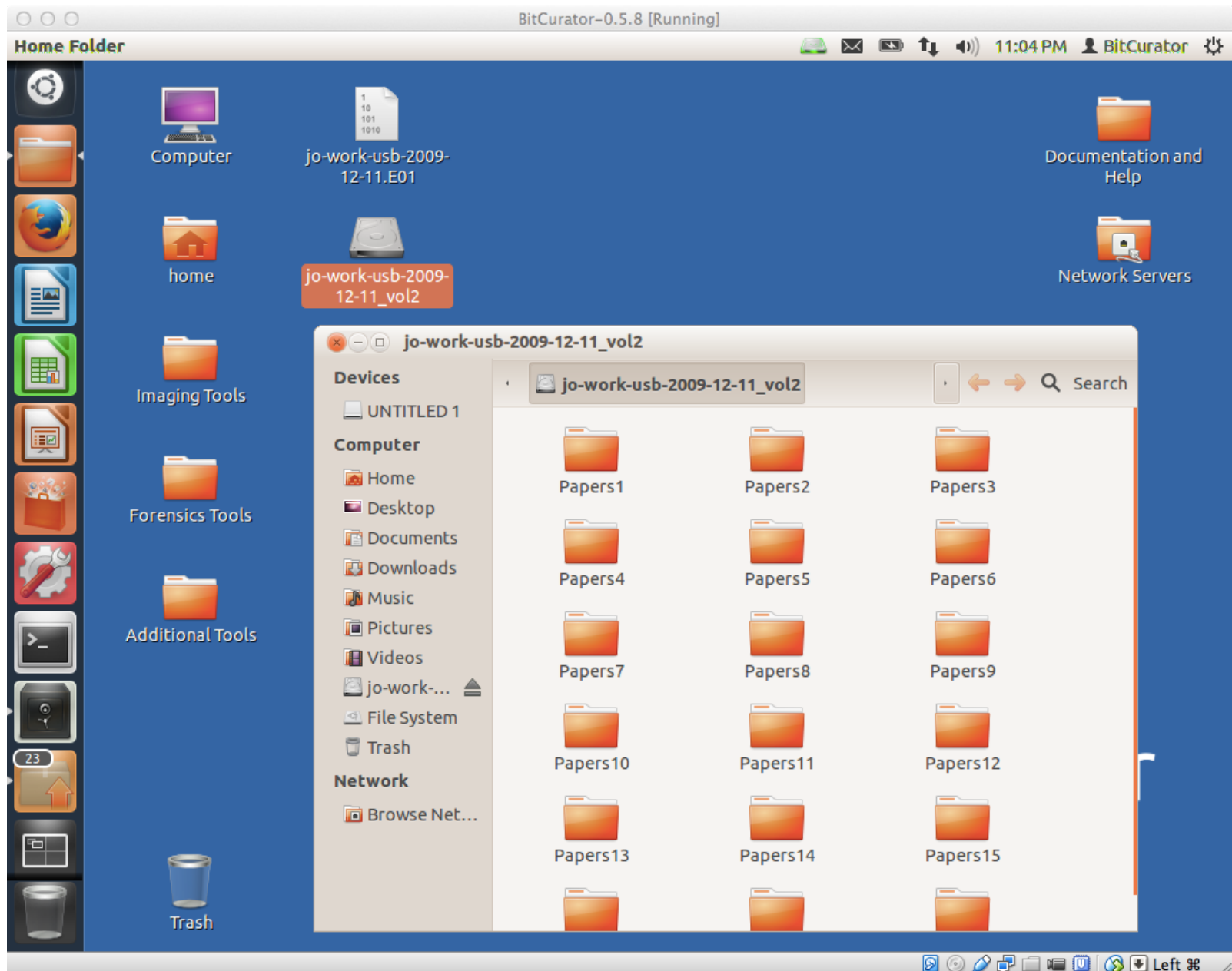
Interacting with disk images



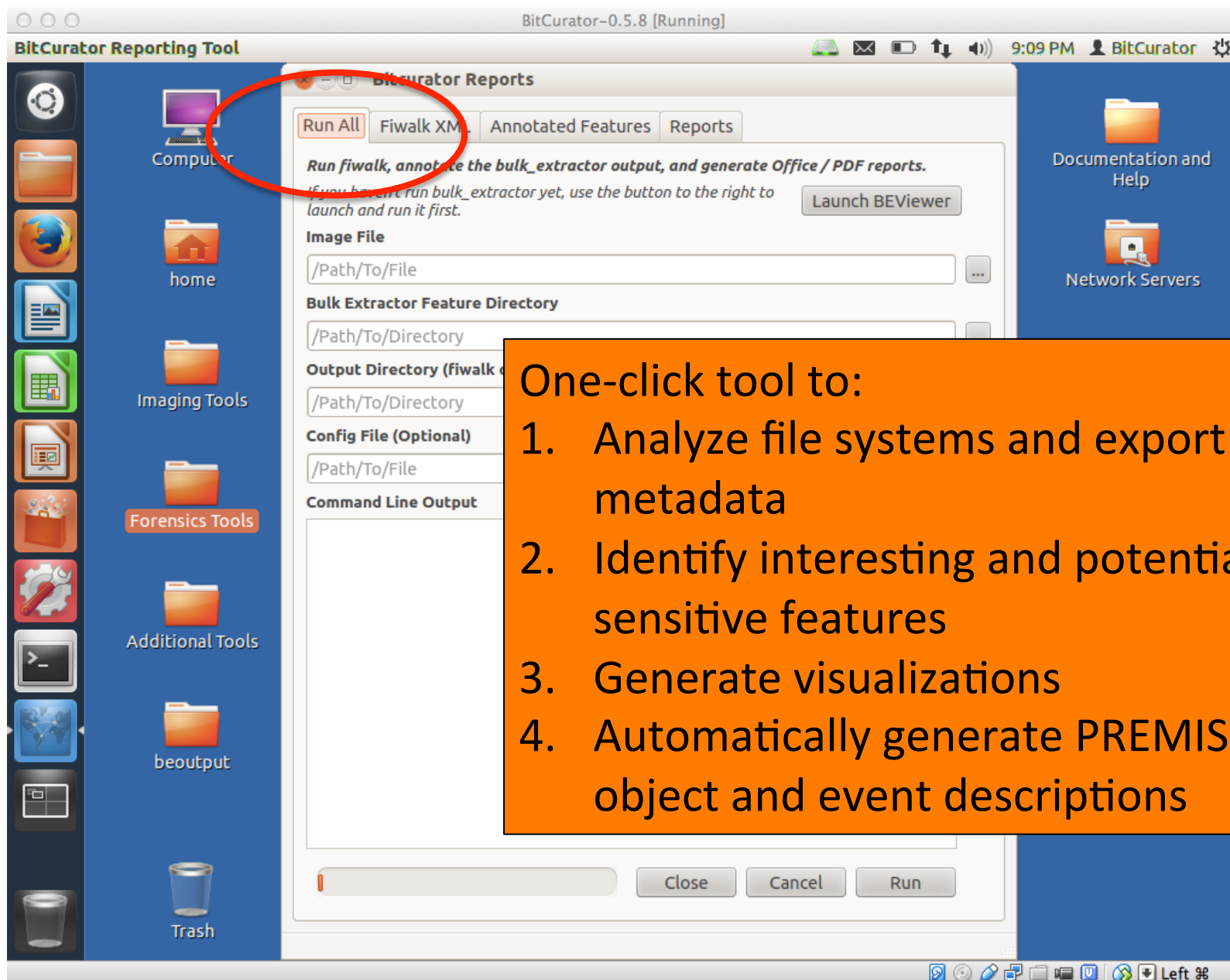
Interacting with disk images



Interacting with disk images



Forensic workflow



PREMIS metadata

BitCurator-0.5.8 [Running]

Mozilla Firefox

file:///home/bc...rts/premis.xml

file:///home/bcadmin/Desktop/bcoutput/reports/premis.xml

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

```
-<premis version="2.0" xsi="http://www.w3c.org/2001/XMLSchema-instance">
  -<object>
    -<objectIdentifier>
      <objectIdentifierType>1ba2e916-7718-11e3-9957-080027fa842e</objectIdentifierType>
      <objectIdentifierValue>/home/bcadmin/Desktop/jo-work-usb-2009-12-11.E01</objectIdentifierValue>
    </objectIdentifier>
    -<event>
      -<eventIdentifier>
        <eventIdentifierType>1ba39af0-7718-11e3-9957-080027fa842e</eventIdentifierType>
        -<eventIdentifierValue>
          E01/home/bcadmin/Desktop/jo-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>1f83cc12</eventIdentifierType>
      -<eventIdentifierValue>
        fiwalk -f -X /home/bcadmin/bcoutput/usb-2009-12-11.E01
      </eventIdentifierValue>
    </eventIdentifier>
    <eventType>File System Analysis</eventType>
  </event>
</premis>
```

Forensic analysis events with UUIDs for each (automatic) step in BitCurator tool (using custom BitCurator PREMIS generation code)

File system metadata...simplified

BitCurator-0.5.8 [Running]

fiwalk-output.xml.xlsx - LibreOffice Calc

File Edit View Insert Format Tools Data Window Help

Calibri 11

C8 $f(x)$ Σ = pdf

	A	B	C	D	E	F	G	H
1	Partition	Filename	Extension	Filesize	Access time	Create time	Modification	MD5 Hash
2	1	Papers3/62372.InsidiousObi.Allan+Jo	pdf	91940	2009-12-10T05:00:00	2009-12-10T19:32:08	2009-11-1	5a6953fd9af383
3	1	Papers3/63572.ControllingIPv4.Amie	pdf	65192	2009-12-10T05:00:00	2009-12-10T19:32:09	2009-11-1	70bd635e9b10f
4	1	Papers3/66312.CacheCoherence.Nelo	pdf					
5	1	Papers3/66976.baronet.Michael+Jone	pdf					
6	1	Papers3/71660.MeshNetworks.Steven	pdf					
7	1	Papers3/73178.SeparatedRobotics.M	pdf					
8	1	Papers3/73190.LANs.Clinton+Vallejo	pdf					
9	1	Papers3/73461.SMPs.Allan+Bearse.p	pdf					
10	1	Papers3/74074.EPOPT.Erik+Drown.p	pdf					
11	1	Papers3/74663.DHCP.Kurt+Tunney.p	pdf					
12	1	Papers3/74927.Database.Richard+D	pdf					
13	1	Papers3/75479.Redundancy.Erik+Tro	pdf					
14	1	Papers3/75825.Agents.Jessie+Calfee	pdf					
15	1	Papers3/77170.P2PArchetypes.Helen	pdf					
16	1	Papers3/79015.IPv7.Earnestine+Rush	pdf	104601	2009-12-10T05:00:00	2009-12-10T19:32:11	2009-11-1	1761dddc8ac7
17	1	Papers3/81007.Smalltalk.Mark+Whit	pdf	54085	2009-12-10T05:00:00	2009-12-10T19:32:11	2009-11-1	3b6a5c8da2532
18	1	Papers3/81784.LocIDPLitvsSkene.Jas	pdf	96018	2009-12-10T05:00:00	2009-12-10T19:32:11	2009-11-1	555f7673696a5
19	1	Papers3/82599.ExpertSystems.Lakis	pdf	68012	2009-12-10T05:00:00	2009-12-10T19:32:12	2009-11-1	4062535ae7bd8
20	1	Papers3/83622.MooresLaw.Thomas+	pdf	63667	2009-12-10T05:00:00	2009-12-10T19:32:12	2009-11-1	b70b8a240c6ae
21	1	Papers3/84160.Write-AheadLogging	pdf	56376	2009-12-10T05:00:00	2009-12-10T19:32:12	2009-11-1	b755683b81c5a
22	1	Papers3/8617.Sou.Sharron+Popejoy	pdf	57563	2009-12-10T05:00:00	2009-12-10T19:32:12	2009-11-1	1221223154836
23	1	Papers3/87310.IPv4.Susan+Hill.pdf	pdf	47110	2009-12-10T05:00:00	2009-12-10T19:32:12	2009-11-1	63497a1839f0b
24	1	Papers3/88094.RAID.William+Brown	pdf	73765	2009-12-10T05:00:00	2009-12-10T19:32:12	2009-11-1	10785254252b0
25	1	Papers3/89922.ActiveNetworks.Mary	pdf	71156	2009-12-10T05:00:00	2009-12-10T19:32:13	2009-11-1	f78a8d47150ccc
26	1	Papers3/89922.ActiveNetworks.Nann	pdf	71193	2009-12-10T05:00:00	2009-12-10T19:32:13	2009-11-1	56011776332dc
27	1	Papers3/91750.LocalIDSplit.Carlene	pdf	51852	2009-12-10T05:00:00	2009-12-10T19:32:13	2009-11-1	11b1ccd01a65b

Per-partition file items with file type, MAC times, MD5 and SHA1 hashes automatically generated via BitCurator GUI

File Object Information

Sheet 1 / 1 PageStyle_File Object Information STD Sum=0 100%

Things can still get messy...



Image credit: <http://www.quickmeme.com/>

...and we know there are some issues

- Main forensic workflow **does not support HFS volumes** (only HFS+ - but we have other tools that can help)
- Disk image mount scripts **fail on more exotic disks** (multiple volumes)
- Really nice GUIs like Autopsy don't run well (or at all) in Ubuntu LTS (12.04) without some major fiddling.
- **VirtualBox doesn't have USB 3.0 support...yet.**
- Distributing large VMs is a **bandwidth hog**. (we've partnered with iBiblio to alleviate this)
- **Preservation metadata** is easy to generate but hard to make useful for those with limited infrastructure
- **Likewise for EAD**
- **Redaction** of private and sensitive information is not a major use case for forensics investigators. We're rolling our own.

We're making it easier to upgrade BitCurator and use modules individually

```
bcadmin@ubuntu: ~/Tools/bitcurator$ ./bc-preinst.sh
```

```
##### C #####  
## ## @ ## ## ##  
## ## ## ##  
## @ ## #### ## ## #####  
## ## ## ## ## ## ## ## ## ## llll #####  
## ## ## ## ## ## ## ## ## ## lllllll ,@  
##### ## ## ## ## ## ## ## ## ## lllllll ,@  
##### ## ## ## ## ## ## ## ## ## ;llll,llll ,@  
## ## ## ## ## ## ## ## ## ## tltt,:ttt ,@  
## ## ## ## ## ## ## ## ## ## tftfl:LLLf ,@  
## ## ## ## ## ## ## ## ## ## fLLCGGGGC ,@  
## ## ## ## ## ## ## ## ## ## LCGGGGGG ,@  
##### ## ## ##### ##### ## ##### CGGGGG ,@  
  
Welcome to the BitCurator Installer!  
Ok, let's get started!
```

It looks like your current working directory is /home/bcadmin/Tools/bitcurator.
We'll be installing some software in /home/bcadmin.

-- Would you like to install git, in order to download BitCurator? -- (y/N)

Scripts with simple YES/NO questions to step you through pulling software from our Git repo, upgrading, and installing dependencies

Phew! Time to go use it!

Software and docs: <http://wiki.bitcurator.net/>

News and blog: <http://www.bitcurator.net/>

Source code: <https://www.github.com/kamwoods/bitcurator>



BitCurator  MITH MARYLAND INSTITUTE FOR
TECHNOLOGY IN THE HUMANITIES



UNC
SCHOOL OF INFORMATION
AND LIBRARY SCIENCE

**The Andrew W. Mellon
Foundation**