Digital Forensics: Advanced

Instructors:
Christopher (Cal) Lee
University of North Carolina at Chapel Hill

Kam Woods
University of North Carolina at Chapel Hill

December 13-14, 2021

SOCIETY OF AMERICAN ARCHIVISTS

About These Slides

Authors
Cal Lee, Kam Woods

Description
These are the slides from Cal Lee and Kam Woods's "Advanced Digital Forensics" class. There are a number of hands-on exercises included. The sample data referenced in these slides is available here: https://github.com/BitCurator/bcc-dfa-sample-data/

Learning object type
Lesson plan/materials

Learning objectives
This learning object might be used in a lesson to satisfy the following learning objectives:

- Identify the appropriate tools to: safely acquire born-digital materials from storage media and other modes of transfer; assist in the appraisal of born-digital materials; scan for sensitive information in born-digital materials; and package born-digital materials for preservation and access.
- Practice using tools in the BitCurator Environment.

Digital Archives Specialist (DAS)

Curriculum and Certification Program offered by SAA:

- Foundational Courses—*must pass 4*
- Tactical and Strategic Courses—*must pass 3*
- Tools and Services Courses—*must pass 1*
- Transformational Courses—*must pass 1*

- **Course examinations are administered online.**

DAS Core Competencies Addressed

- Understand the nature of records in electronic form, including the functions of various storage media, the nature of system dependence, and the effect on integrity of records over time.
- Formulate strategies and tactics for appraising, describing, managing, organizing, and preserving digital archives.
- Integrate technologies, tools, software, and media within existing functions for appraising, capturing, preserving, and providing access to digital collections.
- Curate, store, and retrieve original masters and access copies of digital archives.

https://www2.archivists.org/prof-education/das-curriculum-structure

Agenda

Day 1

- Welcome and introductions
- Motivation and overview
- Technical fundamentals
- Data acquisition considerations
- Potential elements of your own digital forensics lab
- Bit-level treatment of individual files
- Creating and extracting forensic metadata
- BitCurator reporting features
- Other BitCurator environment tools
- Preview of Day 2

Day 2

- Day 1 Postmortem – Questions, Concerns and Insights
- Command-line operations in Linux
- FIDO as an example of a command-line tool
- Regular expressions
- Extracting data from specific types of files: images, office files, email
- Windows artifacts (including the Registry)
- End user access (logistics and technical approaches)
- Incorporating digital forensics into archival workflows
- Challenges, Ethical/legal issues, and donor agreements
- Wrap up and evaluations

Personal Introductions

- Who's teaching you?
- What about you?
  - Who are you (name, institution, job title)?
  - Why are you here (relevance to job, what you hope to get out of the workshop)?
  - What have you done so far to apply digital forensics methods in your institution?

Software You Should Have Installed for the Exercises

- VirtualBox
- VirtualBox Extensions
- BitCurator Virtual Machine
- Exiftool (or on the web via https://exif.tools/)
- Sample data: https://distro.ibiblio.org/bitcurator/samples/saa-dfa-sample-data.zip
- Visit the link above and download this now if you have not already done so!
- Additional tools for Windows users:
  - FTK Imager (Windows only)
  - OSFMount (Windows only)
  - RegRipper (Windows GUI, can also run at the command line in BitCurator environment)

Discussion Scenario

- You've been charged with taking care of data from a prominent community leader who has died unexpectedly
- Her materials include some paper and lots of digital data (on floppies, CDs, and a laptop hard drive)
- What should you do with the floppies?
- CDs?
- Hard drive?

Goals When Acquiring Born-Digital 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 to Apply

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.

Digital Forensics 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	• 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	• Identify personally identifying information, regardless of where it appears • Flag for removal, redaction, closure or restriction

More Product, More (Machine) Processes

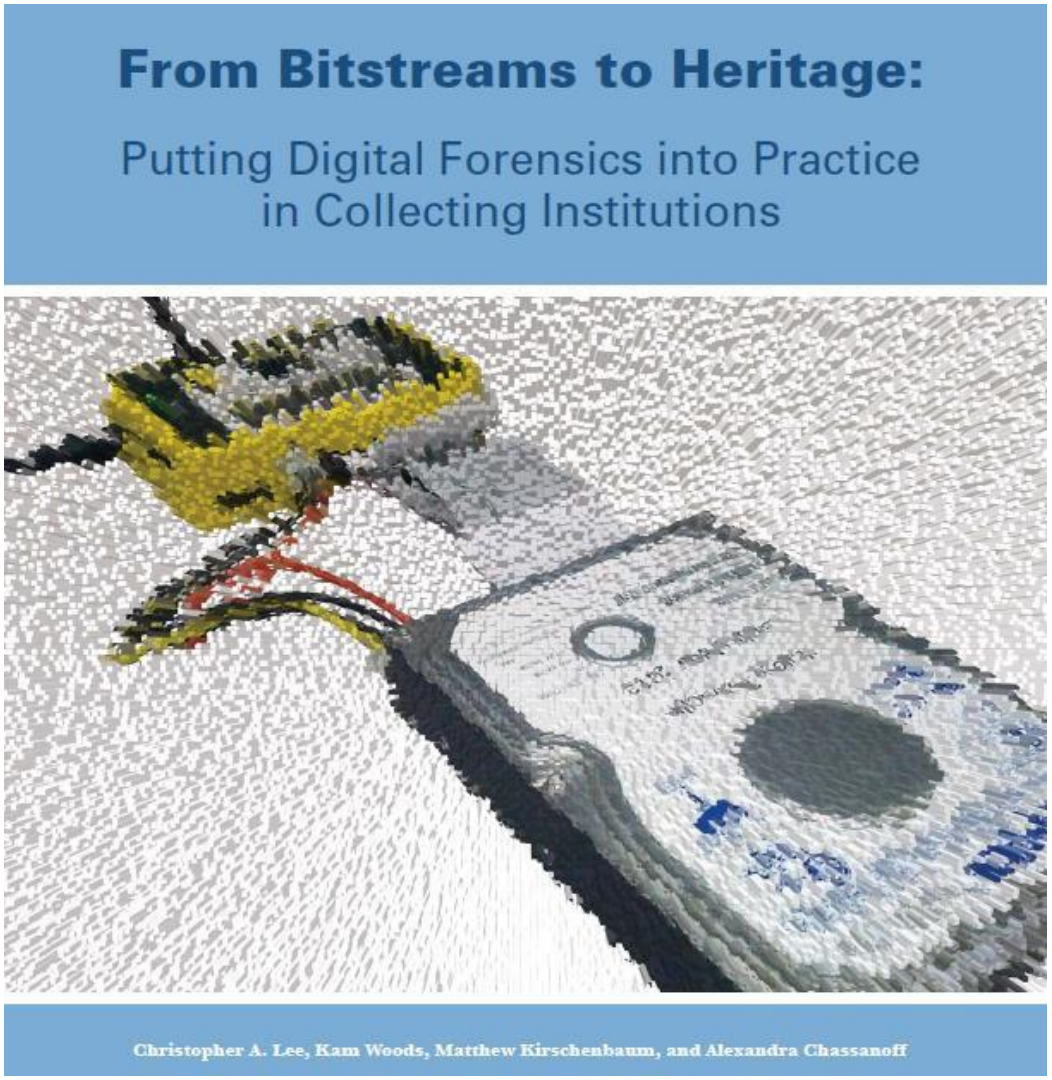
- Archivists need to apply many **more** processes to born-digital records (e.g. integrity checks, metadata extraction, audit trails, characterization)
- The good news is that most of these processes can be performed by **software**

Digital Forensics in Archives

- In recent years, archivists 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

Need for Adaptation of Digital Forensics Tools and Tasks for Archivists

- Existing digital forensics tools provide valuable functionality, but they don't always fit well into primary workflows of archives.
- For example, archives 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)



https://bitcurator.net/wp-content/uploads/sites/1099/2018/08/bitstreams-to-heritage.pdf

After this class, you should be able to:

- Install and operate the BitCurator environment as a virtual machine in VirtualBox
- Explain and recognize different types of metadata stored in common filesystems
- Identify file types based on magic numbers (file signatures)
- Determine potential hardware options for acquiring data from various types of storage media
- Apply common Linux commands at the command line and compose basic regular expressions
- Evaluate disk image format options based on needs and priorities of your institution and collections
- Generate BitCurator reports and use bulk_extractor to identify potentially sensitive data
- Extract and interpret EXIF metadata
- Capture and analyze Windows Registry artifacts using RegRipper
- Determine essential points in your institution's workflows where it will be beneficial to incorporate forensics tools and methods
- Make and justify decisions of professional ethics that emerge when caring for born-digital records
- Recognize technical strategies for providing access

Caveats and Such

- Advanced doesn’t mean "everything we didn’t cover in the Fundamental class"
- There’s much more about digital forensics that we won’t be addressing
- Selective hands-on experience with specific applications
- A license to learn more in the future

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

*Covered in Fundamental class. See also: 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 enc… …ociate paths an…
4	File as “raw” bitstream	Bitstream… …nary values
3	Sub-file data structure	Discrete…
2	Bitstream through I/O equipment	Series of… …media using inp… …trollers, 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

Levels where digital forensics methods and tools can provide a lot of assistance

*Covered in Fundamental class. See also: 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.

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)

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*

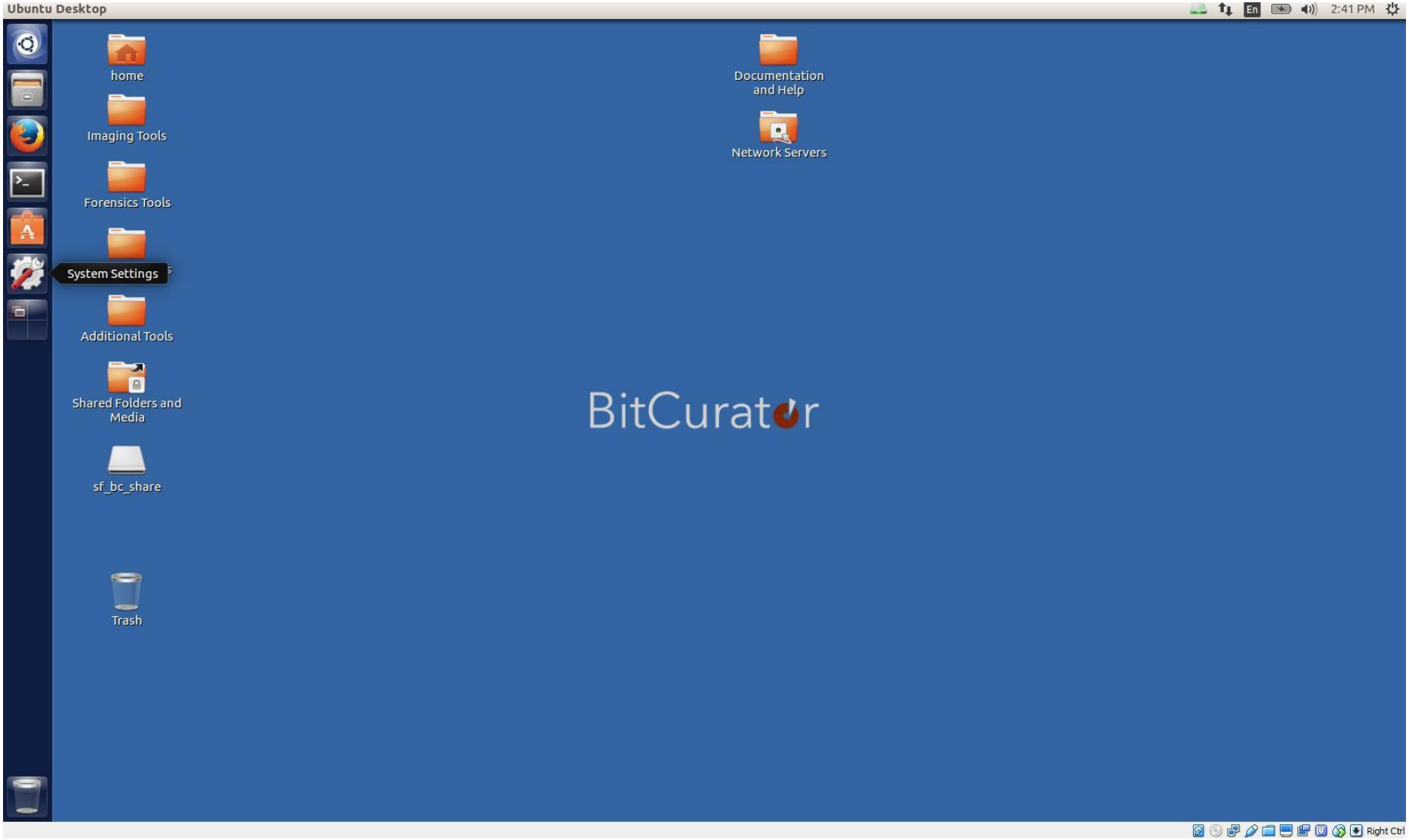
- Bundles, integrates and extends functionality of open source software
- Can be run as:
  - Self-contained environment (based on Ubuntu Linux) running directly on a computer (download installation ISO)
  - Using “bootstrapping” installation scripts to turn any Ubuntu Linux machine into a BitCurator Environment
  - 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: https://bitcurator.net/

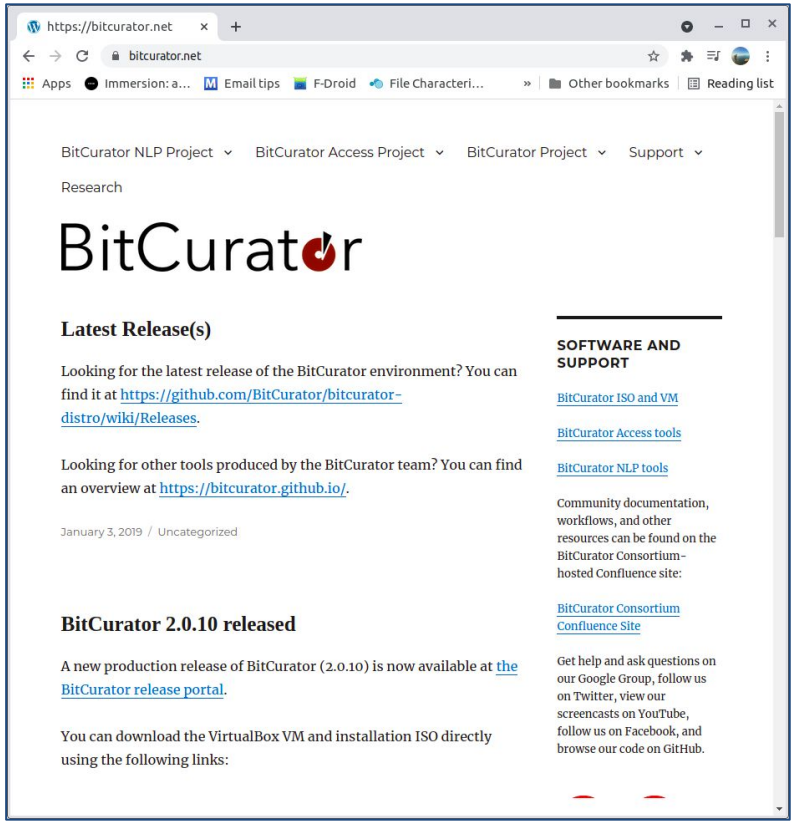
Hands-On Familiarization with VirtualBox and the BitCurator VM*

*For a detailed walk-through of the steps we're following, see the Quickstart Guide:

https://github.com/BitCurator/bitcurator-distro/wiki/Releases#quickstart-guide


Ubuntu Desktop
En
2:41 PM
home
Imaging Tools
Forensics Tools
System Settings
Additional Tools
Shared Folders and Media
sf_bc_share
Trash
Documentation and Help
Network Servers
BitCurator
Right Ctrl

BitCurator Resources

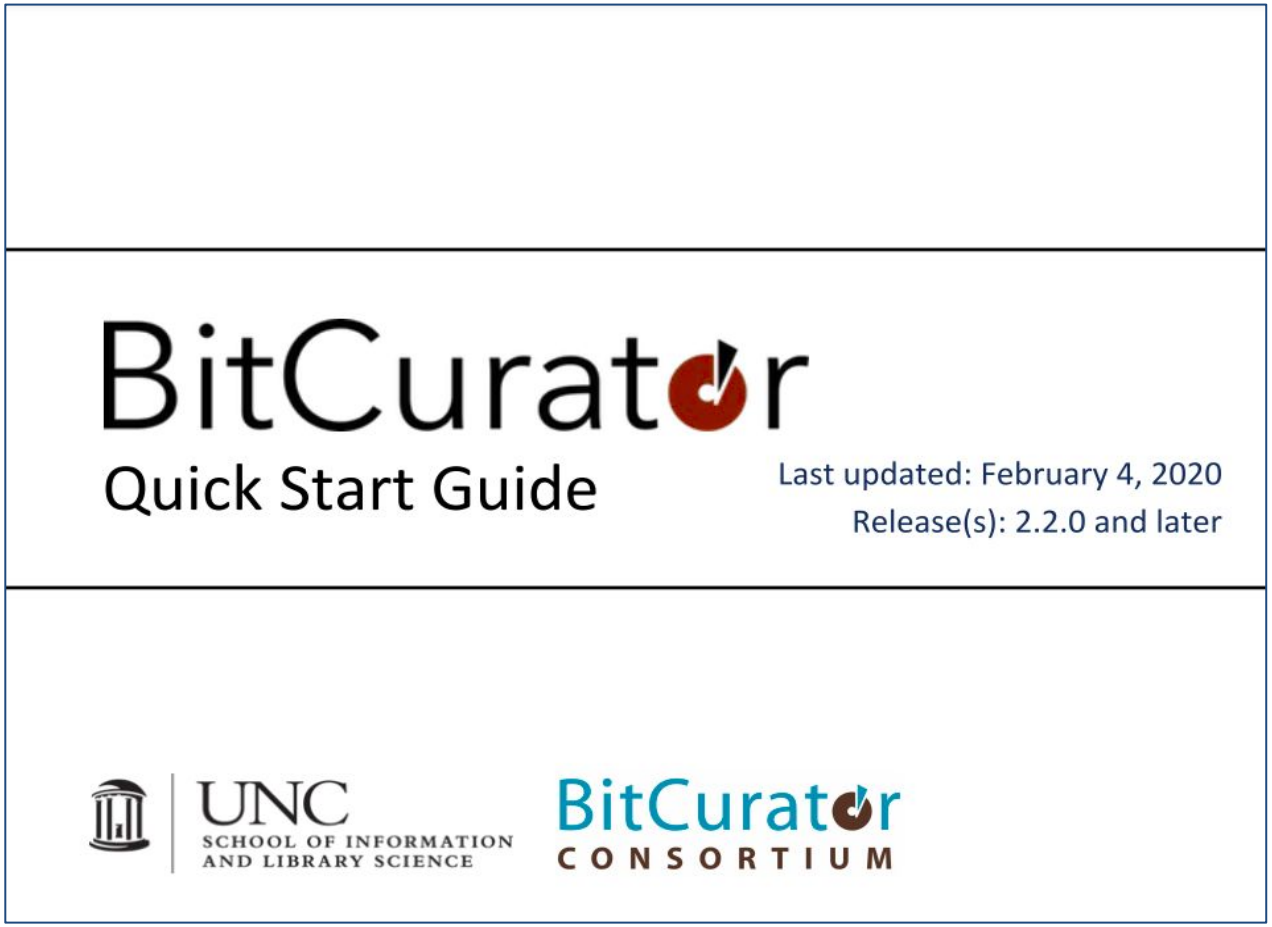


Get the software
Documentation and technical specifications
Screencasts
Google Group
People
Project overview
Publications
News

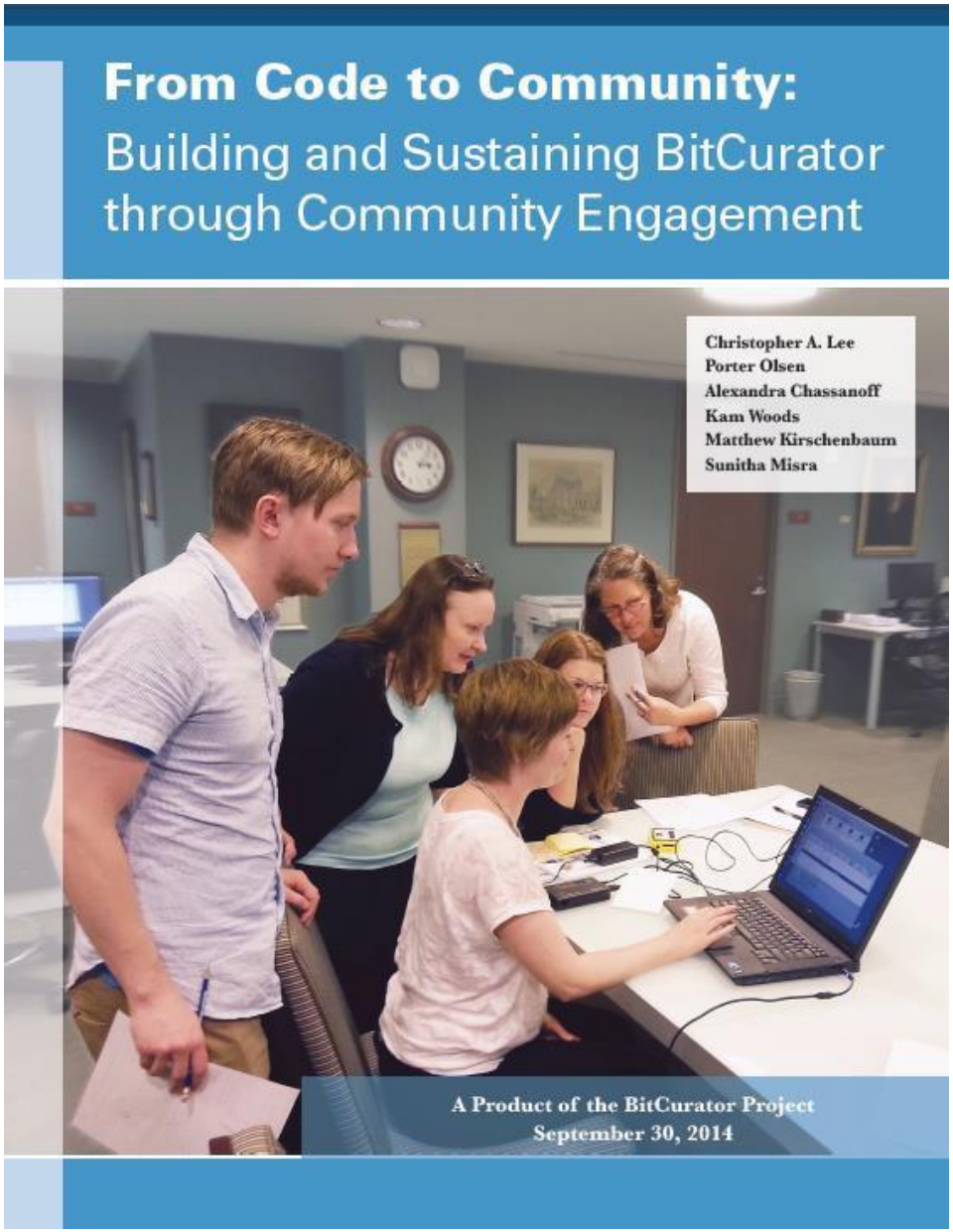
https://bitcurator.net/

Twitter: @bitcurator

Most tasks we will cover in this course are explained in the Quick Start Guide



https://github.com/BitCurator/bitcurator-distro/wiki/Releases#quickstart-guide



https://bitcurator.net/wp-content/uploads/sites/1099/2018/08/code-to-community.pdf

BitCurator Consortium

- Continuing home for hosting, stewardship and support of BitCurator tools and associated user engagement
- Administrative home: Educopia Institute
- Funding based on membership dues
- Software and documentation are free and open source, but membership provides benefits (e.g. support, training, development priority)

https://bitcuratorconsortium.org/

BitCurator CONSORTIUM

Member Login | Search

About Us | Why Digital Forensics | Using BitCurator | Get Involved

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

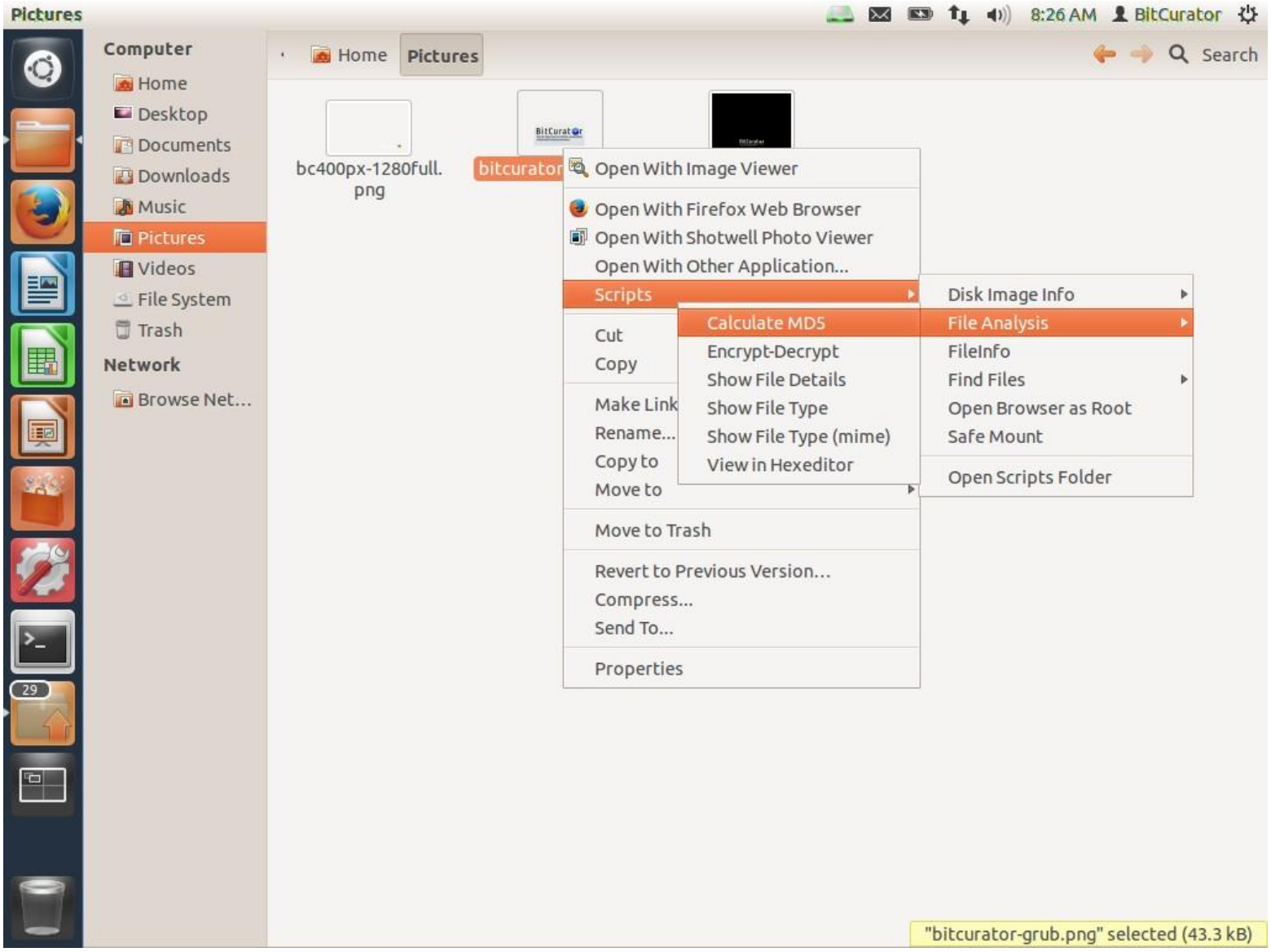
University of

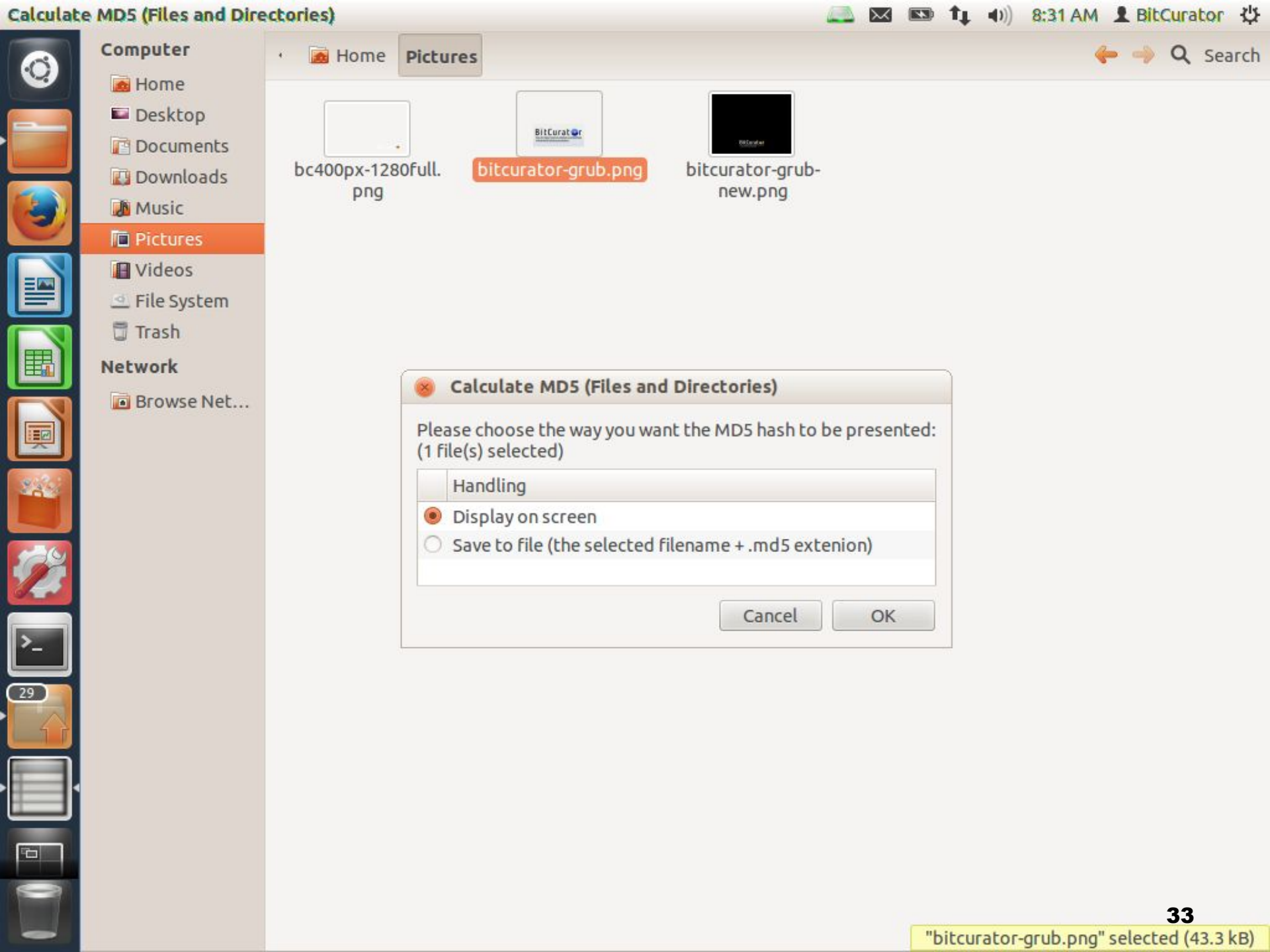
Technical Fundamentals

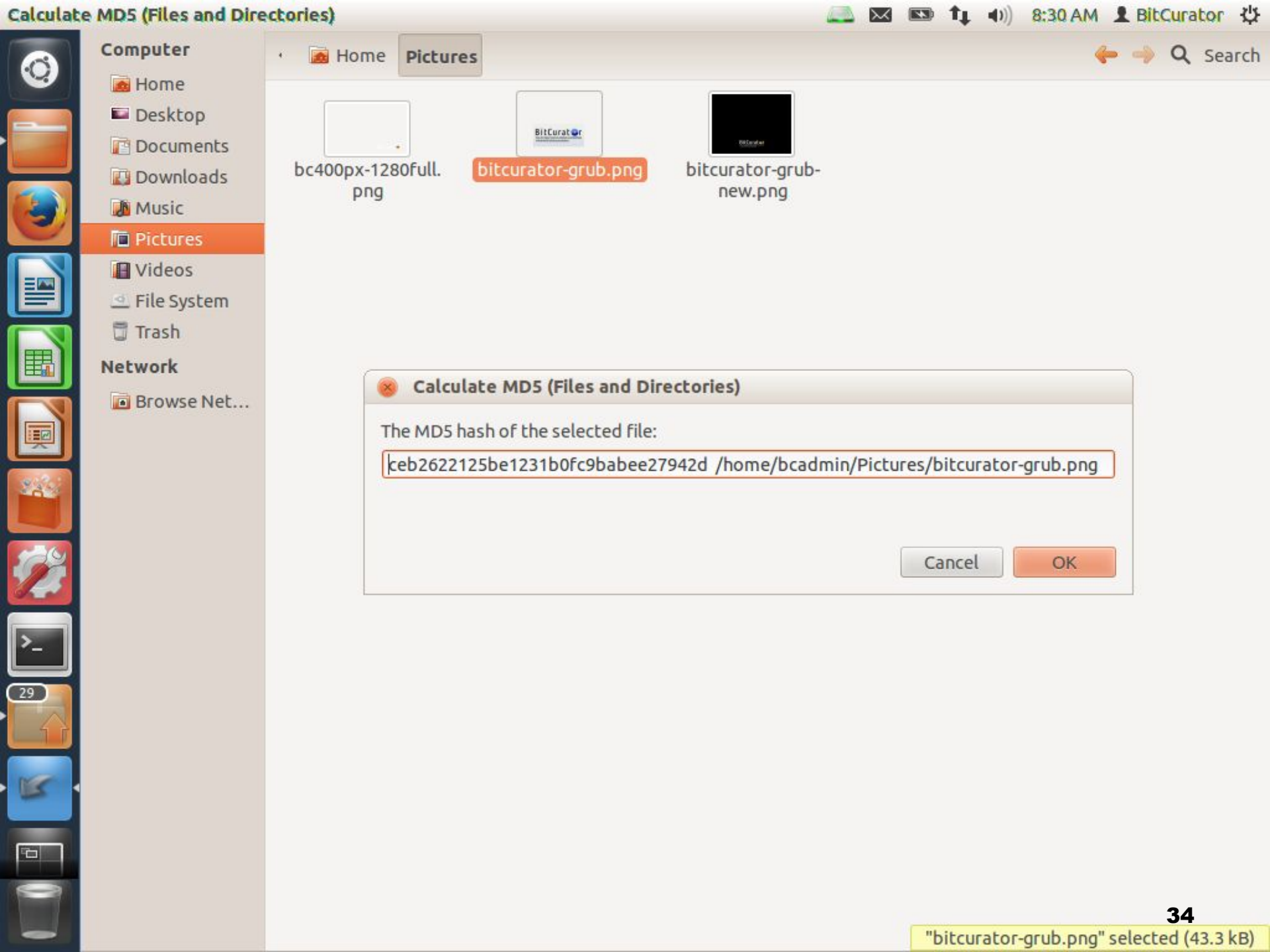
Checksums – Compact Representations of Bitstreams

- A given bitstream, fed into an algorithm, will generate a short string of characters that is **extremely** unlikely to be generated by a different bistream fed into that same algorithm
- Most common = MD5, SHA-1
- Can determine:
  - If bits have changed after a transfer
  - If bits have flipped within a storage environment
  - Whether two different files are identical bitstreams
- A library of hash values can identify "known and notable" (EnCase terminology) files
  - Known – files that can be ignored (e.g. software listed in National Software Reference Library)
  - Notable – specific bitstreams that you're trying to find

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)

Note on MD5/SHA1 - Potential Collisions

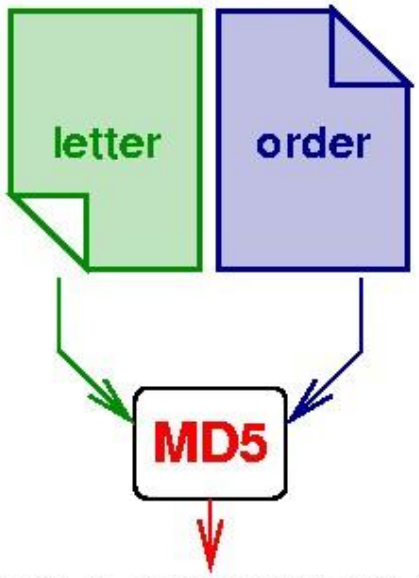
- From a security perspective, MD5 has been "broken" since 2005
- SHA-1 was broken in February 2017
- Someone with malicious intent can create two different bitstreams that result in the same hash (i.e. hash collisions)

Hash Collisions (The Poisoned Message Attack): "The Story of Alice and her Boss"*

Being an intern, Alice does not have any access to secret documents. Not enough for her ...

... tricky Alice decides to fool Caesar. Because Caesar is still relying on the widely used MD5 hash function, she implements the attack from Wang and Yu [WY05] to find MD5 collisions. When she receives her letter of recommendation (on paper), she prepares **two postscript files with the same MD5 hash**:

- One to display the letter of recommendation, and
- a second one, an order from Caesar to grant Alice some kind of a security clearance.



*Stefan Lucks and Magnus Daum,
http://th.informatik.uni-mannheim.de/people/lucks/HashCollisions/
Wayback link:
https://web.archive.org/web/20160713130211/http://th.informatik.uni-mannheim.de/people/lucks/HashCollisions/

Julius. Caesar
Via Appia 1
Rome, The Roman Empire

May, 22, 2005

Order:

Alice Falbala is given full access to all confidential and secret information about GAUL.

Sincerely,

Julius Caesar

Julius. Caesar
Via Appia 1
Rome, The Roman Empire

May, 22, 2005

To Whom it May Concern:

Alice Falbala fulfilled all the requirements of the Roman Empire intern position. She was excellent at translating roman into her gaul native language, learned very rapidly, and worked with considerable independence and confidence.

Her basic work habits such as punctuality, interpersonal deportment, communication skills, and completing assigned and self-determined goals were all excellent.

I recommend Alice for challenging positions in which creativity, reliability, and language skills are required.

I highly recommend hiring her. If you'd like to discuss her attributes in more detail, please don't hesitate to contact me.

Sincerely,

Julius Caesar

Same MD5 Hash: a25f7f0b 29ee0b39 68c86073 8533a4b9

Alternatives to MD5?

SHA (Secure Hash Algorithm)

- Originally developed by the NSA
- Several variants: SHA-0, SHA-1, SHA-2 family
- Early variants (SHA-0, SHA-1) known to be compromised
- Most commonly used is now SHA-256. Can be used to process bitstreams ("messages") up to $3.4 * 10^{38}$ bits (very large!)
- Disadvantage: more time, computing power required to produce

Algorithm and variant		Output size (bits)	Internal state size (bits)	Block size (bits)	Max message size (bits)	Word size (bits)	Rounds	Operations	Collisions found?
SHA-0		160	160	512	$2^{64} - 1$	32	80	add, and, or, xor, rotate, mod	Yes
SHA-1									Theoretical attack (2^{60})[6]
SHA-2	*SHA-256/224*	256/224	256	512	$2^{64} - 1$	32	64	add, and, or, xor, rotate, mod, shift	No
	SHA-512/384	512/384	512	1024	$2^{128} - 1$	64	80		

Source: https://en.wikipedia.org/wiki/Cryptographic_hash_function

Implications of MD5/SHA1 Being "Broken"

- Rarely a concern when hash is used for integrity checks on known items (e.g. verifying that a file was transferred correctly to a repository or that files in storage are still intact)
- Can be a concern if one is relying on a hash as proof of record authenticity – risks can include cases of internal tampering
- There are more robust hash algorithms to address this (SHA-2 family, including SHA-256) so good practice is to generate one of them along with the MD5
- MD5 is still widely used, because it is fast to calculate and still widely supported

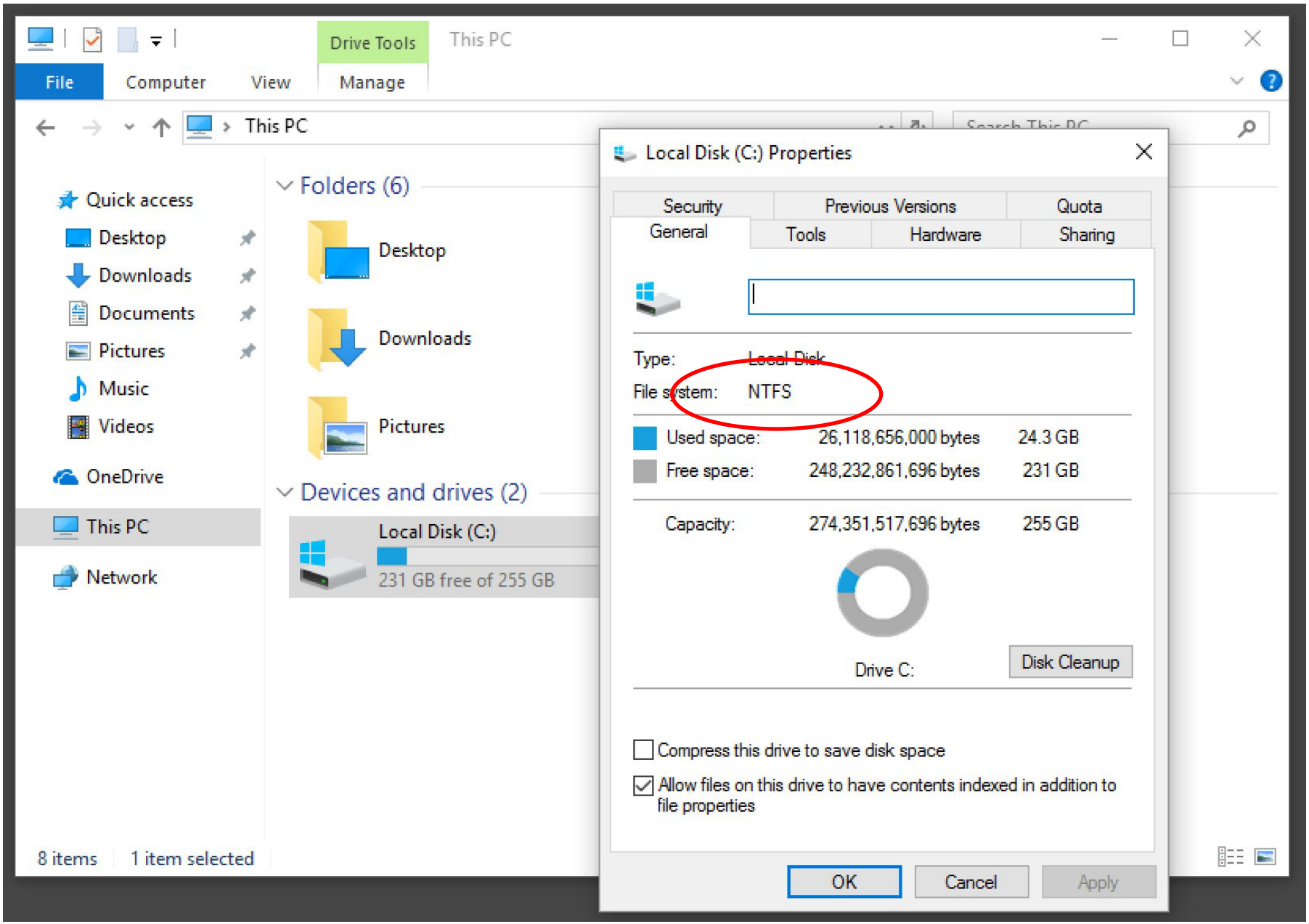
Question:

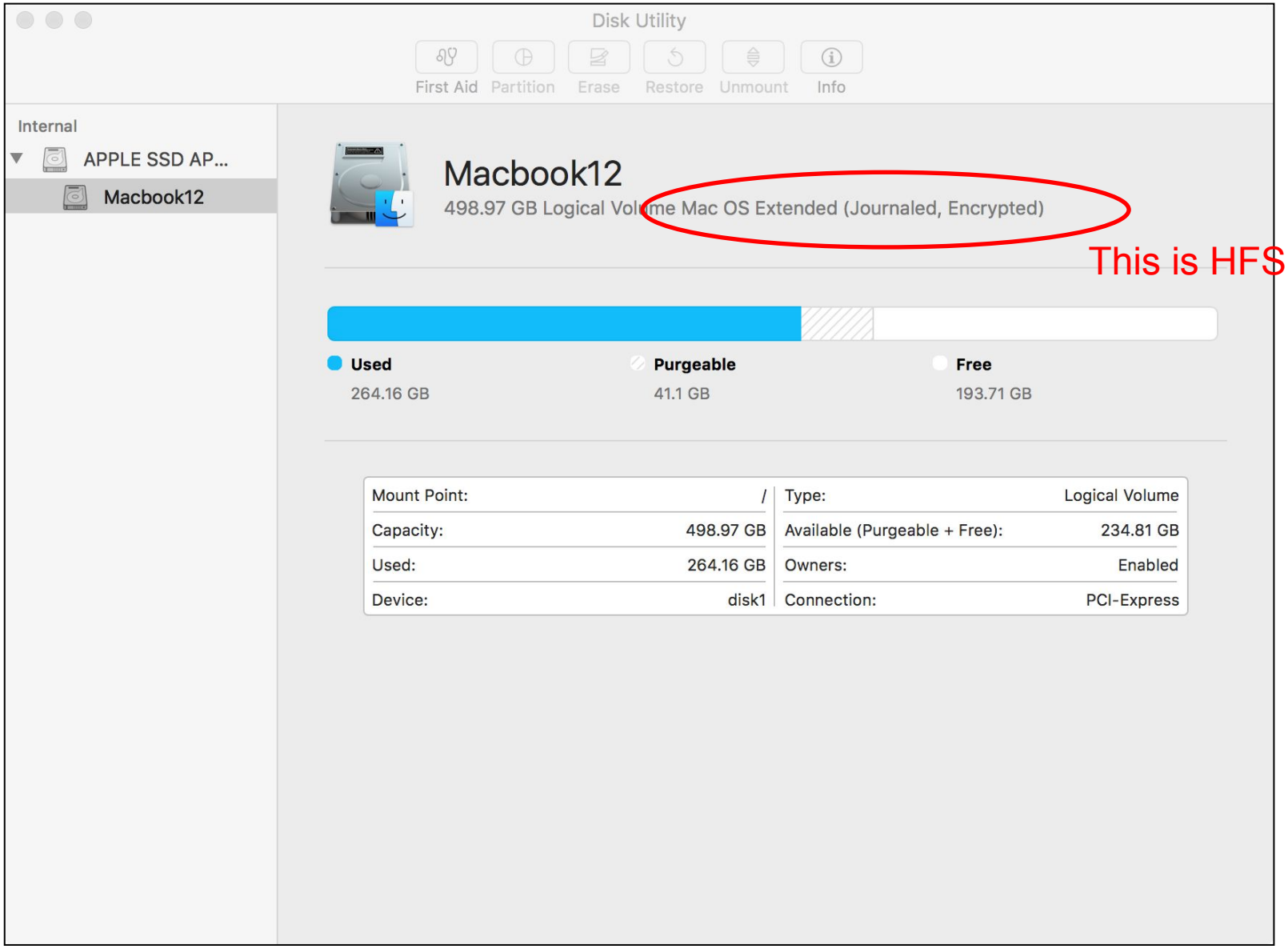
Can you use a cryptographic hash to determine specifically what any given file **contains**?

If not, what could you use?

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)


Drive Tools
This PC
File
Computer
View
Manage
This PC
Quick access
Desktop
Downloads
Documents
Pictures
Music
Videos
OneDrive
This PC
Network
Folders (6)
Desktop
Downloads
Pictures
Devices and drives (2)
Local Disk (C:)
231 GB free of 255 GB
8 items
1 item selected
Local Disk (C:) Properties
Security
Previous Versions
Quota
General
Tools
Hardware
Sharing
Type:
Local Disk
File system:
NTFS
Used space:
26,118,656,000 bytes
24.3 GB
Free space:
248,232,861,696 bytes
231 GB
Capacity:
274,351,517,696 bytes
255 GB
Drive C:
Disk Cleanup
Compress this drive to save disk space
Allow files on this drive to have contents indexed in addition to file properties
OK
Cancel
Apply


Disk Utility
First Aid
Partition
Erase
Restore
Unmount
Info
Internal
APPLE SSD AP...
Macbook12
Macbook12
498.97 GB Logical Volume Mac OS Extended (Journaled, Encrypted)
This is HFS+
Used
264.16 GB
Purgeable
41.1 GB
Free
193.71 GB
Mount Point: /
Type: Logical Volume
Capacity: 498.97 GB
Available (Purgeable + Free): 234.81 GB
Used: 264.16 GB
Owners: Enabled
Device: disk1
Connection: PCI-Express

Name	Operating System(s) Using it as Native File System [often other OSs can also recognize it]
FAT12, FAT16	MS-DOS
FAT32 (VFAT)	Windows 95, 98
exFAT	Windows XP SP2 and later (primary use: USB drives, SD cards)
NTFS	Windows NT, 2000, XP, Server 2003, Server 2008, Vista
MFS	Macintosh System 1-3
HFS (Hierarchical File System)	Macintosh System 4-8
HFS+	Macintosh System 8.1 – 9, OS X 10.0 – 10.11
APFS	macOS 10.12
ext, ext2, ext3, ext4 (Extended File System)	Linux
XFS	Linux, typically Enterprise variants (RHEL)
HPFS (High Performance File System)	OS/2
ISOFS (ISO 9660)	Any OS that reads data from a CD
JFS1 (Journaled File System)	AIX (IBM)
ReiserFS	Several Linux distributions
UFS (Unix File System) aka FFS (Fast File System)	Various flavors of Unix

Name	Operating System(s) Using it as Native File System [often other OSs can also recognize it]
FAT12, FAT16	MS-DOS
FAT32 (VFAT)	Windows 95, 98
exFAT	W
NTFS	W
MFS	M
HFS (Hierarchical File System)	M
HFS+	M
APFS	m
ext, ext2, ext3, ext4 (Extended File System)	Li
XFS	Li
HPFS (High Performance File System)	OS/2
ISOFS (ISO 9660)	Any OS that reads data from a CD
JFS1 (Journaled File System)	AIX (IBM)
ReiserFS	Several Linux distributions
UFS (Unix File System) aka FFS (Fast File System)	Various flavors of Unix

Filesystems you're most likely to encounter

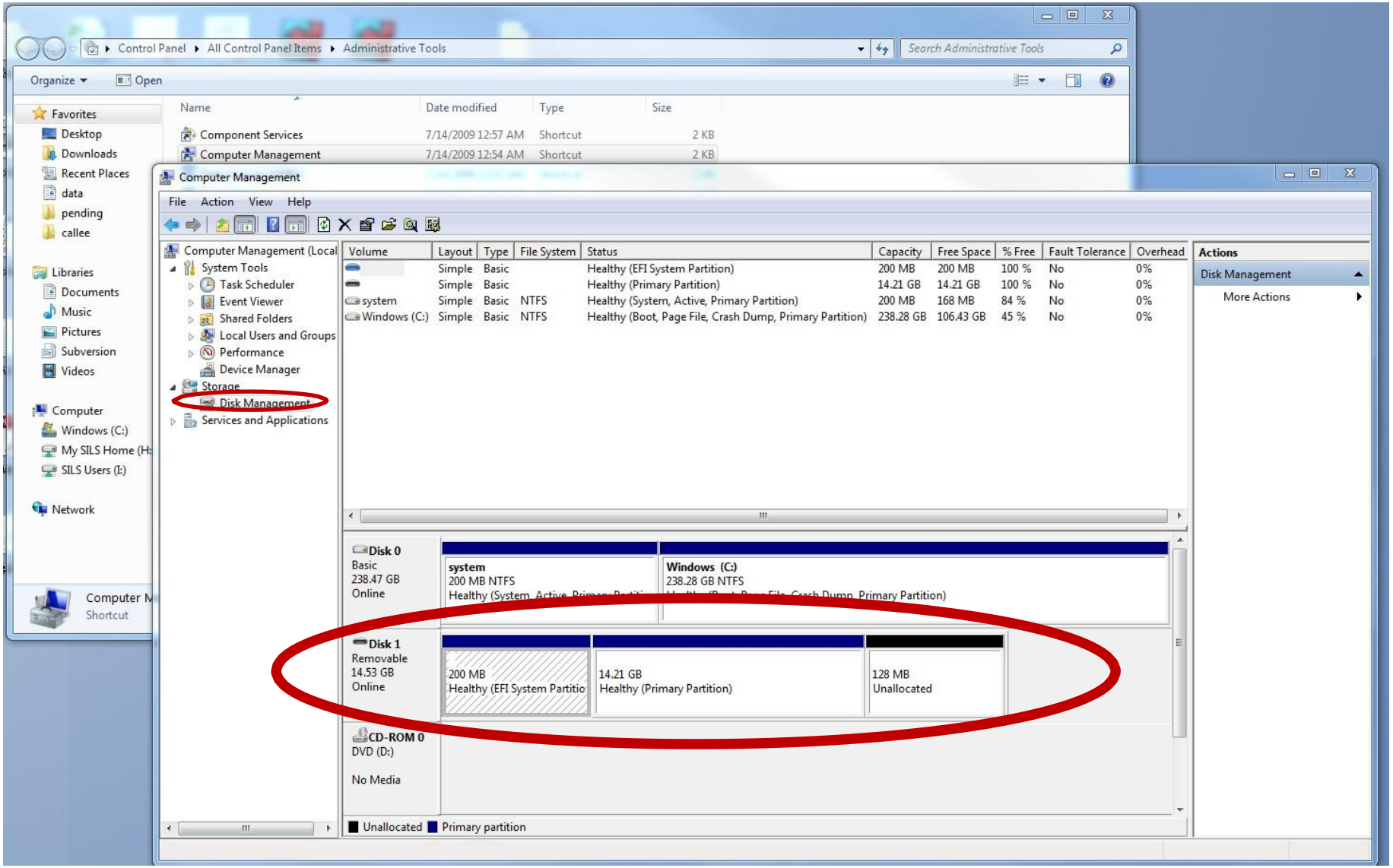
NTFS vs. FAT File System Attributes

- Two disk images are in your zip file, and can also be found at: https://digitalcorpora.s3.amazonaws.com/corpora/scenarios/2009-m57-patents/usb/terry-work-usb-2009-12-11.E01 https://digitalcorpora.s3.amazonaws.com/corpora/drives/nps-2009-ntfs1/ntfs1-gen1.E01
- Load each disk image into a separate instance of FTK Imager (run them side by side to compare what you see) – if you don't have a Windows computer, look on with a partner
- Look at the properties of some files*
- What differences do you notice?

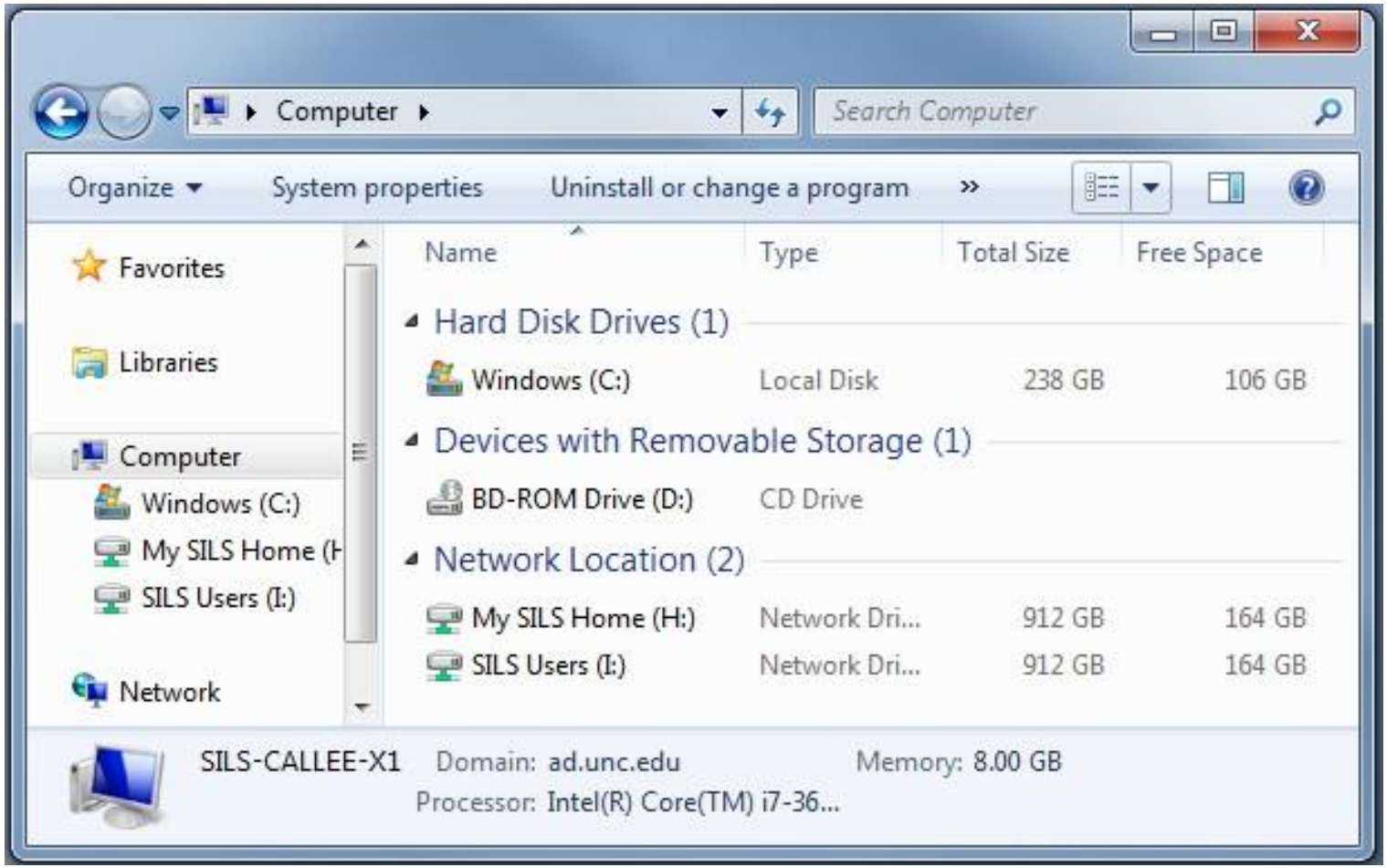
*Properties are shown in the bottom left corner. If you don't see them, go to the View menu at the top and select "Properties." You may need to drag the top of the properties window up to see all of the values.

Connecting a Device vs. Mounting a Filesystem

HFS+ volume is visible through Windows Disk Management after it's connected

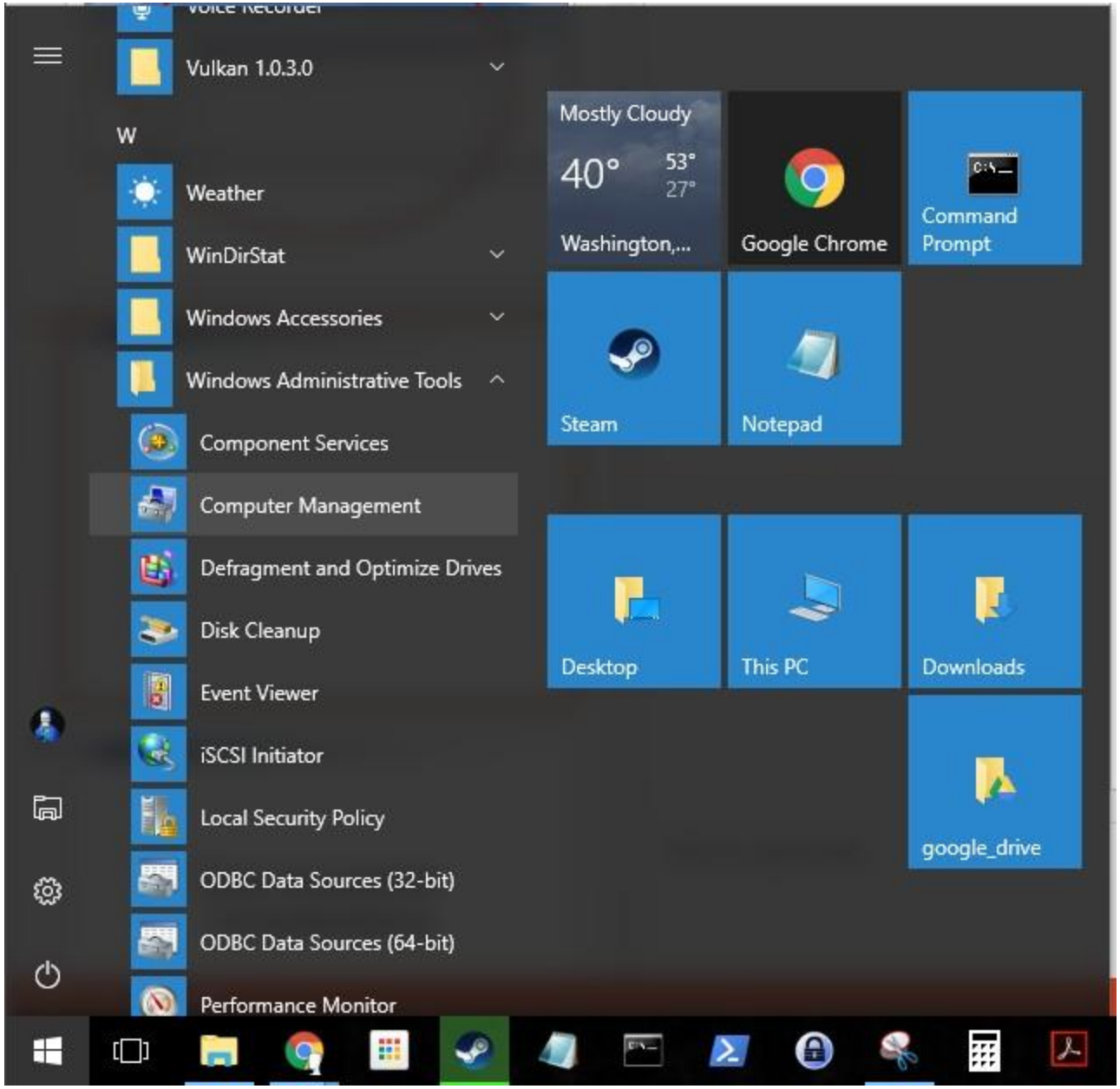


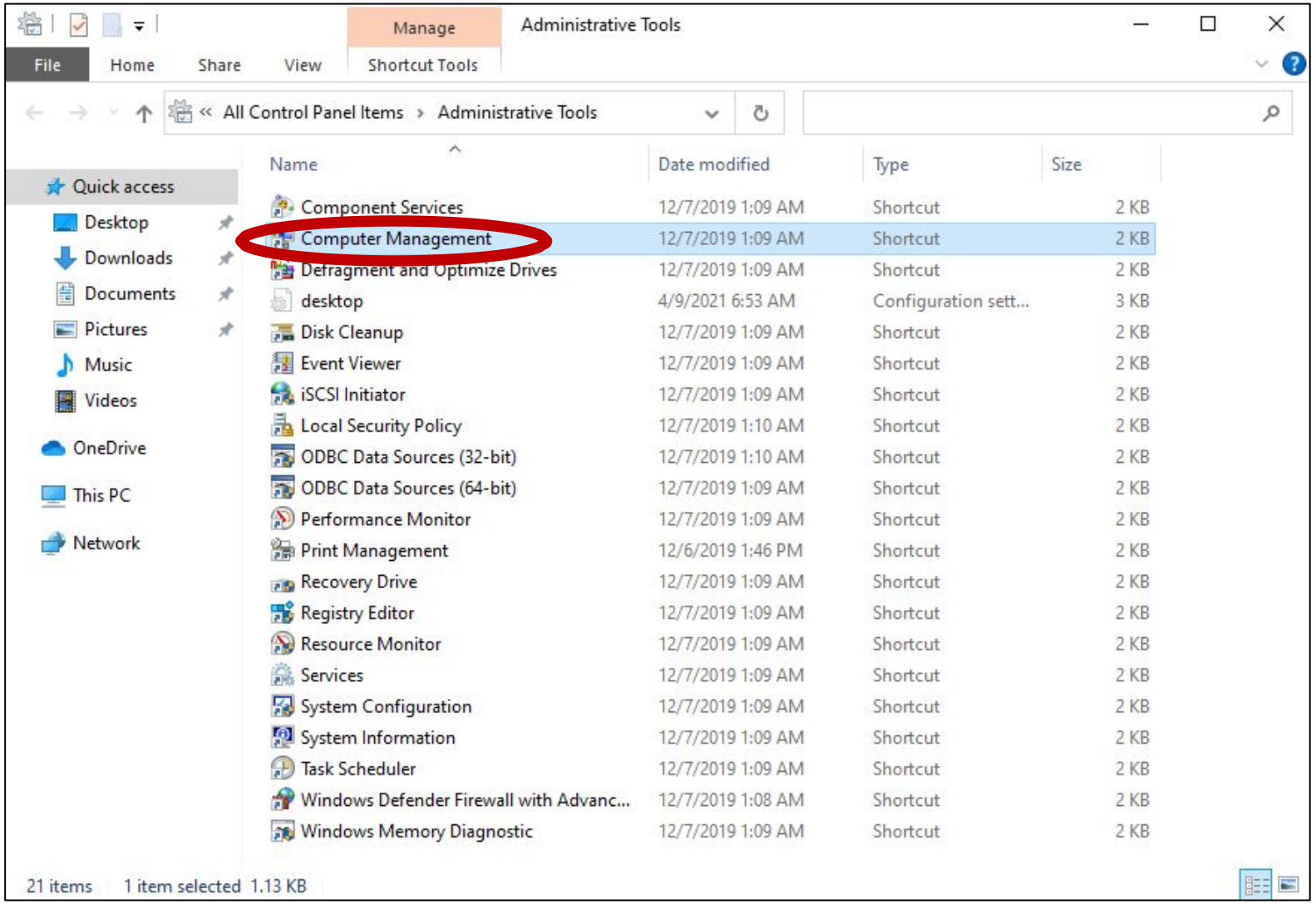
But it's not visible through Windows Explorer, because Windows doesn't know how to mount the file system

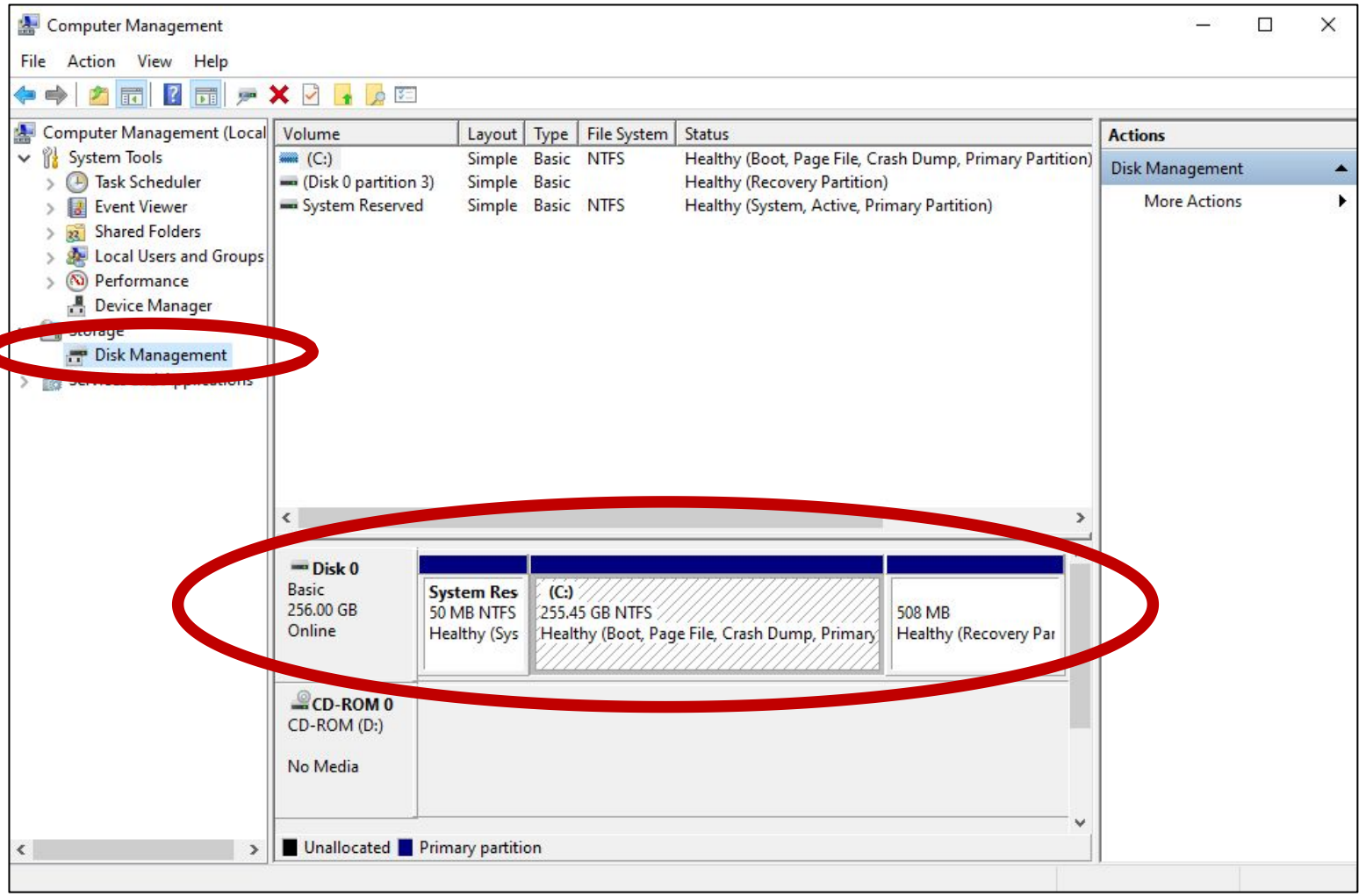


Seeing Attached Devices (Whether or Not They're Mounted) - In Windows

- Control Panel
- Administrative Tools
- Computer Management
- Disk Management


Vulkan 1.0.3.0
W
Weather
WinDirStat
Windows Accessories
Windows Administrative Tools
Component Services
Computer Management
Defragment and Optimize Drives
Disk Cleanup
Event Viewer
iSCSI Initiator
Local Security Policy
ODBC Data Sources (32-bit)
ODBC Data Sources (64-bit)
Performance Monitor
Mostly Cloudy
40°
53°
27°
Washington,...
Google Chrome
Command Prompt
Steam
Notepad
Desktop
This PC
Downloads
google_drive

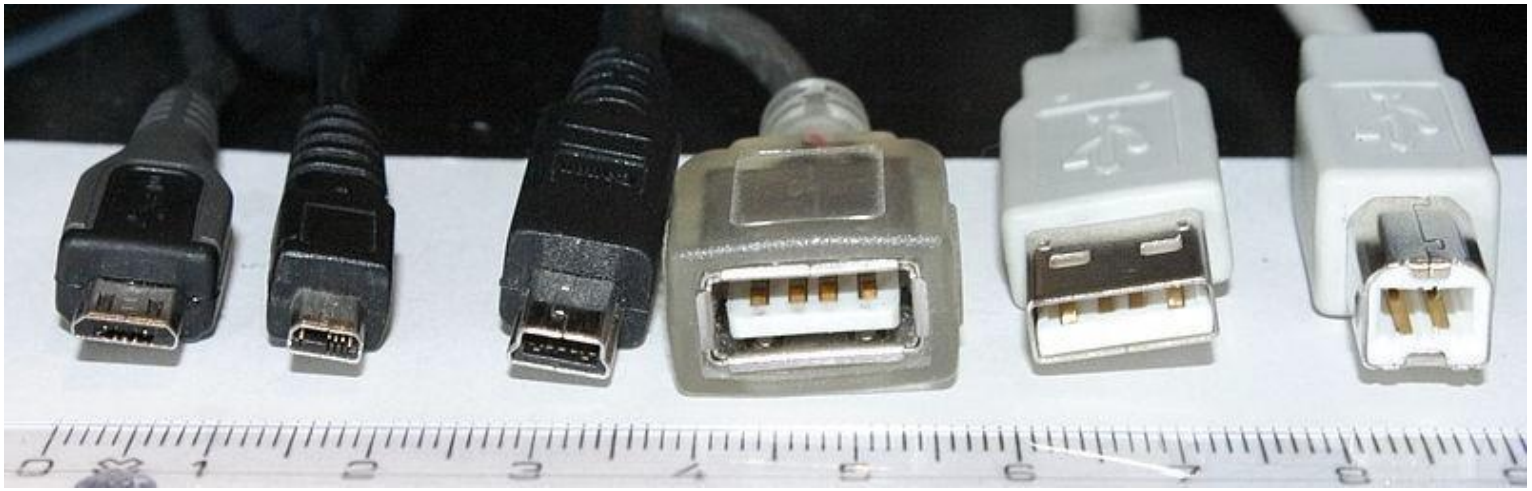

Manage
Administrative Tools
File
Home
Share
View
Shortcut Tools
All Control Panel Items › Administrative Tools
Quick access
Desktop
Downloads
Documents
Pictures
Music
Videos
OneDrive
This PC
Network
Name
Date modified
Type
Size
Component Services 12/7/2019 1:09 AM Shortcut 2 KB
Computer Management 12/7/2019 1:09 AM Shortcut 2 KB
Defragment and Optimize Drives 12/7/2019 1:09 AM Shortcut 2 KB
desktop 4/9/2021 6:53 AM Configuration sett... 3 KB
Disk Cleanup 12/7/2019 1:09 AM Shortcut 2 KB
Event Viewer 12/7/2019 1:09 AM Shortcut 2 KB
iSCSI Initiator 12/7/2019 1:09 AM Shortcut 2 KB
Local Security Policy 12/7/2019 1:10 AM Shortcut 2 KB
ODBC Data Sources (32-bit) 12/7/2019 1:10 AM Shortcut 2 KB
ODBC Data Sources (64-bit) 12/7/2019 1:09 AM Shortcut 2 KB
Performance Monitor 12/7/2019 1:09 AM Shortcut 2 KB
Print Management 12/6/2019 1:46 PM Shortcut 2 KB
Recovery Drive 12/7/2019 1:09 AM Shortcut 2 KB
Registry Editor 12/7/2019 1:09 AM Shortcut 2 KB
Resource Monitor 12/7/2019 1:09 AM Shortcut 2 KB
Services 12/7/2019 1:09 AM Shortcut 2 KB
System Configuration 12/7/2019 1:09 AM Shortcut 2 KB
System Information 12/7/2019 1:09 AM Shortcut 2 KB
Task Scheduler 12/7/2019 1:09 AM Shortcut 2 KB
Windows Defender Firewall with Advanc... 12/7/2019 1:08 AM Shortcut 2 KB
Windows Memory Diagnostic 12/7/2019 1:09 AM Shortcut 2 KB
21 items
1 item selected 1.13 KB


Computer Management
File Action View Help
Computer Management (Local
System Tools
Task Scheduler
Event Viewer
Shared Folders
Local Users and Groups
Performance
Device Manager
Disk Management
Volume Layout Type File System Status
(C:) Simple Basic NTFS Healthy (Boot, Page File, Crash Dump, Primary Partition)
(Disk 0 partition 3) Simple Basic Healthy (Recovery Partition)
System Reserved Simple Basic NTFS Healthy (System, Active, Primary Partition)
Actions
Disk Management
More Actions
Disk 0
Basic
256.00 GB
Online
System Res
50 MB NTFS
Healthy (Sys
(C:)
255.45 GB NTFS
Healthy (Boot, Page File, Crash Dump, Primary
508 MB
Healthy (Recovery Par
CD-ROM 0
CD-ROM (D:)
No Media
Unallocated
Primary partition

Data Acquisition Considerations

Ports and Connectors

- Adoption comes and goes with changes in the industry
- Those that have had wide industry adoption have usually lasted a decade or more
- Important distinction: hardware protocol standards vs. shape of connectors (related but not always the same)



Different USB connectors. From left to right: male Micro USB B-Type, UC-E6 proprietary (not USB), male Mini USB (5-pin) B-type, female A-type, male A-type, male B-type. Shown with a centimeter ruler. Female A-type connector (4th from left) is "upside down" to show the pins.
https://en.wikipedia.org/wiki/File:Usb_connectors.JPG

Adapter Examples

Micro SD to SD

MicroSATA to SATA

NVMe to USB

SATA to IDE

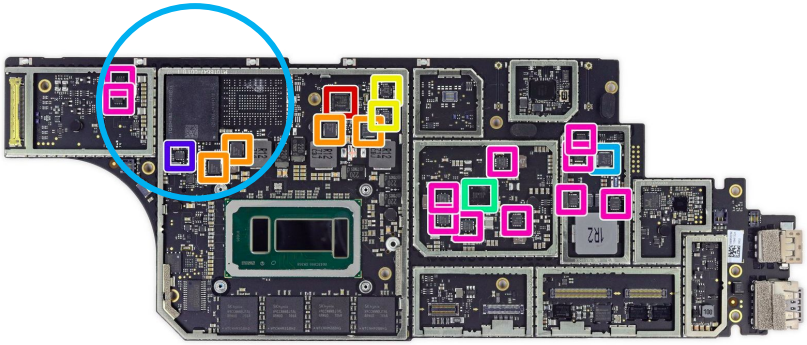
Ethernet to USB

SATA to Molex Power

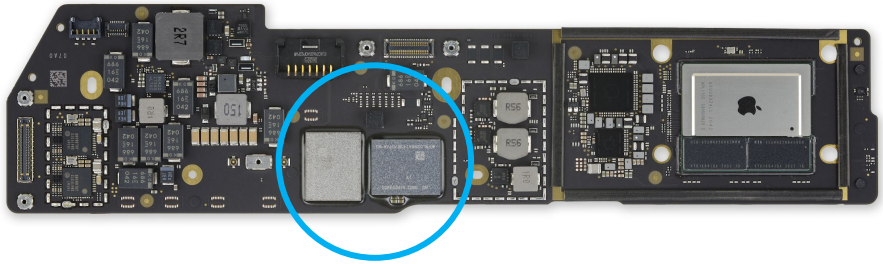
A Note Regarding Modern Laptops (and other devices)

- Some recent laptops, including all Apple laptops post-2019, the Microsoft Surface series of machines, and certain others, have both RAM and storage soldered to the motherboard; these drives cannot* be removed for imaging.

128GB Toshiba storage chip, Microsoft Surface Laptop motherboard (2019)



2 x 128GB Western Digital storage chips, Apple Macbook M1 Air laptop motherboard (2020)



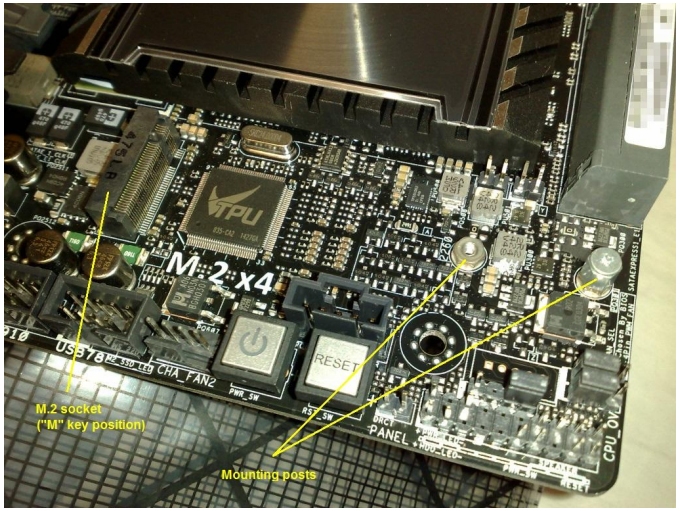
*Except in specialized laboratories.
See https://sumuri.com/how-to-image-an-apple-silicon-mac-with-recon-itr-live/ for an example of a commercial tool used to image Apple Silicon Macs without drive removal.

M.2 (also known as "Next Generation Form Factor")

- Current most common way to connect a mass storage device (e.g. solid state storage) to the inside of many desktop and laptop computers. Cable-less, connects via an edge slot.

- Internal M.2 devices mount to the motherboard or a PCI expansion card. External M.2 devices typically housed in USB enclosures

512GB M.2 Mass Storage Device
https://en.wikipedia.org/wiki/M.2#/media/File:Intel_512G_M2_Solid_State_Drive.jpg

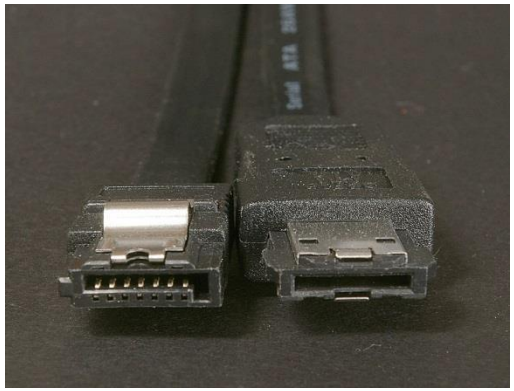


M.2 connector on a motherboard
https://en.wikipedia.org/wiki/M.2#/media/File:M.2_connector_on_a_computer_motherboard.jpg

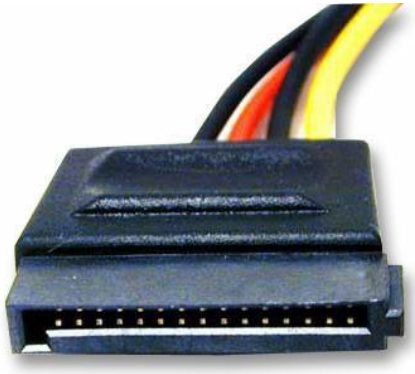
Serial Advanced Technology Attachment (SATA)

- The most common way to connect a mass storage device (e.g. hard drive, solid state drive) to the inside of a computer from 2003 through the mid-2010's.

- eSATA used for external mass storage devices



Data Cables:
SATA on the left, eSATA on the right
https://en.wikipedia.org/wiki/File:SATA2_und_eSATA-Stecker.jpg

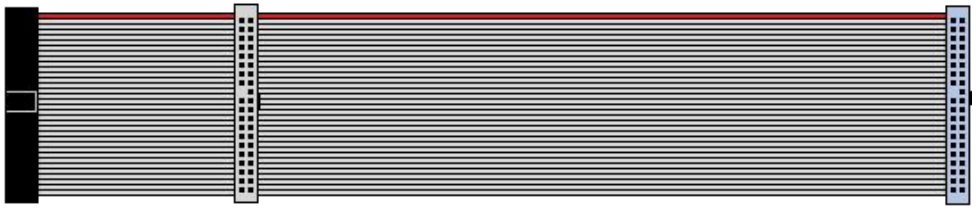


SATA Power Cable

Integrated Drive Electronics (IDE) / AT Attach

- Precursor to SATA
- No longer widely used, but you'll still find it on older drives
- More than one drive can share the same cable, with one being the *master* (device 0) and the other being the *slave* (device 1) – these are set with *jumpers*

https://static.daniweb.com/images/attachments/0/maxtor_jumper.jpg



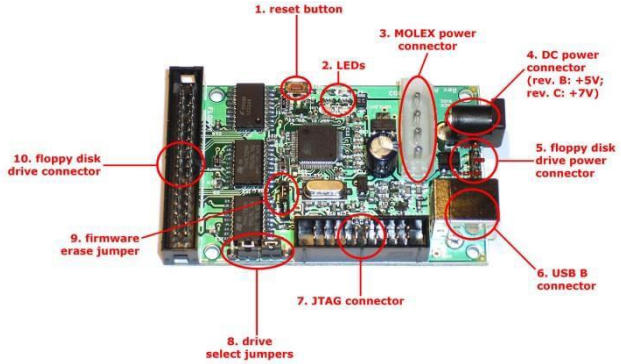
40-pin IDE Ribbon Cable
https://en.wikipedia.org/wiki/File:Nappe.svg

Floppy Disks

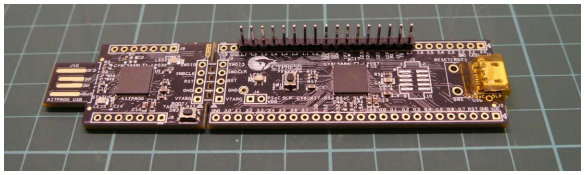
- Physical storage is similar to hard drives (magnetic charges in a spinning disk)
- Various types and sizes, e.g. high density, double density, 3.5 inch, 5.25 inch, 8 inch
- 3.5 inch floppies are relatively easy to read using a USB drive, but older ones are more complicated…

Floppy Controller Hardware

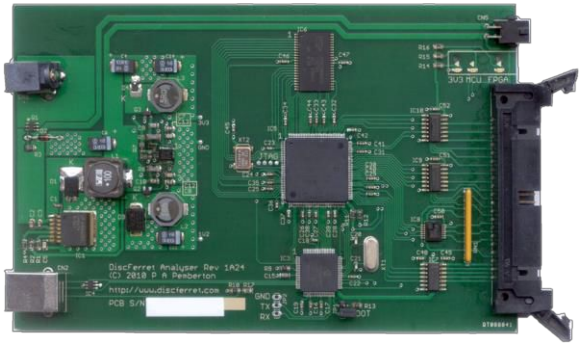
Kryoflux[1]



FluxEngine[2]

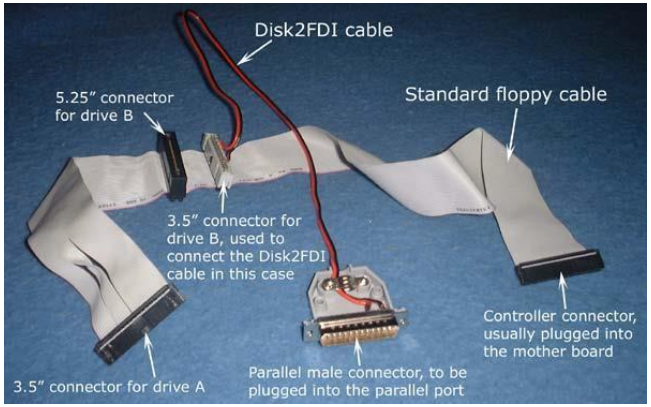


Disc Ferret[3]

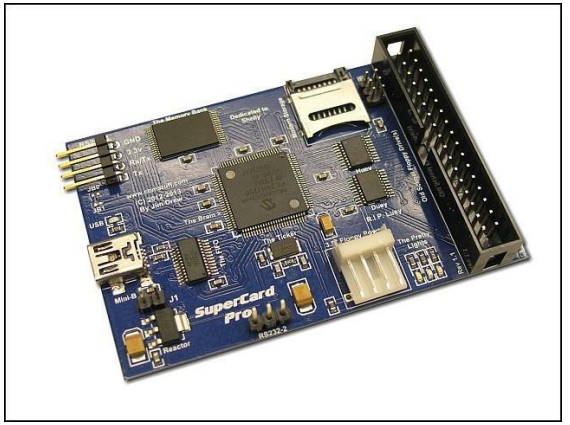


FC 5025[4]

Disk2FDI[5]



SuperCard Pro[6]



1. https://www.kryoflux.com/
2. http://cowlark.com/fluxengine/index.html
3. http://discferret.com/wiki/DiscFerret
4. http://www.deviceside.com/fc5025.html
5. http://disk2fdi.joguin.com/D2FCABLE.htm
6. http://www.cbmstuff.com/proddetail.php?prod=SCP

Common Floppy Formats (Physical)

Many variations over time, often to increase storage density

Floppy disk physical characteristics
(capacity and tracks are nominal, per side)

Size	Density	Tracks	tpi	bpi	Coercivity	Unformatted capacity per side
2½-inch[16][17]	Single	16[16][17]	48[16]			64 KB[16][17]
3½-inch	Double[18]	40[18]	67.5[18]	8650[18]	600 Oe	250 KB
		80	135	8717	600-665 Oe	500 KB
	High	80	135	17434	720-750 Oe	1000 KB
	Extended	80	135	34868	900 Oe	2000 KB
	Triple[12]	240[11]	406.5[11]	36700[11]		6500 KB
5¼-inch	Single/Double	40	48	5876	300 Oe	250 KB
	Double	80	62.5			(Apple FileWare)
	Quad	77	100		300 Oe	500 KB (Micropolis-compatible)
	Quad	80	96	5922	300 Oe	500 KB
	High	80	96	9646	600 Oe	833 KB
8-inch	Single/Double	77	48		300 Oe	1000 KB

Table source: https://en.wikipedia.org/wiki/List_of_floppy_disk_formats

*Coercivity is how resistant the medium is to being demagnetized (larger number means it requires a larger magnetic field to be demagnetized)

Actual bits on most IBM PC-type floppies are encoded using "modified frequency modulation" (MFM)

IBM PC compatibles[19]	8-inch	Single	128	26	77	1	250.25 KB[NB 13][19][20][21]	360	MFM
						2	500.5 KB[NB 13][19][20][21]		
		Double	1024	8		1	616 KB[NB 13][20][21]		
						2	1232 KB[NB 13][19][20][21]		
	5¼-inch	Double	512	8	40	1	160 KB[NB 13]	300	MFM
						2	320 KB[NB 13]		
				9		1	180 KB[NB 13]		
						2	360 KB[NB 13]		
		Quad[NB 15]		8	80	1	320 KB[NB 13]	300	
						2	640 KB[NB 13]		
		High		15	80	2	1200 KB[NB 13]	360	
	3½-inch (90 mm)	Double	512	8	80	1	320 KB[NB 13]	300	MFM
				9			360 KB[NB 13]		
				8		2	640 KB[NB 13]		
				9			720 KB[NB 13]		
		High		18			1440 KB[NB 13]		
				21			1680 KB[NB 13]		
					82		1720 KB[NB 13]		
		Extended		36	80		2880 KB[NB 13]		

Table source: https://en.wikipedia.org/wiki/List_of_floppy_disk_formats

Actual bits on Apple (Macintosh and earlier) may be encoded using "group code recording" (GCR) or "modified frequency modulation" (MFM)

Apple II	5¼-inch	Double	256	13	35	1	113.75 KB	300	GCR
				16		1	140 KB		
	3½-inch (90 mm)	Double	512	Variable (8-12)	80	1	400 KB	394 - 590	GCR
						2	800 KB		
		High	512	18	80	2	1440 KB	300	MFM
Apple Lisa	5¼-inch FileWare	Double	512	Variable (15-22)	46	2	851 KB	218 - 320	GCR
Apple Lisa 2/Macintosh XL	3½-inch (90 mm)	Double	512	Variable (8-12)	80	1	400 KB	394 - 590	GCR
Apple Macintosh						2	800 KB		
		High	512	18	80	2	1440 KB	300	MFM

Table source: https://en.wikipedia.org/wiki/List_of_floppy_disk_formats

Issues with Common Floppy Formats

- Why is this important?
  - Floppy disk drives write the actual bitstream using either GCR or MFM (or other, less common encodings)
  - Drive won't read disks encoded in a non-supported format without specialized hardware (e.g. KryoFlux)
  - Other factors:
    - Some devices (Apple in particular) use "CLV" (constant linear velocity), adjusting rotation speed depending on where the head is writing/reading
    - Others vary angular velocity of the disk based on zones ("zoned CAV").
    - Again, need specialized hardware such as Kryoflux to read these disks on a modern drive

Additional Factors: File Systems Used on Floppies

- Common IBM PC compatible floppies:
  - Typically FAT12 with 512 byte clusters (although there are many less common variations)

Attribute	FAT12
Used For	Floppies and very small hard disk volumes
Size of Each FAT Entry	12 bits
Maximum Number of Clusters	4,086
Cluster Size Used	0.5 KB to 4 KB
Maximum Volume Size	16,736,256 bytes

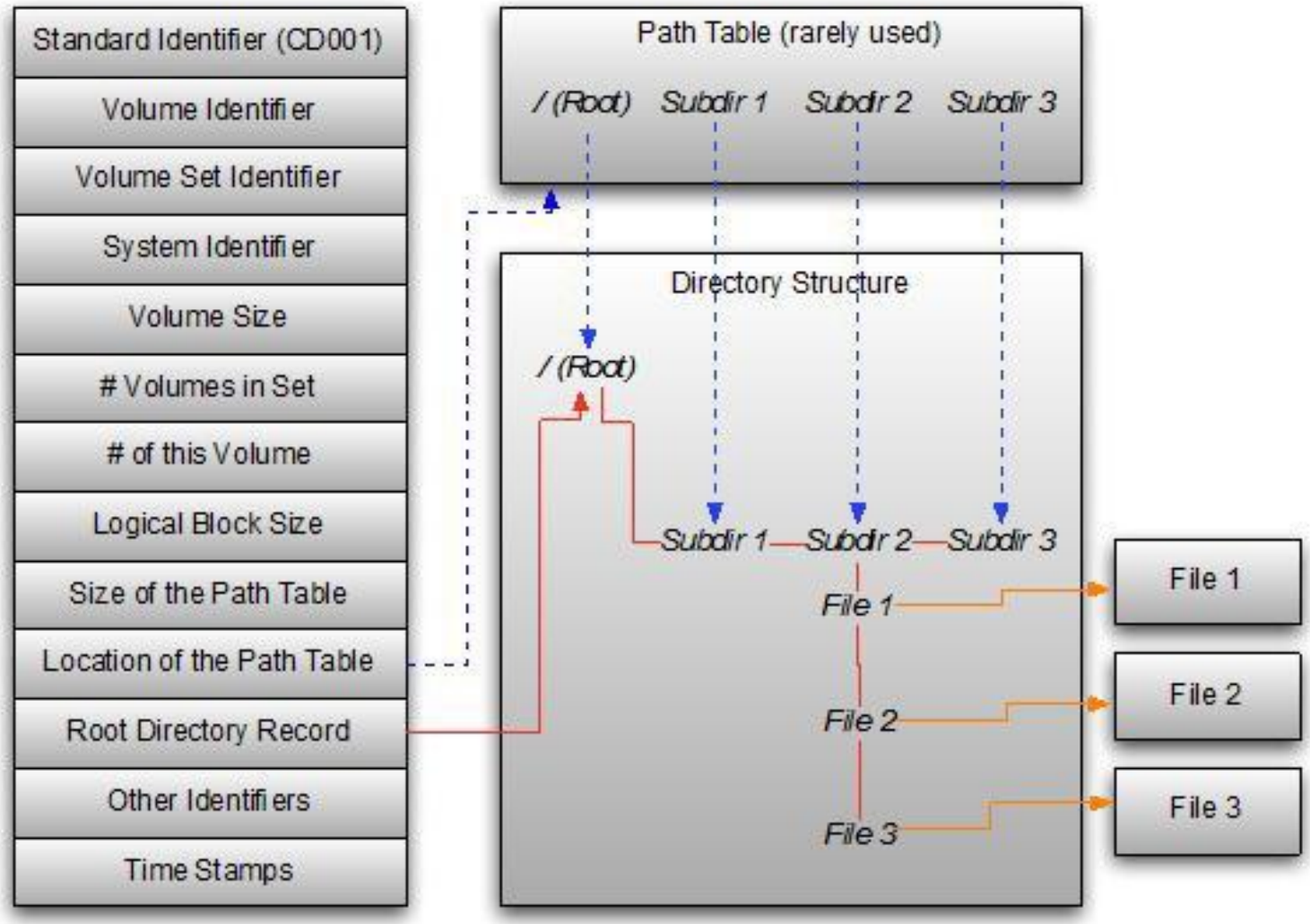
- Common Apple floppies (Macintosh and previous)
  - Apple II: 13 sector disk (5.25”, ProDOS 3.2, 113.75K)
  - Apple II: 16 sector disk (5.25”, ProDOS 3.3, 140K)
  - Apple Macintosh:
    - Double-density 3.5” - Macintosh File System (MFS)
    - Double-density 3.5” - Hierarchical File System (HFS)
    - High-density 3.5” - Hierarchical File System (HFS)

Potential Problems with ISO 9660 Media (e.g. CD-ROMs)

Factor	Complications
Physical damage	May not be visible to the naked eye
Bad length in volume header	Older CD writing tools sometimes miscalculated sector count, so header metadata doesn't match actual length
Incorrect block size	CDs can misidentify as having 512-byte sectors
File truncation	Filesystem may allow you to navigate to files that subsequently appear damaged or won't render at all (files could be truncated or never fully written to disk)
One sector short when written Track at Once (TAO)	TAO disks often represent a length (in volume header) that is one sector short from the actual length

Not all tools are designed to recognize or address these issues

CD-ROM File System Structure



Note that there are structures, such as the Path Table, that are not generally used, but may contain metadata. For more info on issues in managing optical media, see: https://kamwoods.net/publications/woodsbrownarch09.pdf

Creating Disk Images of CDs

- Cdrdao – primarily for ripping audio CDs (addressing issues such as the TAO one discussed above)

- Often sufficient to use dd form CD-ROMs
  - In the BitCurator VM, CD-ROM drive should appear as a device called /dev/sr0
  - Command to acquire:
    bcadmin@ubuntu:~$ **dd if=/dev/sr0 of=FILENAME1.iso**

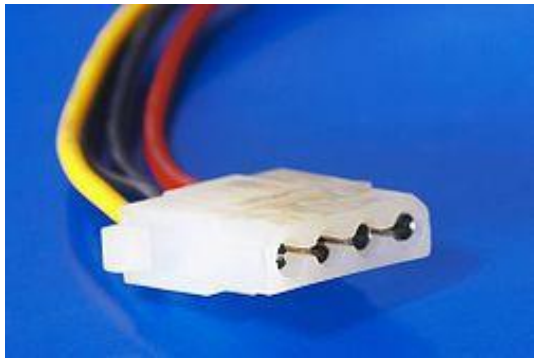
Dealing with Disk Images from CDs

- To modify the file system, you can:
  - Mount the disk image
  - Use mkisofs to create a new ISO file system -
    bcadmin@ubuntu:~$ mkisofs -r -o FILENAME2.iso /media/sr0

- If you created an EWF image of a CD:
  - BitCurator mounting scripts can't determine whether the underlying disk image data is a raw (dd) image or an ISO
  - To mount the image, first use ewfexport (command-line tool) to pull out the raw (dd) image, then rename it as an ISO

Two Important Considerations for Internal Media that are Used as External Media

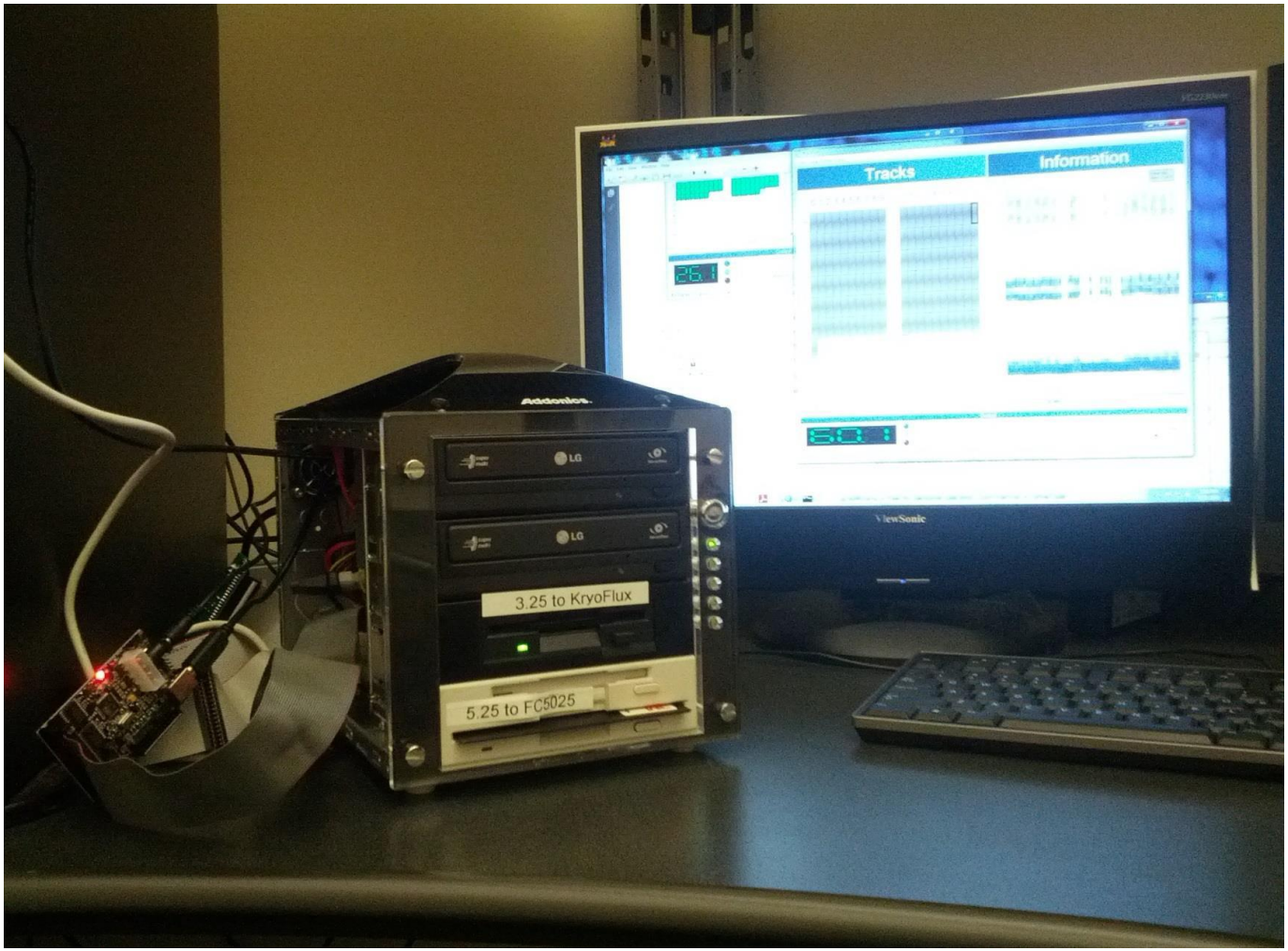
- Power - internal drive needs different connector (often Molex), not the kind that plugs into the wall
- Cooling – when pulled from the computer, you've also separated the drive from the fan, so you should often add an external one to ensure cooling



https://en.wikipedia.org/wiki/File:Molex_female_connector.jpg

https://www.tigerdirect.com/applications/SearchTools/item-details.asp?EdpNo=1648567

Kryoflux installed and running in a mini jukebox



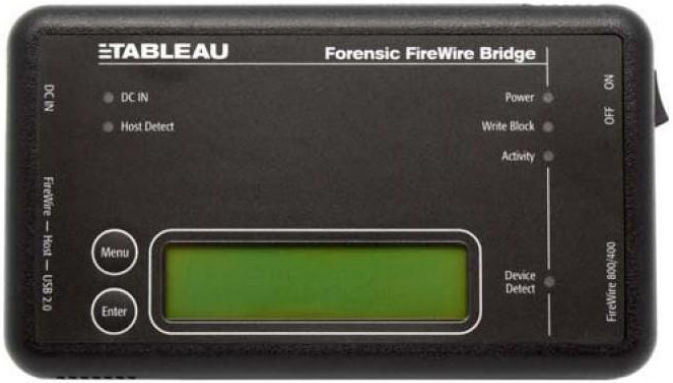
*Adapted from a Mini JukeBox setup designed by the National Library of Australia

Write Blocking – One-Way Street 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



Hard Disk Write Blocker

A WiebeTech SATA/IDE write-blocker

Hard Disk Write Blocker

This end connects to your computer using USB 2.0/3.0, eSATA, or FireWire. The cables are in the box.

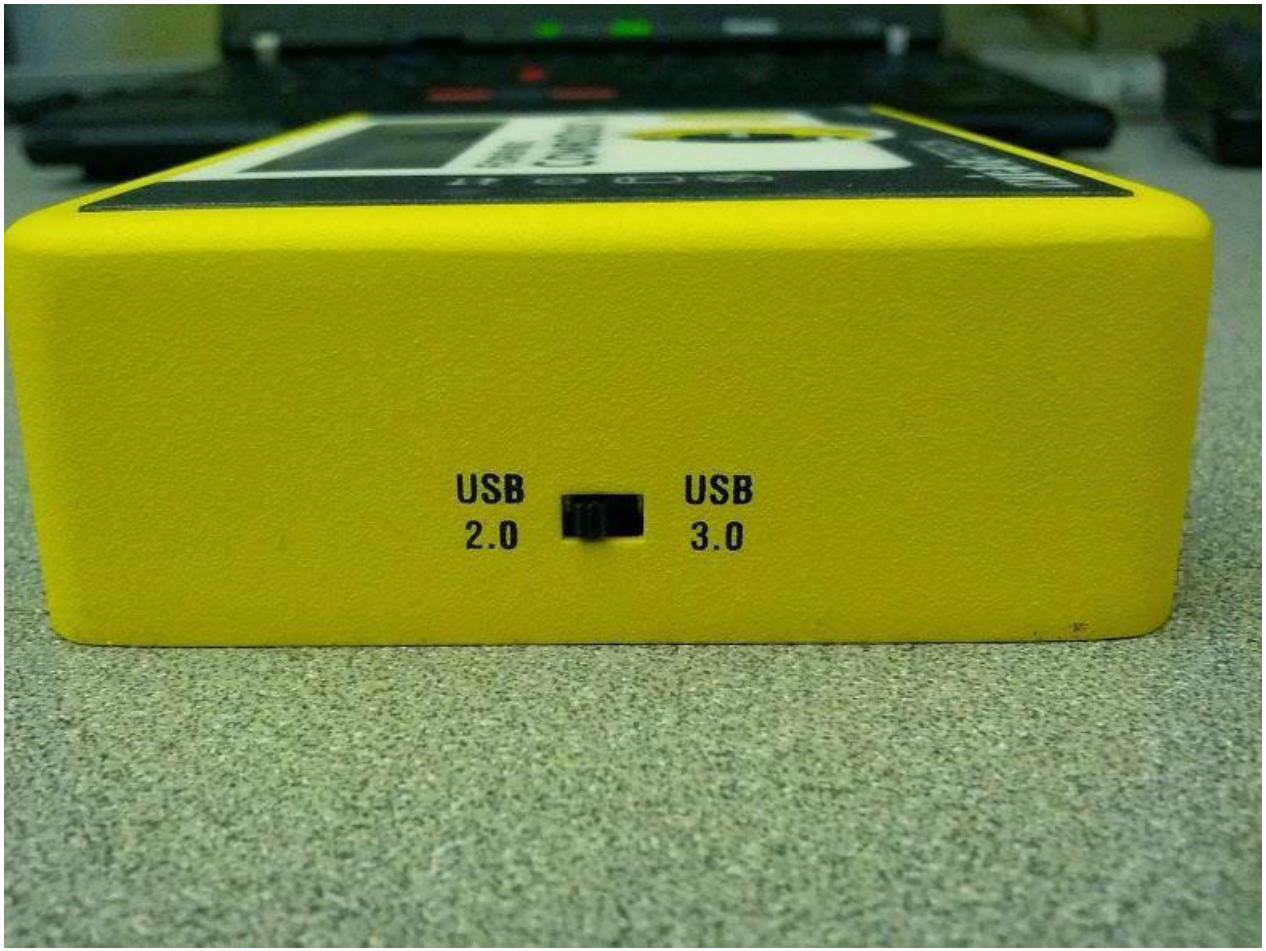
Hard Disk Write Blocker

This end connects to the drive you want to image using SATA or IDE. There are power cables in the box that can connect to either drive.

Hard Disk Write Blocker

This end powers the write blocker itself. There's a power supply in the box.

Hard Disk Write Blocker



The WiebeTech Combodock V5 supports both USB 2.0 and USB 3.0. USB 3 is faster, but may not always be compatible with all computers.

Hard Disk Write Blocker

The ComboDock has been connected to an IDE drive, a host computer, and powered on. It will allow you to select write-blocked or read/write. “Enter” enables write-blocking.

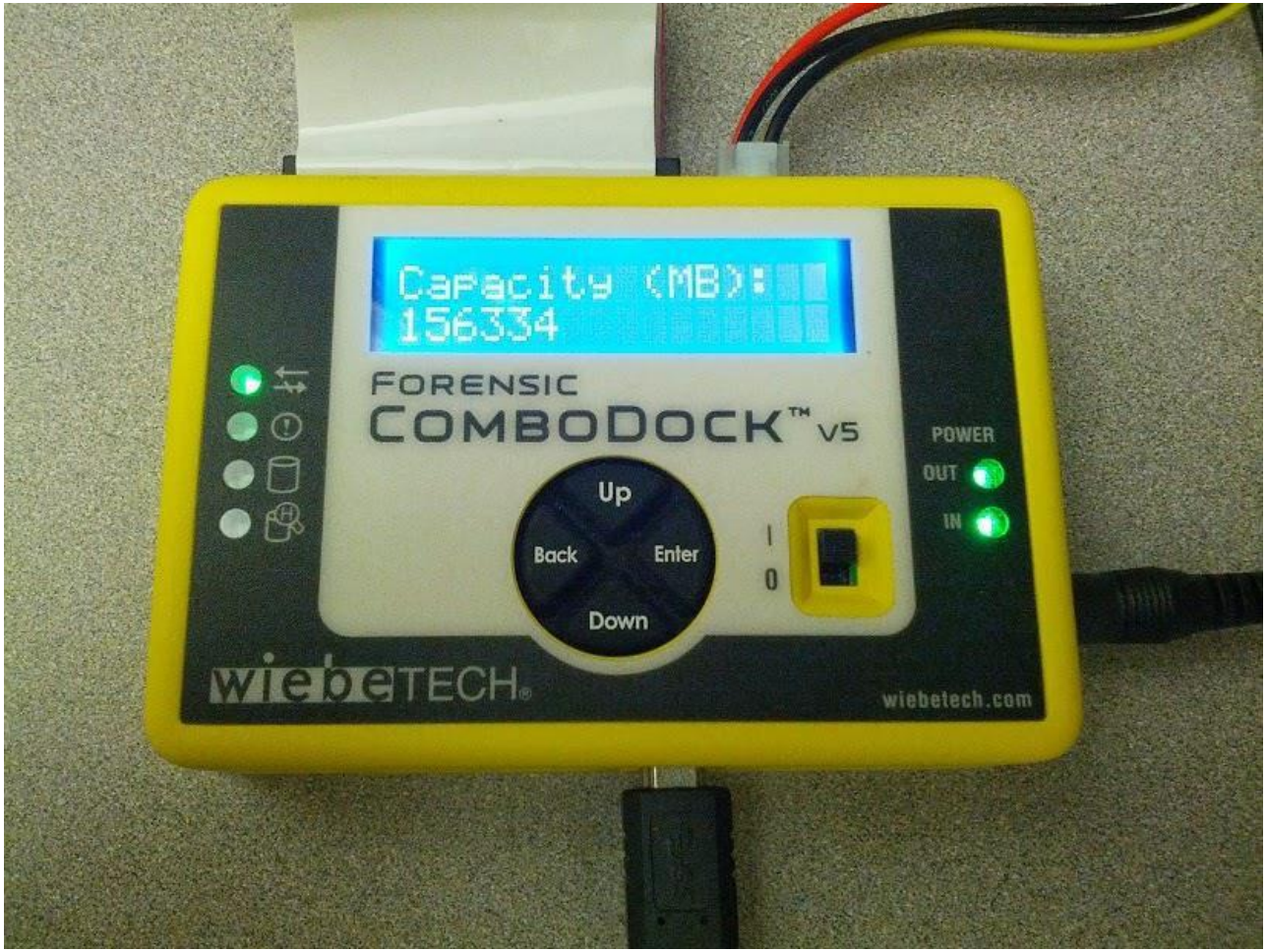
Hard Disk Write Blocker

The drive is now fully powered, and you can use the dock to examine some metadata…

Hard Disk Write Blocker

This drive has just been powered on, and it's registering as 14 degrees Celsius (57 Fahrenheit). This is well below room temperature, but it will quickly rise without a fan.

Hard Disk Write Blocker



This drive is indicating a raw capacity of 156 thousand megabytes (about 150GB).

Hard Disk Write Blocker

The name of the manufacturer is embedded in the metadata.

Why might this be important?

Hard Disk Write Blocker

The model number is next. This number is *not* unique, but common to many drives.

Hard Disk Write Blocker

The serial number, however, *is* unique.

A USB Write Blocker

USB Write Blocker

Above is a USB write blocker manufactured by Tableau.

USB Write Blocker

This end has a power connector for the blocker, and a USB ‘b’ port to connect to the host. The FireWire port is just for updating the write blocker firmware. It has no other use.

USB Write Blocker

This end connects to the USB device. This particular write-blocker will *not* recognize USB floppy drives, only USB flash drives and USB hard disks. And look, a power switch.

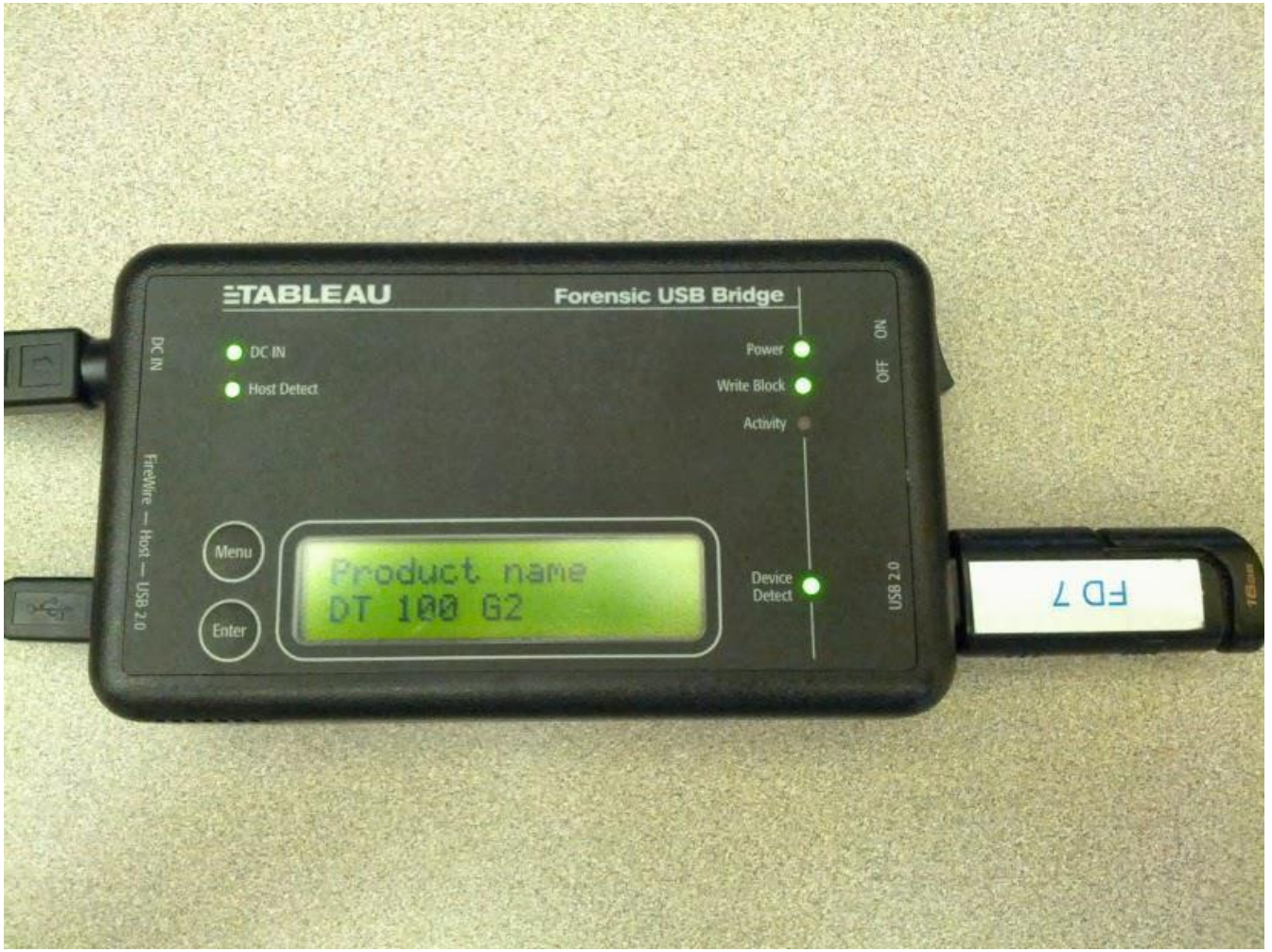
USB Write Blocker

Here's what it looks like when everything is plugged in. Plug everything in before turning the write-blocker on.

USB Write Blocker

Once you turn it on, the “Host Detect” and other lights will light up. If “Host Detect” doesn’t light up, the write blocker can’t see your computer. Something has gone wrong.

USB Write Blocker



Different write blockers may expose different metadata. This write blocker displays the product name, in addition to the manufacturer.

Host-Based Write Blockers

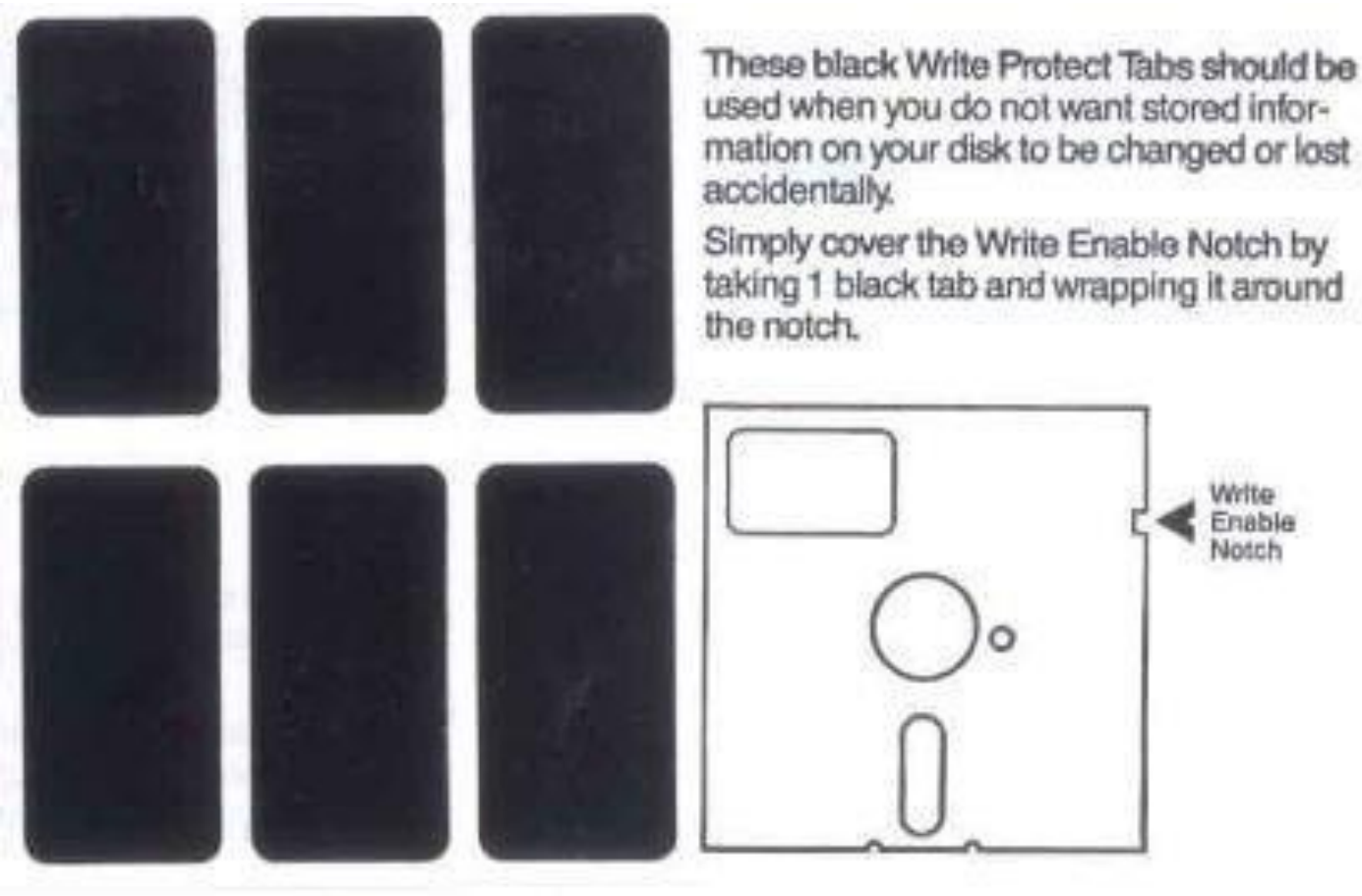
FRED UltraBay

- The “UltraBay” on the FRED provides write blocking for a range of interfaces.
  - USB (top left, next to power switch)
  - SCSI (right of USB) [no longer included in newer FREDs]
  - PATA/IDE (below SCSI)
  - SATA (left of IDE)
- It also includes a MOLEX power connector. There’s a cable in the toolbox that converts this to SATA power, if required.
- The ports in the white box on the right (top picture) are NOT write-blocked.

In earlier FREDs:

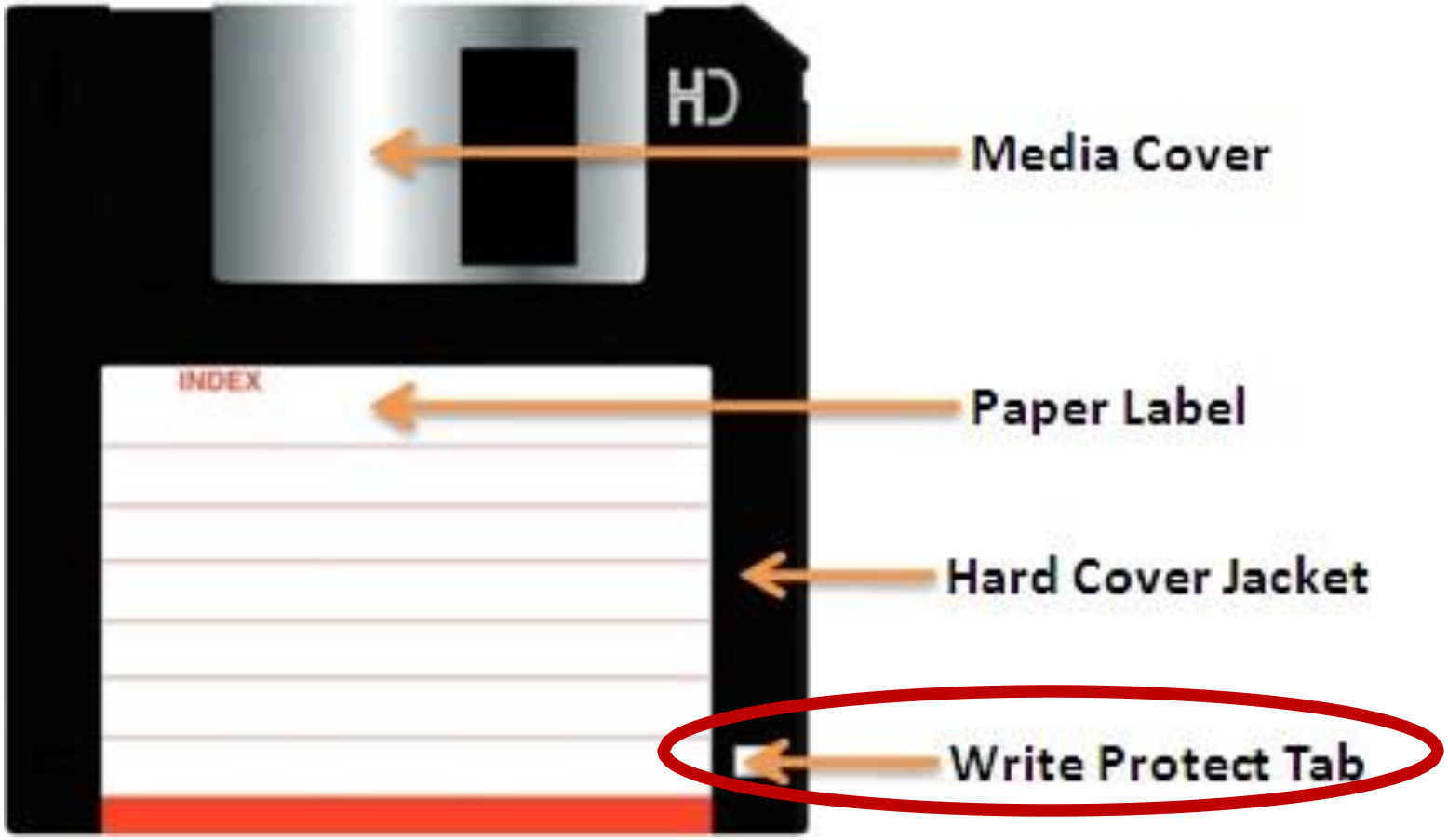
On newer FREDs:

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



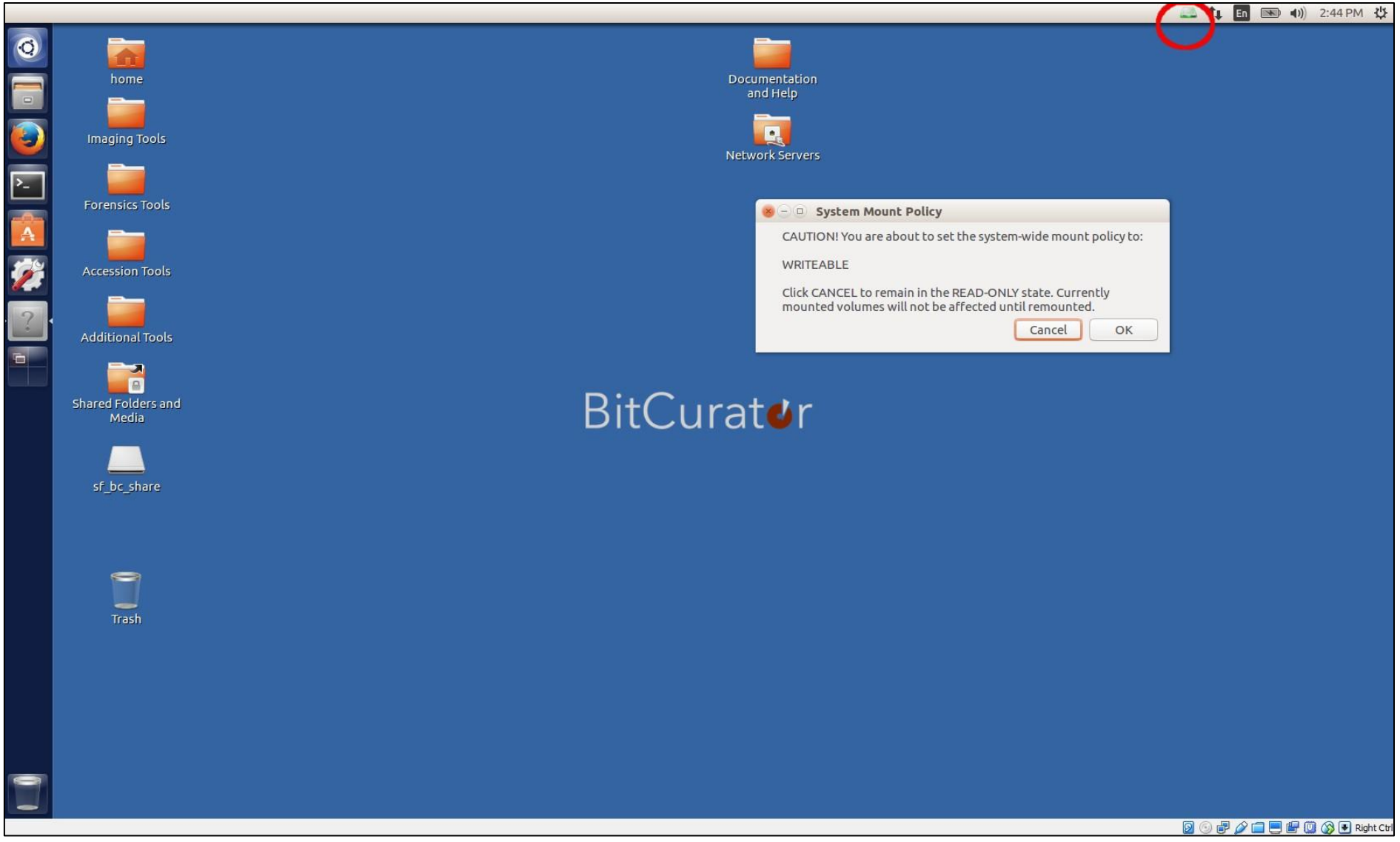
https://en.wikipedia.org/wiki/File:Floppy_tabs_3x2.jpg

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



Source: http://www.techmint.info/2009/09/security-write-protecting-floppy-disks.html
Current: https://web.archive.org/web/20100125050630/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



Other Potential System Changes to Reduce Risk of Writing to File Systems

- Macintosh: Disk Arbitrator "will block the mounting of file systems to avoid mounting as read-write and violating the integrity of the evidence" https://github.com/aburgh/Disk-Arbitrator

- Windows: See "Digital Forensics: How to configure Windows Investigative Workstations" https://www.sans.org/blog/digital-forensics-how-to-configure-windows-investigative-workstations/

Potential Elements of your own Digital Forensics Lab

FRED Options from Digital Intelligence

FRED

The Forensic Standard

Learn more » Buy »

FRED DX

Dual CPU Performance

Learn more » Buy »

FRED SR

High Capacity

Learn more » Buy »

FRED L

Forensic Laptop

Learn more » Buy »

https://digitalintelligence.com/products/overview

Mini Jukebox

Configuration #2 for Manuscripts

3 x Double DVD Drive units (MyBorg-006> MyBorg-007)

- 1 x Storage Tower (black) with 4SATA Multilane connector installed ST4SAML-B, 4 bay aluminium unit

- 2 x Plextor PX-800A DVD-RW Super Multi Drive Drives - Drive Letter Mapping W & X

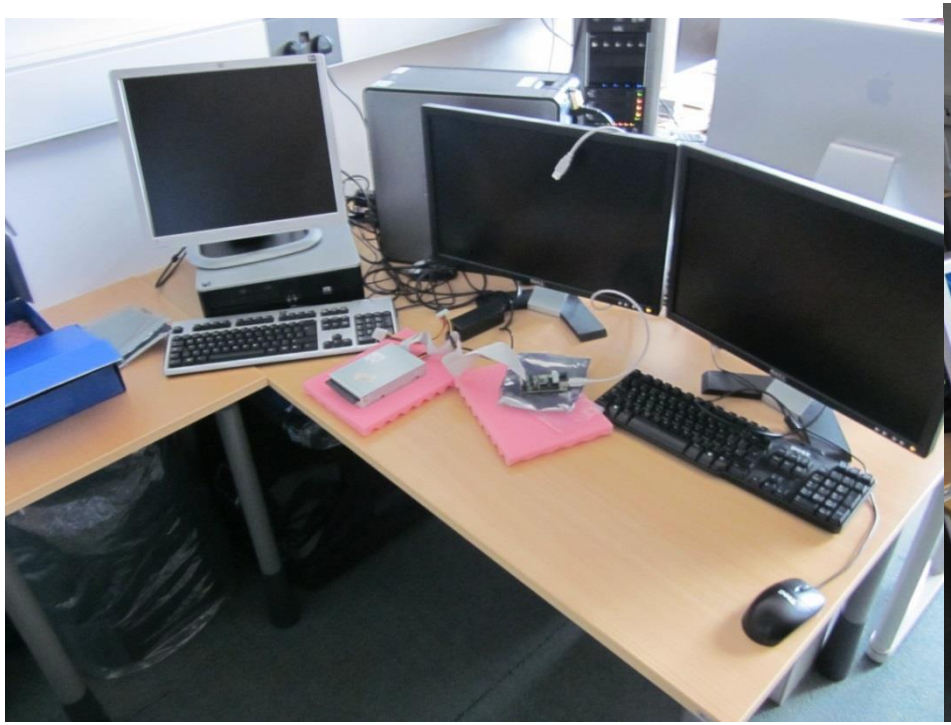
- 1 x 3.5″ floppy disk Drive Letter Mapping A or B

- 1 x Western Digital WD5001ABYS CAVIAR RE2/ 500GB Hard Drive Letter Mapping Z

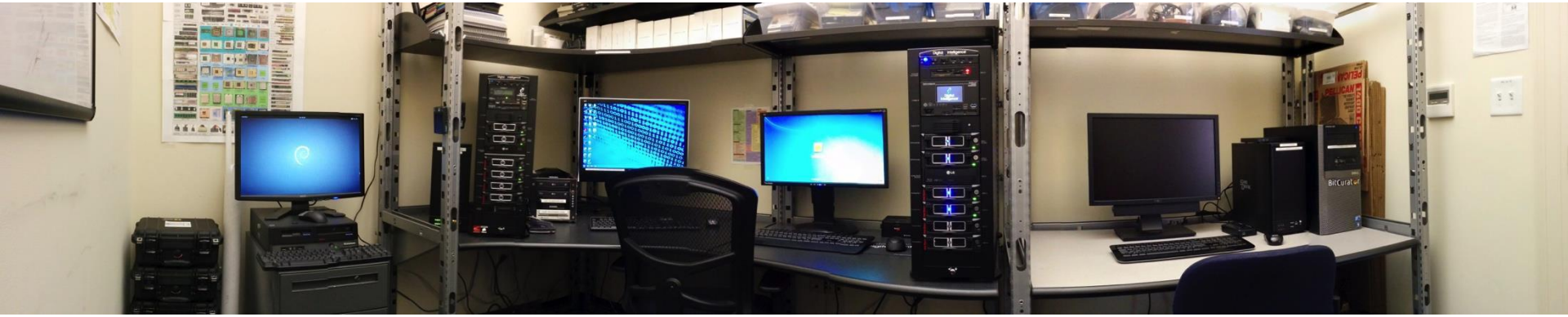
"Prometheus Component Installation Guide Pre-install requirements and methodology." National Library of Australia. October 22, 2008.

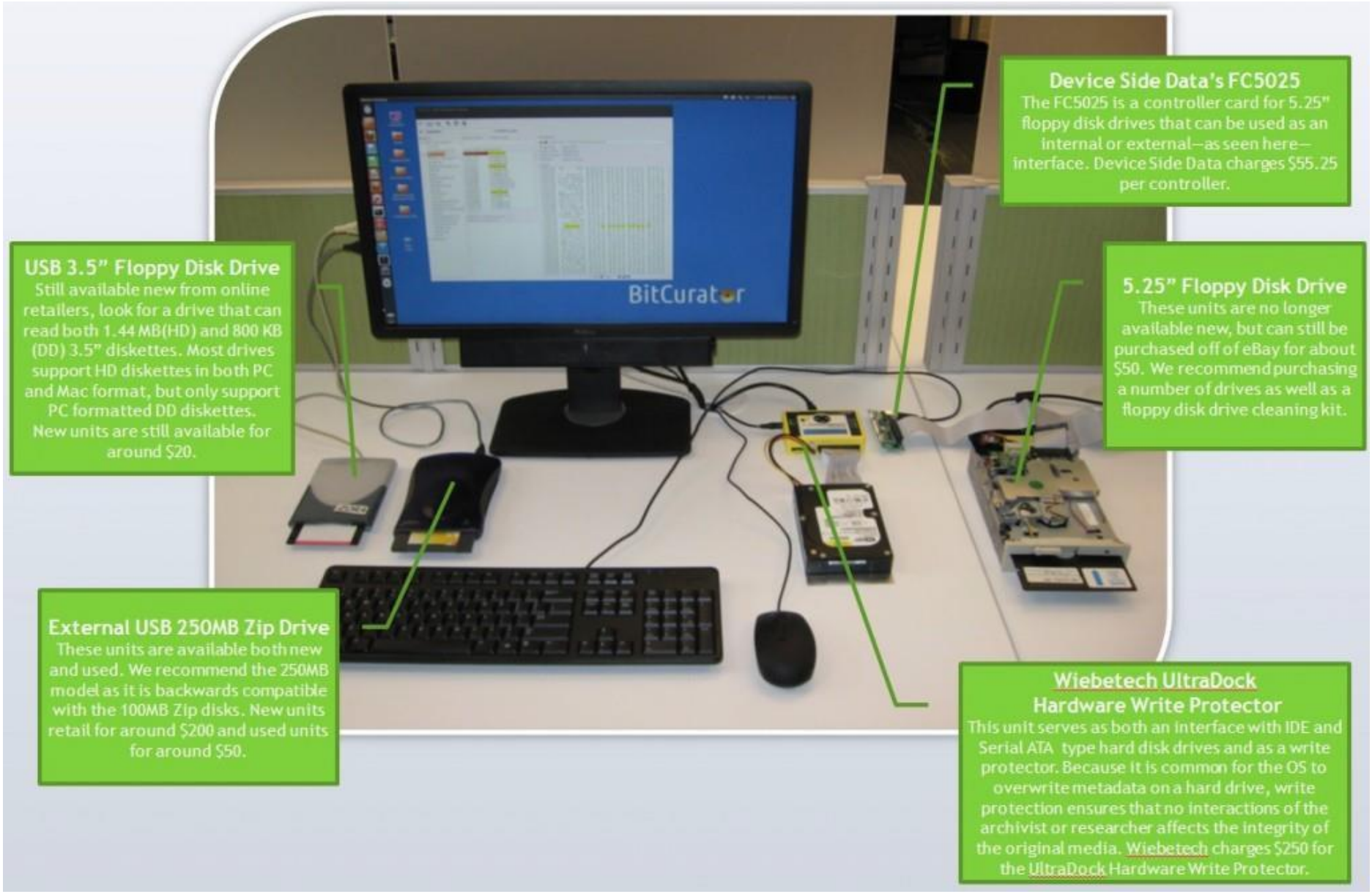
Stanford University Libraries and Academic Information Resources (SULAIR)

The British Library (UK)



School of Information and Library Science at UNC Chapel Hill, North Carolina





https://bitcurator.net/2013/08/02/building-a-digital-curation-workstation-with-bitcurator-update/

Outfitting a Born-Digital Archives Program (Ben Goldman, Penn State University)

https://practicaltechnologyforarchives.org/issue2_goldman/

Useful Resource - Mediapedia

Mediapedia

Home | Search | Advanced Search | Items Listing | Find All Items | Glossary

Item Details

Main Image: Top / Obverse — 010c.jpg

Additional Image: Bottom / Reverse — 011c.jpg

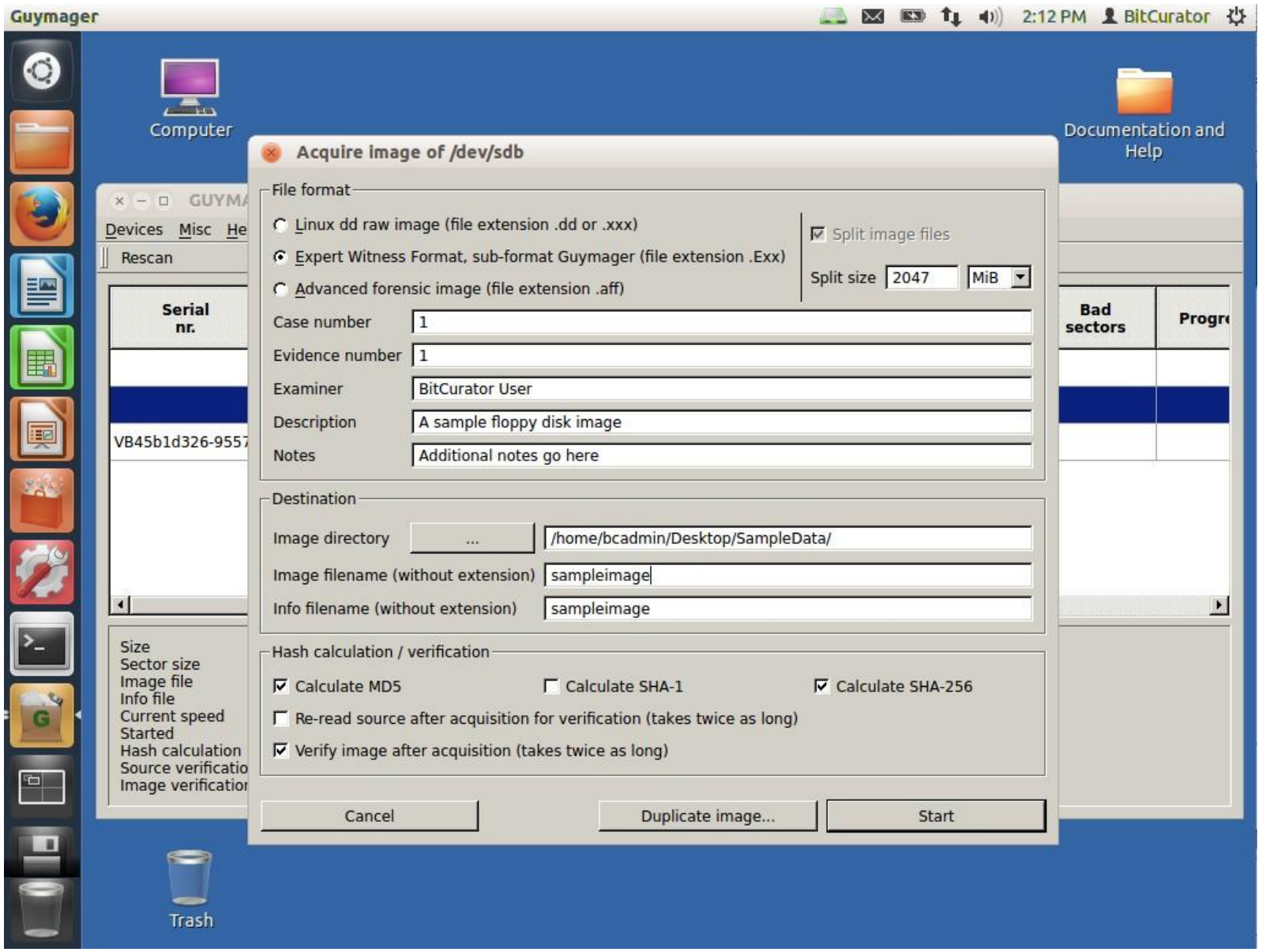
Product Name	CompacTape I (a.k.a. DLT I Tape) TK-50 ½" Data Cartridge
Name of Holotype	DLT tape (formerly CompacTape)
Product Code/Number	SKU 376706
Manufacturer	3M - Scotch - Imation
Genre(s)	Data
Carrier Type	tape cartridge
Process Type	magnetic

https://mediapedia.nla.gov.au/

Creating Exact Copies of Data from Media – Disk Images

- 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

Creating a Disk Image in Guymager



Examples of Disk Image Formats

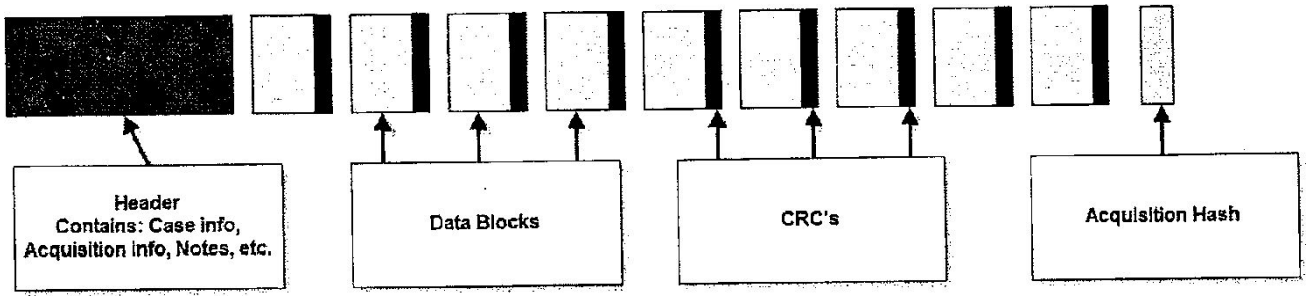
- RAW and Split RAW (RAW stored across multiple files)
- Advanced Forensics Format (AFF) [no longer recommended]
- EnCase Evidence File (.E01)
- ISO (for CD-ROM)
- IMG (floppy or sometimes CD-ROM)

RAW (dd)

- Copies of the raw media data. Often split into smaller chunks to make them more manageable and so that the resulting images can fit onto limited filesystems and media such as FAT or DVD/CDROM.
- Advantages:
  - Very simple, use simple tools to manipulate the image.
  - Image can be easily split for storage and transport on removable media
  - Output can be piped to other applications for immediate processing
- Disadvantages:
  - Can be very large (no compression). Zipped raw images cannot be operated on directly with regular tools (efficiently perform arbitrary seeks).
  - Often too large to store on FAT formatted media
  - No metadata other than filenames, no hashes.
  - No checksumming on files – not robust
    - Missing segments (for example from scratched CD/DVD – can sometimes be overwritten with 0's).
    - Overwritten data (unrecoverable – no checksums on small blocks in file).

Expert Witness Format (Encase)

- Evidence file consists (in order) of: Acquisition information, Data Block, CRC (cyclic redundancy check), acquisition hash (MD5)
- Can be split for storage, transport
- CRC computed for every 32K block; balance between integrity and speed, also makes it very difficult to tamper with the evidence file (1 in 4 billion chance of collision)
- Cannot be manipulated with simple (open source UNIX) tools; support reverse engineered in libewf
- Previously limited to 2GB size
- Largely proprietary
- Has been reverse engineered by Joachim Metz in libewf (used in open source tools that read EWF) - https://github.com/libyal/libewf



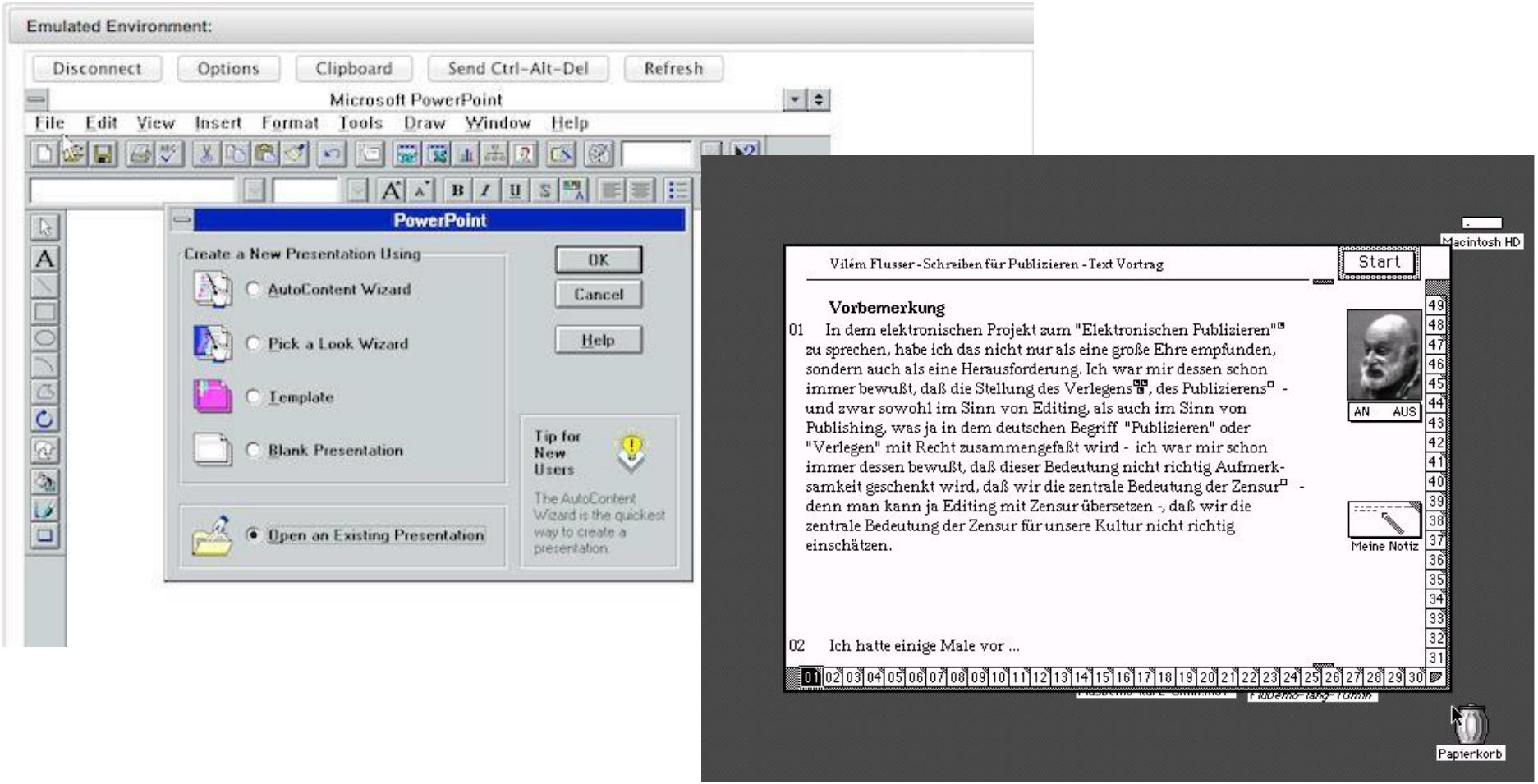
ISO Images (.iso extension) for CD-ROM and DVD

- Similar to raw, but can't contain
  - multiple tracks
  - audio or video tracks
- Don't contain control headers or error correction fields (raw can include these)
- Filesystem usually will be either ISO 9660 (CD-ROM) or UDF (DVDs)

Accessing Data in Disk Images

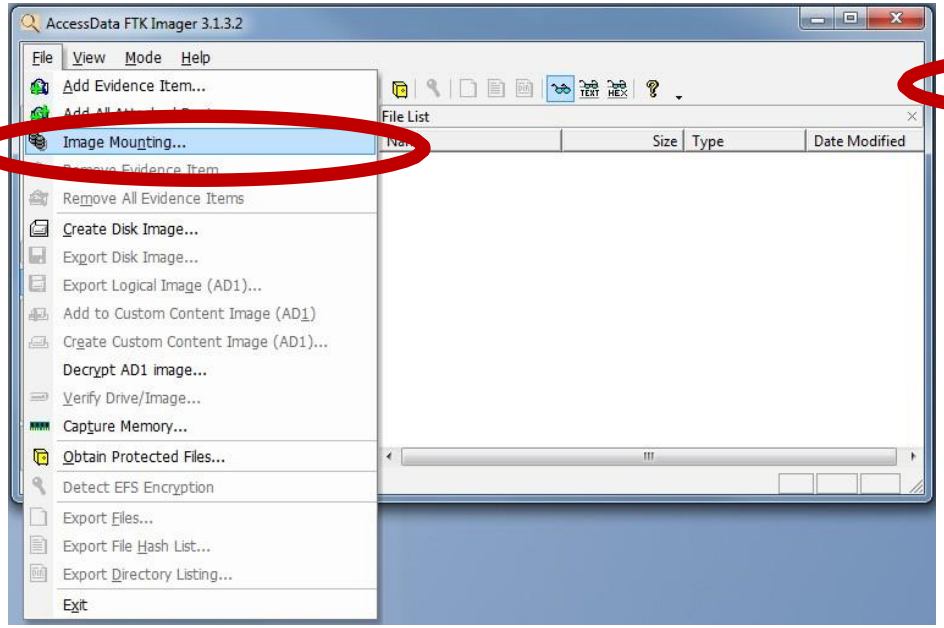
- Virtualization and emulation
- Mounting the original filesystem
- Accessing (but not mounting) disk images using forensics software
- Two options discussed later for end user access:
  - Remote, dynamic access to disk image contents
  - Cross-drive analysis

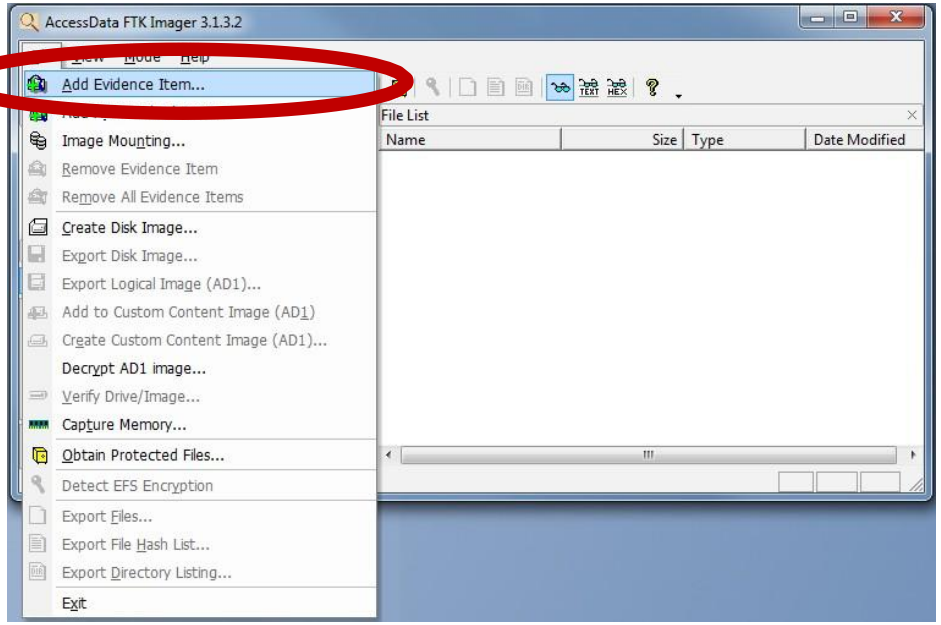
Emulation as a Service



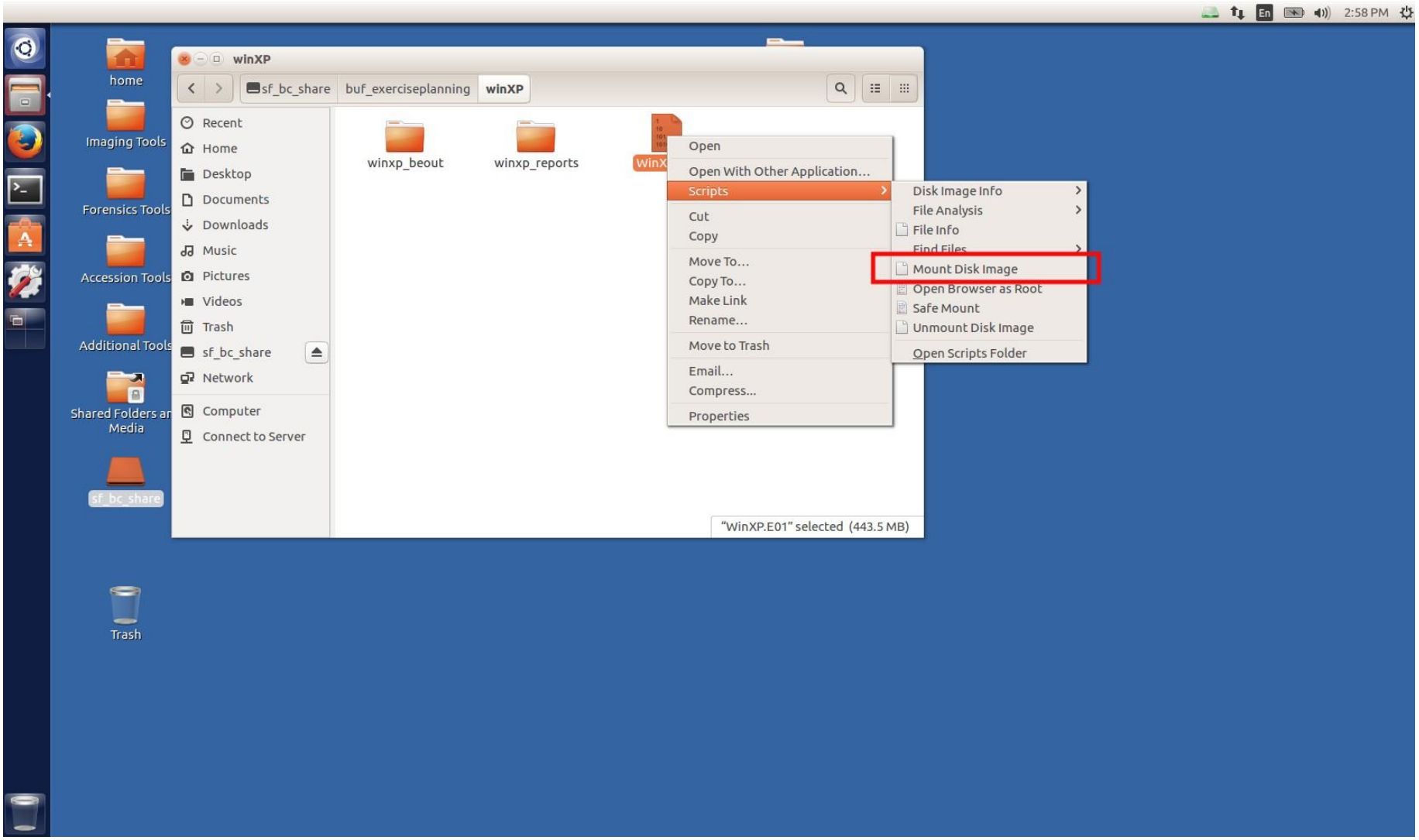
https://eaasi.gitlab.io/program_docs/intro-emulation-workshop/06-eaasi/index.html

What's the difference between the two options shown in FTK Imager below?

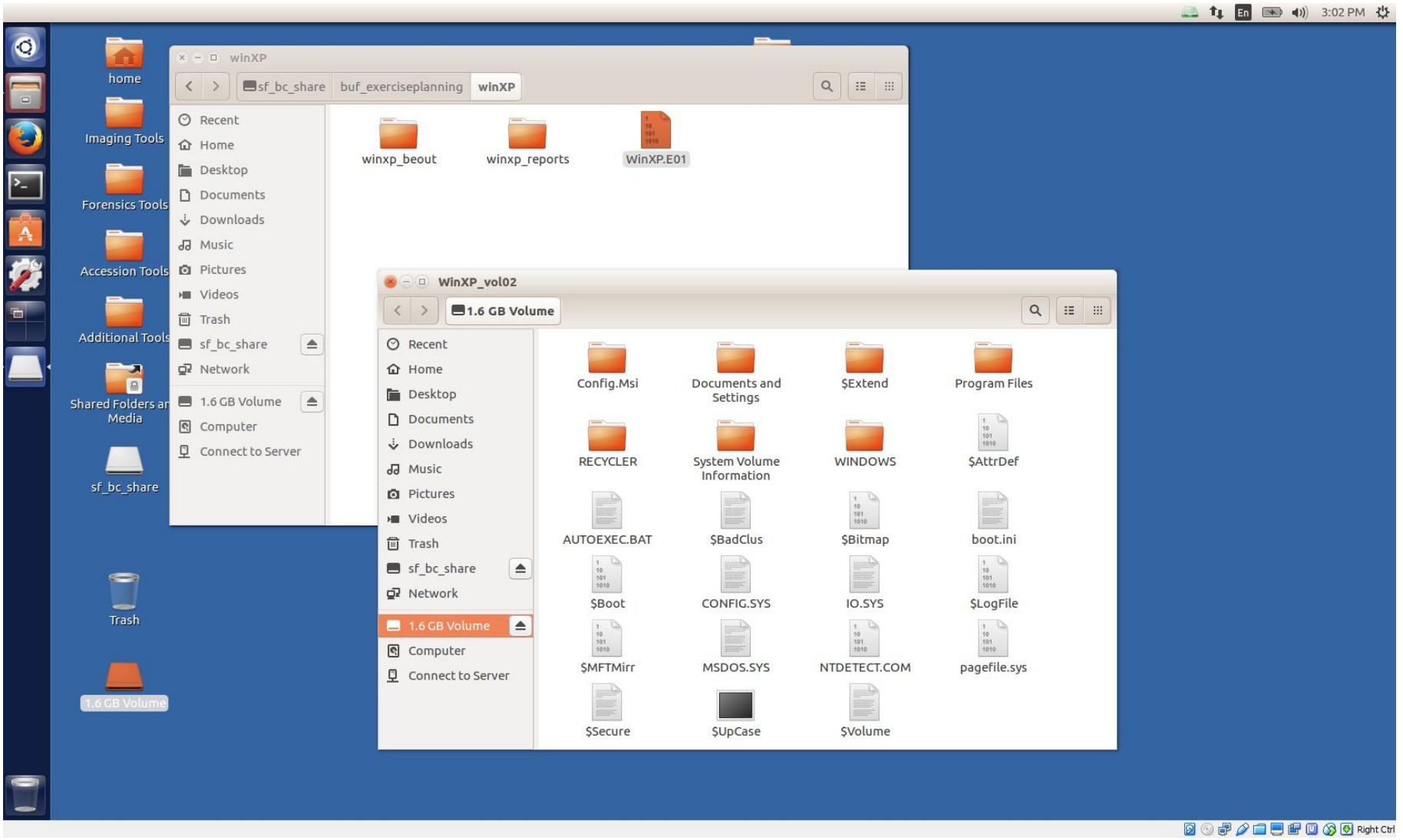




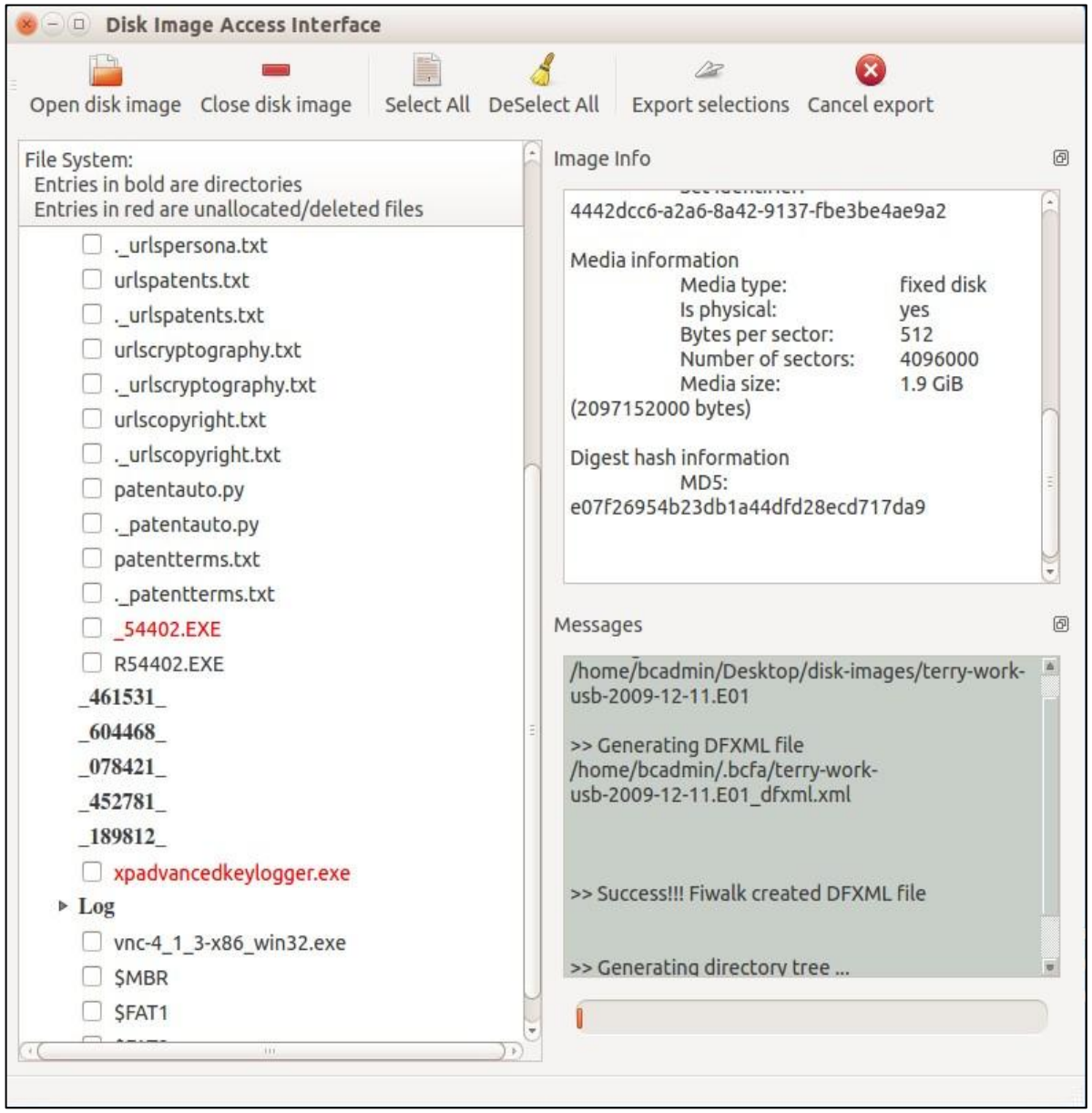
Mounting a Disk Image to Browse the Contents



Mounting a Disk Image to Browse the Contents



Exporting Selected Files from a Disk Image



Exercise: Multiple Views into Disk Image Files

- Resources we'll be using (also in the zip file you downloaded):
  1. ISO file - http://www.ils.unc.edu/callee/25.iso
  2. IMG file – https://distro.ibiblio.org/bitcurator/lab/something.img
  3. OSFMount (Windows only)
  4. FTK Imager (Windows only)
  5. BitCurator Environment
- Step 1 – Mount the ISO and IMG files using **OSFMount**
- Step 2 – Find the drives using **Windows Explorer** and investigate their contents
- Step 3 – Open **FTK Imager** and add both images as evidence items, and explore what we see in the drives
- Step 4 – Use the **BitCurator environment** to mount the disk images [Right click on image file, then select: Scripts > Mount Disk Image]
- Step 5 – Use the **BitCurator environment** to select files within the images to export [Use Forensics Tools > BitCurator Disk Image Access]

Bit-Level Treatment of Individual Files

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

Hex Dump Tools

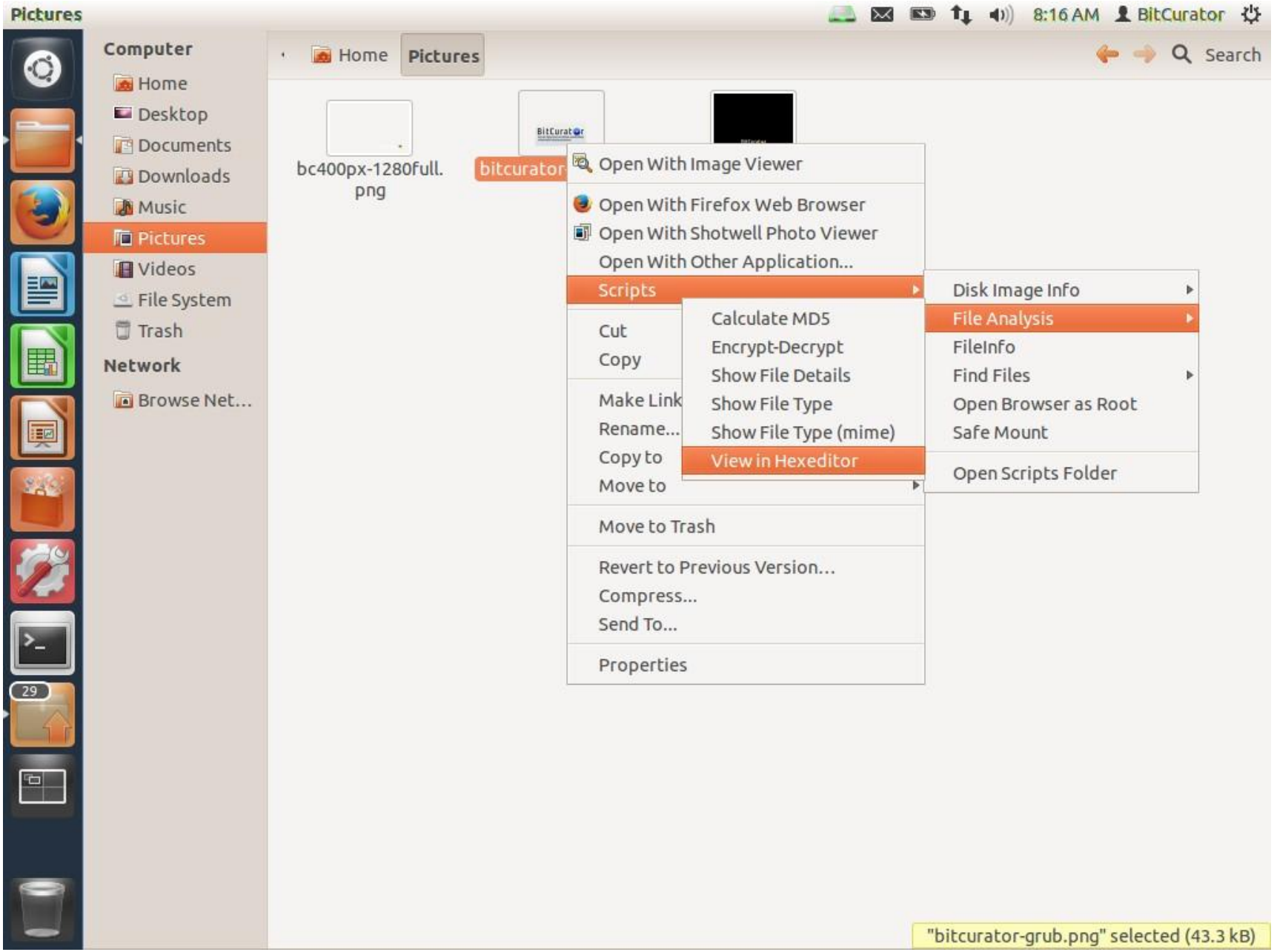
- Many free or inexpensive tools available for download, e.g. Cygnus Hex Editor, Hex Workshop, HexAssistant, MiniDumper, Hex Fiend (Mac), GHex (Linux)*

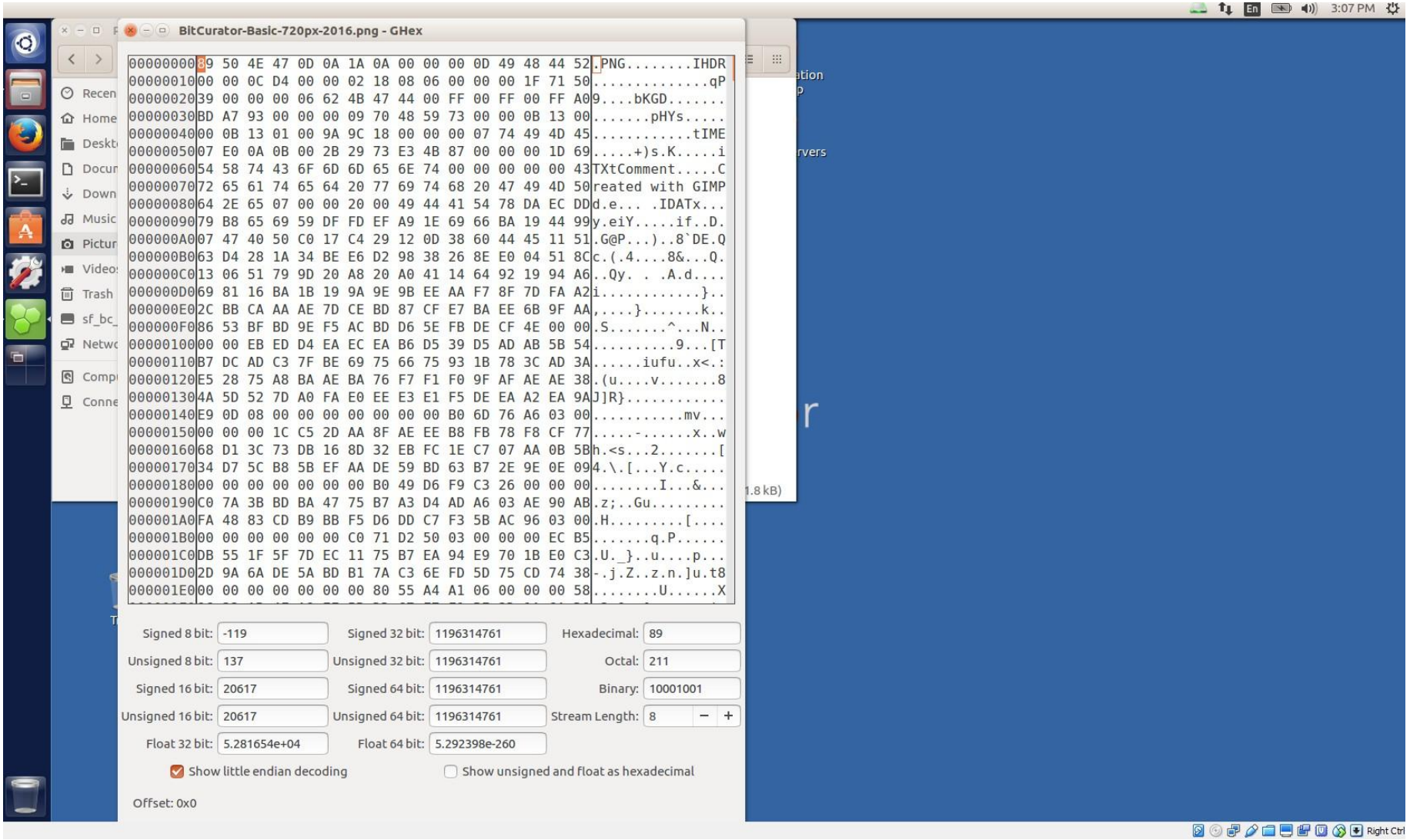
- BitCurator environment has a built-in hex editor (GHex)

  Online tool: https://hexed.it/

* See https://en.wikipedia.org/wiki/Comparison_of_hex_editors

In the BitCurator environment:




BitCurator-Basic-720px-2016.png - GHex
00000000 89 50 4E 47 0D 0A 1A 0A 00 00 00 0D 49 48 44 52 .PNG........IHDR
00000010 00 00 0C D4 00 00 02 18 08 06 00 00 00 1F 71 50 ....qP
00000020 39 00 00 00 06 62 4B 47 44 00 FF 00 FF 00 FF A0 9....bKGD.......
00000030 BD A7 93 00 00 00 09 70 48 59 73 00 00 0B 13 00 ....pHYs.....
00000040 00 0B 13 01 00 9A 9C 18 00 00 00 07 74 49 4D 45 ....tIME
00000050 07 E0 0A 0B 00 2B 29 73 E3 4B 87 00 00 00 1D 69 .....+)s.K.....i
00000060 54 58 74 43 6F 6D 6D 65 6E 74 00 00 00 00 00 43 TXtComment.....C
00000070 72 65 61 74 65 64 20 77 69 74 68 20 47 49 4D 50 reated with GIMP
00000080 64 2E 65 07 00 00 20 00 49 44 41 54 78 DA EC DD d.e... .IDATx...
00000090 79 B8 65 69 59 DF FD EF A9 1E 69 66 BA 19 44 99 y.eiY.....if..D.
000000A0 07 47 40 50 C0 17 C4 29 12 0D 38 60 44 45 11 51 .G@P...)..8`DE.Q
000000B0 63 D4 28 1A 34 BE E6 D2 98 38 26 8E E0 04 51 8C c.(.4....8&...Q.
000000C0 13 06 51 79 9D 20 A8 20 A0 41 14 64 92 19 94 A6 ..Qy. . .A.d....
000000D0 69 81 16 BA 1B 19 9A 9E 9B EE AA F7 8F 7D FA A2 i............}..
000000E0 2C BB CA AA AE 7D CE BD 87 CF E7 BA EE 6B 9F AA ,....}.......k..
000000F0 86 53 BF BD 9E F5 AC BD D6 5E FB DE CF 4E 00 00 .S.......^...N..
00000100 00 00 EB ED D4 EA EC EA B6 D5 39 D5 AD AB 5B 54 ....9...[T
00000110 B7 DC AD C3 7F BE 69 75 66 75 93 1B 78 3C AD 3A ......iufu..x<.:
00000120 E5 28 75 A8 BA AE BA 76 F7 F1 F0 9F AF AE AE 38 .(u....v.......8
00000130 4A 5D 52 7D A0 FA E0 EE E3 E1 F5 DE EA A2 EA 9A J]R}............
00000140 E9 0D 08 00 00 00 00 00 00 00 B0 6D 76 A6 03 00 ....mv...
00000150 00 00 00 1C C5 2D AA 8F AE EE B8 FB 78 F8 CF 77 .....-......x..w
00000160 68 D1 3C 73 DB 16 8D 32 EB FC 1E C7 07 AA 0B 5B h.<s...2.......[
00000170 34 D7 5C B8 5B EF AA DE 59 BD 63 B7 2E 9E 0E 09 4.\.[...Y.c.....
00000180 00 00 00 00 00 00 00 B0 49 D6 F9 C3 26 00 00 00 ....I...&...
00000190 C0 7A 3B BD BA 47 75 B7 A3 D4 AD A6 03 AE 90 AB .z;..Gu.........
000001A0 FA 48 83 CD B9 BB F5 D6 DD C7 F3 5B AC 96 03 00 .H.........[....
000001B0 00 00 00 00 00 00 C0 71 D2 50 03 00 00 00 EC B5 ....q.P......
000001C0 DB 55 1F 5F 7D EC 11 75 B7 EA 94 E9 70 1B E0 C3 .U._}..u....p...
000001D0 2D 9A 6A DE 5A BD B1 7A C3 6E FD 5D 75 CD 74 38 -.j.Z..z.n.]u.t8
000001E0 00 00 00 00 00 00 00 80 55 A4 A1 06 00 00 00 58 ....U......X
Signed 8 bit: -119
Signed 32 bit: 1196314761
Hexadecimal: 89
Unsigned 8 bit: 137
Unsigned 32 bit: 1196314761
Octal: 211
Signed 16 bit: 20617
Signed 64 bit: 1196314761
Binary: 10001001
Unsigned 16 bit: 20617
Unsigned 64 bit: 1196314761
Stream Length: 8
Float 32 bit: 5.281654e+04
Float 64 bit: 5.292398e-260
Show little endian decoding
Show unsigned and float as hexadecimal
Offset: 0x0
3:07 PM
Right Ctrl

Identifying File Types

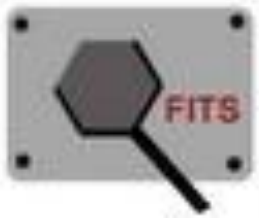
- Magic numbers and file signatures
- File extensions
- Metadata stored in file system
- MIME types

Magic Numbers and File Signatures

- Distinct string or pattern that is found within files of a given type (most often in the header)
- Most effective searches for magic numbers often involve regular expressions (e.g. grep) in order to indicate multiple variations of a pattern
- Utilities that use this: file (Unix), TrID, DROID, FITS
- Examples:

File Format	Hex	ASCII
DOC	D0 CF 11 E0 A1 B1 1A E1	ÐÏà¡±á
JPG	FF D8 FF	ÿØÿ
PDF	25 50 44 46 2D 31 2E	%PDF-1.
ZIP	50 4B 03 04	PK..

File Information Tool Set (FITS) https://code.google.com/p/fits

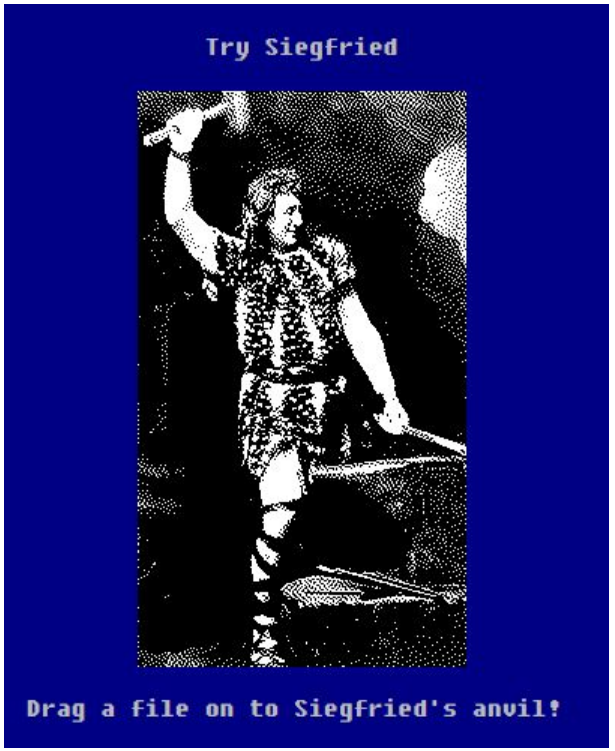


- FITS “identifies, validates, and extracts technical metadata for various file formats. It wraps several third-party open source tools, normalizes and consolidates their output, and reports any errors. FITS was created by the Harvard University Library Office for Information Systems for use in its Digital Repository Service (DRS).”
- Tools currently bundled into it:
  - Jhove
  - Exiftool
  - National Library of New Zealand Metadata Extractor
  - DROID
  - FFIdent
  - File Utility (windows)
- FITS may no longer be available as a part of BitCurator due to JRE incompatibilities

Siegfried

http://www.itforarchivists.com/siegfried/

- Signature-based file format identification tool
  - PRONOM file format signatures (National Archives of UK) (default)
  - MIME-info file format signatures (freedesktop.org)
  - FDD file format signatures (Library of Congress)
- Unlike FITS, does not have validation built in, and fewer extraction tools, but much lighter weight.
- Has a lot of customization for output
  - CSV
  - YAML (text)
  - DROID CSV
  - JSON
  - stdout



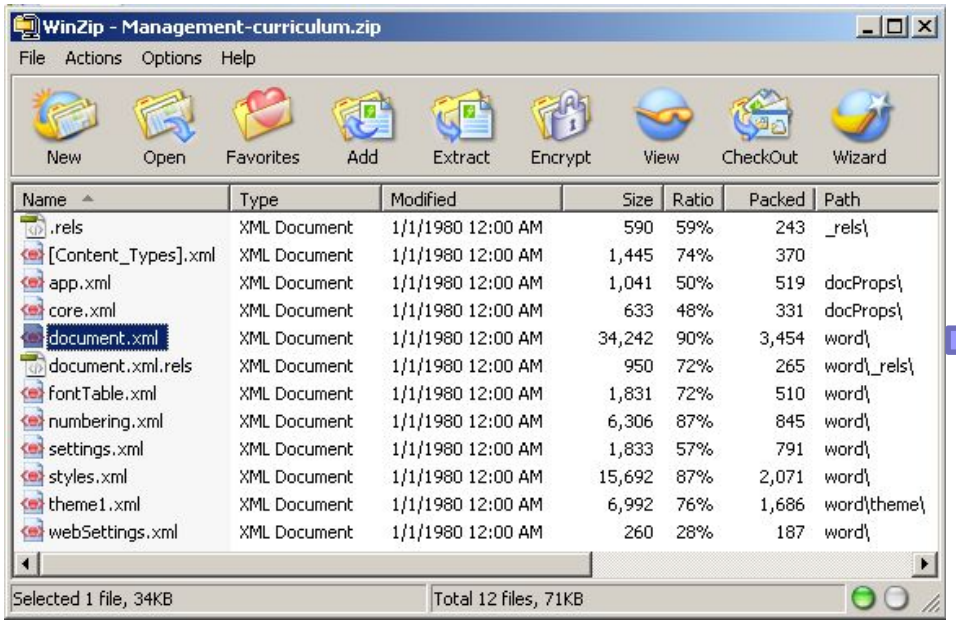
Brunnhilde

https://github.com/tw4l/brunnhilde

- Reporting companion for Siegfried
  - Requires Siegfried (but running Brunnhilde also runs Siegfried)
  - Command-line and GUI (we'll be using the CLI version later)
- Reports generated
  - HTML (human readable)
  - Siegfried CSV
  - Directory tree
  - Other CSVs extracted from Siegfried logs (e.g., warnings, unidentified files)
- Can run other processes too, but not required
  - Virus scan
  - bulk_extractor
  - Disk image processing

File Extensions

- Changing file extension usually changes default application OS uses to open (i.e. associates with) the file
- The "8.3" (eight characters, followed by three-character extension) limit in the past – based on FAT – resulted in many creative uses of the extension part of the file name (e.g. reports1.994, april-94.rpt)
- Convention is often still to use only three letters
- No authority for standardizing use, so three-letter extensions are often shared by many formats
- Security risks associated with trusting the file extension to be accurate – malicious code masquerading as another type of file (e.g. viruses sent as email attachments)




```
1  <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2  <w:document xmlns:ve="http://schemas.openxmlformats.org/markup-compatibility/2006"
3    xmlns:o="urn:schemas-microsoft-com:office:office"
4    xmlns:r="http://schemas.openxmlformats.org/officeDocument/2006/relationships"
5    xmlns:m="http://schemas.openxmlformats.org/officeDocument/2006/math"
6    xmlns:v="urn:schemas-microsoft-com:vml"
7    xmlns:wp="http://schemas.openxmlformats.org/drawingml/2006/wordprocessingDrawing"
8    xmlns:w10="urn:schemas-microsoft-com:office:word"
9    xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main"
10   xmlns:wne="http://schemas.microsoft.com/office/word/2006/wordml">
11   <w:body>
12     <w:p w:rsidR="00CD2E84" w:rsidRDefault="00015E33" w:rsidP="00015E33">
13       <w:pPr>
14         <w:jc w:val="center"/>
15         <w:rPr>
16           <w:b/>
17         </w:rPr>
18       </w:pPr>
19       <w:r>
20         <w:rPr>
21           <w:b/>
22         </w:rPr>
23         <w:t>INLS 585: Management for Information Professionals</w:t>
24       </w:r>
25     </w:p>
26     <w:p w:rsidR="00015E33" w:rsidRDefault="00015E33" w:rsidP="00015E33">
27       <w:r>
28         <w:rPr>
29           <w:i/>
30         </w:rPr>
31         <w:t>Texts:</w:t>
32       </w:r>
33     </w:p>
34     <w:p w:rsidR="00015E33" w:rsidRDefault="00015E33" w:rsidP="00015E33">
35       <w:pPr>
36         <w:spacing w:after="0" w:line="240" w:lineRule="auto"/>
37       </w:pPr>
38       <w:r>
39         <w:t xml:space="preserve">Robbins, S.P. and D.A. </w:t>
40       </w:r>
```

MIME types ("Content-type", "internet media type")

- Widely adopted and recognized by applications
- Based on two-level hierarchy (e.g. text/html, application/octet-stream, image/tiff)
- Major advantage is official registration of MIME types through a central authority

MIME types

Name	MIME Type / Internet Media Type	File Extension	More Details
3D Crossword Plugin	application/vnd.hzn-3d-crossword	.x3d	IANA: 3D Crossword Plugin
3GP	video/3gpp	.3gp	Wikipedia: 3GP
3GP2	video/3gpp2	.3g2	Wikipedia: 3G2
3GPP MSEQ File	application/vnd.mseq	.mseq	IANA: 3GPP MSEQ File
3M Post It Notes	application/vnd.3m.post-it-notes	.pwn	IANA: 3M Post It Notes
3rd Generation Partnership Project - Pic Large	application/vnd.3gpp.pic-bw-large	.plb	3GPP
3rd Generation Partnership Project - Pic Small	application/vnd.3gpp.pic-bw-small	.psb	3GPP
3rd Generation Partnership Project - Pic Var	application/vnd.3gpp.pic-bw-var	.pvb	3GPP
3rd Generation Partnership Project - Transaction Capabilities Application Part	application/vnd.3gpp2.tcap	.tcap	3GPP
7-Zip	application/x-7z-compressed	.7z	Wikipedia: 7-Zip

https://www.iana.org/assignments/media-types/media-types.xhtml

National Software Reference Library (NSRL)

- The NSRL (https://www.nsrl.nist.gov) includes a library of hashes of files associated with a large number of software tools developed over the past few decades. See the product list at:

  https://www.nsrl.nist.gov/RDS/rds_2.41/ProdList.txt

- Various third-party tools can be used to access the NSRL. There's a web interface available at:

  https://www.hashsets.com/home/

  (Navigate to Hash Set Engines > National Software Reference Library > SEARCH BY NAME / MD5). But it often generates invalid results, so the following instructions are based on running a command-line tool instead.

Using NSRL Hash Sets to Investigate System Files

- Find a directory from your computer that contains system files.
  - For Windows, a good place to look is in Computer > Local Disk (C:) > Program Files. For example, you could select Program Files > 7-Zip.
  - On a Mac, look in /Applications/ and select a specific folder
- Move the contents of the directory to a new folder called system-files on your host computer's desktop.
- Navigate to your shared folders [Desktop > Shared Folders and Media] in the BitCurator environment and copy the folder system-files to the desktop of the BitCurator environment
- Use md5deep to create a set of md5 hashes of the files in the system-files folder, then pipe the output into nsrllookup to generate lists of known and unknown hashes:
  - *md5deep –r ~/Desktop/system-files | nsrllookup –s nsrllookup.com -K known-hashes.txt -U unknown-hashes.txt*
- What is the above command doing?
- Look at the contents of the two files:
  - type known-hashes.txt
  - type unknown-hashes.txt

For Your Reference: Running NSRL Lookup in Windows

- Visit: https://rjhansen.github.io/nsrllookup/ and download the Windows binary (64-bit).
- Open the .zip file and extract the executable to your desktop.
- Visit: https://github.com/jessek/hashdeep/releases and download md5deep-4.4.zip.
- Open the .zip file and extract md5deep64.exe to your desktop.
- Open a command prompt window (in the start box, type “cmd”). Navigate to your desktop (cd Desktop).
- Type: nsrllookup –help
- Same commands as in previous slide but *use quotation marks around the file path in the command.*

Exercise: Using PRONOM

The PRONOM technical registry contains information about a wide variety of file formats, including versioning information. You can find it at https://www.nationalarchives.gov.uk/PRONOM/Default.aspx. PRONOM has an online search feature that can be used to view the registry.

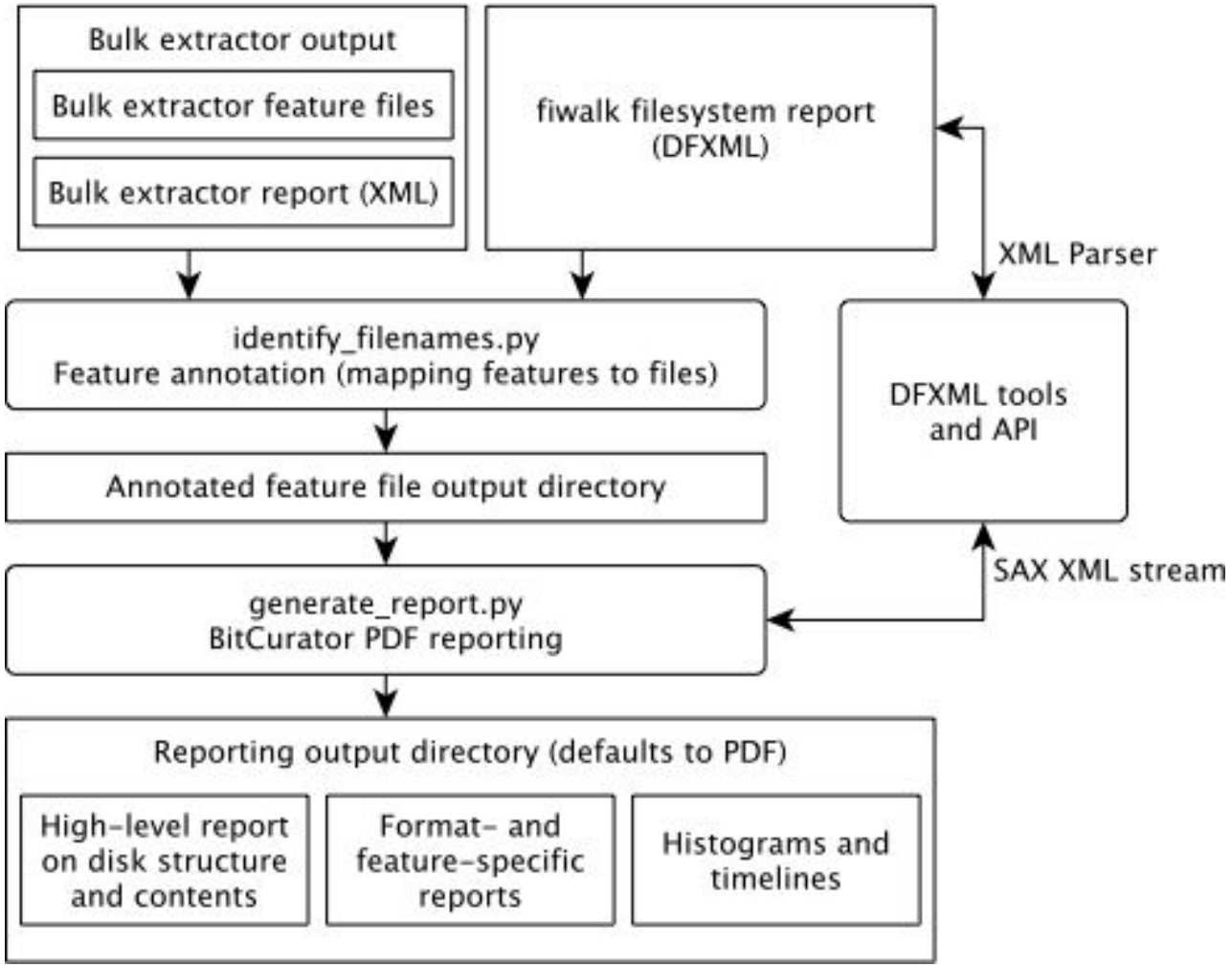
Click on “Search PRONOM” and navigate to the “File Format” tab. Clicking on the first search button (under “1. File Formats”) will allow you to view all of the entries in the registry.

DROID incorporates information from PRONOM. It also uses file magic and file format extensions to provide a “best effort” at identifying file types. If you’d like to know more about DROID, you can find a quick demonstration video at: https://vimeo.com/24718678

Note: We’ll see DROID output in the Siegfried exercise later.

Creating and Extracting Forensic Metadata

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.

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

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.

Code
Issues 8
Pull requests 0
Pulse
Graphs

HTTPS clone URL
https://github.com/c
You can clone with HTTPS or Subversion.
Clone in Desktop
Download ZIP

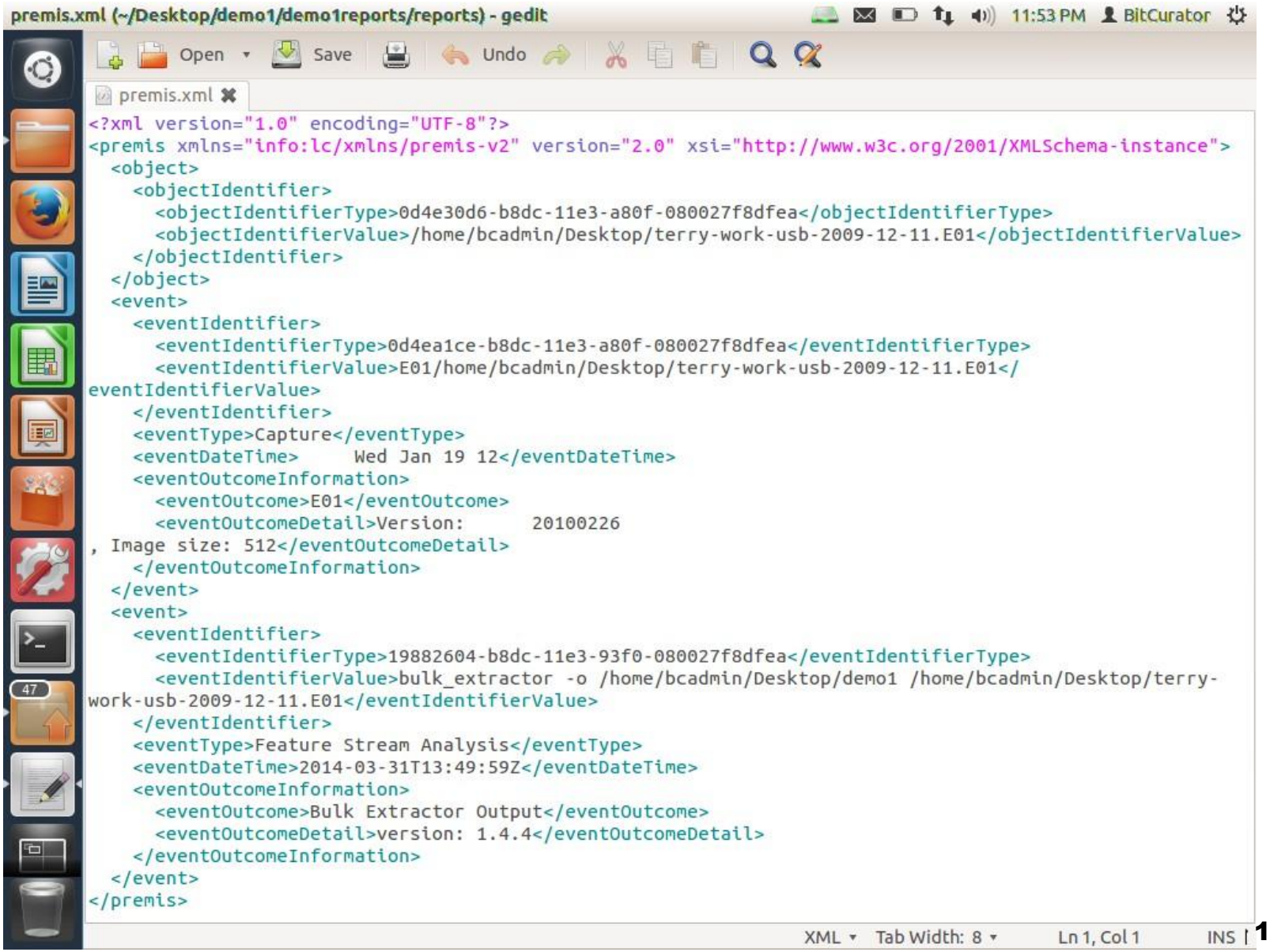
https://github.com/dfxml-working-group/dfxml_schema

Operationalizing Original Order - Filesystem Metadata Output from fiwalk*

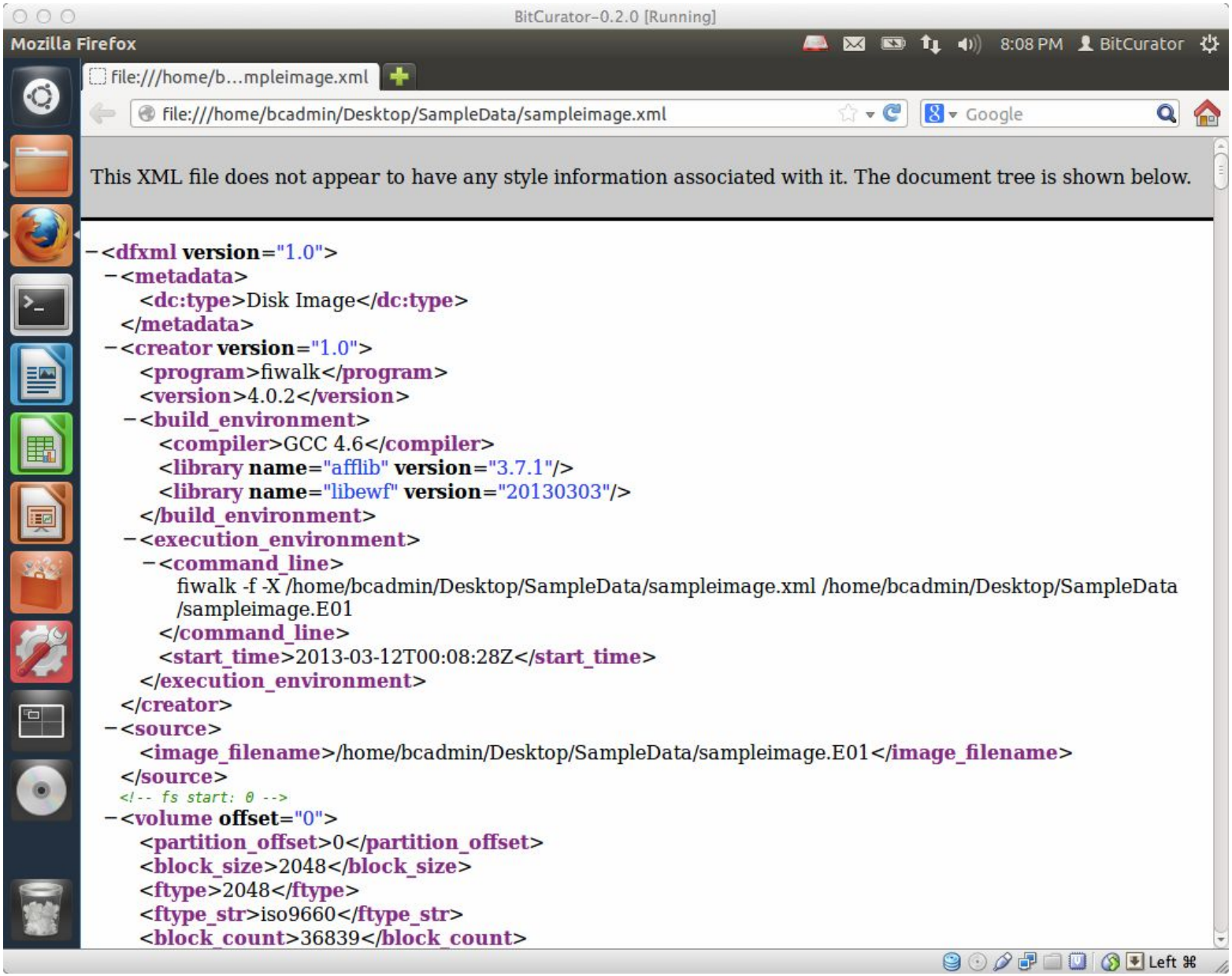
```
<fileobject>
      <filename>Documents and Settings/All Users/Documents/
                My Pictures/Sample Pictures/Blue hills.jpg
      </filename>
      ...
      <filesize>28521</filesize>
      <alloc>1</alloc>
      <used>1</used>
      <inode>6245</inode>
      ...
      <uid>0</uid>
      <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

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

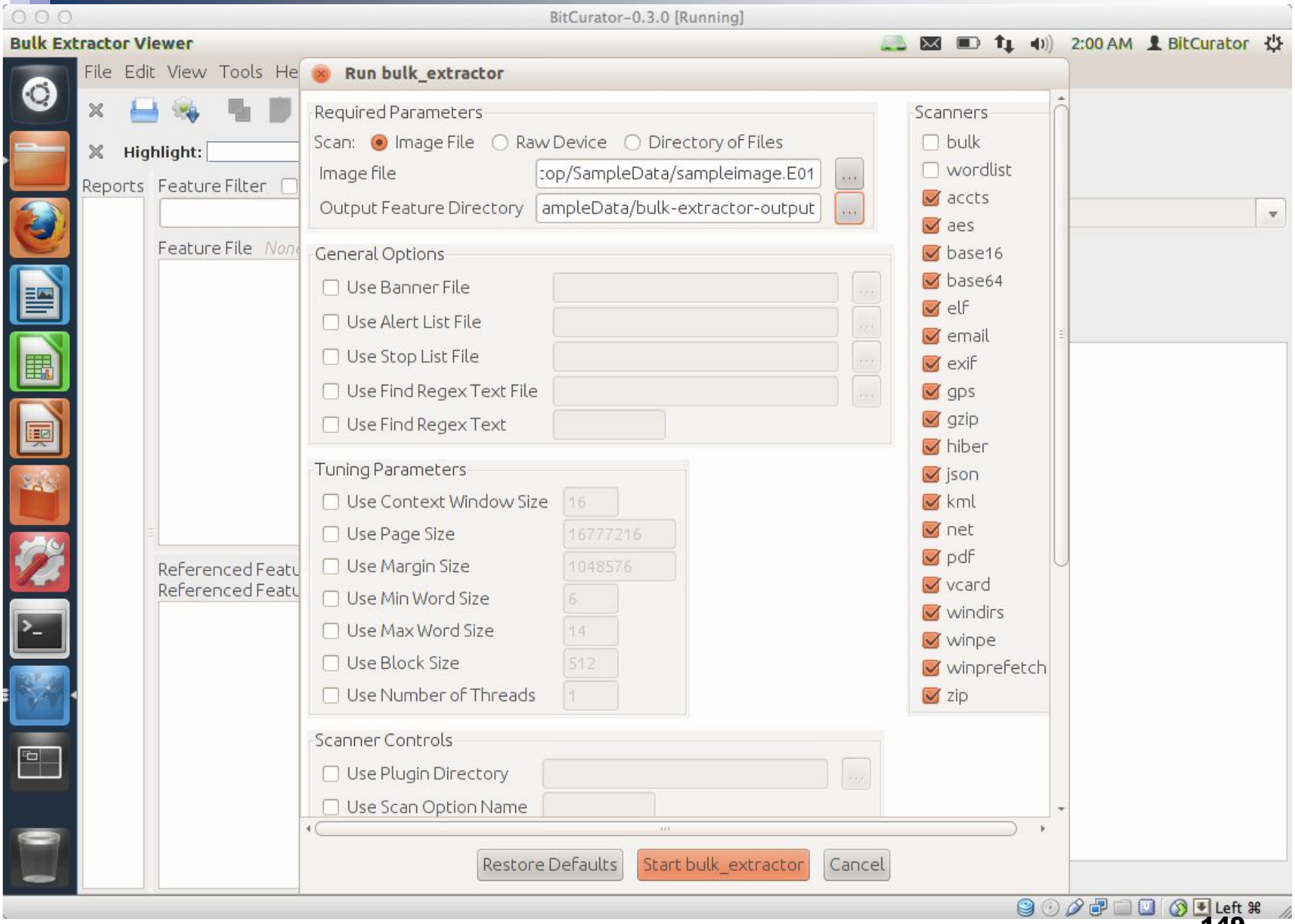


Provenance – DFXML Output from fiwalk

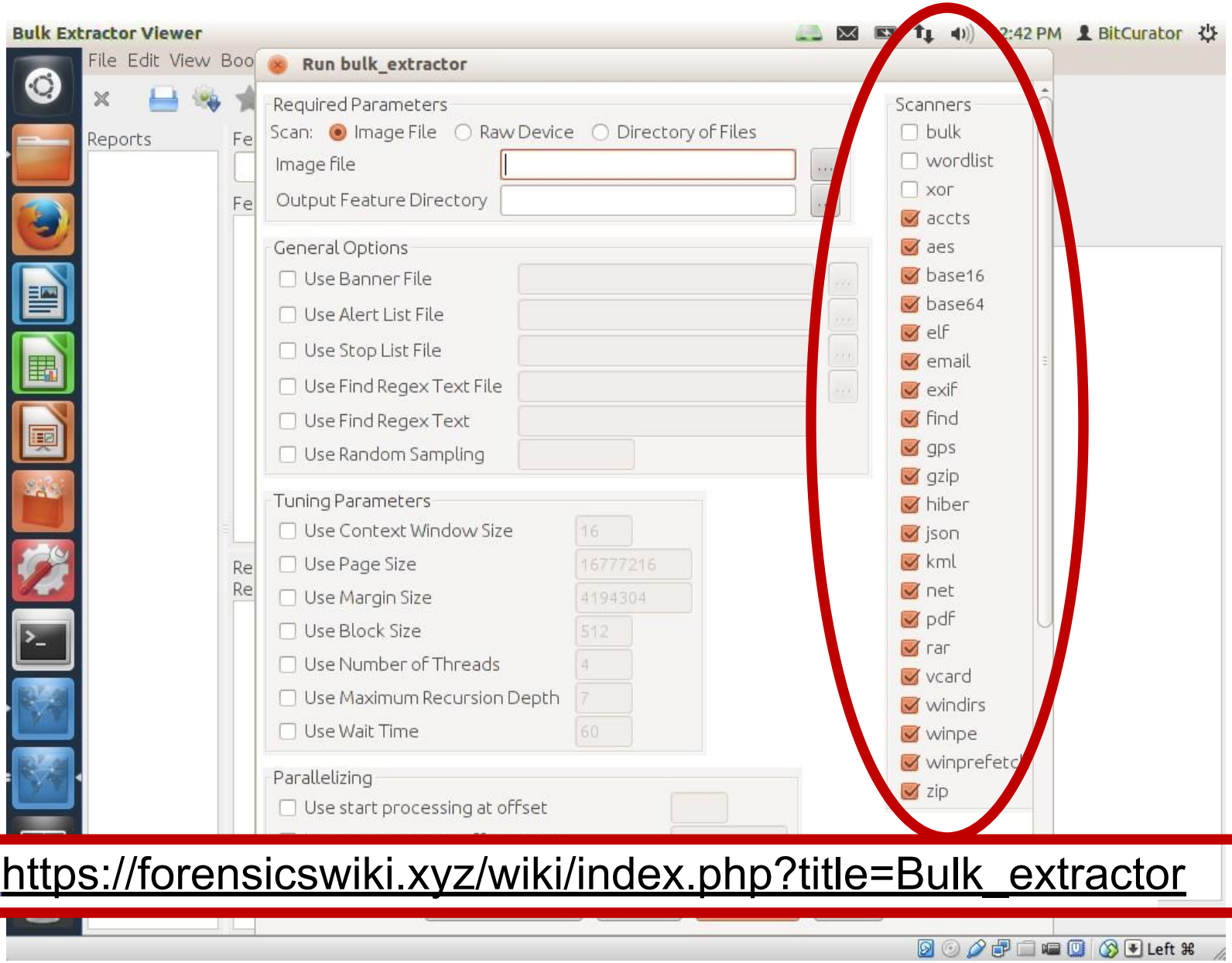


Identifying “Features” of Interest in Disk Images or Directories

Bulk Extractor

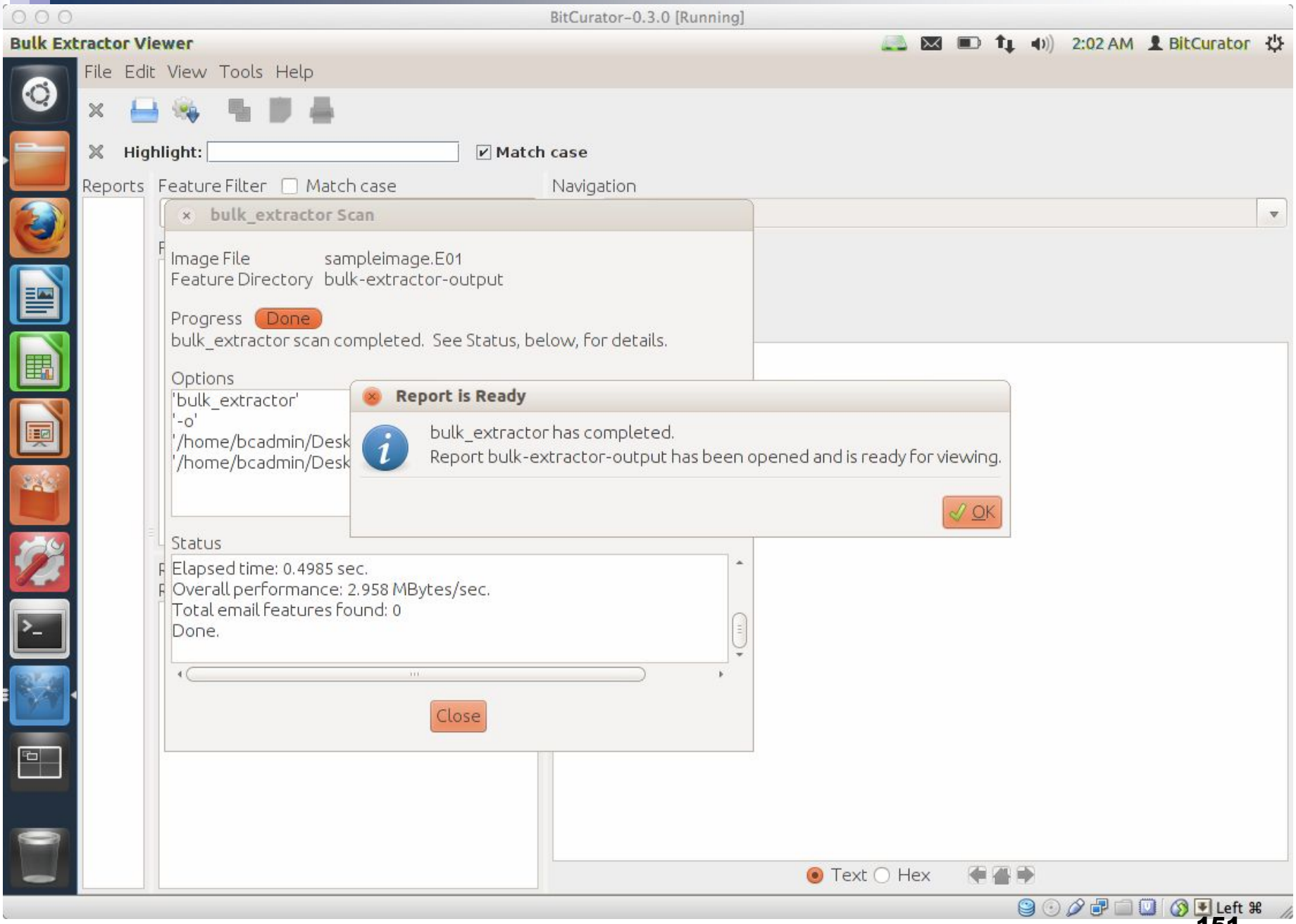

BitCurator-0.3.0 [Running]
Bulk Extractor Viewer
2:00 AM
BitCurator
File Edit View Tools He
Highlight:
Reports
Feature Filter
Feature File
Referenced Featu
Referenced Featu
Run bulk_extractor
Required Parameters
Scan: Image File Raw Device Directory of Files
Image file
:op/SampleData/sampleimage.E01
Output Feature Directory
ampleData/bulk-extractor-output
General Options
Use Banner File
Use Alert List File
Use Stop List File
Use Find Regex Text File
Use Find Regex Text
Tuning Parameters
Use Context Window Size
16
Use Page Size
16777216
Use Margin Size
1048576
Use Min Word Size
6
Use Max Word Size
14
Use Block Size
512
Use Number of Threads
1
Scanner Controls
Use Plugin Directory
Use Scan Option Name
Scanners
bulk
wordlist
accts
aes
base16
base64
elf
email
exif
gps
gzip
hiber
json
kml
net
pdf
vcard
windirs
winpe
winprefetch
zip
Restore Defaults
Start bulk_extractor
Cancel
Left ⌘

Bulk Extractor* – Identifying Potentially Sensitive Information

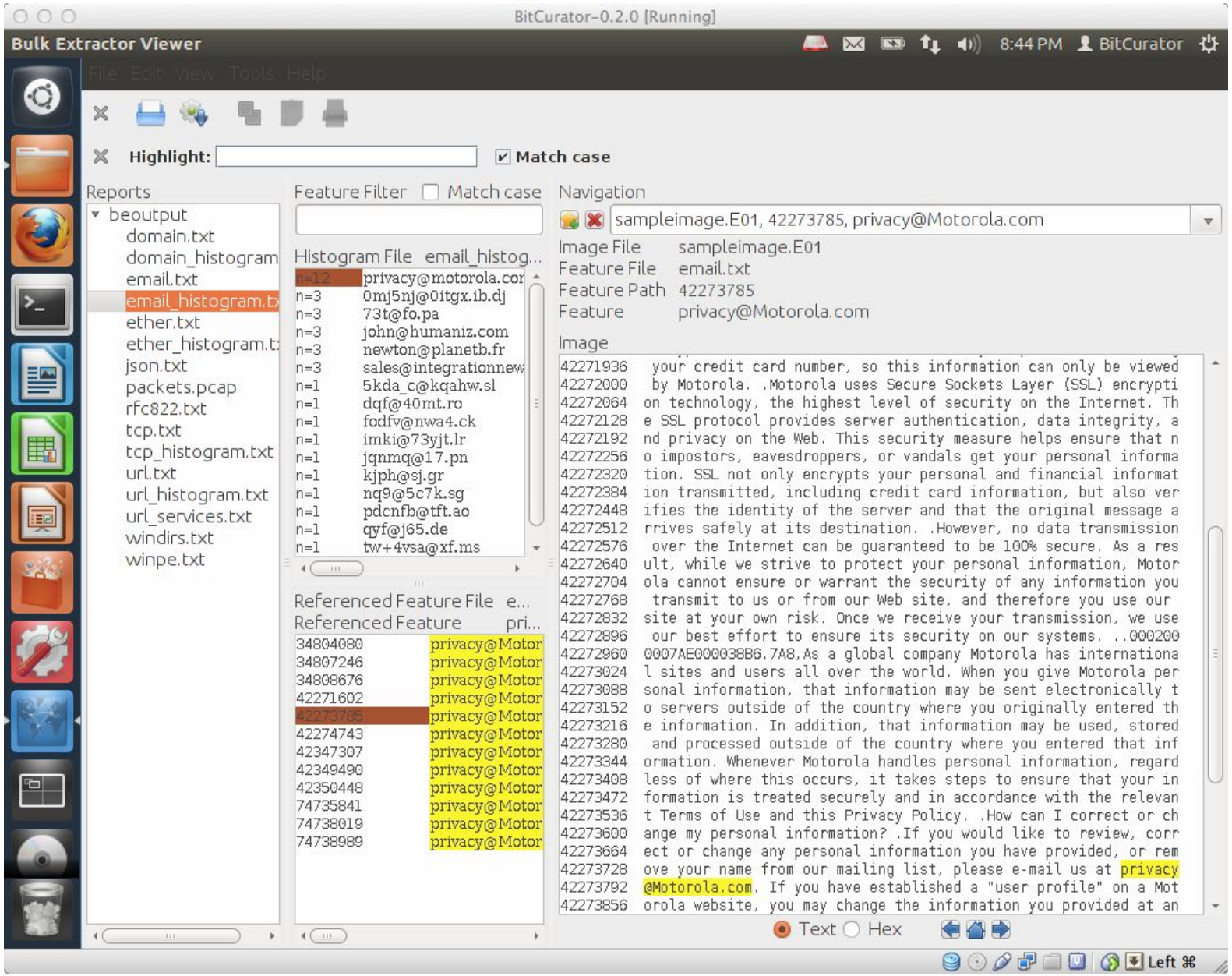


See: https://forensicswiki.xyz/wiki/index.php?title=Bulk_extractor

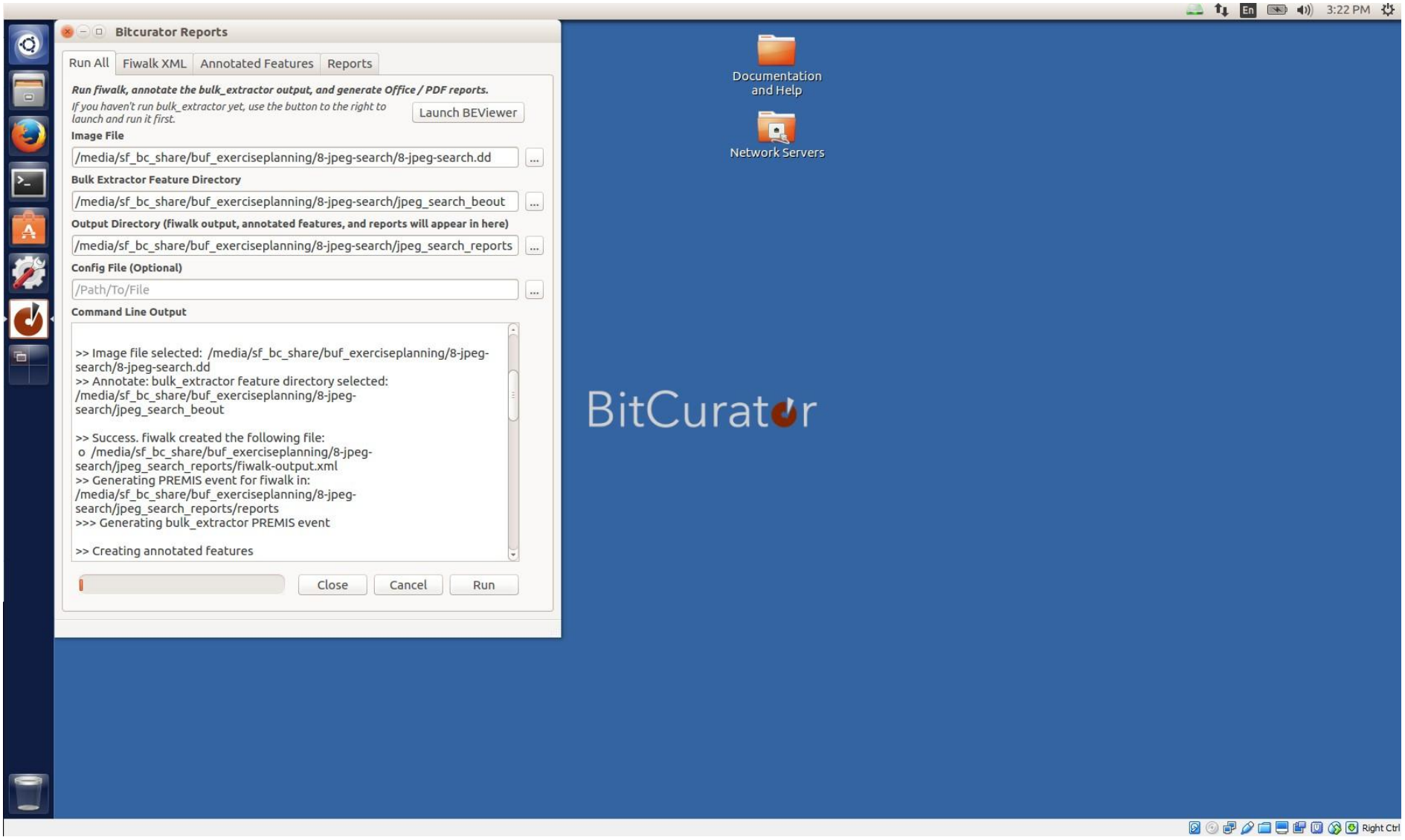
*Developed by Simson Garfinkel


BitCurator-0.3.0 [Running]
Bulk Extractor Viewer
2:02 AM
BitCurator
File Edit View Tools Help
Highlight:
Match case
Reports
Feature Filter
Match case
Navigation
bulk_extractor Scan
Image File sampleimage.E01
Feature Directory bulk-extractor-output
Progress
Done
bulk_extractor scan completed. See Status, below, for details.
Options
'bulk_extractor'
'-o'
'/home/bcadmin/Desk
'/home/bcadmin/Desk
Report is Ready
bulk_extractor has completed.
Report bulk-extractor-output has been opened and is ready for viewing.
OK
Status
Elapsed time: 0.4985 sec.
Overall performance: 2.958 MBytes/sec.
Total email features found: 0
Done.
Close
Text
Hex
Left ⌘

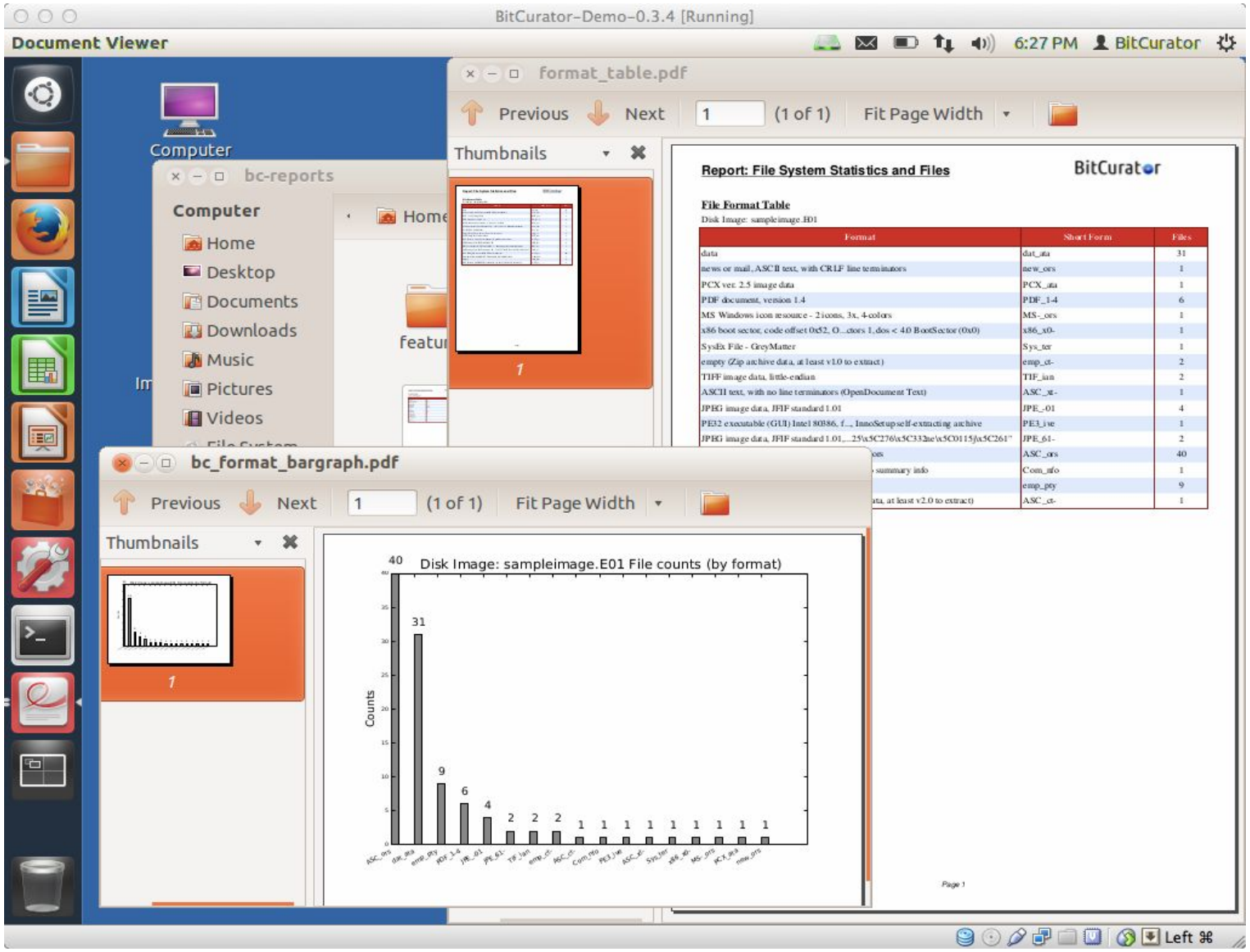
Histogram of Email Addresses (Specific Instances in Context on Right)



BitCurator Reporting Tool



Various Specialized BitCurator Reports



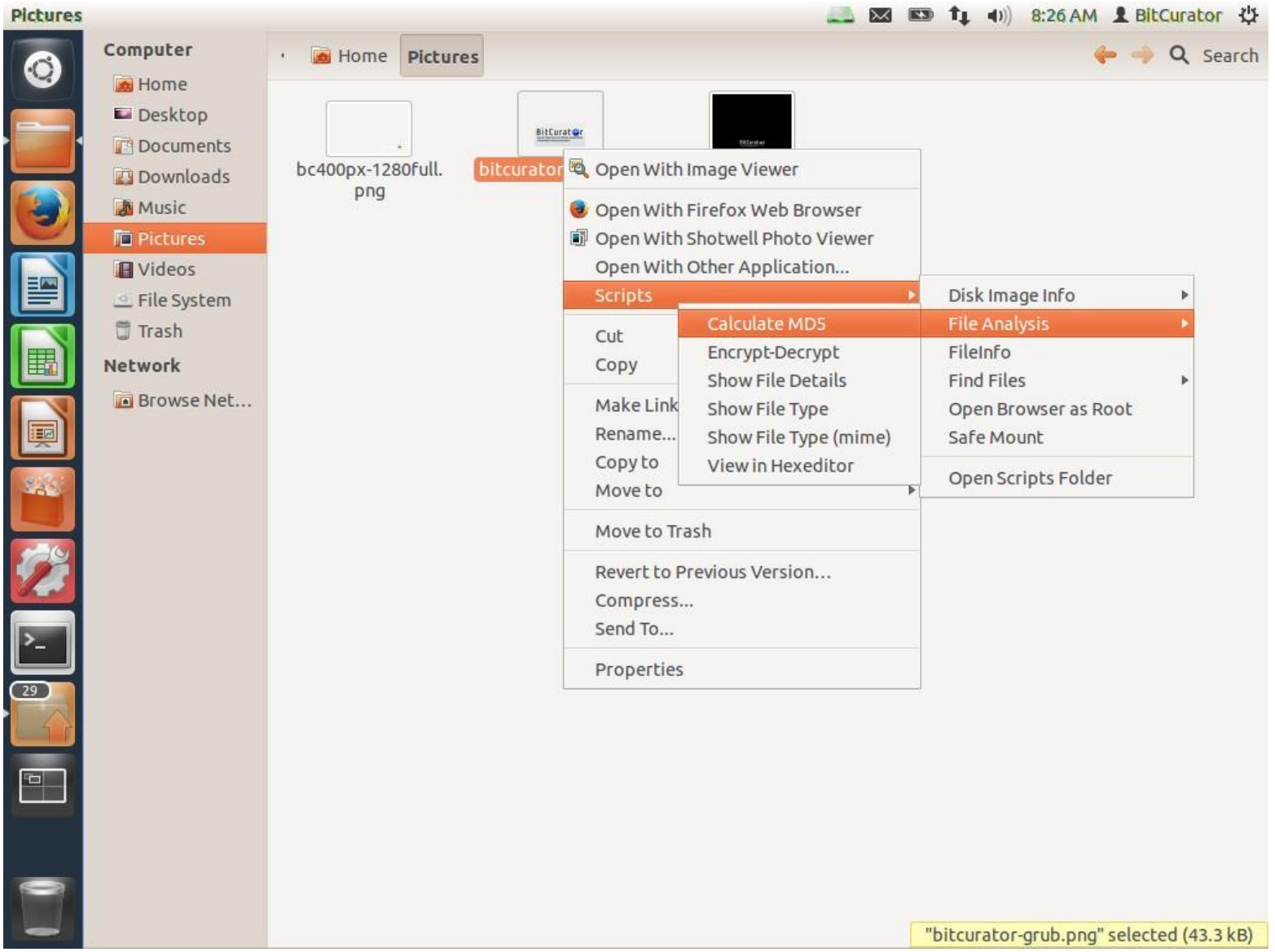
Detail: Specialized BitCurator Reports

File	Content
bc_format_bargraph.pdf	histogram of file formats found on the volume
bulk_extractor_report.pdf	high-level overview of feature locations on disk
fiwalk_deleted_files.pdf	shows paths to any deleted materials found in a given partition
fiwalk-output.xml.xlsx	Excel converted DFXML output (file system metadata)
fiwalk_report.pdf	high-level overview of file system characteristics
format_table.pdf	long-form file format names for formats shown in bar graph
premis.xml	PREMIS preservation metadata

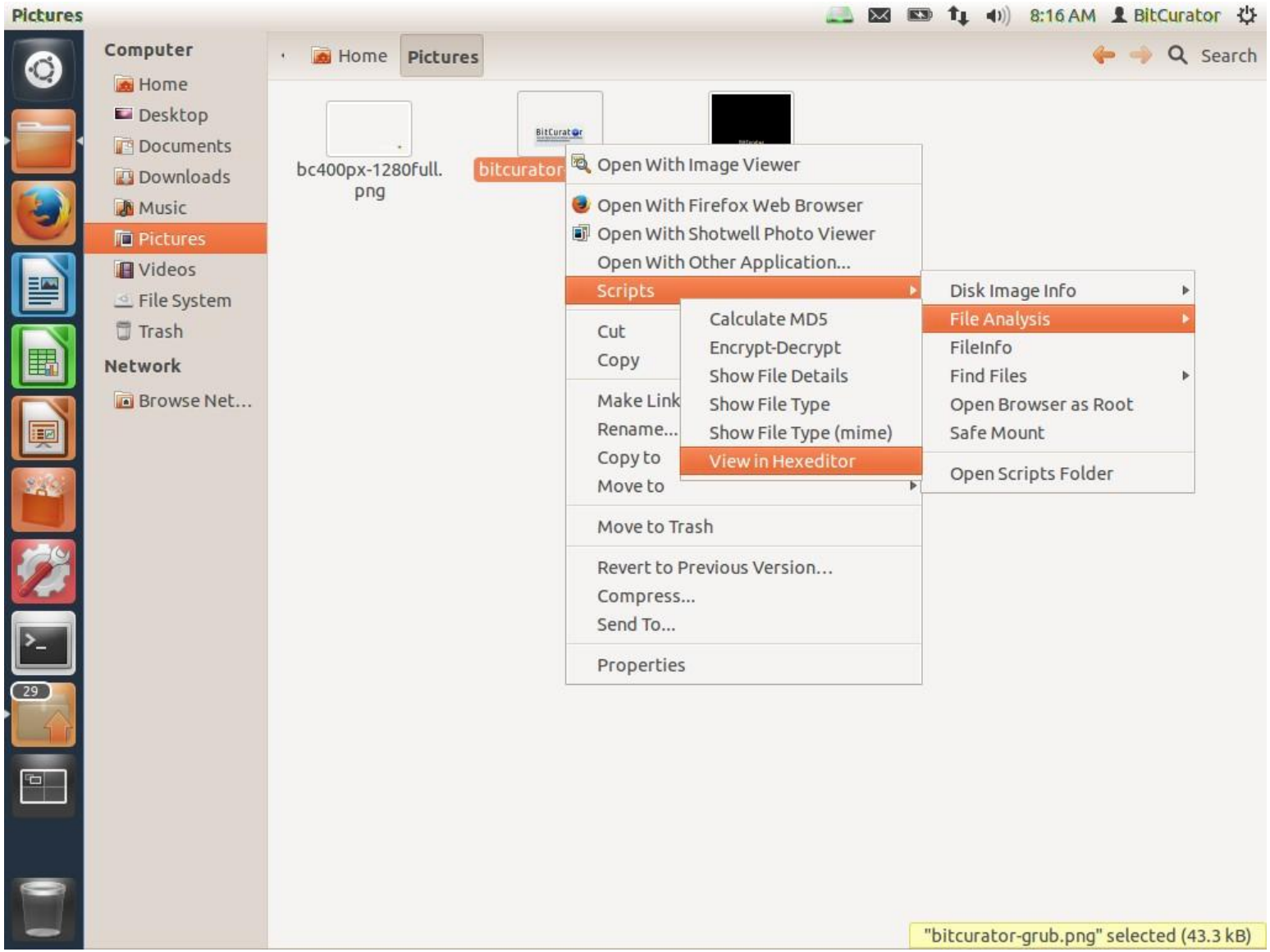
Nautilus Scripts

- Scripts that can be run using Nautilus (GNOME file manager)
- Most provide more convenient access (right click and menu selection) to functions performed by applications that could also be run directly

Right-click on file or directory and create MD5



Quick access to a hex view



Other functionality to meet 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 (I addition to Guymager	dd, dcfldd, ddrescue, cdrdao (for audio CDs)
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

Other functionality to meet user needs

Function	Tool(s)
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	GTK Hash, md5deep, md5sum
Compare hashes of files to hashes in the National Software Reference Library (NSRL) of known system files	nsrllookup
View and edit bytestreams (hex editor)	Bless Hex Editor, GHex

Command Line Operations

- Opens up many more possibilities, such as:
  - stringing tools together
  - performing batch operations
  - changing parameters from their default values
  - using tools that are only available through the command line (no GUI)

Some Considerations

- Role of pipes – feed output from one process into another process
- Switches – settings that can be applied to a command (e.g. -a, -r)
- Argument – a specific piece of data that is processed by a program (e.g. a variable or fixed value)
- Regular expressions – used to find patterns (more on this later)
- Text created in Windows and Unix, even though they're both ASCII, will encode new lines differently, so you may need to translate using a tool such as **dos2unix** or **unix2dos**.

Some Important Commands and Tasks

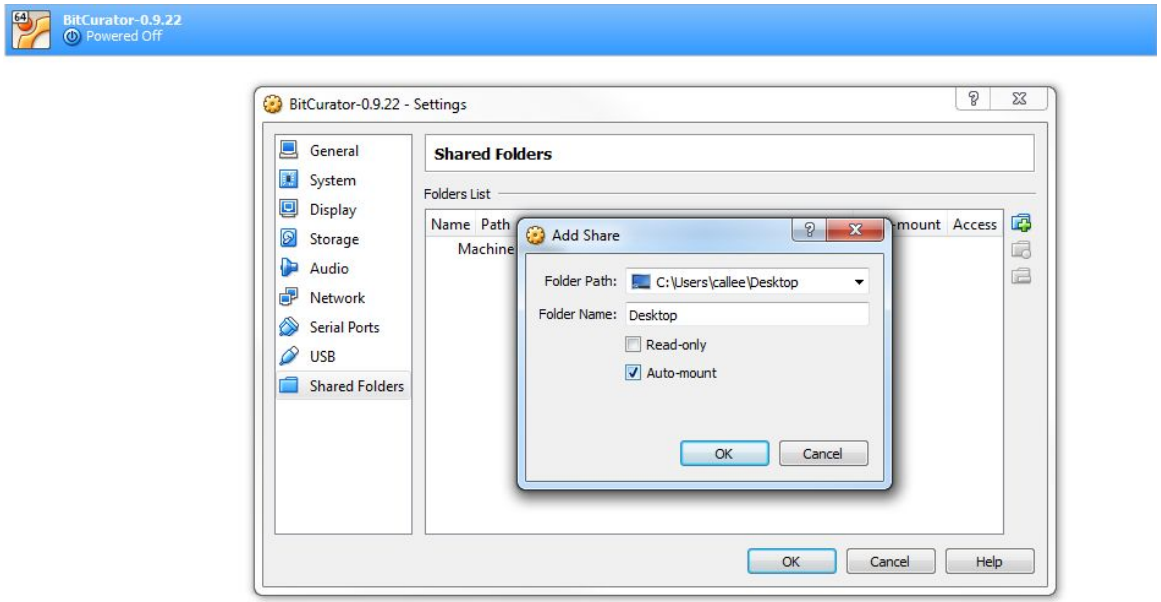
- **mkdir** – make a directory
- **cd** – change the directory that you're in ["cd .." goes to the parent of the current directory]
- **ls** – list contents of a directory
- **md5sum** – generate cryptographic hashes
- **cat** – output content of a text file (can be concatenation of contents of two files)
- **file** – determine file types based on magic numbers
- **strings** – matches patterns in the text (ASCII) parts of a file (file can be binary)
- **diff** – compare two files
- **hexdump** – very basic (non-GUI) hex viewer

General Unix/Linux CLI Tips

- **man** – manual page that explains how to run a command or some other technical information (e.g. ascii page)
- **control-z** – quit currently running program
- **clear** – clear the screen (hide text from previous commands)
- **Up arrow** – cycles through previous commands, so you can rerun (or adapt) them
- **Tab** – hit this key after you've started typing a string that the operating system can fill in for you (e.g. a long file name). Hitting tab multiple times will cycle through available options.

Exercise: Basic Linux Commands

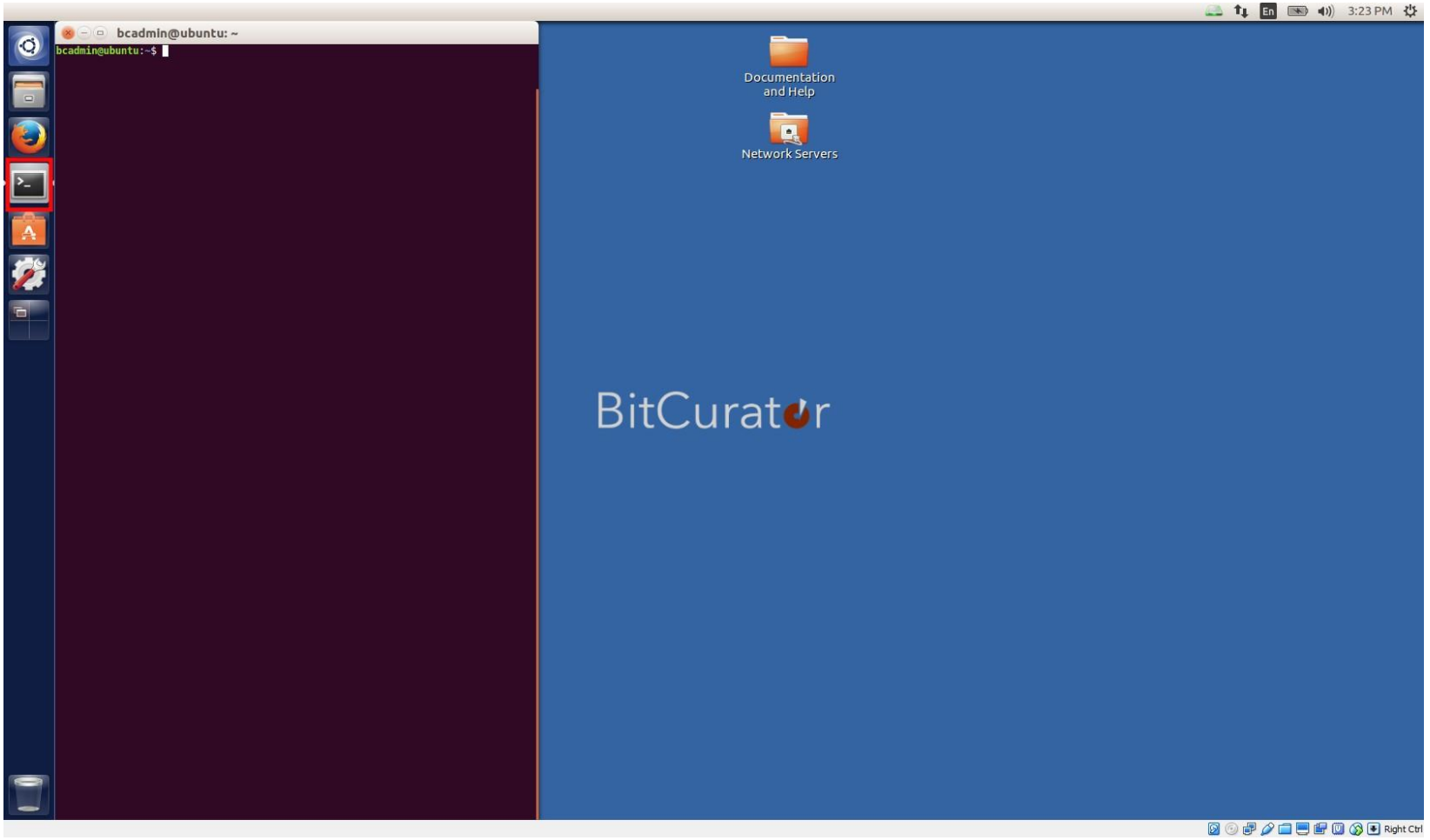
- The saa-dfa-sample-data.zip file you downloaded earlier contains a folder of sample files named **file_ident_ex**
- If you haven't done this already, add shared folder to BitCurator VM, pointing to the desktop of the host



- Move the **file_ident_ex** folder to the BitCurator VM desktop

Exercise: Basic Linux Commands

Open a terminal in the BitCurator environment (using the Terminal icon in the dock)



Exercise: Basic Linux Commands

Command	Reason/Explanation
pwd	Show the directory you're currently in
ls	List the contents of the current directory
cd Desktop	Change the current directory to Desktop
ls	List the contents of the current directory
unzip files.zip	Decompress and unpack content of files.zip
ls	List the contents of the current directory
cd files	Change the current directory to files
ls	List the contents of the current directory
md5sum [file name of first file] > firsthash	Create a hash of a file and output it to a text file
less firsthash	Display the content of the output to the screen
Control-z	Stop the "less" program
md5sum [file name of second file] > secondhash	Create a hash of a second file and output it to a text file
cat firsthash secondhash > bothhashes	Combine the context of the two output files
more bothhashes	Display the content of the output to the screen
most bothhashes	Display the content of the output to the screen (follow instructions for adding it), then run this command again

Gives you the right administrative permissions

sudo apt-get install most

Uses Advanced Packaging Tool to get the program

Exercise: Basic Linux Commands

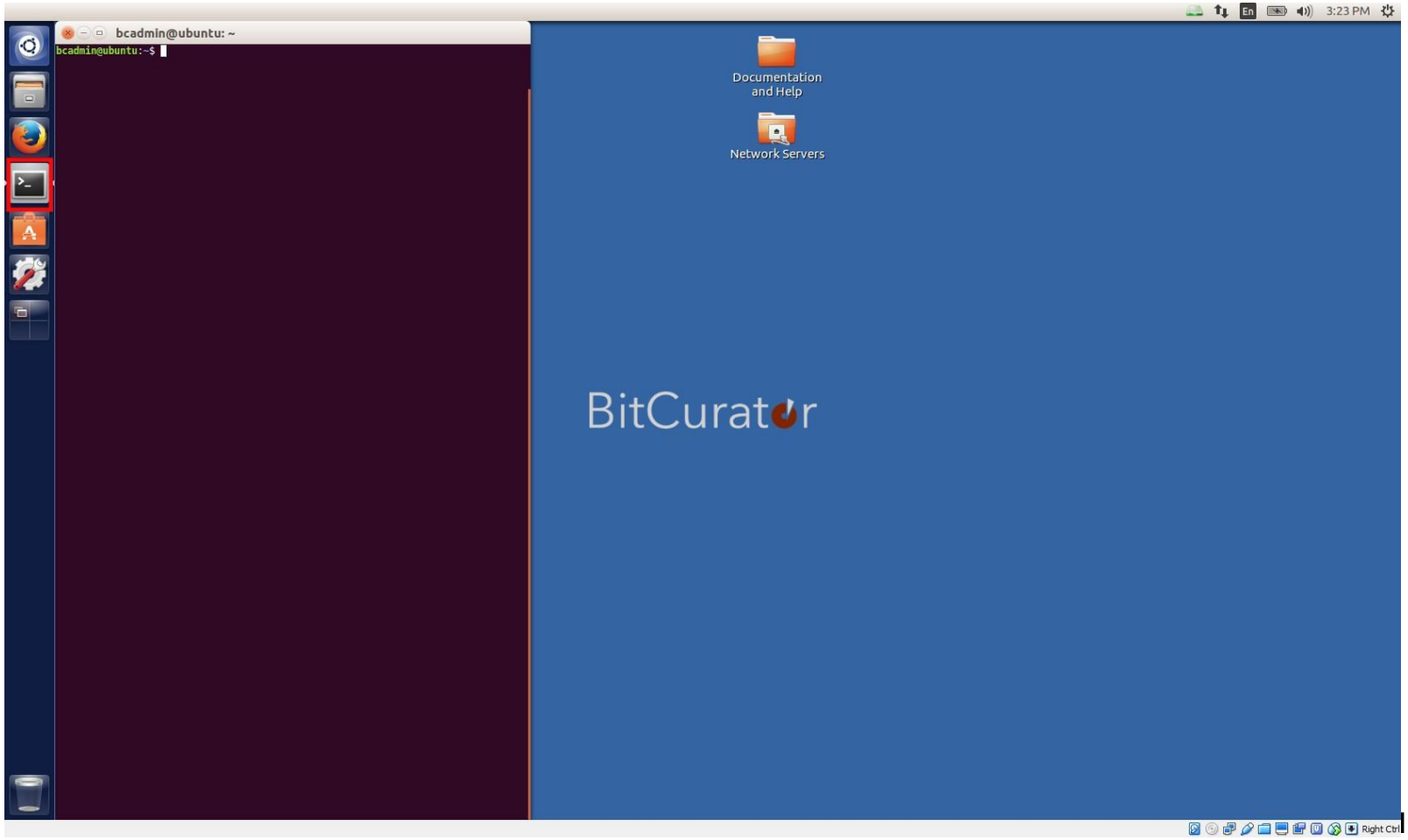
Command	Reason/Explanation
rm firsthash	Delete (remove) firsthash file
rm secondhash	Delete (remove) secondhash
ls	List the contents of the current directory
hexdump [file name] -C | less	Show hex dump of a given file [-C switch shows the standard view of hex on left and ASCII on right]
Use up and down arrows	Navigate within the hex view of the file's content
:q	Quit the "less" program

Exercise: Siegfried and Brunnhilde

- Over the next few slides, we will run Siegfried and Brunnhilde in over the same set of files in several different ways
- Goals
  - Generate characterization and related technical metadata
  - Illustrate how the data can be configured for different uses
  - Identify decision points when data are unclear
- Source files to analyze: file_ident_ex directory in the the zip file you downloaded and extracted earlier

~~Installing Siegfried and Brunnhilde~~

- ~~Start up the BitCurator VM (if it's not already running)~~
- ~~Open a Terminal Window~~

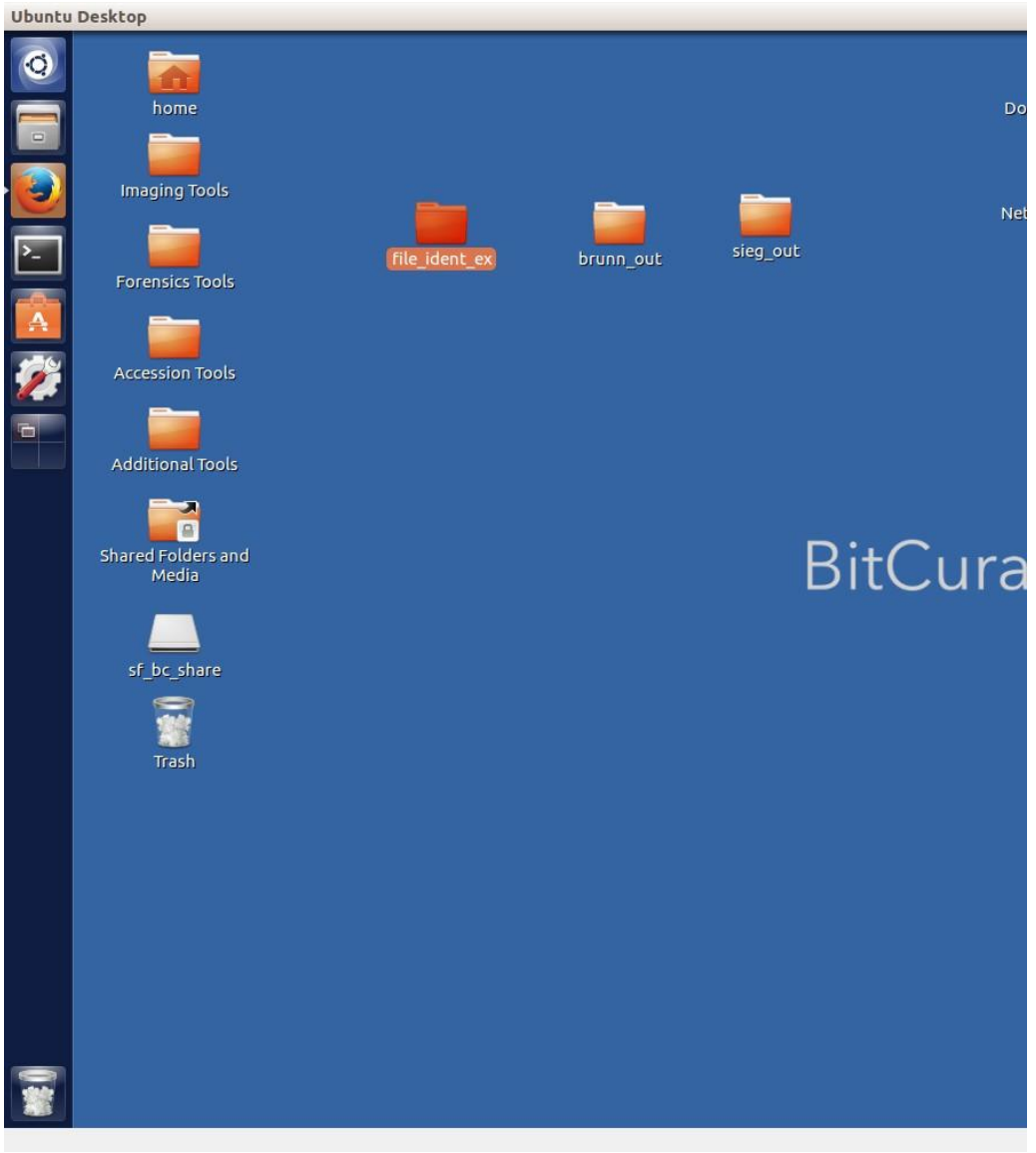


Installing Siegfried and Brunnhilde

- Enter the following commands in Terminal:
  - wget -qO - https://bintray.com/user/downloadSubjectPublicKey?username=bintray | sudo apt-key add –
  - echo "deb http://dl.bintray.com/siegfried/debian wheezy main" | sudo tee -a /etc/apt/sources.list
  - sudo apt-get update && sudo apt-get install Siegfried
  - sudo pip install brunnhilde
- Note: there is a text file in the sample files that includes these commands if you want to copy/paste them

Running Siegfried and Brunnhilde I

- Ensure you have extracted the saa-dfa-sample-data.zip file
- Create a shared folder (we'll step you through this)
- Start up the BitCurator VM
- Create "sieg-out" and "brunn-out" folders on the desktop
- Drag "file_ident_ex" folder from the shared saa-dfa-sample-data directory to the Desktop



Running Siegfried and Brunnhilde II

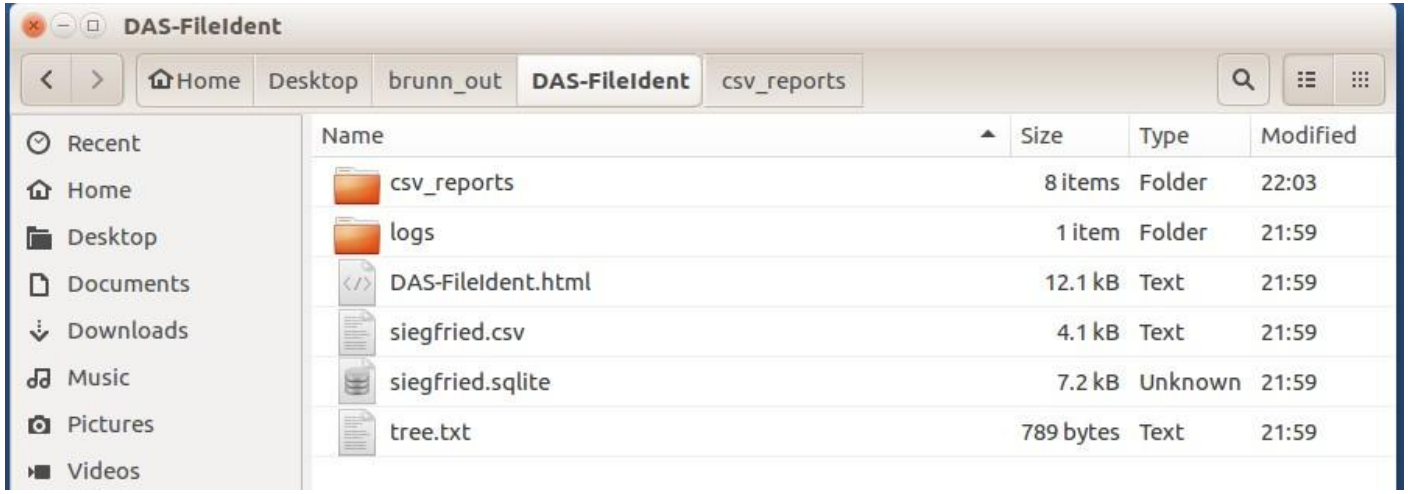
- Open a Terminal window
- At the prompt, enter the following command:
  - **sf ~/Desktop/file_ident_ex/ > ~/Desktop/sieg_out/sieg_out.yaml**
- Open the sieg_out directory and look around
  - What did the command do?
  - What does the file tell you?
  - How would you characterize the data presented in the file?
  - Does anything strike you as odd? Particularly useful?

Running Siegfried and Brunnhilde III

- In the same Terminal window, enter the following commands:
  - **sf -droid ~/Desktop/file_ident_ex/ > ~/Desktop/sieg_out/sieg_out-droid.csv**
  - **sf -json ~/Desktop/file_ident_ex/ > ~/Desktop/sieg_out/sieg_out-json.json**
- Open the sieg_out directory and look around
  - What did the commands do?
  - How do these files differ from the one created on the previous slide?
  - Between the three output files, which do you think is most useful presentation of the data? (Hint: there may be more than one answer)
  - Do any of these files strike you as particularly useful? Particularly worthless?

Running Siegfried and Brunnhilde IV

- In the same Terminal window, enter the following command:
  - **brunnhilde.py -w ~/Desktop/file_ident_ex/ ~/Desktop/brunn_out/DAS-FileIdent**
- Open the brunn_out directory and look around
  - Did Brunnhilde perform any tasks over and above Siegfried?
  - How do the Brunnhilde output files differ from those generated by Siegfried?
  - Inspect csv_reports. How would you characterize what you see here?
  - Are the files here that Brunnhilde and Siegfried found problematic? What conclusions might you draw from them?
  - Is there information that Brunnhilde highlighted that you missed in Siegfried's output?



Regular Expressions

- What is a regular expression, or regex?
  - Simply a pattern for matching bits of text.
- What are regex's useful for?
  - Three things:
    1. *Matching*: Does this text contain a pattern?
    2. *Replacement:* Replace some part of the text with other text
    3. *Extraction:* Yanking out a bit of the text to use somewhere else.

Regular Expressions

- Regular expressions may contain ordinary letters, numbers, and *a few special symbols* that allow you to match a wide range of patterns with a small amount of syntax.
- A regex needs to be interpreted by a program (such as a Perl or Python script) or by an application (such as the Forensic Toolkit)
- Regular expression syntax may look different in different languages and programs.

Regular Expressions – Special Characters

Character	Meaning
.	Match anything
+	Match one or more occurrences
*	Match zero or more occurrences
^	Match only at the start of the text
$	Match only at the end of text
\w	Match an alphanumeric word
\d	Match a number
\s	Match any whitespace
\S	Match anything **except** whitespace

Regular Expressions - Examples

Expression	Explanation
r+	Match the letter 'r' one or more times
t*	Match the letter 't' zero or more times
\d	Match any single digit (shorthand for the expression [0-9]
\dd	Match any pair of digits (alternatively, [0-9][0-9])

Regular Expressions – More Complex Examples

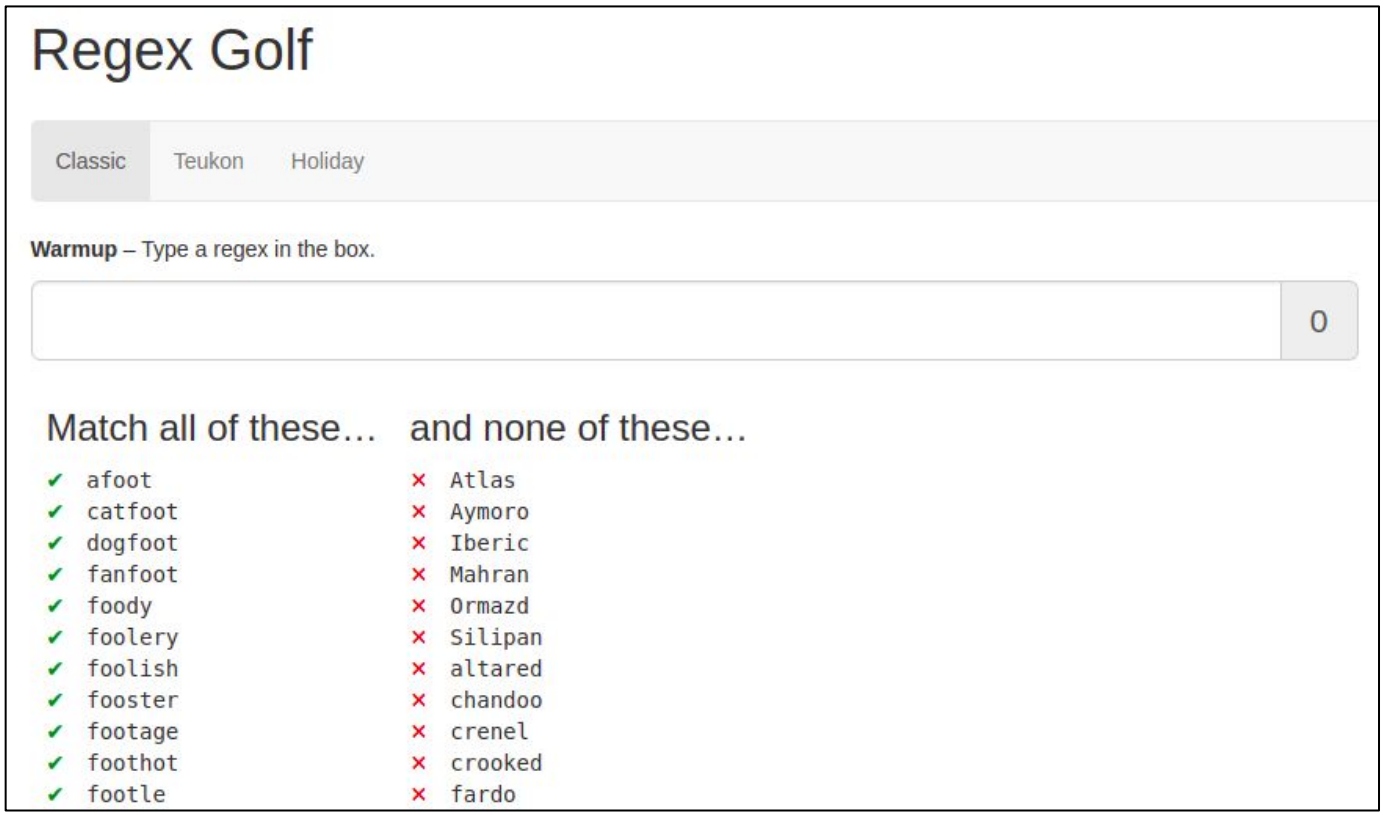
Expression	Explanation
\d+\.\d+	Any two-digit decimal number (e.g. 5.0, 15.2345)
\d+(\.\d+)*	Any integer or decimal number (e.g. 5, 5.0, 15.2345)
(\d\d\d\d)-(\d\d)-(\d\d)	Match any date strings of the form YYYY-MM-DD

What's "wrong" with the second and third regex patterns shown here? What else might they match?

Regular Expressions in FTK

- Tools for building regular expressions one part at a time
- Various default regular expressions that you can use or adapt

For Fun on Your Own – Regex Golf!



Try it at https://alf.nu/RegexGolf

Extracting Data From Specific Types of Files

Exchangeable Image File Format (EXIF)

- Possible tags:

https://exiftool.org/TagNames/EXIF.html

Camera manufacturer	Canon
Camera model	Canon EOS 1200D
Author	Praveen. P
Exposure time	1/60 sec (0.016666666666667)
F-number	f/11
ISO speed rating	200
Date and time of data generation	22:29, 22 November 2018
Lens focal length	41 mm
Show extended details	

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
```

*https://exiftool.org/TagNames/EXIF.html (Also available through the BitCurator environment)

Exiftool Exercise in BitCurator

- Start up the BitCurator VM
- Download one or more pictures to your desktop that you'd like to examine
- Options for viewing EXIF:
  1. ~~PyEXIFToolGUI:~~
     - ~~Navigate to Desktop > Forensics Tools > PyEXIFToolGUI~~
     - ~~Open the tool and select File > Load Images~~
     - ~~Let's also add some GPS coordinates: Select Edit Data, enter the values, then select Save to Selected Image(s) [Make sure that the image is selected]~~
  2. File info menu:
     - Navigate to the image file
     - Right click on it and select Scripts > File Info > Meta Information [Pick EXIF Data]
  3. exiftool at the command line:
     - Open a command prompt window
     - Navigate to where you stored the image (e.g. cd Desktop)
     - Type: *exiftool [Filename]*
     - Note: You can scroll up and down by using Shift + Page Up/Page Down, or you can invoke the command as *exiftool [Filename]* | *less* (type "q" to quit)

Optional Exiftool Exercise (Windows or Mac)

- Download one or more pictures to your desktop that you'd like to examine
- Download and unzip the latest Windows executable (or Mac package) from: https://exiftool.org/

- Save exiftool(-k).exe to your desktop
- Change file name to: exiftool(-k -a -u -g1 -w txt).exe [NOTE: This is changing the parameters for running the software – same as if you were to add these switches at the command line. This trick might not work on a Mac, but you can always issue the commands directly.]

- -k = pause the program before terminating
- -a = allow extraction of duplicate tags
- -u = extract unknown tags
- -g1 = organize output by tag group
- -w = write output text file

- Drag and drop pictures onto the exiftool icon and examine the results
- Change file name to: exiftool(**-X** -k -a -u -g1 -w **xml**).exe
- Drag and drop pictures onto the exiftool icon and examine the results
- For more about exiftool, see: https://github.com/exiftool/exiftool

Stripping Metadata From Images

Social Media site/system	Summary	Displays correctly?		Displays 4Cs?	Save As embedded?			Download embedded?		
500px - www.500px.com Tested in late 2015	Some embedded metadata fields are shown, all correctly, but not the rights-relevant 4C fields. Metadata preserved in SaveAs file. Compared to 2013: SaveAs preserves metadata now = improvement	Exif	IPTC	IPTC	Exif	IPTC IIM	IPTC XMP	Exif	IPTC IIM	IPTC XMP
BEHANCE - www.behance.net Tested in late 2015	All rights-relevant fields and more are shown, all corectly. Embedded metadata is preserved in the SaveAs and the downloaded image file. Compared to 2013: not tested then	Exif	IPTC	IPTC	Exif	IPTC IIM	IPTC XMP	Exif	IPTC IIM	IPTC XMP
Dropbox - www.dropbox.com Tested in late 2015	No embedded metadata shown. Embedded metadata only preserved in the downloaded image file but not in the SaveAs. Compared to 2013: also SaveAs files preserved metadata then = decline	Exif	IPTC	IPTC	Exif	IPTC IIM	IPTC XMP	Exif	IPTC IIM	IPTC XMP
EyeEm - www.eyeem.com Tested in late 2015	No embedded metadata shown. SaveAs file was downscaled and all metadata was stripped off. Compared to 2013: not tested then	Exif	IPTC	IPTC	Exif	IPTC IIM	IPTC XMP	Exif	IPTC IIM	IPTC XMP
Facebook - www.facebook.com Tested in late 2015	No embedded metadata shown. SaveAs file preserved Copyright Notice and Creator in IIM, anything else is stripped off. Surprise: 2 IIM fields contain data generated by Facebook. Compared to 2013: at least 2 fields in IIM survive now = slight improvement	Exif	IPTC	IPTC	Exif	IPTC IIM	IPTC XMP	IPTC	IPTC IIM	IPTC XMP
Flickr FREE account- www.flickr.com Tested in late 2015	Some embedded metadata fields are shown, all correctly, but not all rights-relevant 4Cs. Embedded metadata is stripped off SaveAs files but preserved in downloaded files. Compared to 2013: plus = any downloded file preserves metadata now; minus = even high resolution SaveAs file does not preserve it now.	Exif	IPTC	IPTC	Exif	IPTC IIM	IPTC XMP	Exif	IPTC IIM	IPTC XMP
Google Photo - photos.google.com Tested in late 2015	Some embedded metadata fields are shown, all correctly, but not all rights-relevant 4Cs. SaveAs works only for downscaled files - only Exif metadata is preserved. Downloaded files preserved all metadata. Compared to 2013/Google+ photos: SaveAs file gets IIM and XMP metadata stripped off now = decline	Exif	IPTC	IPTC	Exif	IPTC IIM	IPTC XMP	Exif	IPTC IIM	IPTC XMP
Img.ly - www.img.ly Tested in late 2015	No embedded metadata shown. Embedded metadata is preserved in the high resolution/original size SaveAs image file but stripped off in a downscaled file. Compared to 2013: the loss of metadata in downscaled images was not tested in 2013.	Exif	IPTC	IPTC	Exif	IPTC IIM	IPTC XMP Exif IPTC IIM IPTC XMP	Exif	IPTC IIM	IPTC XMP
Instagram - instagram.com Tested in late 2015	Tested using the Instagram iOS app v 6.4.1: No embedded metadata fields are shown. No retrieval of image files possible. Compared to 2013: then SaveAs was possible - with stripped off metadata.	Exif	IPTC	IPTC	Exif	IPTC IIM	IPTC XMP	Exif	IPTC IIM	IPTC XMP
Joomeo - www.joomeo.com Tested in late 2015	Some embedded metadata fields are shown, all correctly, but not the rights-relevant 4Cs. Embedded metadata preserved in the downloaded image files. Compared to 2013: more embedded metadata were shown then, including 4Cs = slight decline	Exif	IPTC	IPTC	Exif	IPTC IIM	IPTC XMP	Exif	IPTC IIM	IPTC XMP
LINKED IN 2015 - www.linkedin.com Tested in late 2015	No embedded metadata shown. Only embedded Exif fields are preserved in SaveAs files. Compared to 2013: not tested then.	Exif	IPTC	IPTC	Exif	IPTC IIM	IPTC XMP	Exif	IPTC IIM	IPTC XMP
Pictify - www.pictify.com Tested in late 2015	No embedded metadata shown. No retrieval of image files possible. Compared to 2013: then SaveAs was possible - with stripped off metadata.	Exif	IPTC	IPTC	Exif	IPTC IIM	IPTC XMP	Exif	IPTC IIM	IPTC XMP
Pinterest - www.pinterest.com	No embedded metadata shown. Embedded metadata preserved in high resolution/original size images, but IIM and XMP metadata is									

https://www.embeddedmetadata.org/social-media-test-results.php

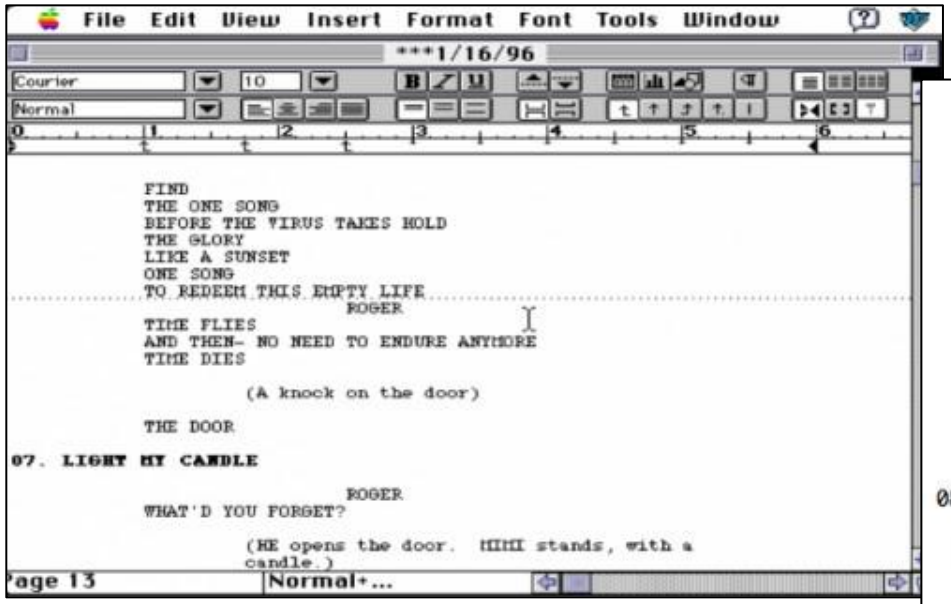
Office Documents

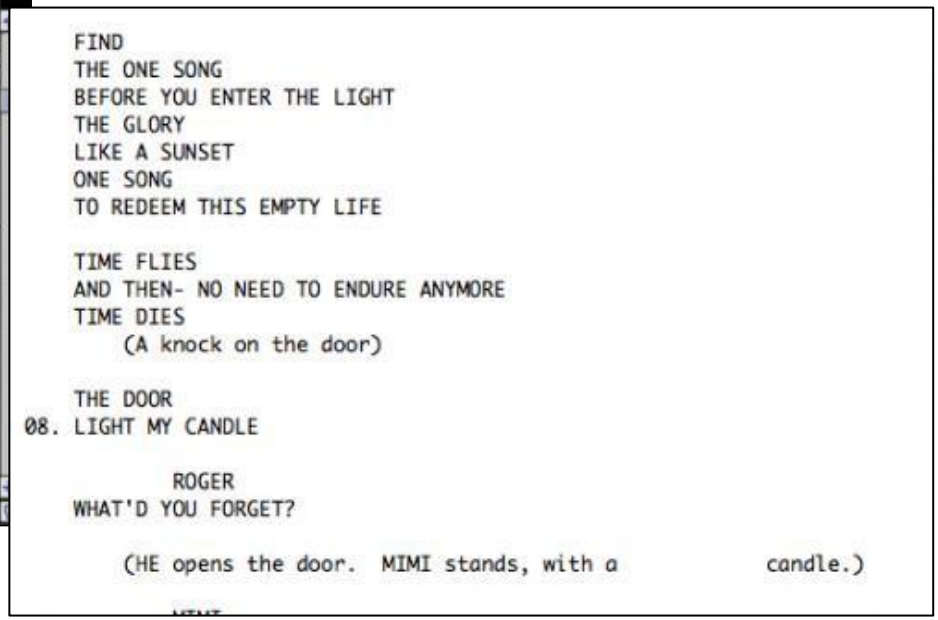
- Are the “new” office formats (ODF and OOXML) better or worse for forensics?
- What kinds of information can you get out of them?
- What sorts of approaches might you take to view and/or extract the information?

Office Documents – PPTX File Example

- Your zip file contains a document named “The NDSA Levels of Digital Preservation.pptx”
  (also at http://library.harvard.edu/sites/default/files/The%20NDSA%20Levels%20of%20Digital%20Preservation_3.pptx)
- Change the file extension to .zip
- Open it with 7-Zip or WinZip
- Extract all the files
- Examine the contents of the resulting directory
  - Can you find a thumbnail of the first slide?
  - Where are the slides stored?
  - Where are embedded images stored?
  - Can you determine who created the file?

Jonathan Larson Fast Save Example





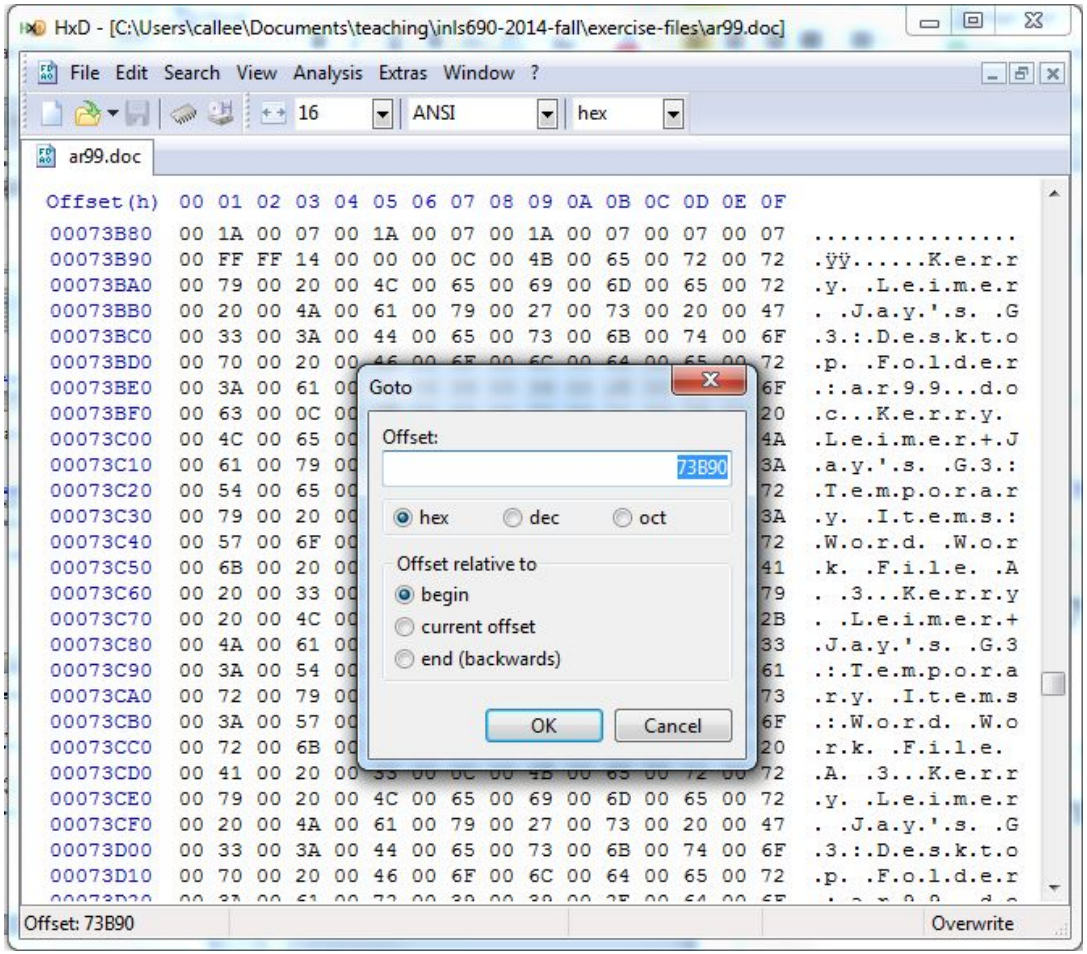

```
00028b60  09 09 09 2a 2a 2a 31 2f  31 36 2f 39 36 4f 55 52  |...***1/16/96OUR|
00028b70  20 57 45 44 44 49 4e 47  4f 4e 20 54 48 45 20 53  | WEDDINGON THE S|
00028b80  4f 46 41 53 4f 46 41 54  48 45 20 56 49 52 55 53  |OFASOFATHE VIRUS|
00028b90  20 54 41 4b 45 53 20 48  4f 4c 44 4d 45 45 54 20  | 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****|
```

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

Hidden Data Exercise – Using a Hex Editor

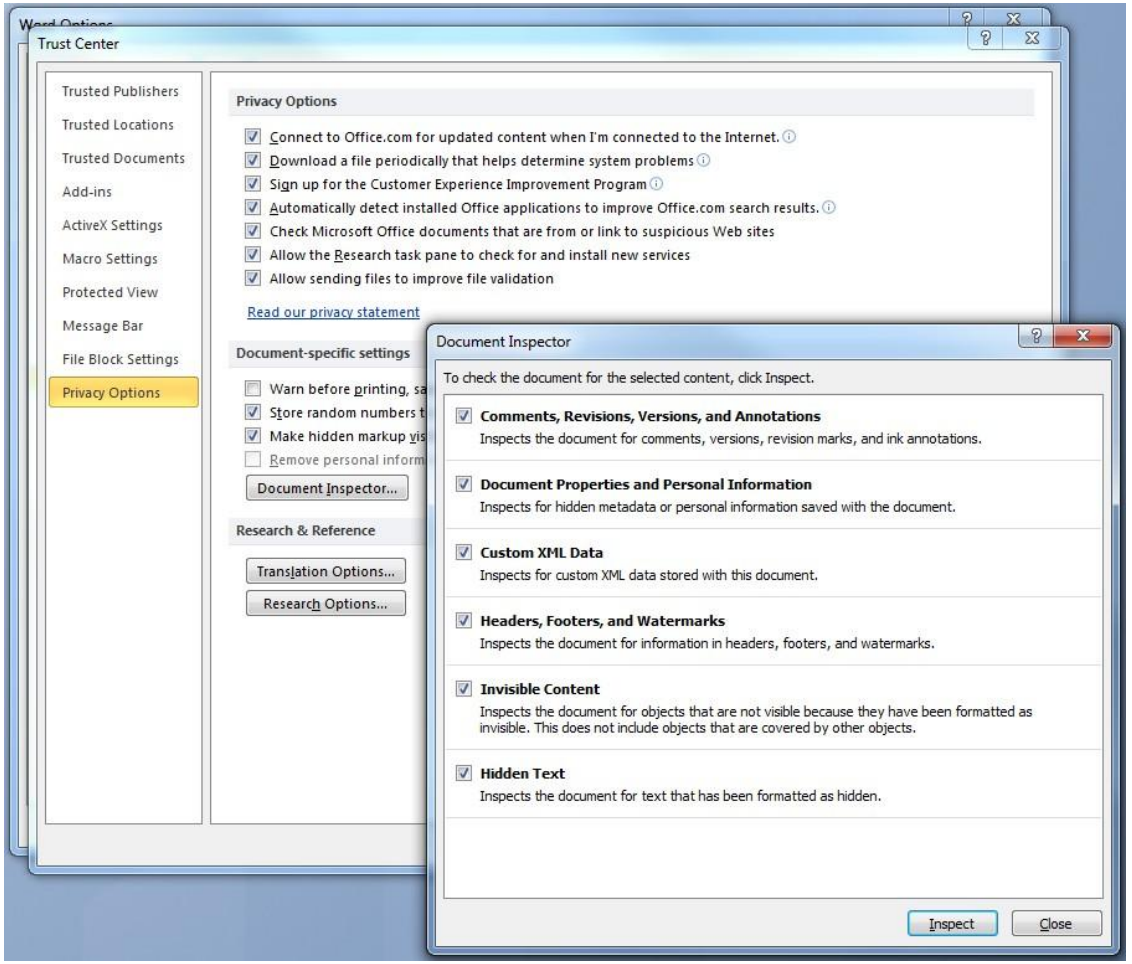
Your zip file contains a document named ar99.doc (also at: http://web.archive.org/web/20000816164723/http://microsoft.com/msft/ar99/downloads/ar99.doc)

- Open the file in HxD (or upload to https://hexed.it)
- Go to offset 73B90 (use Search > Goto or just Control+G)
- What do you see there?
- What does it tell you about the document?



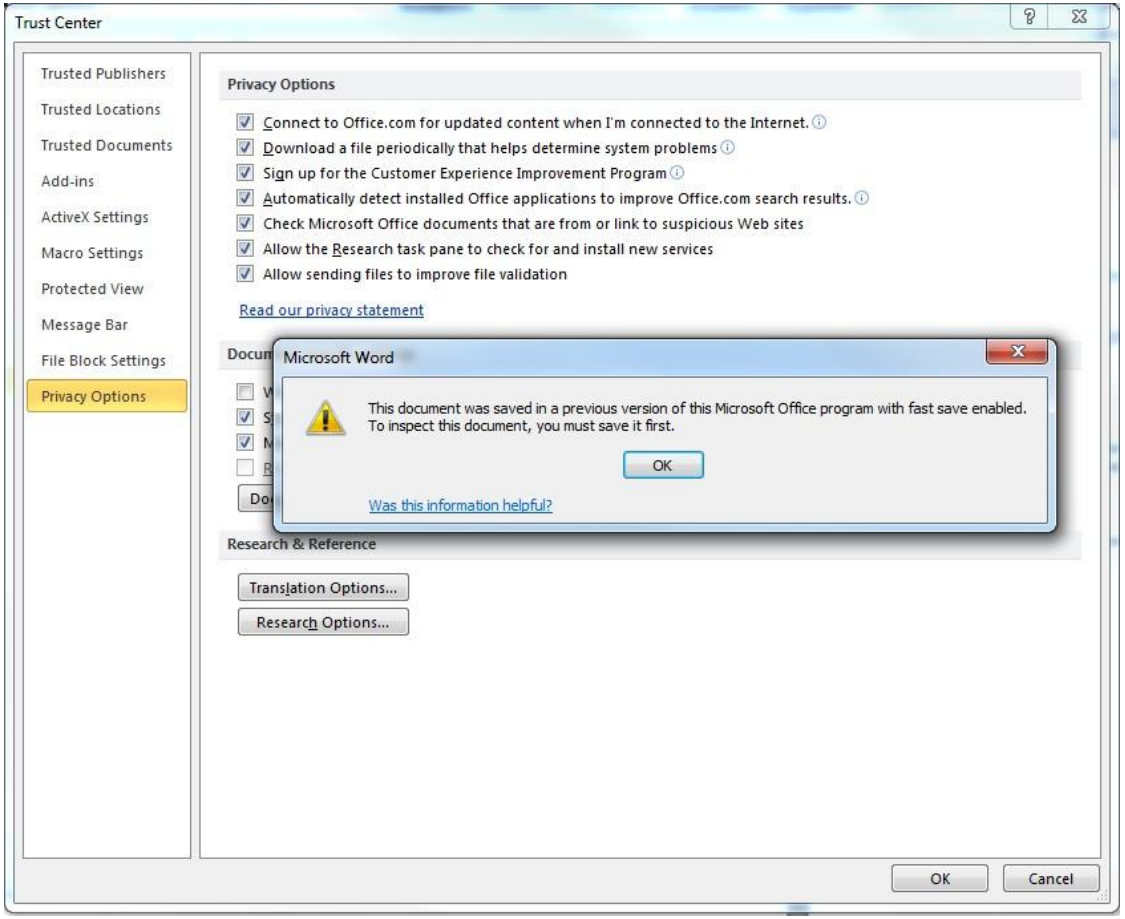
Hidden Data Exercise – Inspection in MS Word

- Do the following:
  - Open it in Word – what is it?
  - If prompted to do so at the top, select “Enable Editing”
  - Select: File > Options > Trust Center > Trust Center Settings…
  - Then Privacy Options > Document Inspector > Inspect



Hidden Data Exercise – Inspection in MS Word

- Are you prompted with this?
- Why do you think this is?
- If you see this, click OK, then save the document
- Run Document Inspector again
- What does it tell you?

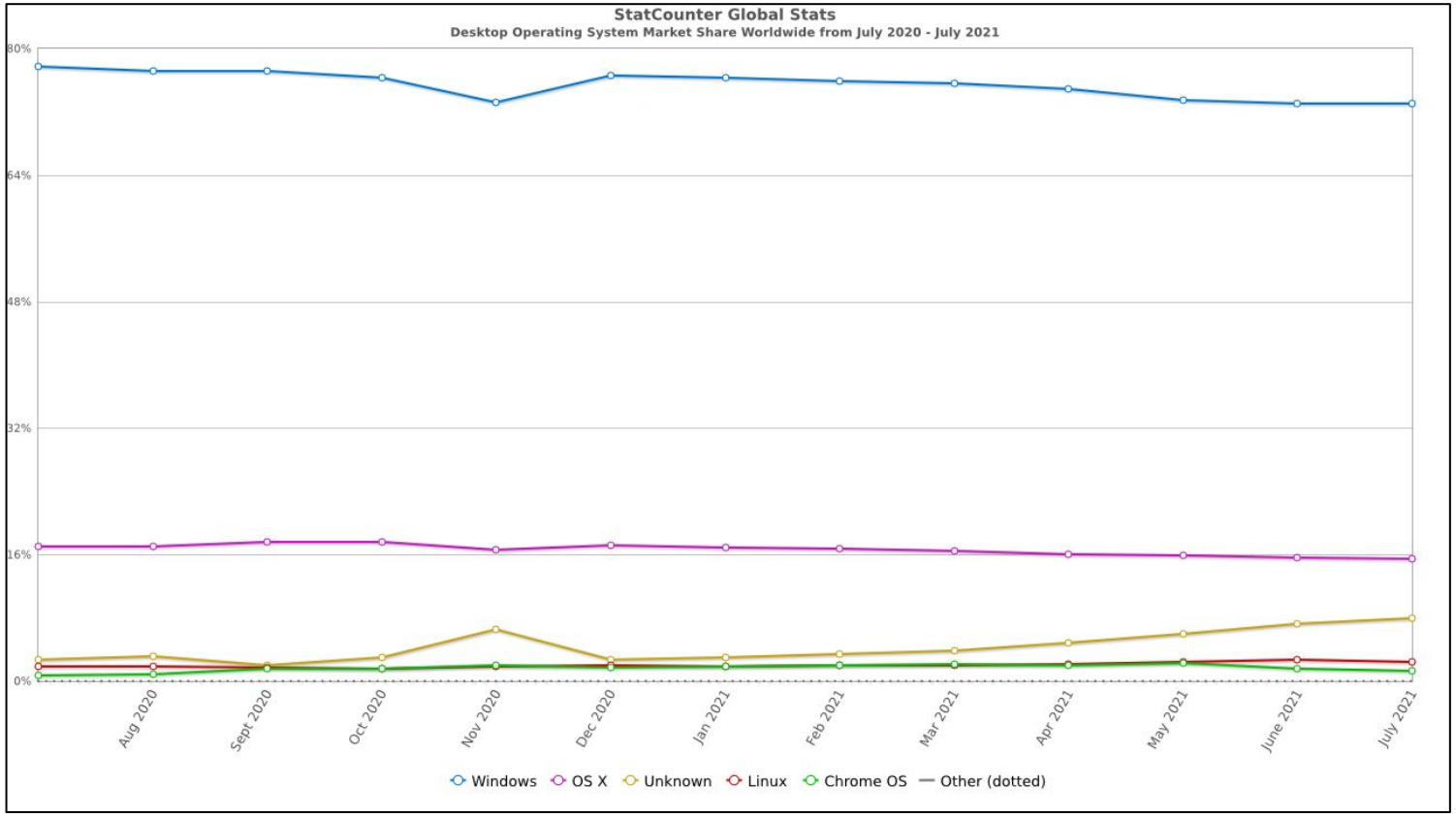


Email

- What's in an email header?
- Which parts of the header would be of most interest to you as someone responsible for managing and preserving a collection that includes email?
- Which parts of the header would be of most interest to future researchers?

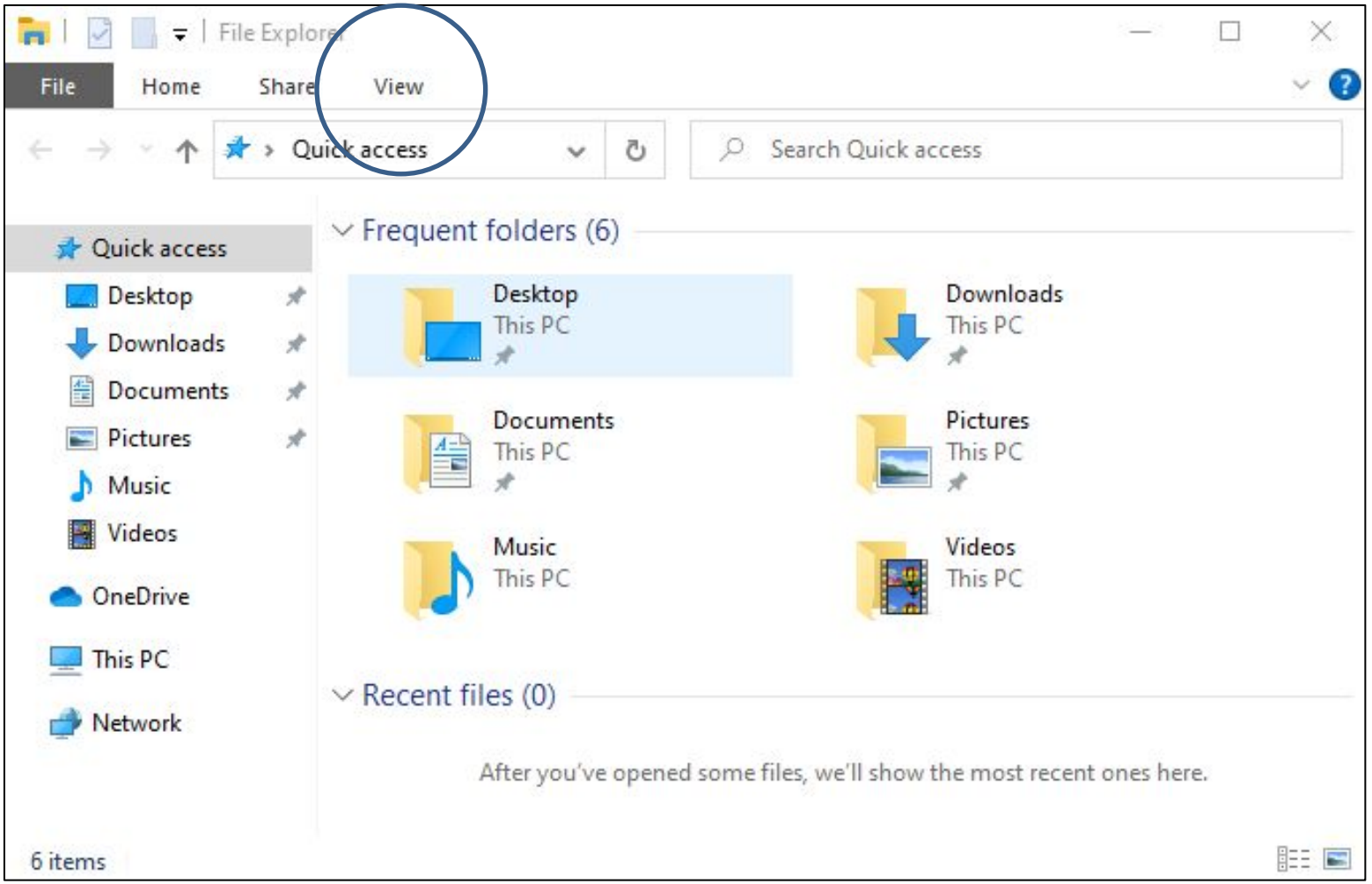
Windows Artifacts

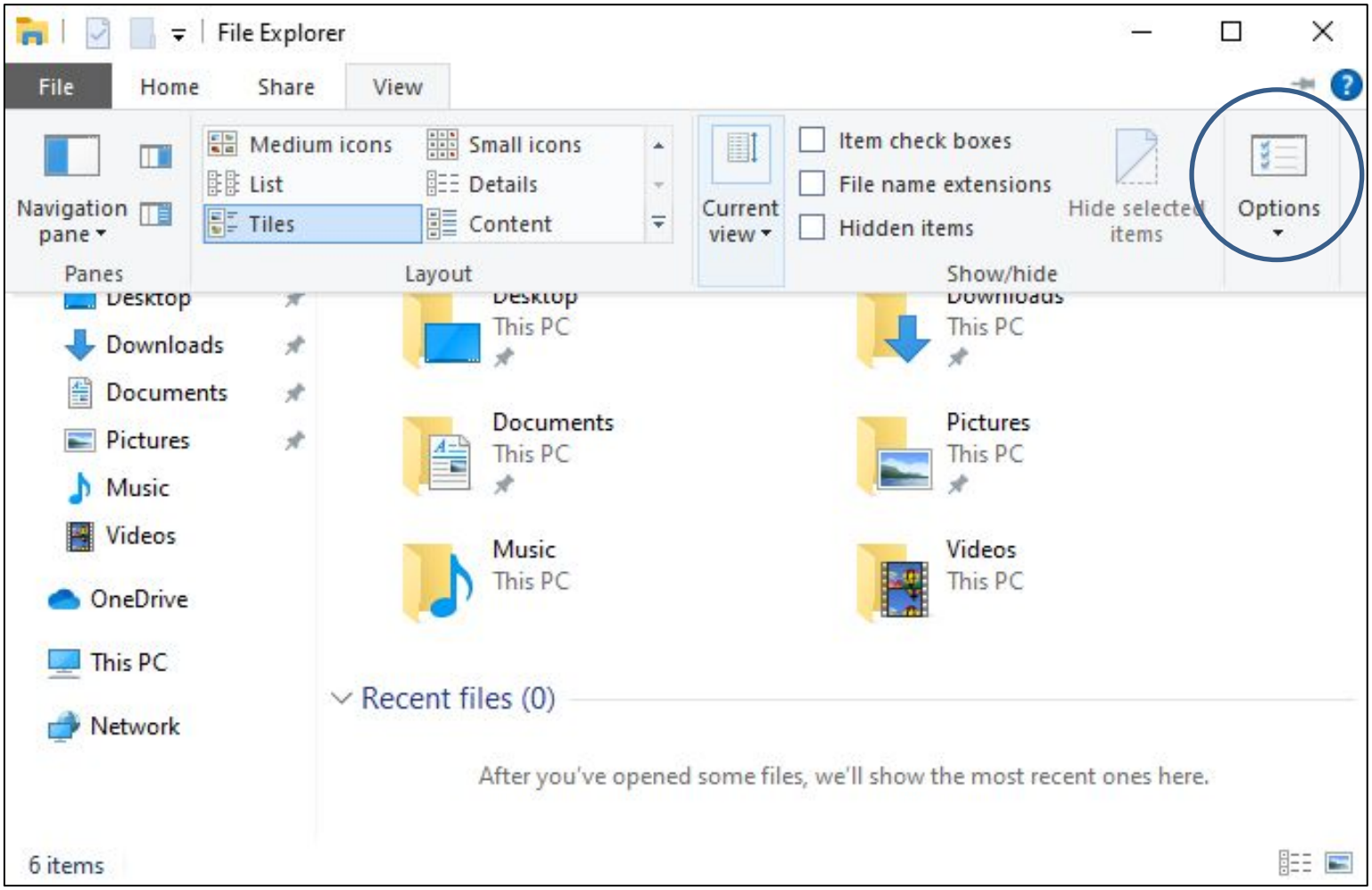
Desktop Operating System Market Share

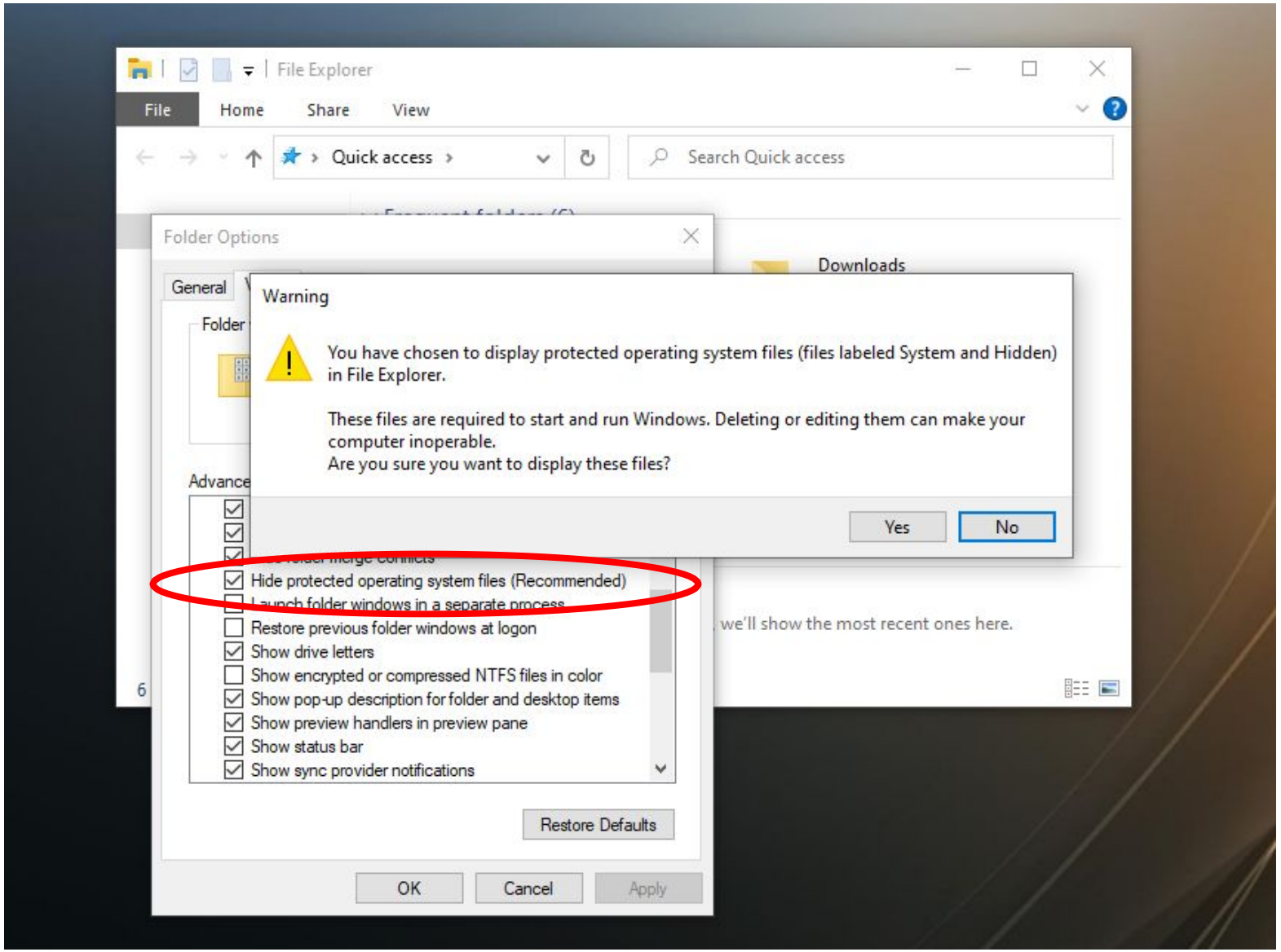


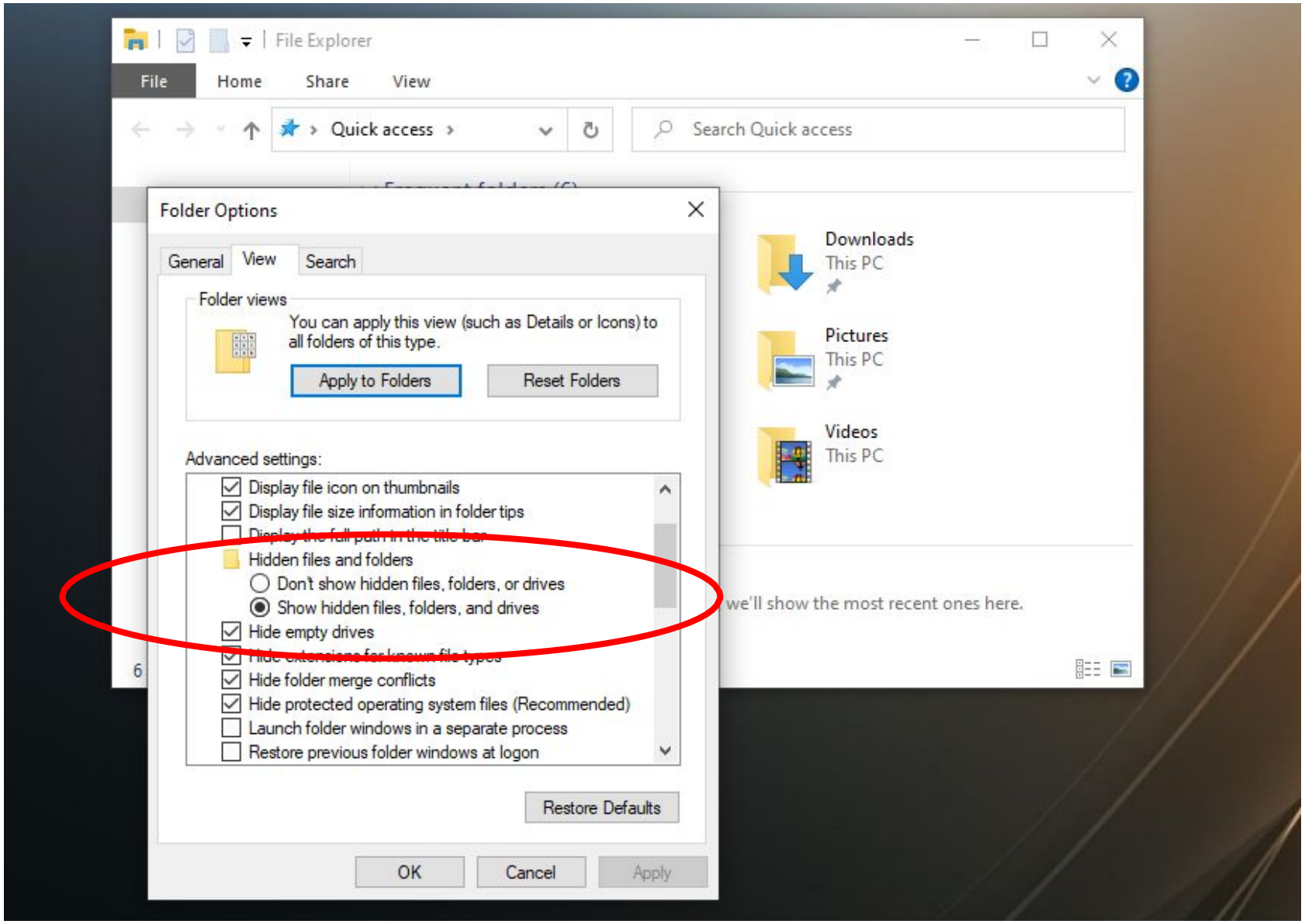
https://gs.statcounter.com/os-market-share/desktop/worldwide

Let’s make sure you can see all the files on your computer.


File Explorer
File
Home
Share
View
Quick access
Search Quick access
Quick access
Desktop
Downloads
Documents
Pictures
Music
Videos
OneDrive
This PC
Network
Frequent folders (6)
Desktop
This PC
Downloads
This PC
Documents
This PC
Pictures
This PC
Music
This PC
Videos
This PC
Recent files (0)
After you've opened some files, we'll show the most recent ones here.
6 items


File Explorer
File
Home
Share
View
Navigation pane
Panes
Medium icons
Small icons
List
Details
Tiles
Content
Layout
Current view
Item check boxes
File name extensions
Hidden items
Hide selected items
Options
Show/hide
Desktop
Downloads
Documents
Pictures
Music
Videos
OneDrive
This PC
Network
Desktop
This PC
Downloads
This PC
Documents
This PC
Pictures
This PC
Music
This PC
Videos
This PC
Recent files (0)
After you've opened some files, we'll show the most recent ones here.
6 items


File Explorer
File
Home
Share
View
Quick access
Search Quick access
Downloads
Folder Options
General
Warning
You have chosen to display protected operating system files (files labeled System and Hidden) in File Explorer.
These files are required to start and run Windows. Deleting or editing them can make your computer inoperable.
Are you sure you want to display these files?
Yes
No
Hide protected operating system files (Recommended)
Restore previous folder windows at logon
Show drive letters
Show encrypted or compressed NTFS files in color
Show pop-up description for folder and desktop items
Show preview handlers in preview pane
Show status bar
Show sync provider notifications
Restore Defaults
OK
Cancel
Apply
we'll show the most recent ones here.


File Explorer
File
Home
Share
View
Quick access
Search Quick access
Downloads
This PC
Pictures
This PC
Videos
This PC
we'll show the most recent ones here.
Folder Options
General
View
Search
Folder views
You can apply this view (such as Details or Icons) to all folders of this type.
Apply to Folders
Reset Folders
Advanced settings:
Display file icon on thumbnails
Display file size information in folder tips
Hidden files and folders
Don't show hidden files, folders, or drives
Show hidden files, folders, and drives
Hide empty drives
Hide folder merge conflicts
Hide protected operating system files (Recommended)
Launch folder windows in a separate process
Restore previous folder windows at logon
Restore Defaults
OK
Cancel
Apply

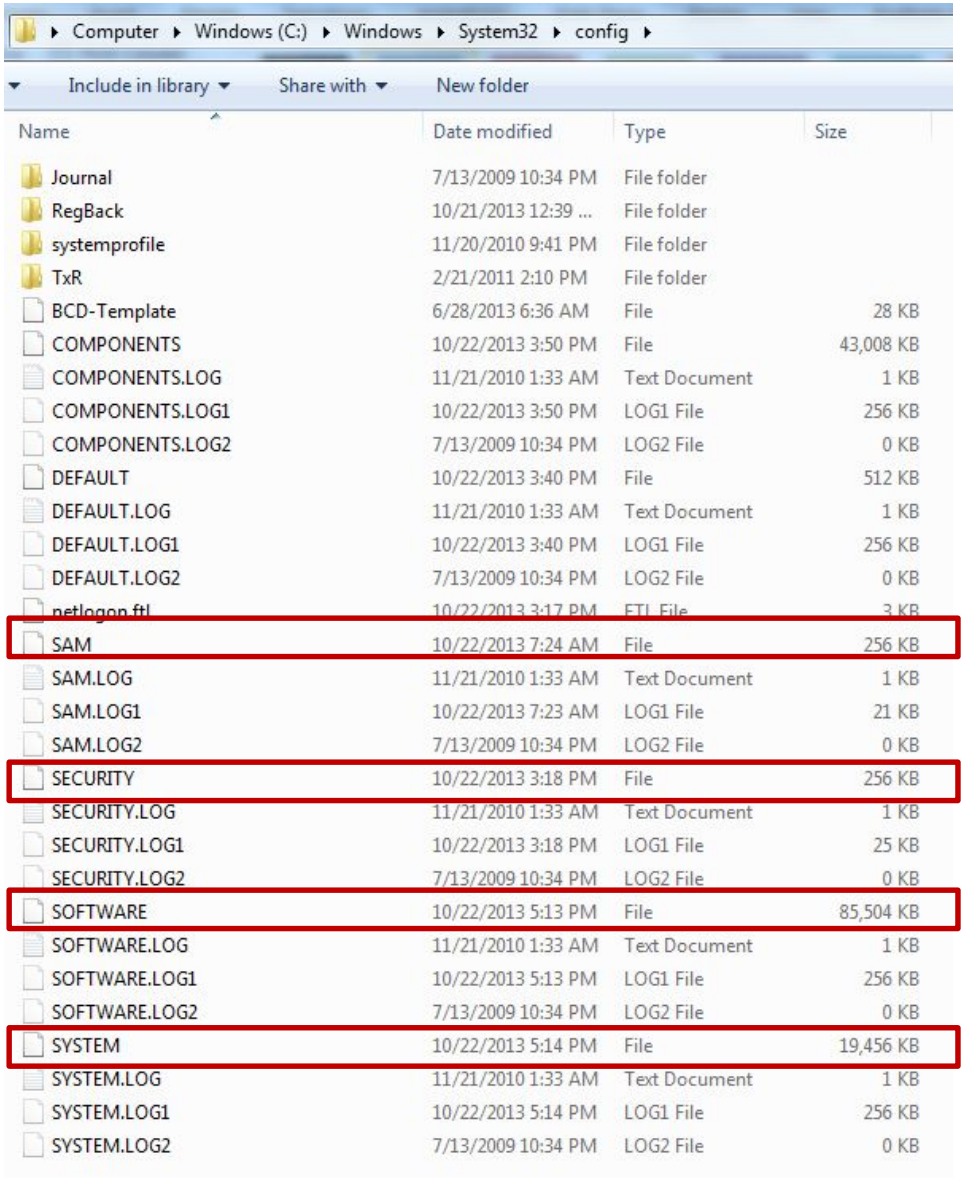
Windows Registry

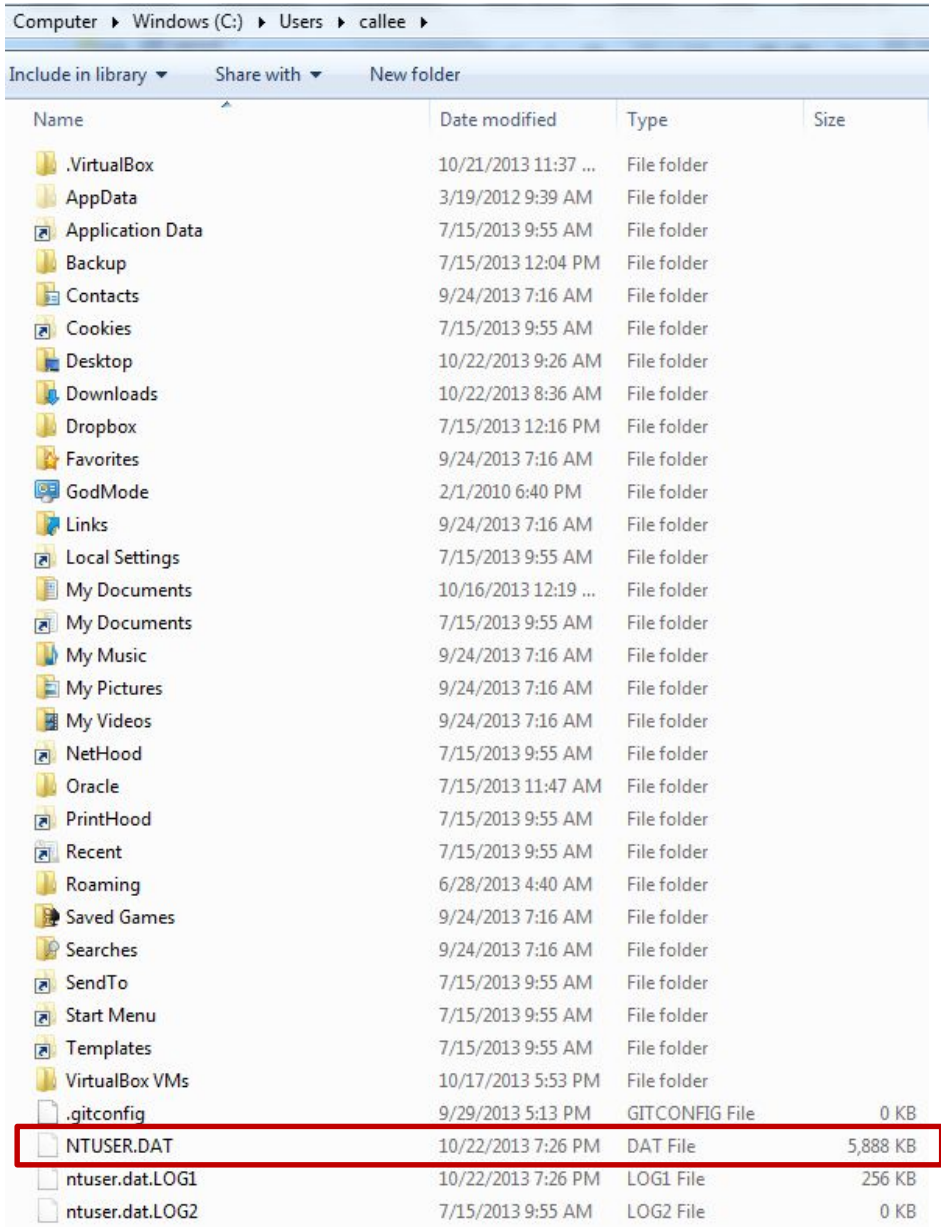
- Information about:
  - Applications installed
  - Application settings
  - Hardware installed
  - Hardware settings
  - User interface and system preferences
  - User accounts
  - Locations of files and recent activities, e.g. Most Recently Used (MRU)
  - Lots of online activities, e.g. usernames and passwords, browsing and search query history

Five Main Registry Files

File	Description
NTUSER.DAT	One for each user account, includes information such as Most Recently Used (MRU) file lists, desktop settings, default application behaviors
SAM (Security Accounts Manager)	User account information (including passwords) and security settings
SECURITY	User and group security policies, e.g. which accounts can load device drivers, get remote access to the machine
SOFTWARE	Information about all install programs, including settings and directory paths
SYSTEM	Windows systems settings, such as drive letter mappings, storage volume information, system boot profile, last known good configuration, system name, Windows setup information, hardware profile information

Where are They Located?

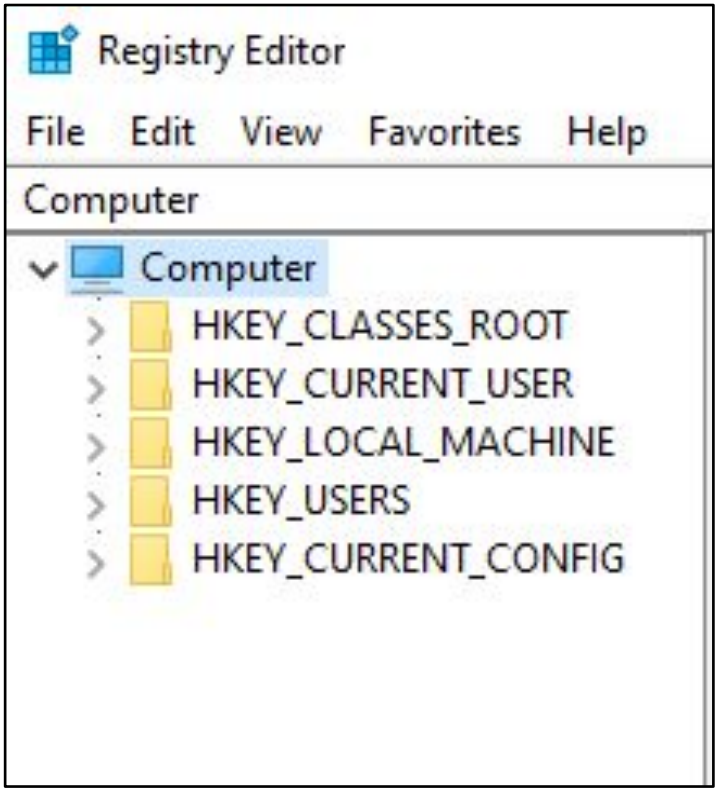




Registry Hives

Structure:

Hive
- Key
  - Subkey
    - Value



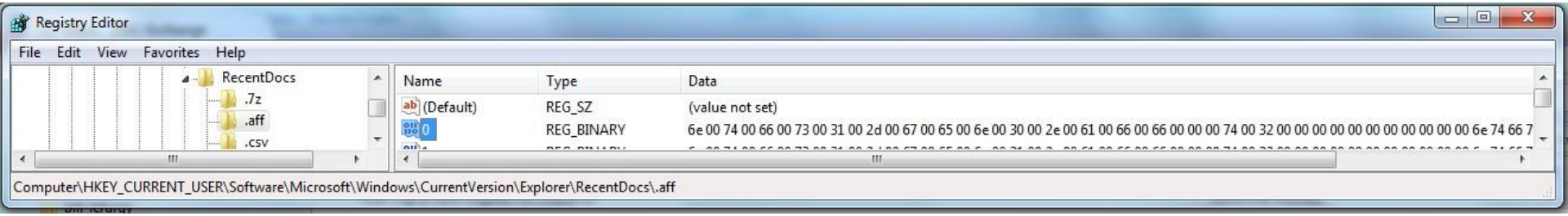
Example:
HKEY_CURRENT_USER\Software\Microsoft\Windows\CurrentVersion\Explorer\RecentDocs

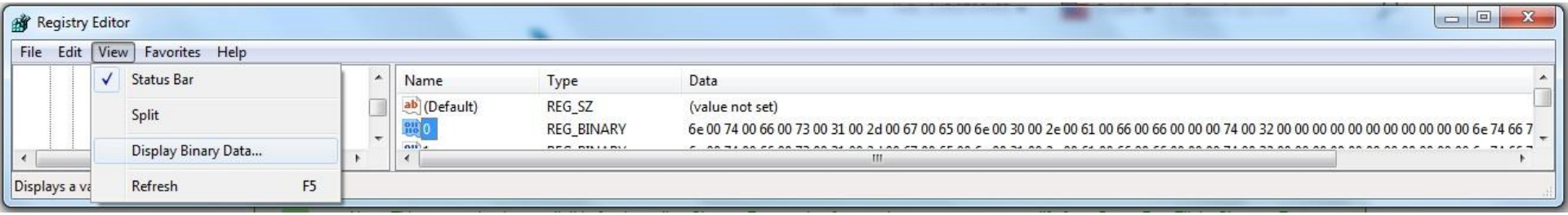
What do you think this is?

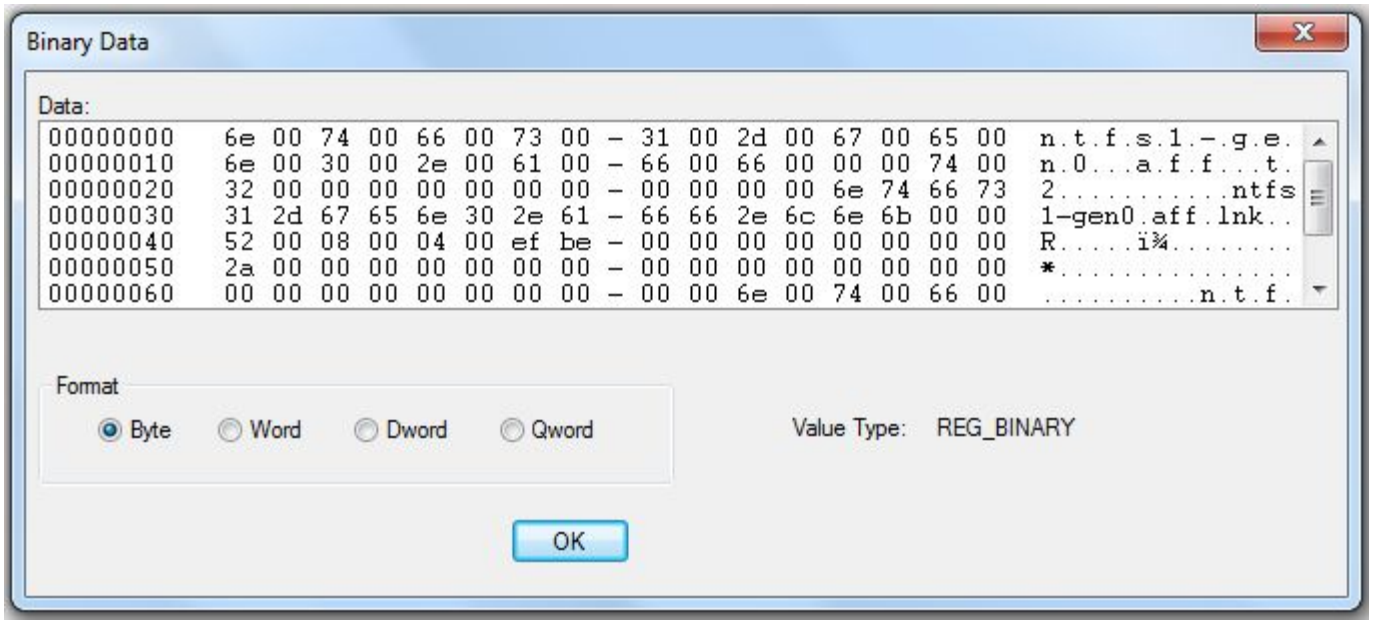
Registry Hives

Name	Description
HKEY_CLASSES_ROOT	Just points to HKEY_LOCAL_MACHINE\Software\Classes
HKEY_CURRENT_USER	**User setting information, which is generated dynamically from HKEY_USERS when a user logs into Windows**
HKEY_LOCAL_MACHINE	**Hardware and software settings that are specific to this computer but shared across users (generated at startup from SYSTEM.DAT)**
HKEY_USERS	**Information about each of the user accounts on the computer, e.g. desktop settings, default software behaviors - generated at startup from NTUSER.DAT files, and when user logs out of applications or out of Windows, data are written back to the ntUSER.DAT files**
HKEY_CURRENT_CONFIG	Just points to HKEY_LOCAL_MACHINE\Config

Question: Where would you find these registry hives on a disk image? (Hint: This is a trick question)


Registry Editor
File Edit View Favorites Help
RecentDocs
.7z
.aff
.csv
Name
Type
Data
(Default)
REG_SZ
(value not set)
0
REG_BINARY
6e 00 74 00 66 00 73 00 31 00 2d 00 67 00 65 00 6e 00 30 00 2e 00 61 00 66 00 66 00 00 00 74 00 32 00 00 00 00 00 00 00 00 00 00 00 6e 74 66 7
Computer\HKEY_CURRENT_USER\Software\Microsoft\Windows\CurrentVersion\Explorer\RecentDocs\.aff


Registry Editor
File Edit View Favorites Help
Status Bar
Split
Display Binary Data...
Refresh F5
Name
Type
Data
(Default)
REG_SZ
(value not set)
0
REG_BINARY
6e 00 74 00 66 00 73 00 31 00 2d 00 67 00 65 00 6e 00 30 00 2e 00 61 00 66 00 66 00 00 00 74 00 32 00 00 00 00 00 00 00 00 00 00 00 6e 74 66 7


Binary Data
Data:
00000000 6e 00 74 00 66 00 73 00 - 31 00 2d 00 67 00 65 00 n.t.f.s.1.-.g.e.
00000010 6e 00 30 00 2e 00 61 00 - 66 00 66 00 00 00 74 00 n.0...a.f.f...t.
00000020 32 00 00 00 00 00 00 00 - 00 00 00 00 6e 74 66 73 2.....ntfs
00000030 31 2d 67 65 6e 30 2e 61 - 66 66 2e 6c 6e 6b 00 00 1-gen0.aff.lnk..
00000040 52 00 08 00 04 00 ef be - 00 00 00 00 00 00 00 00 R.....ï¾.....
00000050 2a 00 00 00 00 00 00 00 - 00 00 00 00 00 00 00 00 *.....
00000060 00 00 00 00 00 00 00 00 - 00 00 6e 00 74 00 66 00 .....n.t.f.
Format
Byte
Word
Dword
Qword
Value Type: REG_BINARY
OK

Registry Hive Value Types

Type	Description
REG_BINARY	Raw binary data displayed as hexadecimal*
REG_DWORD	32-bit unsigned integer (4 bytes)
REG_EXPAND_SZ	Variable-length string, usually in UTF-16 (Unicode)
REG_FULL_RESOURCE_DESCRIPTOR	Series of nested arrays used by a hardware device, binary data displayed as hexadecimal*
REG_LINK	Symbolic link to another registry key (Unicode)
REG_MULTI_SZ	Ordered list of strings (multi-string value), usually in UTF-16
REG_NONE	No specific type – displayed as hexadecimal*
REG_QWORD	64-bit integer (8 bytes)
REG_RESOURCE_LIST	Series of nested arrays used by a hardware device, binary data displayed as hexadecimal*
REG_RESOURCE_REQUIREMENTS_LIST	Series of nested arrays used by a hardware device, binary data displayed as hexadecimal*
REG_SZ	Fixed-length text string, usually in UTF-16

* Can be opened and viewed in a hex editor

Security ID (SID)

- One assigned to each user account
- Associated with various resources, including files, folders and Recycling Bins

SID Example

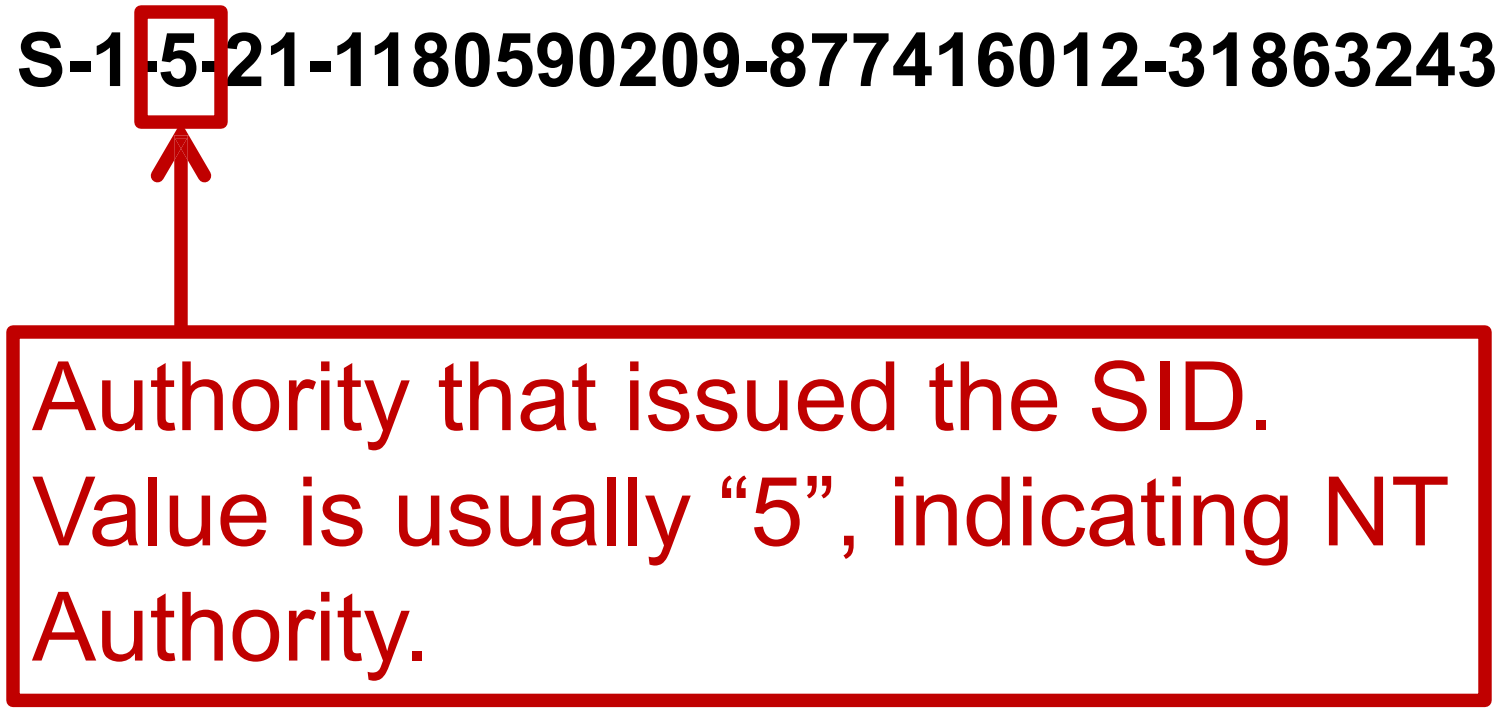
S-1-5-21-1180590209-8774160 12-3186324384-1002

S-1-5-21-1180590209-8774160 12-3186324384-1002

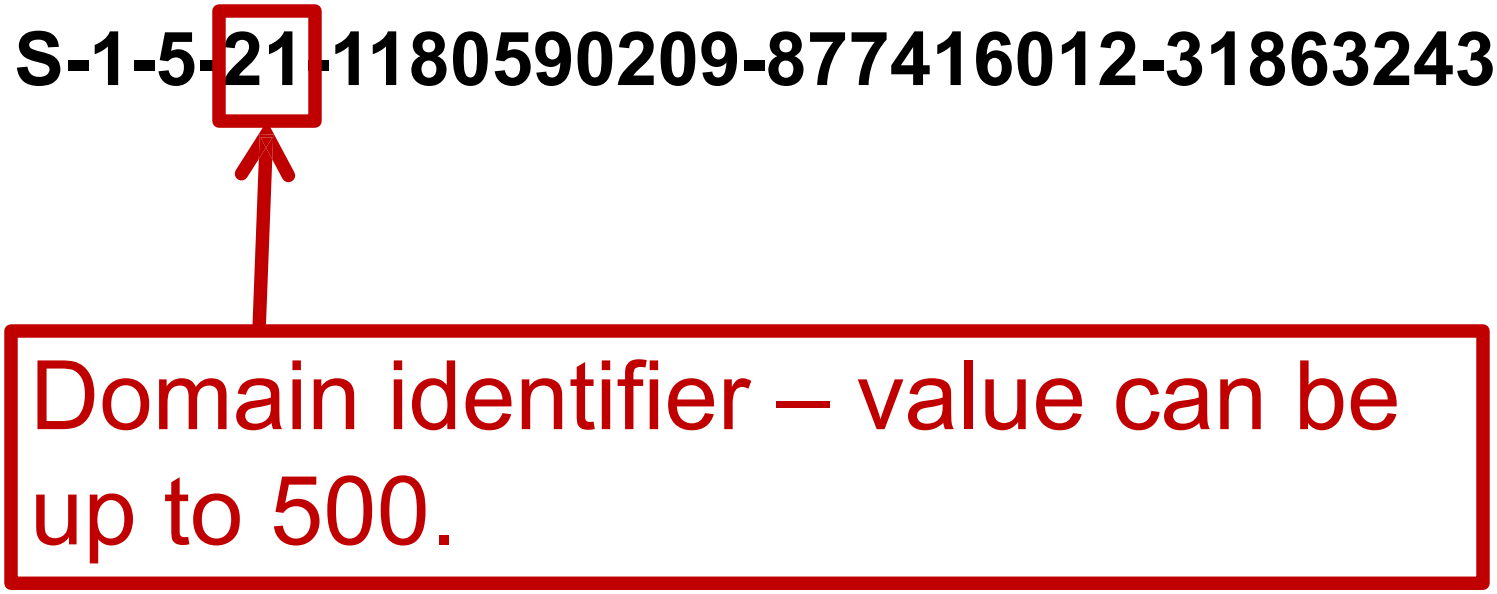
Always an “S”, indicating that this is an SID.

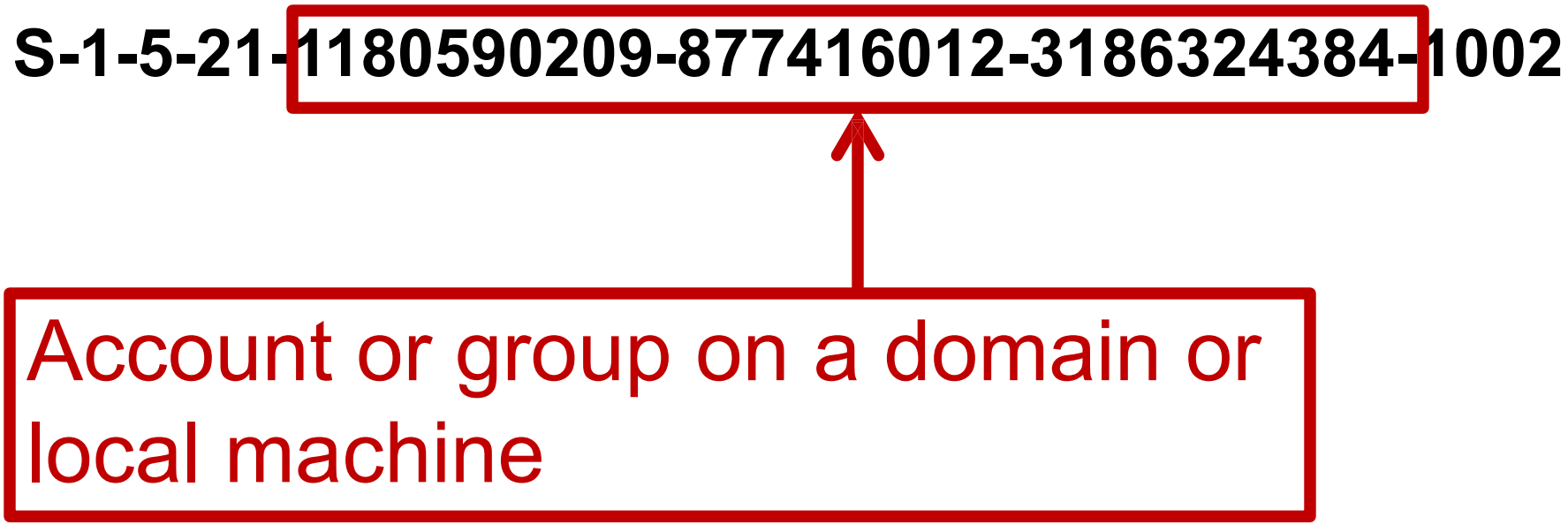
S-1-5-21-1180590209-8774160 12-3186324384-1002

S-1-5-21-1180590209-8774160 12-3186324384-1002



S-1-5-21-1180590209-8774160 12-3186324384-1002




S-1-5-21-1180590209-8774160 12-3186324384-1002
Account or group on a domain or local machine

S-1-5-21-1180590209-8774160 12-3186324384-1002

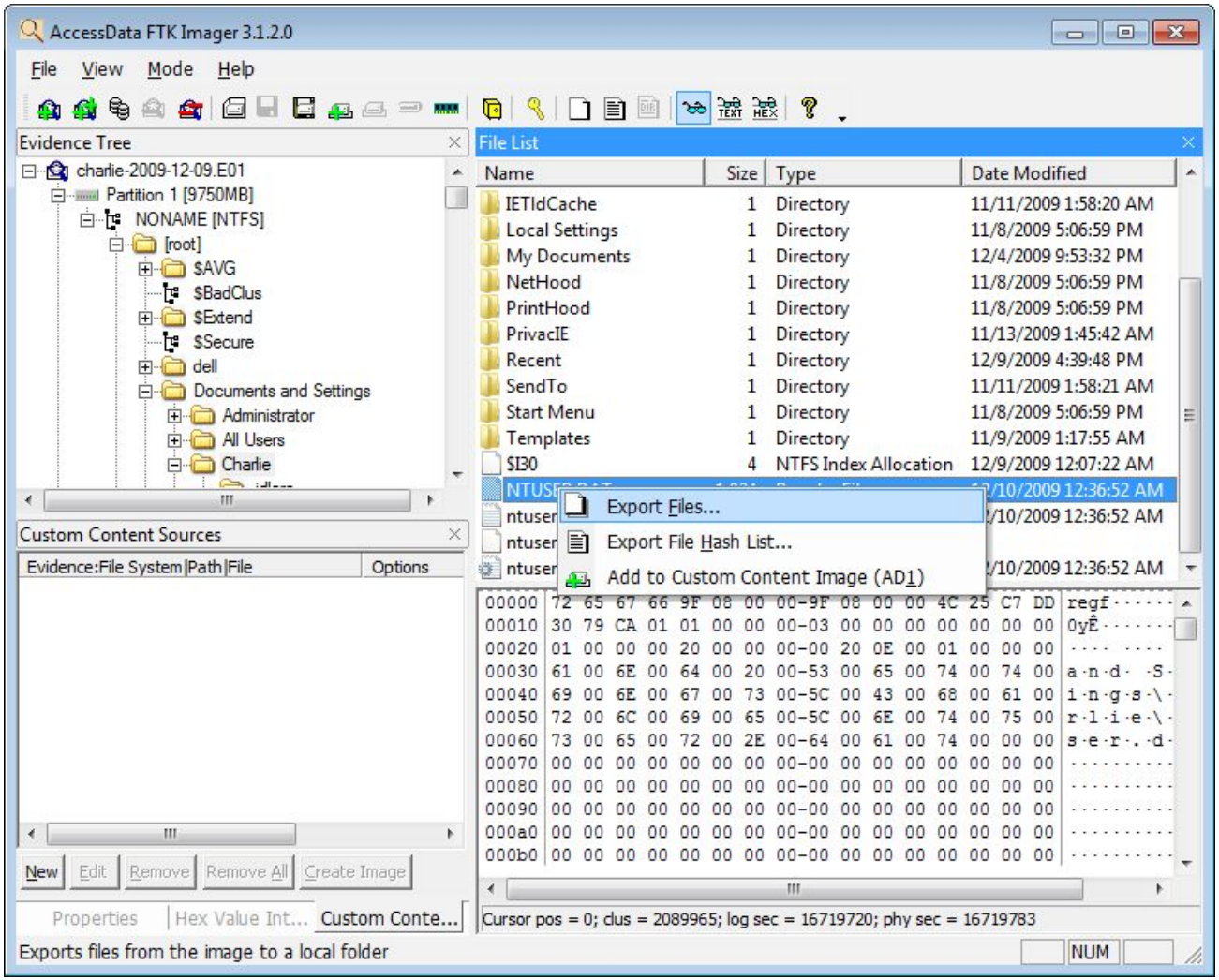
Relative Identifier (RID), designating a specific user in the SAM file. Those below 1000 are default accounts (e.g. 500 = Administrator), and those 1000 or above are created for specific groups or users.

Examining an NTUSER.DAT File

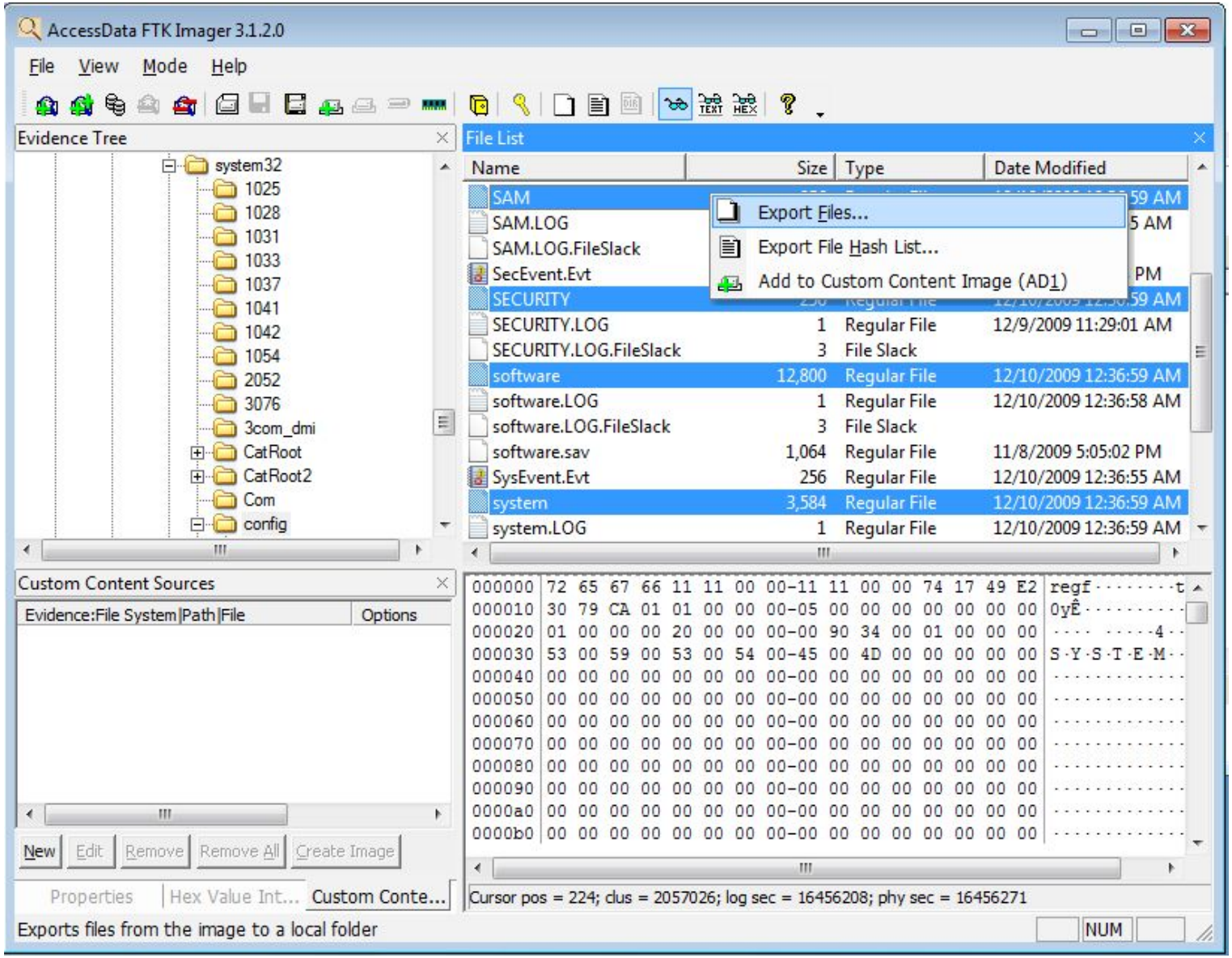
- The files in registry (a folder within the zip file you downloaded earlier) were extracted from a full-drive (including the operating system) disk image

- The following is an example of how these files can be extracted using FTK Imager

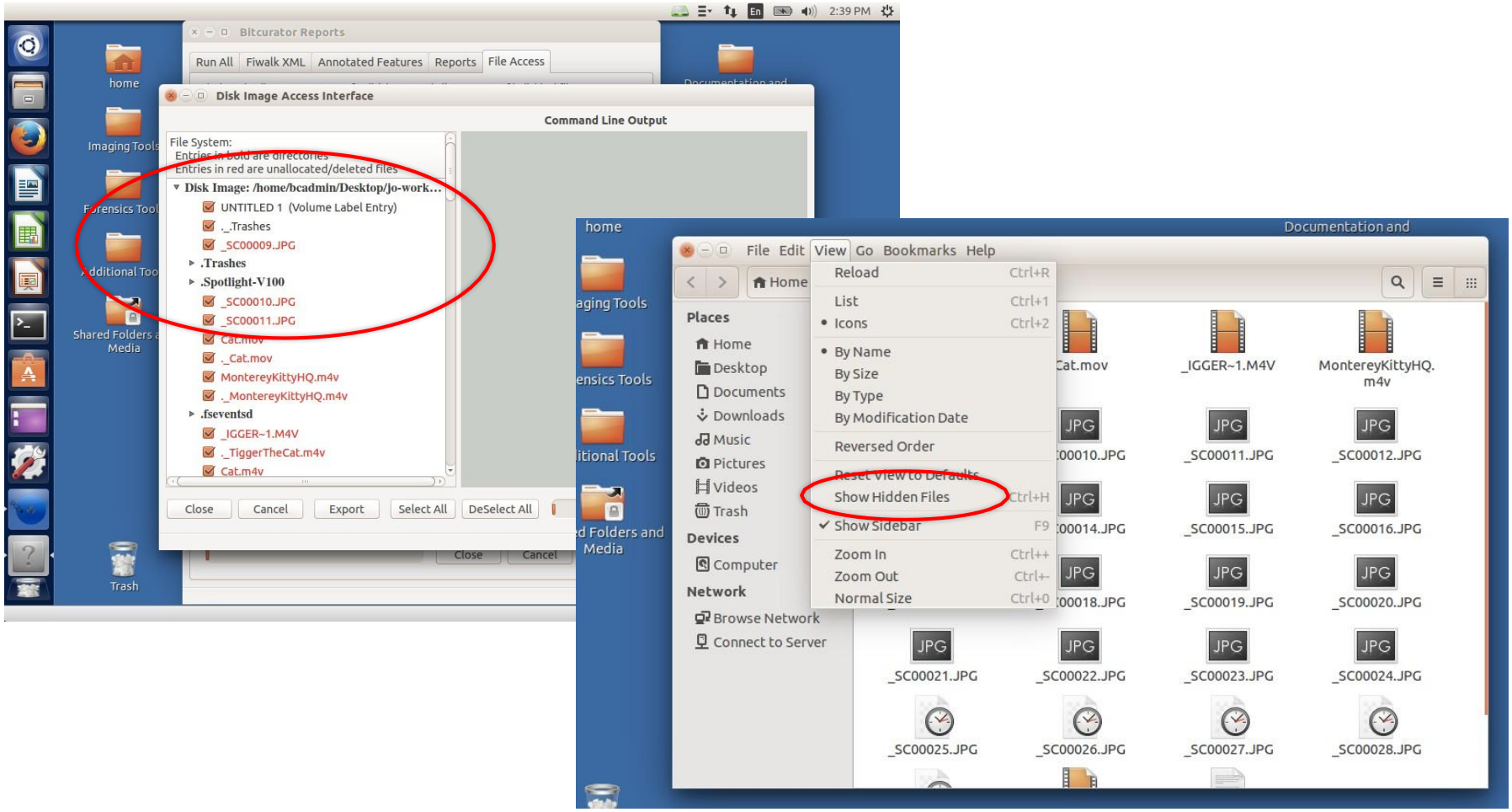
- Navigate to: Partition 1 > [root] > Documents and Settings > Charlie > NTUSER.DAT
- Right click on NTUSER.DAT and select Export Files.



Then export the other four registry files from Windows\System32\config

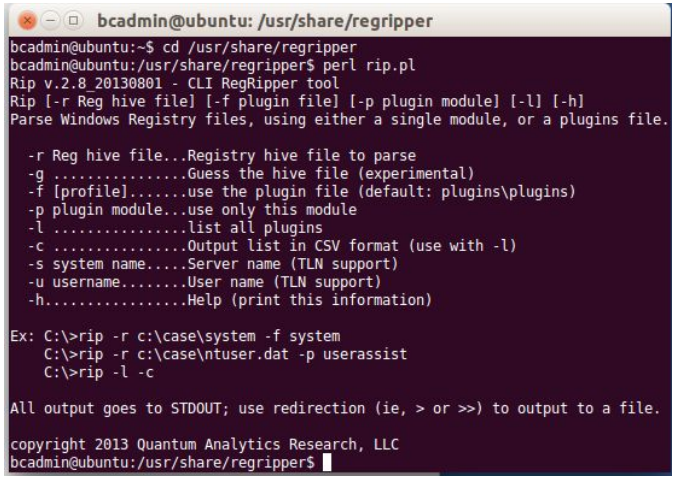


Perform these same tasks in the BitCurator environment



RegRipper Instructions - BitCurator

- Navigate to Forensics Tools, and click on the RegRipper icon
- NOTE: **IGNORE** examples that it presents, because they use commands and syntax for Windows, not Linux
- Issue each of the following commands:



perl rip.pl -r ~/Desktop/sample-data/registry/NTUSER.DAT > ~/Desktop/ntuser-report -f ntuser

perl rip.pl -r ~/Desktop/sample-data/registry/SAM > ~/Desktop/sam-report -f sam

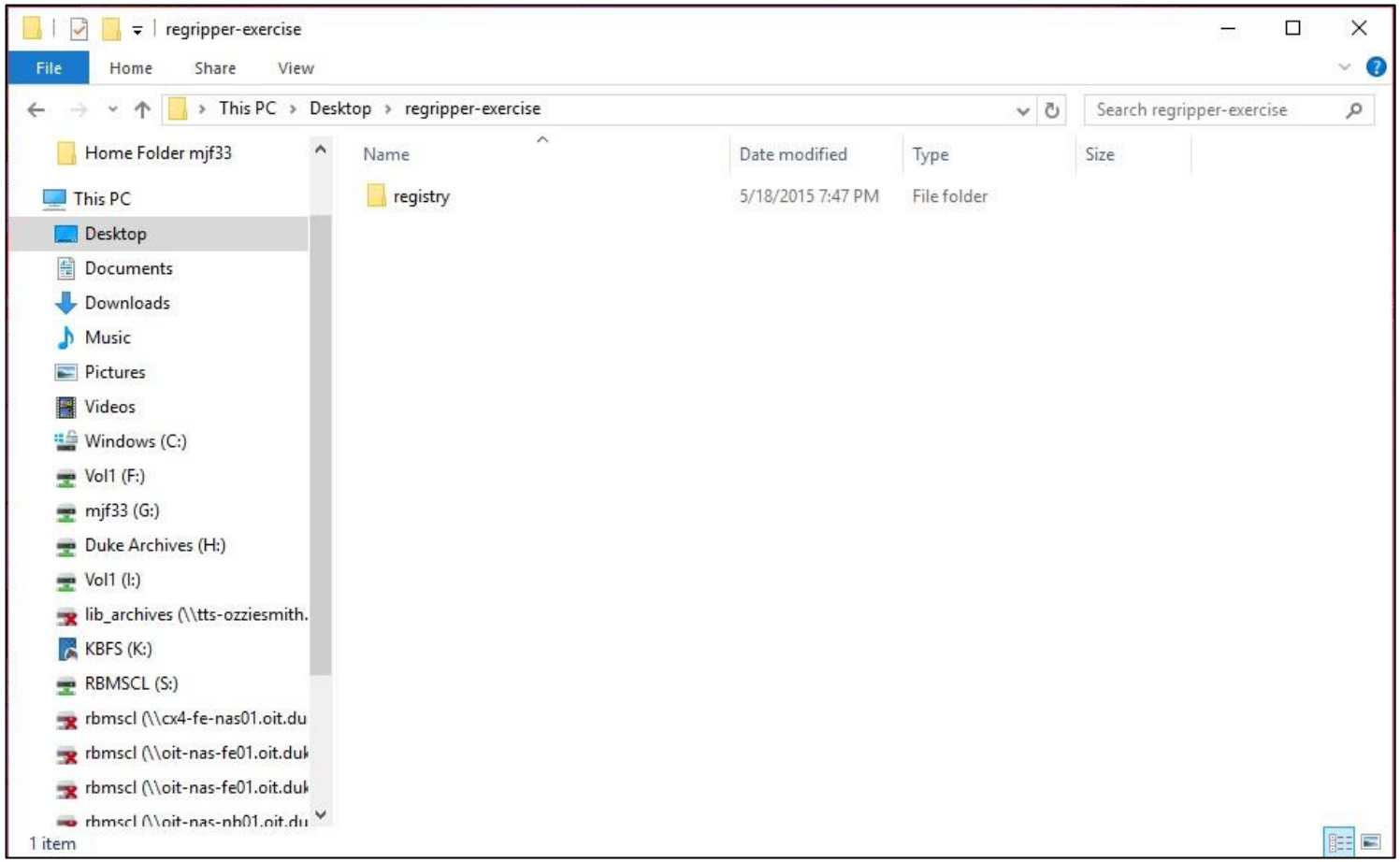
perl rip.pl -r ~/Desktop/sample-data/registry/SECURITY > ~/Desktop/security-report -f security

perl rip.pl -r ~/Desktop/sample-data/registry/SOFTWARE > ~/Desktop/software-report -f software

perl rip.pl -r ~/Desktop/sample-data/registry/SYSTEM > ~/Desktop/system-report -f system

*Enter each command in its entirety before hitting enter (line breaks above are simply to fit the text onto the slide, not ones that you should type yourself). Remember that the up arrow and tab can save you time when typing commands.

RegRipper Instructions – Windows I



- Create a folder called regripper-exercise on your desktop
- Find the registry directory in the folder you extracted from the saa-das-sample-data zip file earlier
- Copy this folder to the regripper-exercise folder on your Desktop

RegRipper Instructions – Windows II

- Navigate to saa-dfa-sample-data\RegRipper3.0
- Run rr.exe (it may simply appear as rr) by double clicking it.
- The next set of steps will be run 5 times - once for each of the files in regripper-exercise\registry
- Next to the Hive File window, select Browse
  - Navigate to regripper-exercise\registry and select the first Hive File
  - E.g., NTUSER.DAT
- Next to Report File, select Browse
  - Create a new file in regripper-exercise that corresponds to the Hive File above
  - E.g., NTUSER_report.txt
- ~~In the Profile dropdown, select the appropriate profile~~
  - ~~E.g., ntuser~~ profile selection is not required in RegRipper 3.0
- Select Rip It.
- Repeat the above steps for SAM, SECURITY, SOFTWARE, and SYSTEM

RegRipper Output Questions

Examine ntuser-report.txt	**Are you able to identify files that the user recently opened? If so, what were they? Can you determine what the most recently opened files of specific types (e.g. txt) were?**
Examine sam-report.txt	**How many accounts were there on the computer that is represented in the disk image? What is the Relative Identifier (RID) for the user account you're examining? What other interesting information can you gain from the SAM report about this user account and how might you use that information?**
Examine security-report.txt	**What is the Machine SID for the computer represented in the disk image? Why would you want to know this? How does it relate to the RID that you identified above?**
Examine software-report.txt	**Identify three different applications that were installed on the computer and the file paths where the applications were stored.**
Examine system-report.txt	**Find the devclass output. What does this output tell you? How might this information be useful?**

RegRipper Output Discussion – ntuser-report

- Are you able to identify the files that the user recently opened? If so, what were they?
  - How did you go about finding this information?
  - What line number(s) points to this information?
- Can you determine what the most recently open files of specific types (e.g. txt) were?
  - How did you go about finding these?
  - What line numbers have this information?
- Look at lines 1109-1117 - what type of information are you looking at?
- Is there any other information you find particularly compelling in this report?
- What might you do with this information?

RegRipper Output Discussion – sam-report

- How many accounts were there on the this computer?
  - How did you go about finding this information?
  - What line number(s) points to this information?
- What was the Relative Identifier (RID) for the user account you're examining?
  - How did you go about finding this?
- How many logins did Pat make on this machine?
- Is there any other information you find particularly compelling in this report?
- What might you do with this information?

RegRipper Output Discussion – security-report

- What is the Machine SID for the computer represented here?
  - How did you go about finding this information?
  - What line number(s) points to this information?
- Why would you want to know this information
- How does this relate to the RID in the previous report?

security-report.txt

```
1  auditpol v.20121128
2  (Security) Get audit policy from the Security hive file
3
4  auditpol
5  Policy\PolAdtEv
6  LastWrite Time Sun Nov  8 15:34:54 2009 (UTC)
7
8  Length of data: 44 bytes.
9  0x00000000: 00 fa 07 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
10 0x00000010: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
11 0x00000020: 00 00 00 00 00 00 00 00 09 00 00 00              ............
12 **Auditing is NOT enabled.
13 ----------------------------------------
14 lsasecrets v.20100219
15 (Security) TEST - Get update times for LSA Secrets
16
17
18 Domain secret - $MACHINE.ACC
19 Error: Can't call method "get_value" on an undefined value at
   C:\Users\mjf33\Desktop\das_stuff\das-forensics-flash-drive-files-excluding-vbox-bitcurator-slides-20160513\
20
```

RegRipper Output Discussion – software-report

- Identify three different applications that were installed on this computer
  - How did you go about finding this information?
  - What line number(s) points to this information?

- Why would you want to know this information?

- How might it aid description?

software-report.txt

```
1   Launching appinitdlls v.20130425
2   appinitdlls v.20130425
3   (Software) Gets contents of AppInit_DLLs value
4
5   AppInit_DLLs
6   Microsoft\Windows NT\CurrentVersion\Windows
7   LastWrite Time Fri Nov 20 18:55:34 2009 (UTC)
8     AppInit_DLLs : {blank}
9     LoadAppInit_DLLs : 1
10  *LoadAppInit_DLLs value globally enables/disables AppInit_DLLS.
11  0 = disabled (default)
12
13  Wow6432Node\Microsoft\Windows NT\CurrentVersion\Windows not found.
14  Analysis Tip: The AppInit_DLLs value should be blank; any DLL listed
15  is launched with each user-mode process.
16  ----------------------------------------
17  apppaths v.20120524
18  (Software) Gets content of App Paths subkeys
19
20  App Paths
21  Microsoft\Windows\CurrentVersion\App Paths
```

RegRipper Output Discussion – system-report

- Find the devclass output
- What does this output tell you?
- How might this information be useful?

system-report.txt

```
ControlSet001\Control\Session Manager\AppCertDlls not found.
---------------------------------------
appcompatcache v.20130425
(System) Parse files from System hive Shim Cache

Signature: 0xdeadbeef
WinXP, 32-bit
C:\Program Files\AVG\AVG9\avgsrmax.exe
ModTime: Mon Dec  7 23:47:51 2009 Z
UpdTime: Tue Dec  8 17:52:13 2009 Z
Size   : 361752 bytes

C:\Program Files\RealVNC\VNC4\winvnc4.exe
ModTime: Thu Oct 16 01:13:58 2008 Z
UpdTime: Tue Dec  8 01:45:23 2009 Z
Size   : 439632 bytes

C:\RAM\mddbak.exe
ModTime: Sat Nov 14 01:07:38 2009 Z
UpdTime: Sun Dec  6 16:10:51 2009 Z
Size   : 95104 bytes

C:\Program Files\Java\jre6\bin\jqsnotify.exe
ModTime: Sun Oct 11 12:17:34 2009 Z
```

Viewing and Copying Registry Information if You're Running the Original Environment

- What if you’re logged in to the original computer? How might you get information out of the registry?
- What if you wanted to replicate that registry information on another computer?
- Hint: There are tools built into Windows for this.

Restore Points

- Snapshots of Registry hives and some other essential system (including .EXE, .INI, .LNK) files. They're created:
  - when there are major system changes, e.g. installing software
  - at regularly scheduled intervals
  - if the user manually creates one
- Let's look at some restore points: Start Button > All Programs > Accessories > System Tools > System Restore [or just "System Restore" in the Start box]

Examining the Recycle Bin

1. In the start menu box, type **cmd**
2. Type: **cd c:\$recycle.bin**
   (What is this doing?)
3. Type **dir /a**
   (What is this doing?)
4. Type **dir *.* /s**
   (What is this doing?)
5. Put one or more files into the Recycle Bin
   (by moving there or by deleting)
6. Repeats steps 2-4. What do you see now?

A Brief Discussion of Mac Forensics

- No Registry, so where is all the good stuff stored?

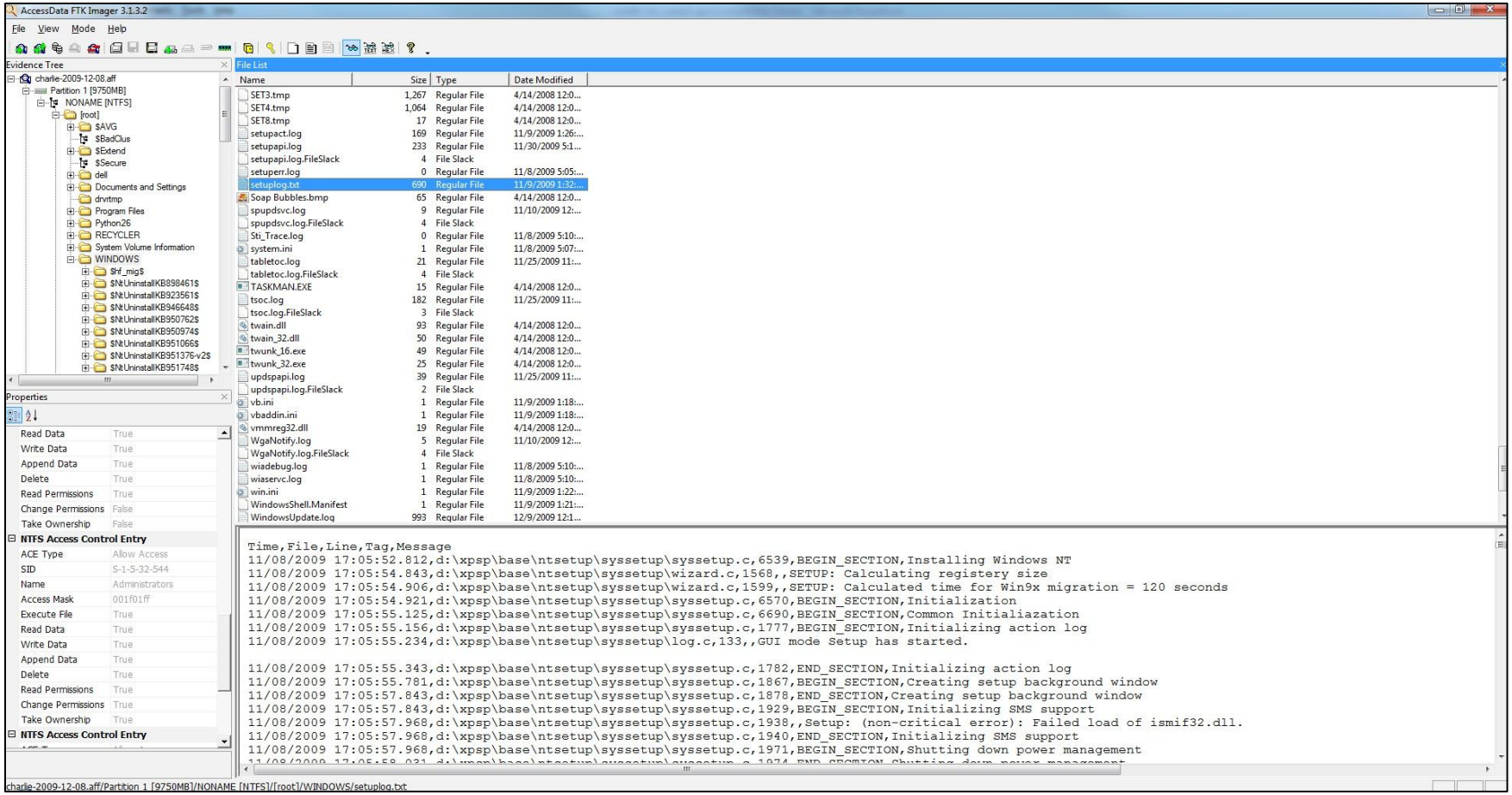
- See: https://forensicswiki.xyz/wiki/index.php?title=Mac_OS_X_10.9_-_Artifacts_Location but note that this is information a snapshot in time; artifact locations tend to change between versions of macOS.

Archival Importance and Role of SID

- If the volume is NTFS, you can find the SID associated with a specific file
- If you also have registry files from the original computer (particularly SAM.DAT), you can get information associated with that SID, such as the name of the user/group, last time he/she logged in, and various other account details

setuplog.txt

- See disk image example below: Partition 1 > [root] > WINDOWS > setuplog.txt



- What do you see in this file?
- What information could be useful for digital curation?
- When/how might you use it?

End User Access Scenarios*

- Virtualization and emulation
- Mounting the original filesystem
- Accessing (but not mounting) disk images using forensics software
- Remote, dynamic access to disk image contents
- Cross-drive analysis

*Note: The first three were discussed earlier

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

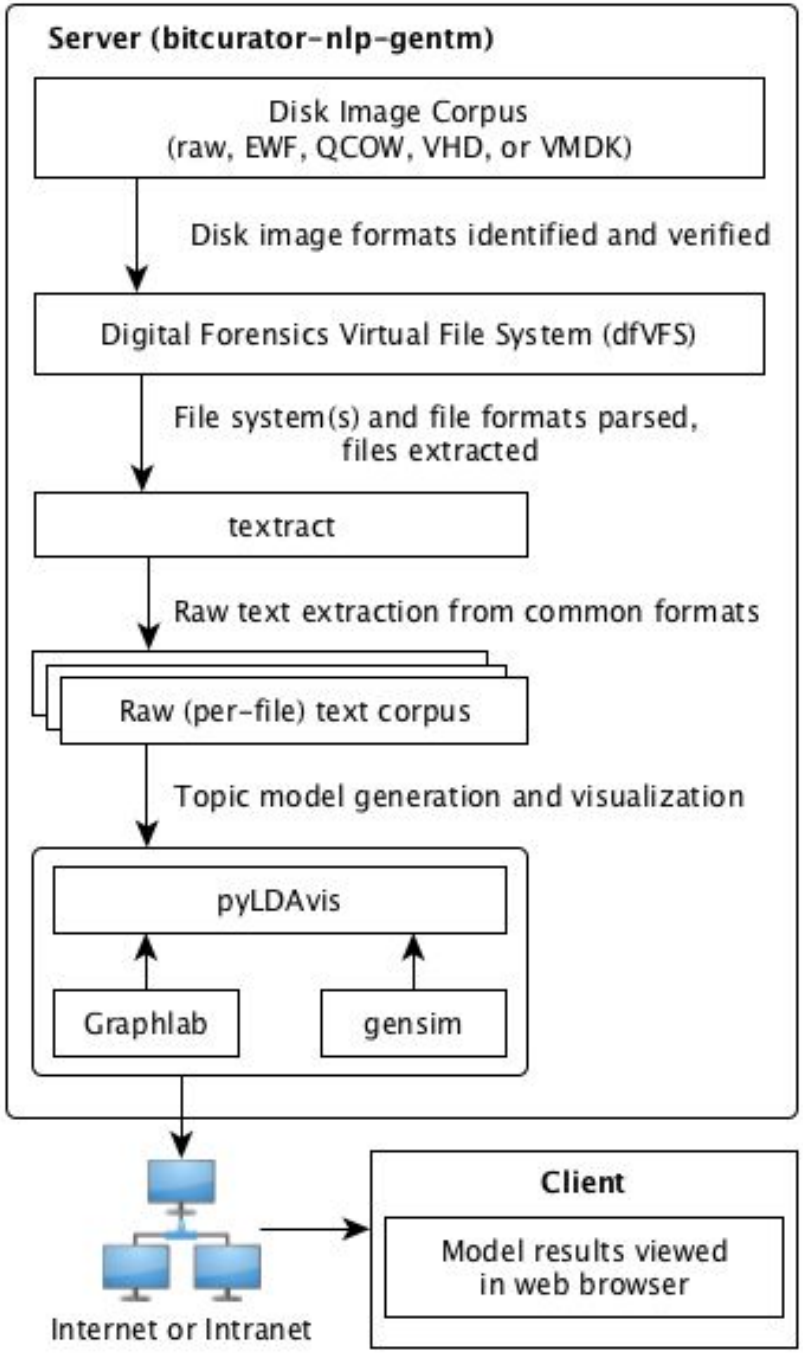
BitCurator Access Redaction Tools

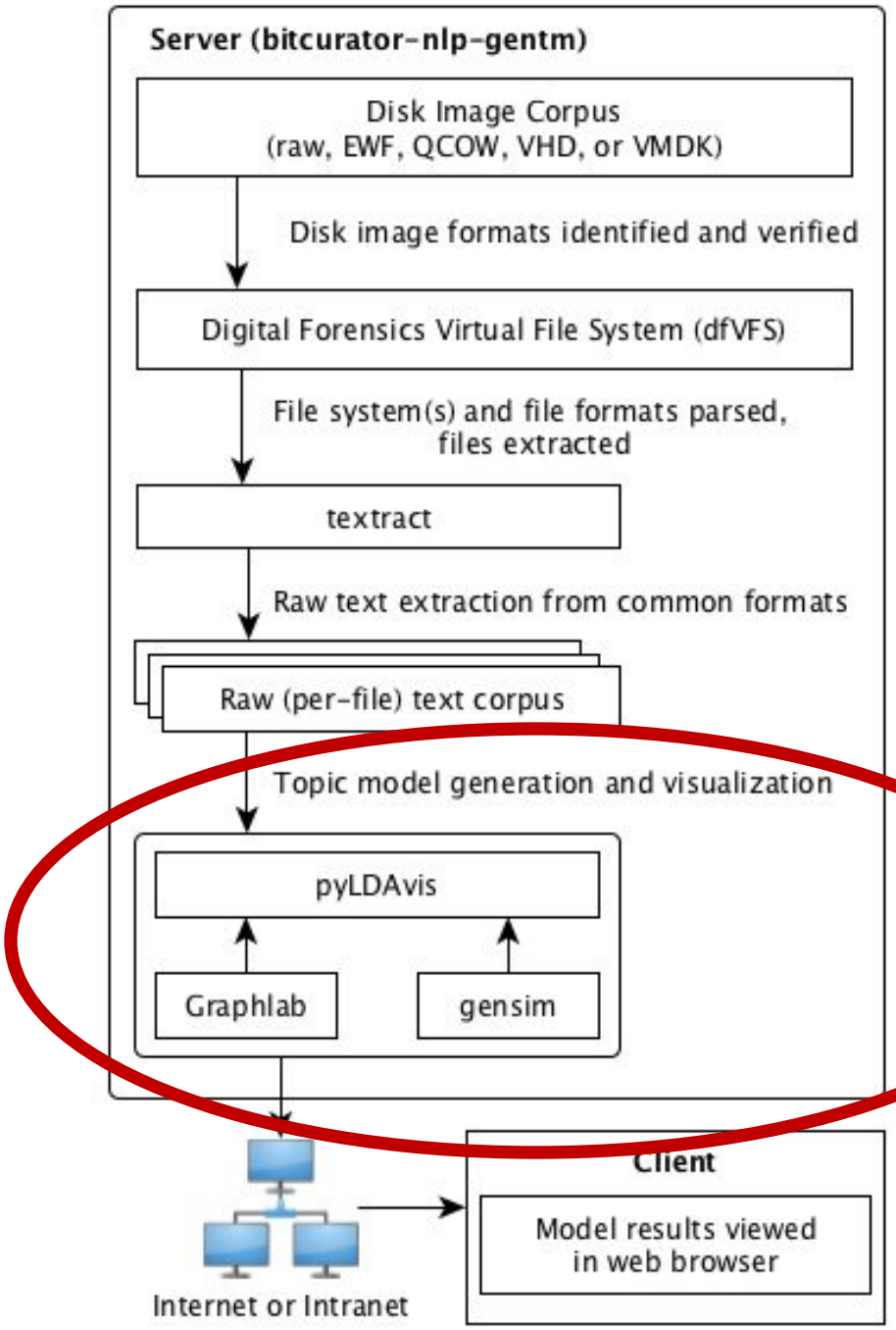
- Software to redact strings and byte sequences identified in disk images
- Three types of redaction actions:
  - SCRUB (overwrite the bytes in the target with zeroes),
  - FILL (overwrite by filling with a given character),
  - FUZZ (altering the content of a binary, so it can no longer run).
- Best used through a command-line interface but also include a graphic user interface (GUI) that supports the same functions
- Python API allowing institutions to develop custom redaction facilities using open-source tools including lightgrep

https://github.com/bitcurator/bitcurator-access-redaction

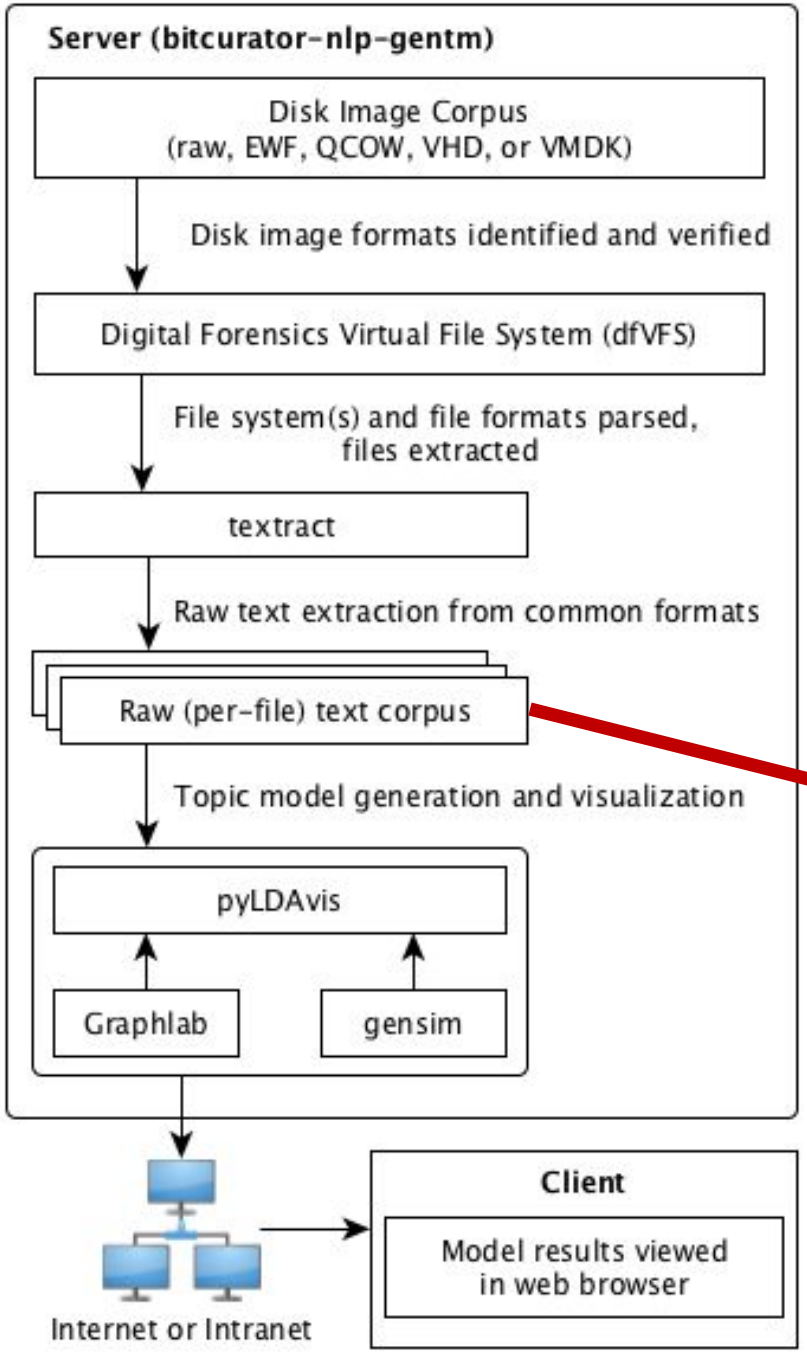
BitCurator NLP

- Funded by Andrew W. Mellon Foundation: October 1, 2016 – September 30, 2018
- Develop software for collecting institutions to extract, analyze, and produce reports on features of interest in text extracted from born-digital materials
- Use existing natural language processing software libraries to identify and report on those items likely to be relevant to ongoing preservation, information organization, and access activities
- May include entities (e.g. persons, places, and organizations), potential relationships among entities (e.g. appear together within documents or set of documents), and topic models to provide insight into how concepts are naturally clustered within the documents.


Server (bitcurator-nlp-gentm)
Disk Image Corpus
(raw, EWF, QCOW, VHD, or VMDK)
Disk image formats identified and verified
Digital Forensics Virtual File System (dfVFS)
File system(s) and file formats parsed,
files extracted
textract
Raw text extraction from common formats
Raw (per-file) text corpus
Topic model generation and visualization
pyLDAvis
Graphlab
gensim
Internet or Intranet
Client
Model results viewed
in web browser

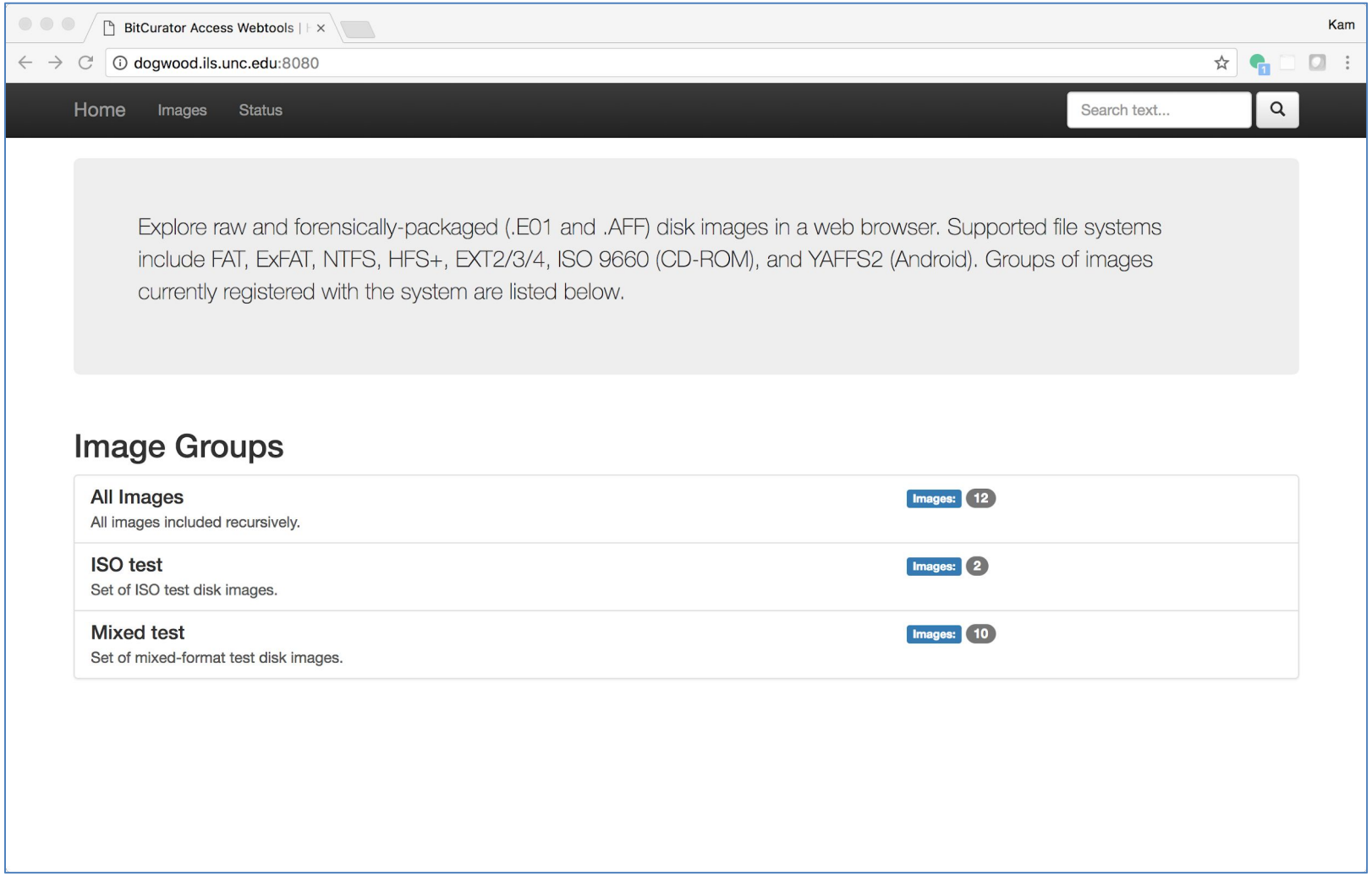


This is showing topic modeling, which we'll look at in more detail soon.

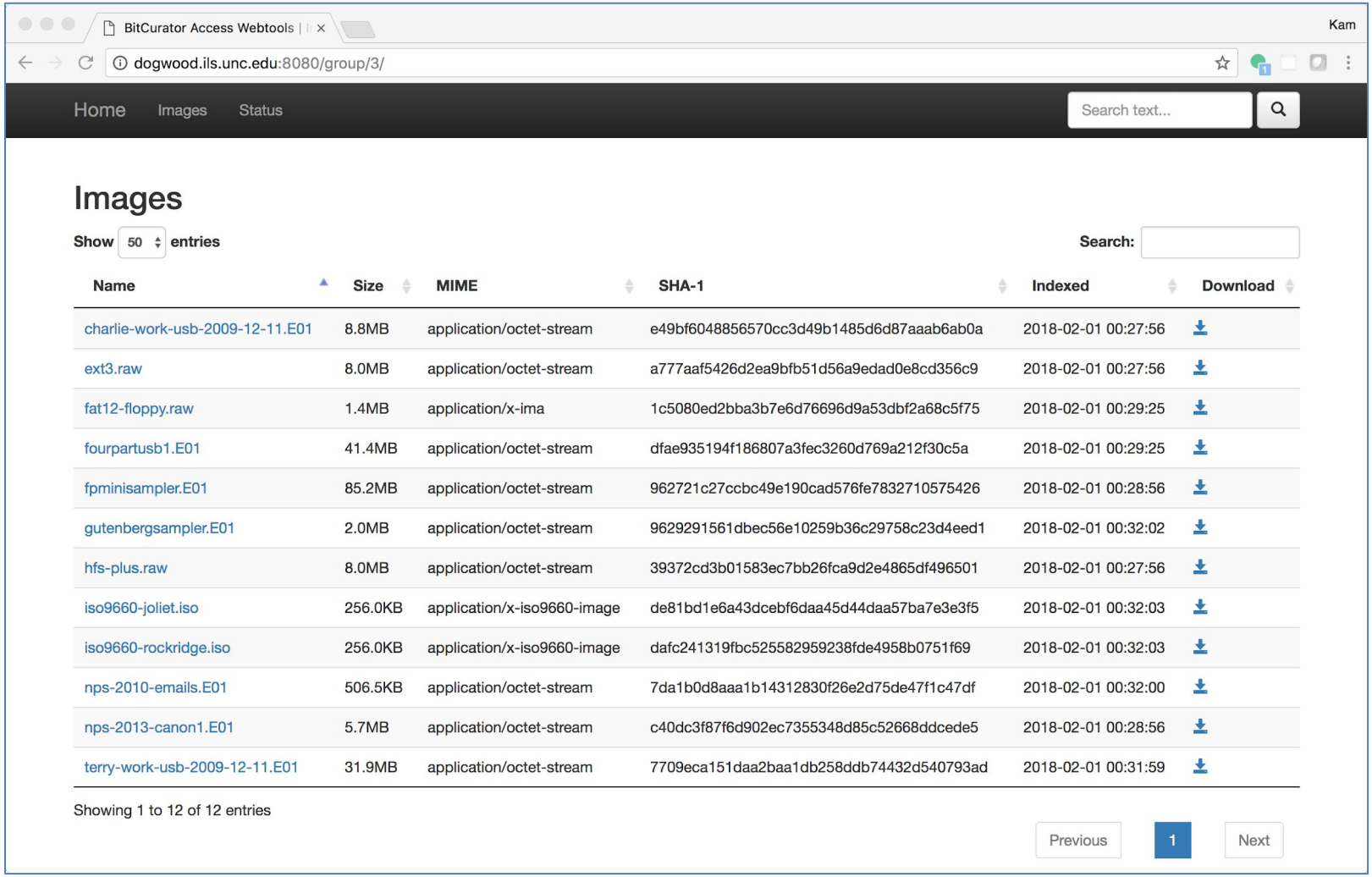


Another path at this point is to feed the text into spacy for named-entity recognition

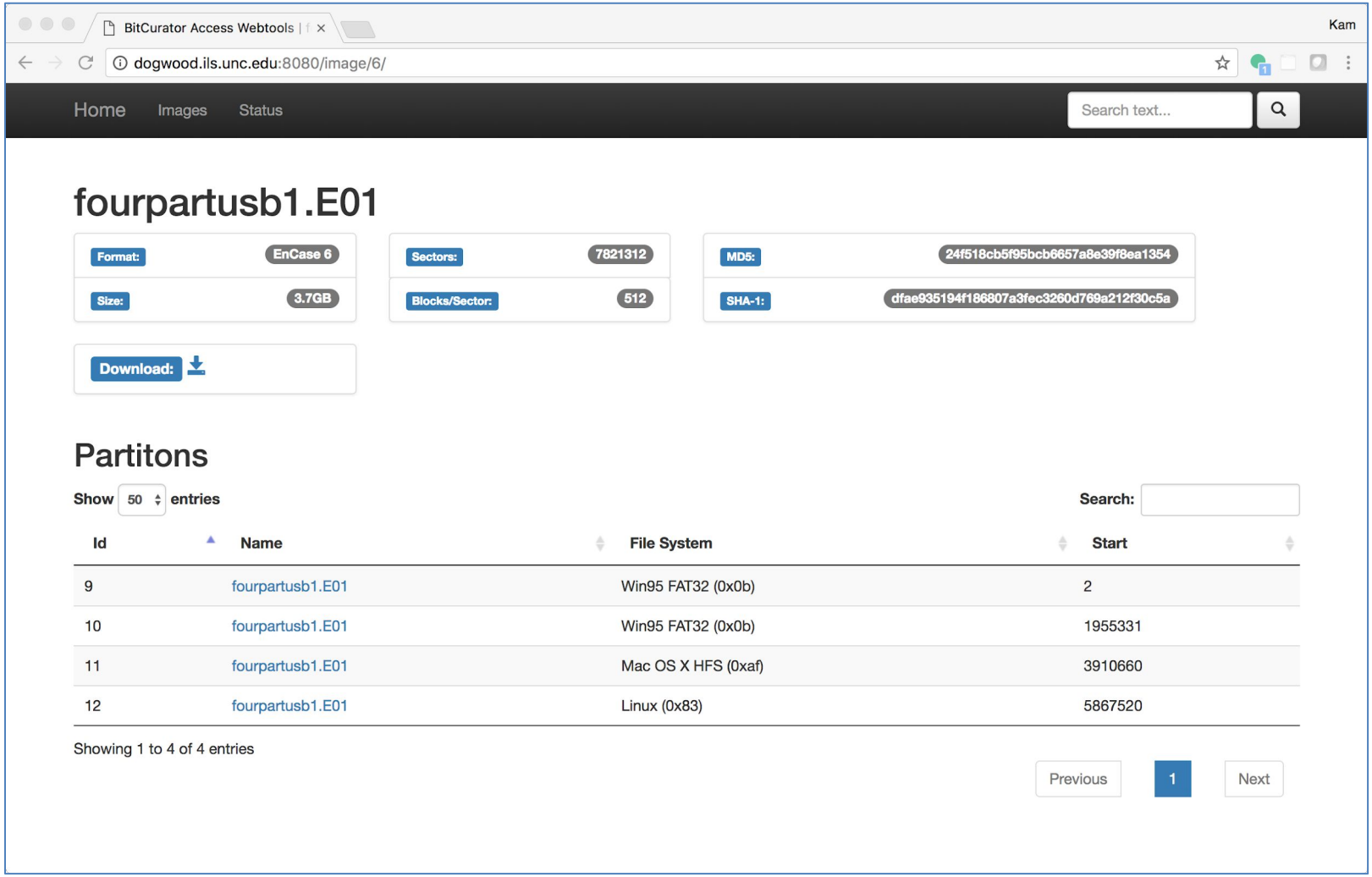
BitCurator Access Webtools



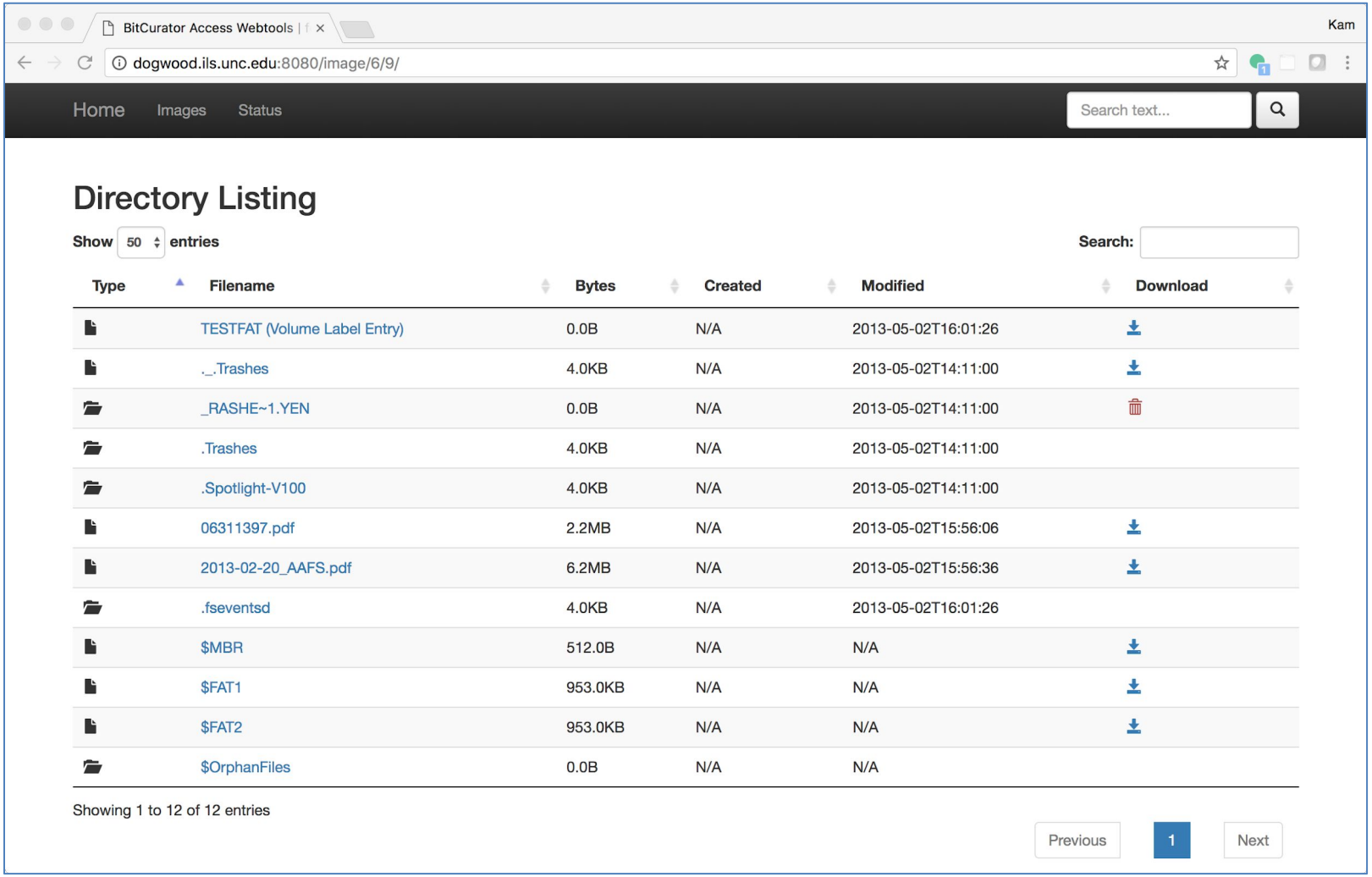
BitCurator Access Webtools



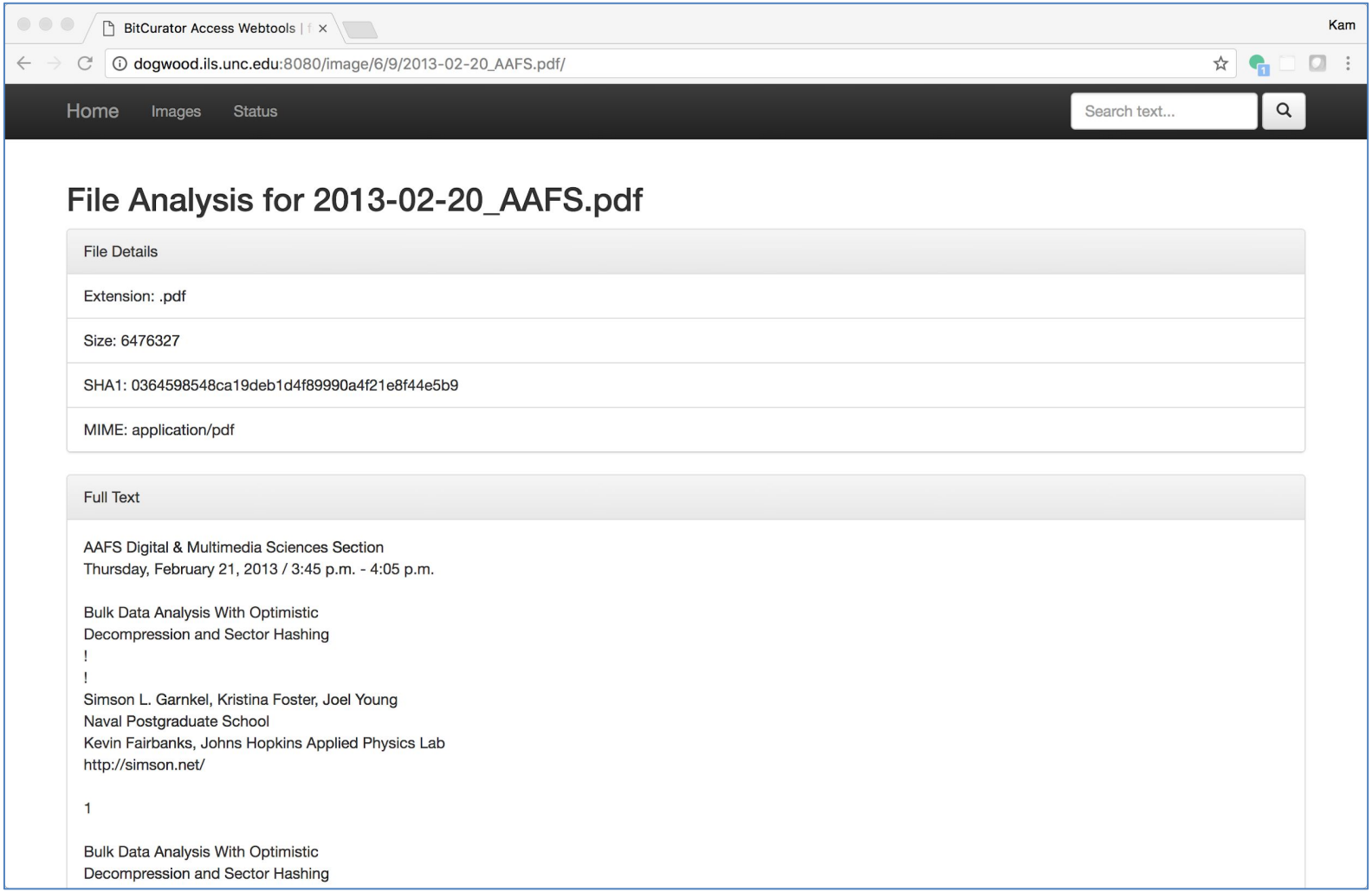
BitCurator Access Webtools



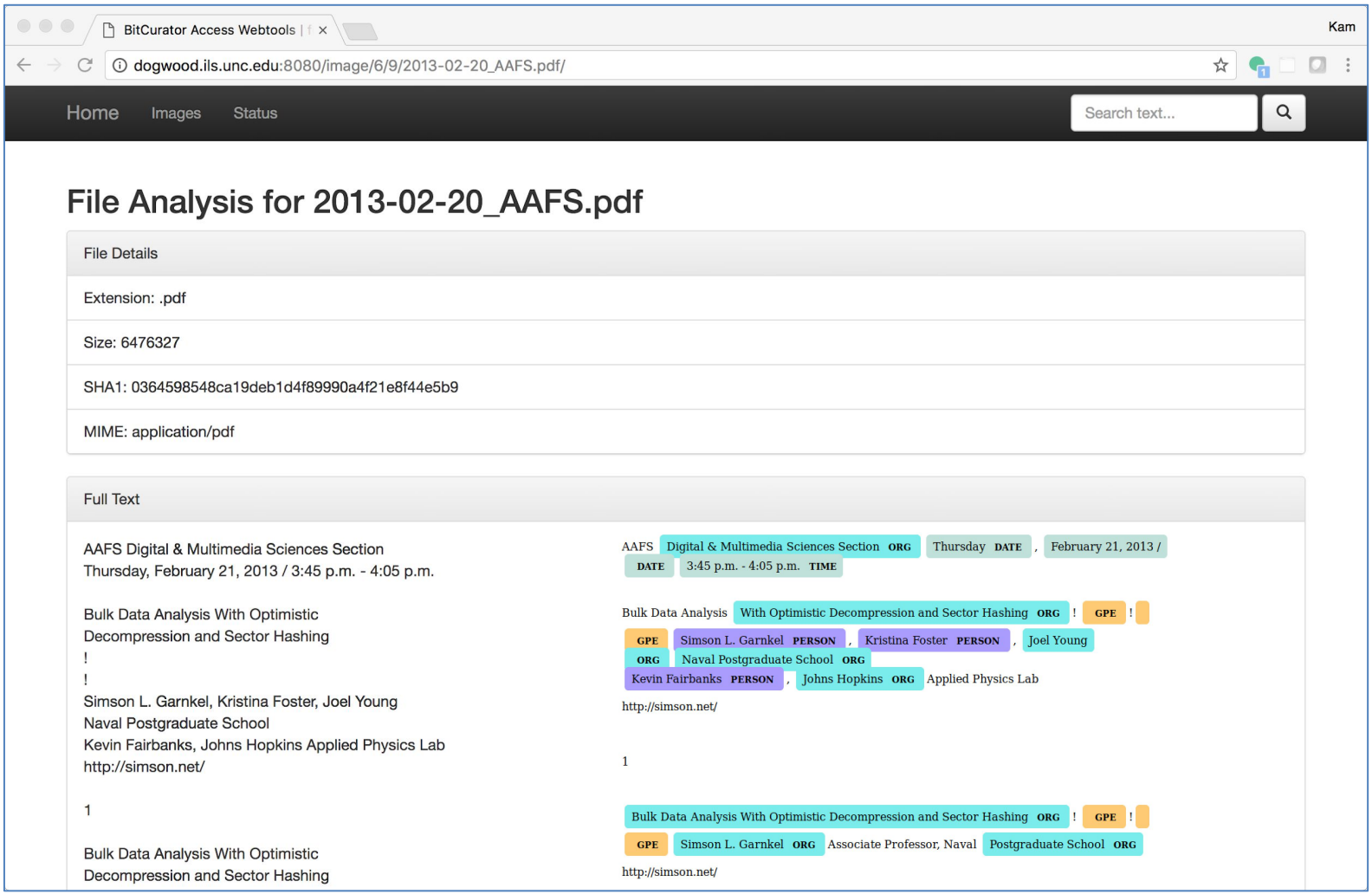
BitCurator Access Webtools



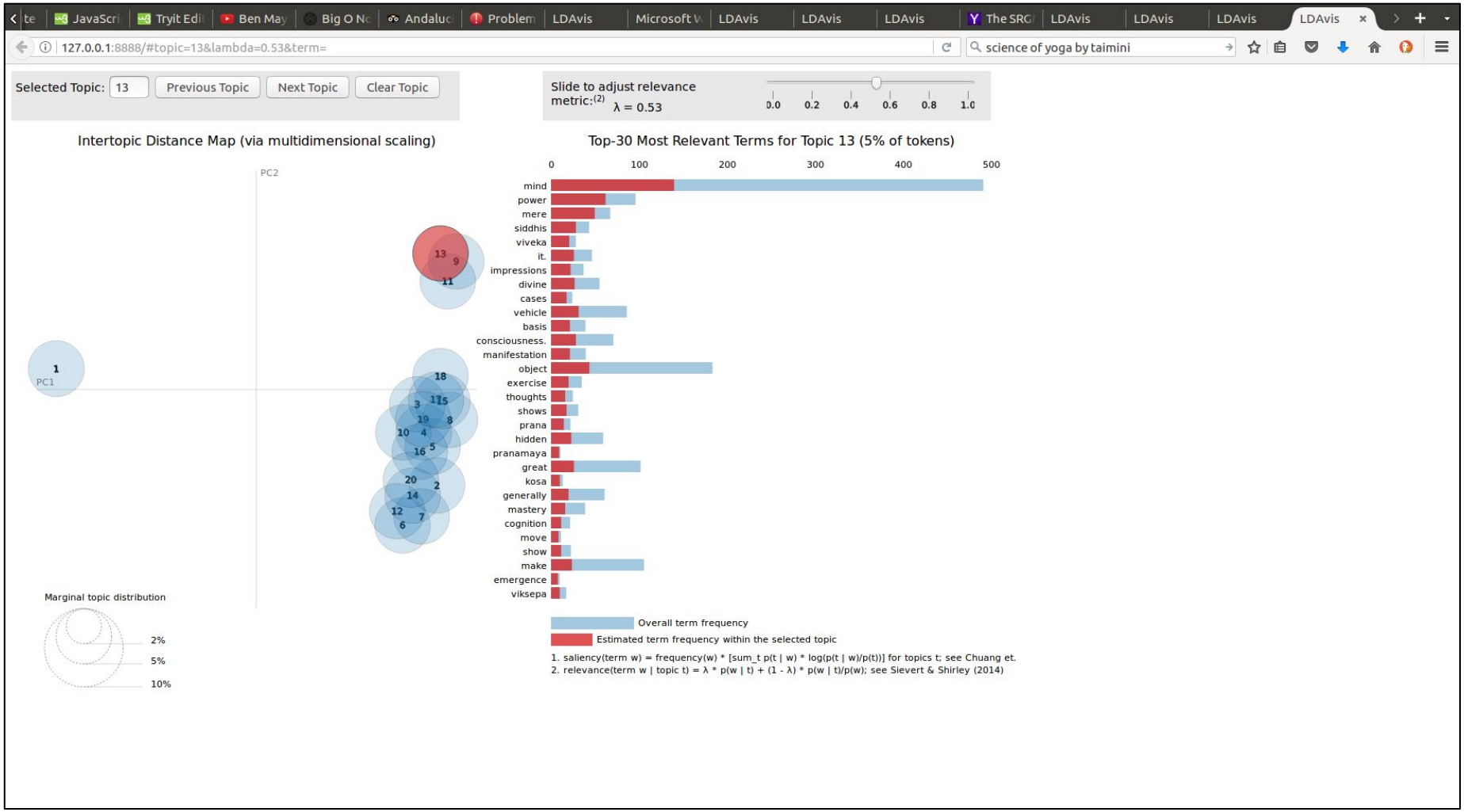
BitCurator Access Webtools



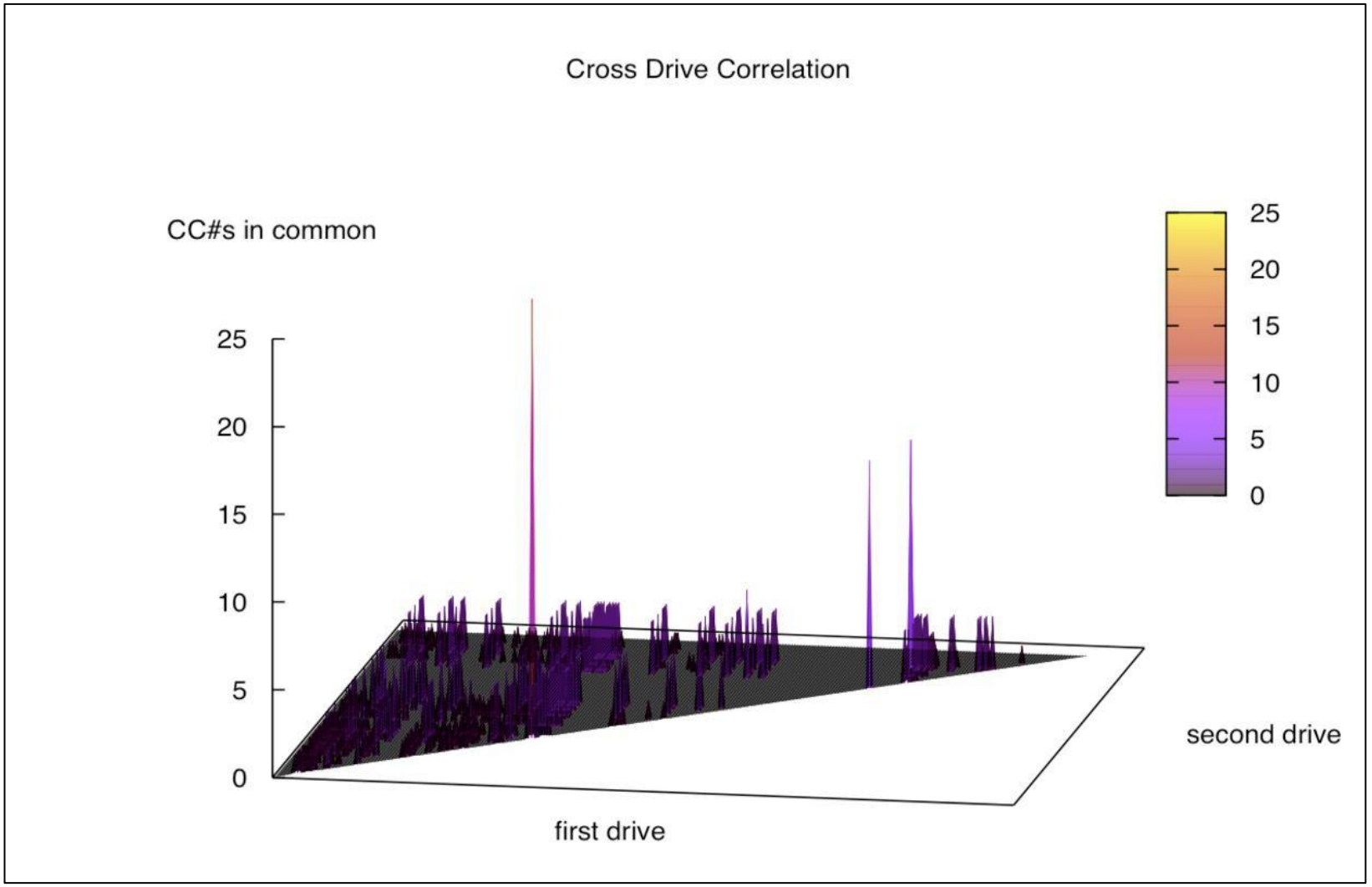
BitCurator Access Webtools



Topic Modeling in bitcurator-nlp-gentm (using pyLDAvis



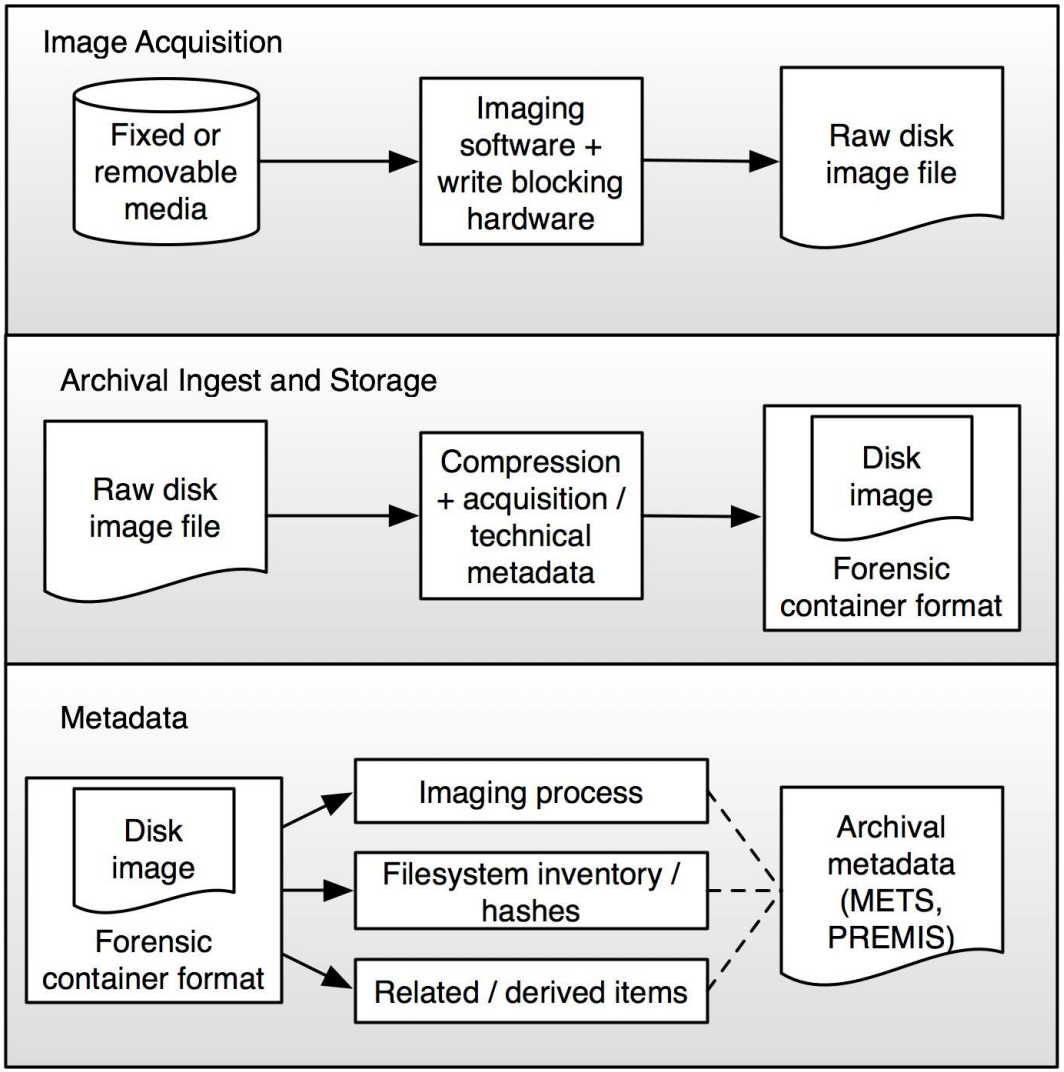
Forensic Feature Extraction and Cross-Drive Analysis



Source: Simson L. Garfinkel, “Forensic Feature Extraction and Cross-Drive Analysis,” Digital Forensics Research Workshop, August 15, 2006.

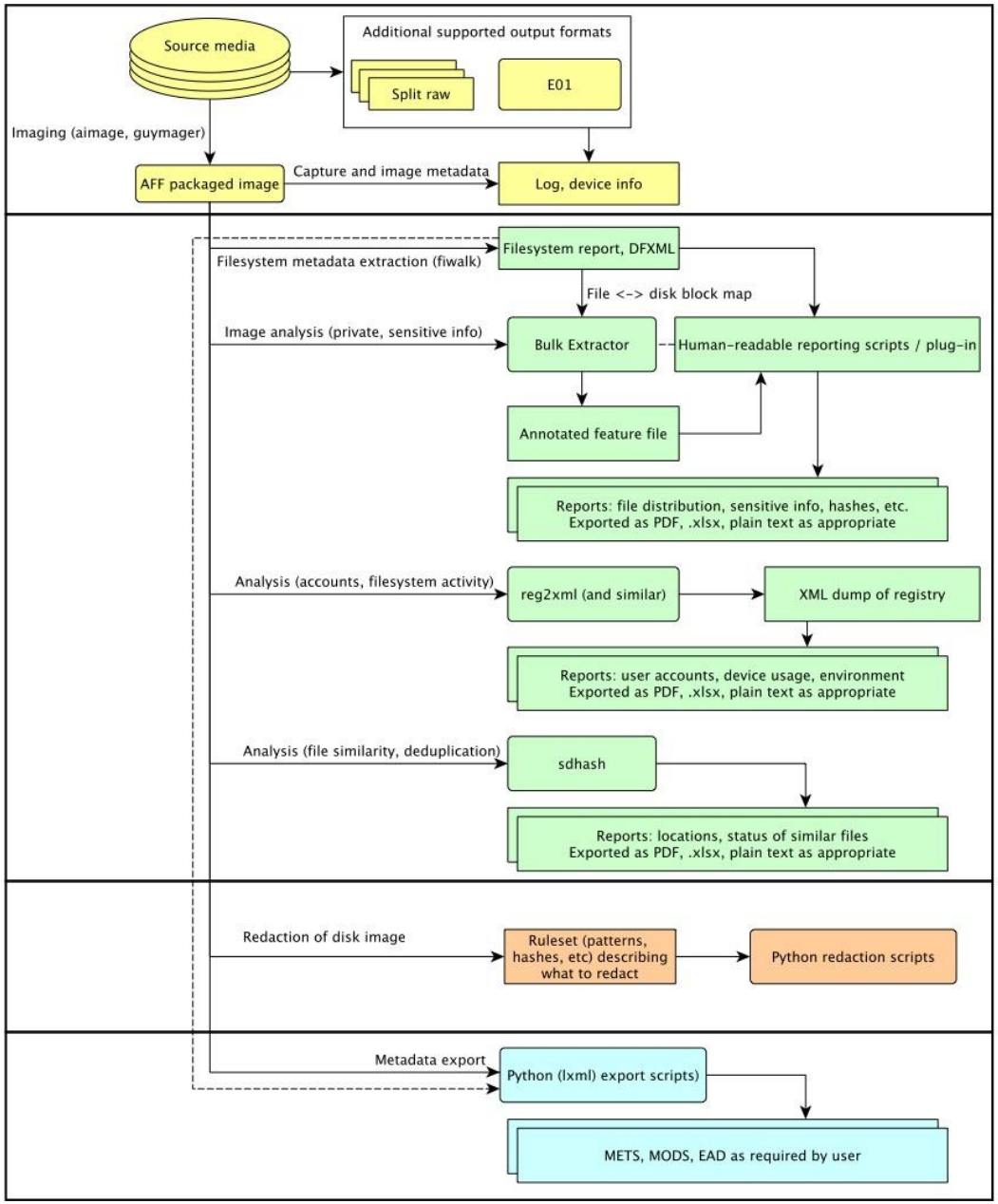
Incorporating digital forensics into archival workflows

Storage Media Acquisition and Handling Profile for Digital Repositories*



*Woods, Kam, Christopher A. Lee, and Simson Garfinkel. "Extending Digital Repository Architectures to Support Disk Image Preservation and Access." In *JCDL '11: Proceeding of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*, 57-66. New York, NY: ACM Press, 2011.

BitCurator-Supported Workflow



Acquisition

Reporting

Redaction

Metadata export

- **Acquisition**
- **Reporting**
- **Redaction**
- **Metadata Export**

See: https://bitcurator.net/

Five Sources of Workflow Examples

Martin J. Gengenbach, "'The Way We Do it Here': Mapping Digital Forensics Workflows in Collecting Institutions," A Master's Paper for the M.S. in L.S degree. August 2012.
https://web.archive.org/web/20170526011942/http://digitalcurationexchange.org/system/files/gengenbach-forensic-workflows-2012.pdf

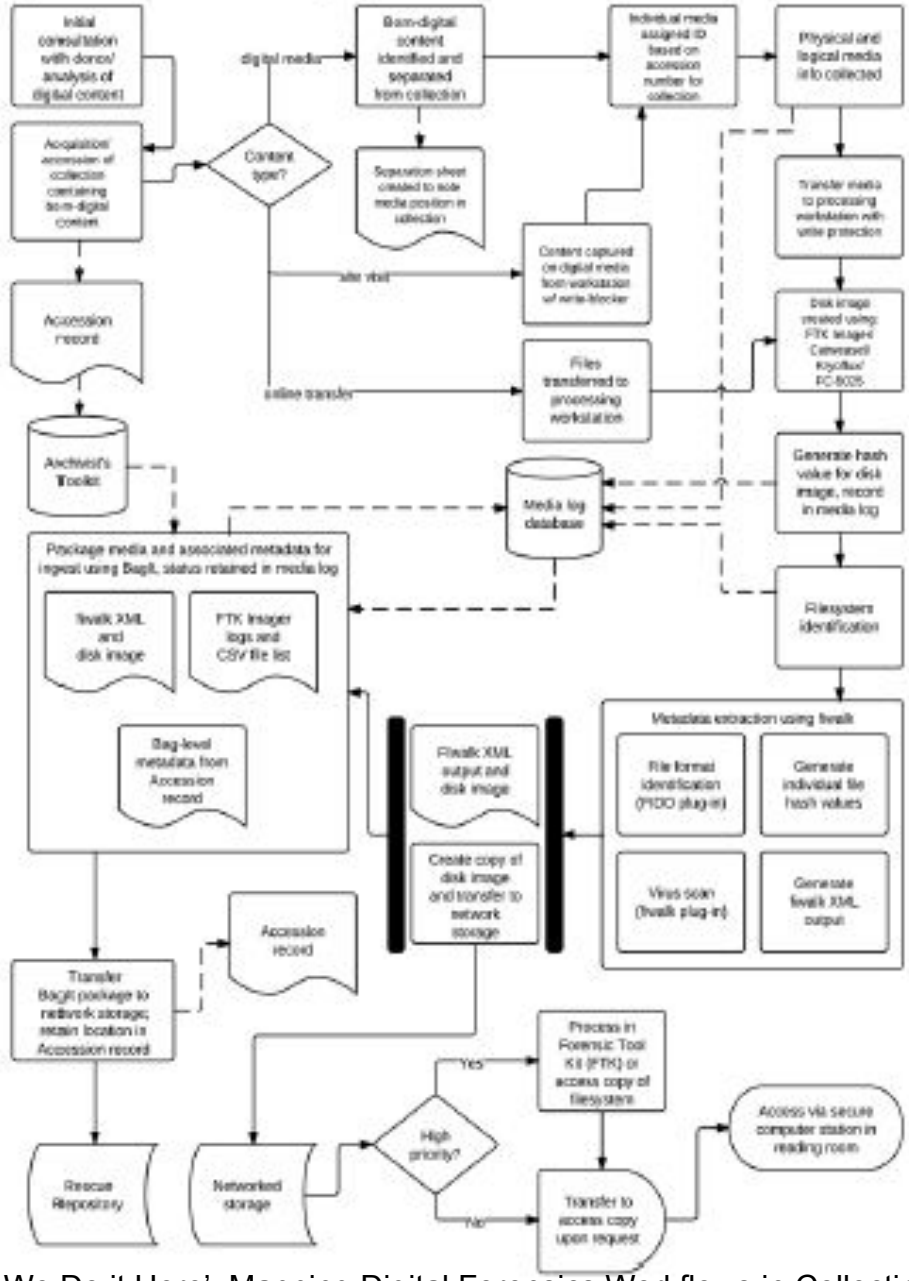
AIMS Work Group, "AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship," January 2012.
https://dcs.library.virginia.edu/files/2013/02/AIMS_final.pdf

Digital Sustainability Lab – Massachusetts Institute of Technology
https://web.archive.org/web/20160408225012/http://www.dpworkshop.org/sites/default/files/DCM-Pipeline_28Apr2015.pdf

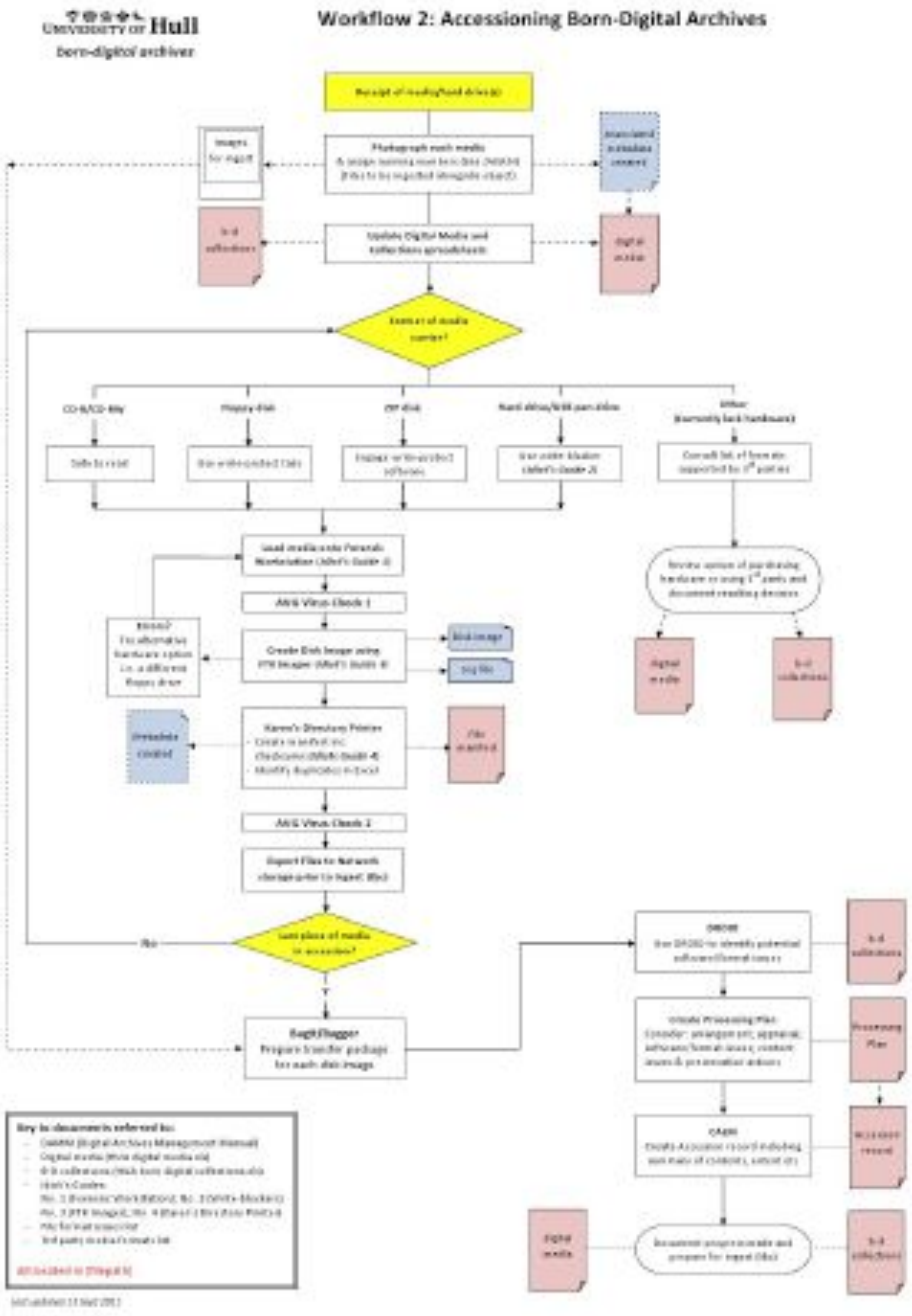
Workflows, BitCurator Consortium
https://bitcuratorconsortium.org/workflows

OSSArcFlow Project - https://educopia.org/research/ossarcflow

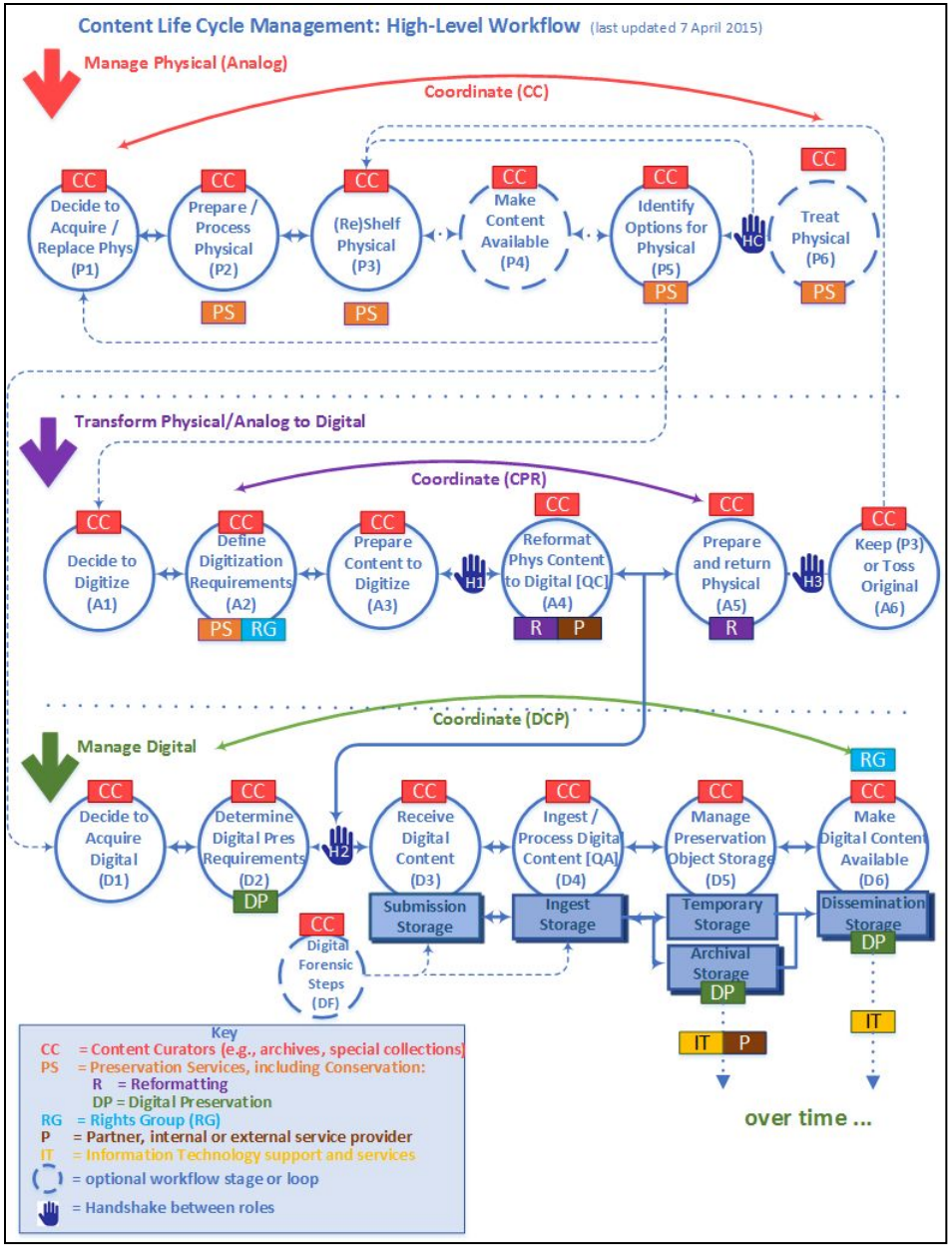
Figure 1. Beinecke Rare Book and Manuscript Library, Yale University



Martin J. Gengenbach, "'The Way We Do it Here': Mapping Digital Forensics Workflows in Collecting Institutions," A Master's Paper for the M.S. in L.S degree. August, 2012.



AIMS Work Group, “AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship,” January 2012.



Kari Smith, Massachusetts Institute of Technology.
https://web.archive.org/web/20160408225012/http://www.dpworkshop.org/sites/default/files/DCM-Pipeline_28Apr2015.pdf

BitCurator CONSORTIUM

Member Login | Search

About Us | Why Digital Forensics | Using BitCurator | Get Involved | Events

Using BitCurator

- Getting Started
- Documentation
- Workflows
- Videos

Not a member?

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

Workflow

The following workflows depict the step-by-step processes BitCurator Consortium members follow to acquire, process, describe, and store the born-digital materials in their collections. Most of these resources are only accessible to members. Learn more about the benefits of membership.

If you are interested in adding a workflow to our listing, please contact us.

Title	Contributor	Release Date
Processing Workflow	The University of Maryland, Libraries	2016 March 22
Princeton University Archives (Members Only)	Princeton University	2015 June 30
Penn State Born Digital (Members Only)	Penn State University	2014 July 29
Duke University Archives	Duke University	2012 August 12
Beineke Rare Books and Manuscripts Library	Yale University	2012 August 12
Maryland Institute for Technology in the Humanities	The University of Maryland, MITH	2012 August 12
University of North Carolina, Chapel Hill, Archives	University of North Carolina Chapel Hill, SILS	2012 August 12
University of Virginia Libraries	University of Virginia	2012 August 12
Yale University, Manuscripts and Archives	Yale University	2012 August 12

https://bitcuratorconsortium.org/workflows

EDUCOPIA INSTITUTE

Search

About Us | Communities | Research | Consulting | Events | Downloads | Contact Us

Research

- Continuing Education
  - Nexus
  - Mapping the Landscapes
- Digital Preservation
  - Aligning National Approaches to Digital Preservation (ANADP)
  - Chronicles
  - Distributed Digital Preservation (DDP)
  - Electronic Theses and Dissertations
  - Identifying Continuing Opportunities for National Collaboration (ICONC)
  - OSSArcFlow
  - News on the Margins
- Scholarly Communication
  - Chrysalis
  - Developing A Curriculum to Advance Library-Based Publishing
- Incubating Programs and Ideas

Digital Preservation | OSSArcFlow

OSSArcFlow

Contact:
Katherine Skinner

Additional Documents:
OSSArcFlow proposal

Investigating, Synchronizing, and Modeling a Range of Archival Workflows for Born-Digital Content

Project Abstract

The Educopia Institute, in collaboration with the University of North Carolina at Chapel Hill School of Information and Library Science (UNC SILS), LYRASIS, and Artefactual, Inc., are investigating, synchronizing, and modeling a range of workflows to increase the capacity of libraries and archives to curate born digital content. These archival workflows will incorporate three leading open source software (OSS) platforms—BitCurator, Archivematica, and ArchivesSpace—and the project will be designed to generate findings that can be generalizable to settings that are using other platforms and applications.

This project will significantly impact curation practices by increasing our understanding of how institutions of different sizes and types may engage in OSS tool integration and workflow development. Our findings will be used to support a broad range of libraries and archives actively collecting and curating digital content. The knowledge gained by working with multiple institutions of different types and sizes will also broaden field-wide understanding of curation approaches and priorities, and how those impact the use of tools and capabilities in Archivematica, ArchivesSpace, and BitCurator. We expect the empirical findings about institutional needs, as well as formal workflow models, to contribute to digital curation research literature.

This project has been generously funded by the Institute of Museum and Library Services.

Project Outputs

Digital Dossiers

https://educopia.org/research/ossarcflow

Challenges

- Incorporation into LAM workflows, e.g. metadata conventions, connections to collection management systems
- Obsolete storage media and filesystems
- Dealing with large, internally complex data files
- Provision of public access
- Defining and implementing ethical commitments

SWAT (Software and Workstations for Antiquated Technology) Sites

"A community-based approach would use SWAT sites wherein a few self selected institutions acquire and maintain the gear and expertise to read data and transfer content from particular types of obsolete media. The SWAT sites would provide transfer services for institutions that don't have the capacity to read a particular medium (or the SWAT sites might become the likely places to deposit particular types of media)."

Erway, Ricky. "Swatting the Long Tail of Digital Media: A Call for Collaboration." Dublin, Ohio: OCLC Research, 2012.
http://www.oclc.org/research/publications/library/2012/2012-08.pdf

See also:
Ricky Erway and Ben Goldman, "Agreement Elements for Outsourcing Transfer of Born Digital Content," August 2014,
http://oclc.org/research/publications/library/2012/2012-06r.html#agreement

Legal and Ethical Issues

Ethical Dilemmas

- What ethical dilemmas related to born-digital materials have you faced or do you expect to face?
- What would the competing interests or values be?
- How would you decide?

Donor Agreements

- Donor agreements (as of 2012) tend not to address the kinds of issues raised in this class*
- What are the most important issues to resolve with creators/donors?
- What's the right level of detail to include in donor agreements and discussions with potential donors?

*Matthew J. Farrell, "Born-Digital Objects in the Deeds of Gift of Collecting Repositories: A Latent Content Analysis," A Master's Paper for the M.S. in L.S degree. July 2012, https://cdr.lib.unc.edu/indexablecontent?id=uuid:385c4fd9-a403-4ba3-85ac-2ea128400ddb&ds=DATA_FILE

Specific Guidance Documents

- Redwine, Gabriela, Megan Barnard, Kate Donovan, Erika Farr, Michael Forstrom, Will Hansen, Jeremy Leighton John, Nancy Kuhl, Seth Shaw, and Susan Thomas. "Born Digital: Guidance for Donors, Dealers, and Archival Repositories." Washington, DC: Council on Library and Information Resources, 2013.
- Nelson, Naomi L, et al. "Gift/Purchase Agreements." In *Managing Born-Digital Special Collections and Archival Materials*, 122-126. SPEC Kit 329. Washington, DC: Association of Research Libraries, 2012. [Includes donor agreements and policies from Duke University, Bentley Historical Library, and Beinecke Rare Book and Manuscript Library]
- Pyatt, Timothy D. "Deed of Gift Addenda for Collections with Electronic Records." Pennsylvania State University. 2012.

A Guide to Deeds of Gift – Society of American Archivists*

Text added in 2013:

"Be aware that any digital materials that you donate, including computers, computer disks, and other digital storage media, may contain passwords, web browsing history, other users' files, and copies of seemingly deleted files. Whether or not these files are apparent to researchers will depend on the initial method of transfer and on the repository's access policies and procedures for handling digital material, which may change over time as technology evolves. Discuss any concerns about deleted content with the archivist or curator."

*https://www2.archivists.org/publications/brochures/deeds-of-gift

What does it mean for electronically stored information (ESI) to be “accessible”?

"The person responding need not provide discovery of electronically stored information from sources that the person identifies as **not reasonably accessible because of undue burden or cost**." (Rule 45 (d)(1)(D)) (emphasis added)

Judge Shira Scheindlin:

"[t]he more information there is to discover, the more expensive it is to discover all the relevant information until, in the end, 'discovery is not just about uncovering the truth, but also about how much of the truth the parties can afford to disinter.' "

(Zubulake I, 217 F.R.D. at 311 (quoting Rowe Entm't, Inc. v. William Morris Agency, Inc., 205 F.R.D. 421, 423 (S.D.N.Y. 2002)).

Seven-Factor Test from *Zubulake v. UBS Warburg*

1. extent to which the request is specifically tailored to discover relevant information
2. availability of such information from other sources
3. total cost of production, compared to the amount in controversy
4. total cost of production, compared to the resources available to each party
5. relative ability of each party to control costs and its incentive to do so
6. importance of the issues at stake
7. relative benefits to the parties of obtaining the information

(217 F.R.D. at 322)

Zubulake's Five Categories of ESI (Most to Least Accessible)*

- Active, online data
- Near-line data
- Offline storage
- Backup tapes
- Erased, fragmented or damaged data

*See: Lange, Michele C. S., and Kristin M. Nimsger. *Electronic Evidence and Discovery: What Every Lawyer Should Know Now*. 2nd ed. Chicago, IL: Section of Science & Technology Law American Bar Association, 2009. p.75.

Magistrate Judge John Facciola:

"…I am anything but certain that I should permit a party who has failed to preserve accessible information without cause to then complain about the inaccessibility of the only electronically stored information that remains"

(*Disability Rights Council of Greater Wash. v. Wash. Metro. Transit Auth.*, 242 F.R.D. 139 (D.D.C. 2007)).

Rights to Control Information

- Most frequently discussed in library lit is copyright
- Claims can extend far beyond intellectual property rights, as defined by law
- Cultural property, replevin and repatriation
- Right to privacy
- Protection of human subjects in research
- Privileged or protected information (e.g. client-attorney, healthcare, social services, library circulation, source – journalist)
- Right to publicity – individual's protection from unauthorized commercial use of her name, persona, or likeness
- Prevention of misappropriation (including plagiarism)

"If a forensic examiner has complete confidence in his/her conclusions, this is usually an indication that he/she is missing something – there is always uncertainty and all assertions should be qualified accordingly..."

Casey, Eoghan. "Error, Uncertainty, and Loss in Digital Evidence." *International Journal of Digital Evidence* 1, no. 2 (2002).

"Investigators cannot, in general, directly observe digital data and instead they can only observe the data displayed on a monitor or other output device, which is driven by various types of hardware and software. Because the **observation of the data is indirect**, a hypothesis must be formulated that the actual data is equal to the observed data. Testing this hypothesis requires that the hardware and software being used are accurate and reliable. Hypotheses also need to be formulated about the data abstractions that exist and the previous states and events that occurred."*

*Carrier, Brian D. "A Hypothesis-Based Approach to Digital Forensic Investigations." Doctoral Dissertation, Purdue University, 2006. p.11 (emphasis added).

Examples of Potentially Useful Inferences (that could be wrong)

- Name embedded in a MS Word file is the document's author
- Given IP address identifies an individual
- Presence of email addresses on different hard drives indicate correspondence patterns between individuals
- Many common MD5 values across storage locations indicate sharing of files across those locations (context-based filtering can help to address this)
- Last modified date indicates when a document was finalized
- Parts of a page available through the WayBack Machine for a given date represent the parts of the page as available on that date

The "Keyboard Dilemma"

"Even if a document can be traced to a particular computer and/or IP address, how can we identify who was actually at the keyboard composing the document? It is a particular problem in environments where multiple users may have access to the same computer or when users do not have to authenticate themselves to access a particular account."

Chaski, Carole. "The Keyboard Dilemma and Authorship Identification." In *Advances in Digital Forensics III: IFIP International Conference on Digital Forensics, National Center for Forensic Science, Orlando, Florida, January 28-January 31, 2007*, edited by Philip Craiger and Sujeet Shenoi. New York, NY: Springer, 2007. p.133.

Shared Computer Use in the Home*

*Frohlich, David, and Robert Kraut. "The Social Context of Home Computing." In *Inside the Smart Home*, edited by Richard Harper, 127-62. London: Springer, 2003.

Ethics Questions To Consider

1. When acquiring a disk as part of a collection, should you create a bit-level image of the disk, in order to ensure the potential to recreate not only the "payload" data of files but also various forms of information within and below the filesystem?
2. Should you retain "hidden" data in a Word document or only retain what you assume to be the text that the author intended?
3. You're responsible for managing a Microsoft Outlook .pst file over time (including saved and sent messages, calendar items, draft and deleted messages, address book, and viruses). Should you retain the whole .pst file or extract messages and attachments that were sent and received?
4. If a collection documents the life of an individual, how would you determine the appropriate scope for collecting information associated with that person's online presence (e.g. postings, affiliations, profiles, micro-contributions)?
5. If your institution routinely "normalizes" submitted files into designated file formats, are you obligated to ensure that the normalization doesn't violate the intentions of creator or other interested stakeholders? If so, what does this obligate you to do specifically?
6. Someone cropped a set of images in order to remove sensitive parts, but the images still have pixel information and embedded thumbnail reflecting the "removed" parts. How should you approach the management of the images?

Lessons and Insights

- Digital forensics has arrived for archival processing
- Introduction of digital forensics doesn't dictate specific policies or practices
- The disk image is the cornerstone of many forensics methods
- "Taking bitstreams seriously" can have major advantages
- Disk images afford new access scenarios

To Learn More About Available Software

Forensics Wiki. https://forensicswiki.xyz/page/Main_Page

BitCurator Environment. https://bitcurator.net

BitCurator Software Overview. https://bitcurator.github.io/

Community Owned digital Preservation Tool Registry (COPTR)
https://coptr.digipres.org/Main_Page

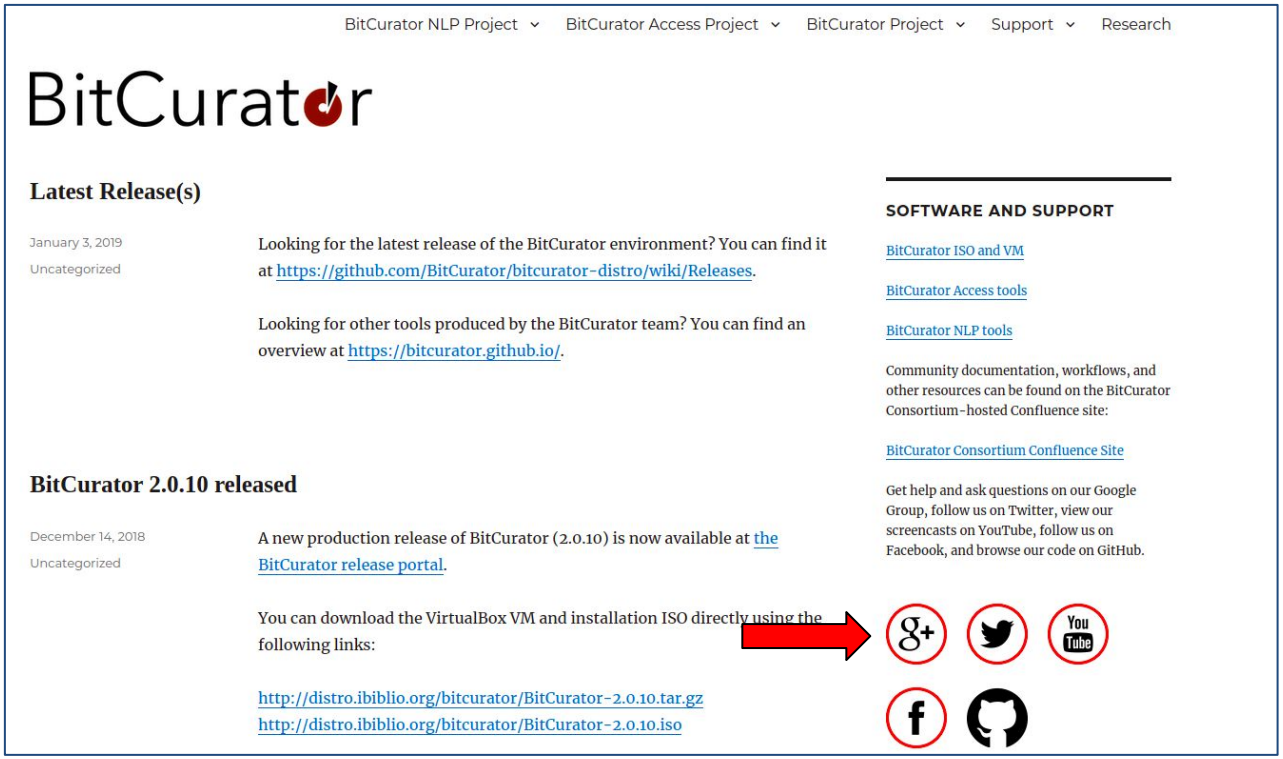
Information Guides on Tools for Electronic Records. Minnesota State Archives.
https://www.mnhs.org/preserve/records/electronicrecords/erpreserveres.php

Lifecycle Tools for Archival Email Stewardship.
https://docs.google.com/spreadsheets/d/1V1N22xnr5e0EbDlZWx58bjYO6rkrMrYH9wGX9-CK8c4/

Tools for processing, managing, and preserving electronic records. University of Minnesota.
https://www.lib.umn.edu/dp/guides

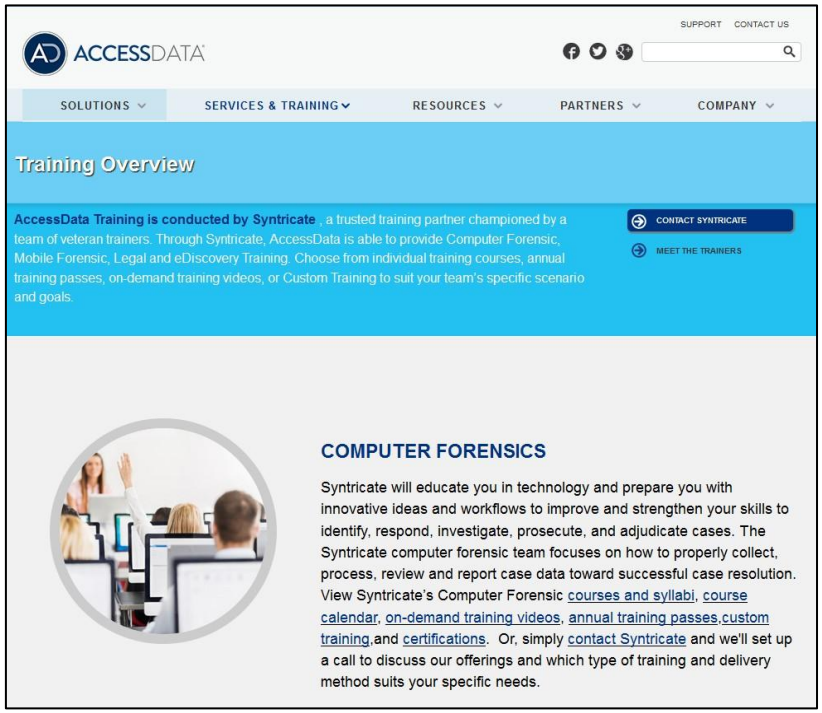
Online Forums

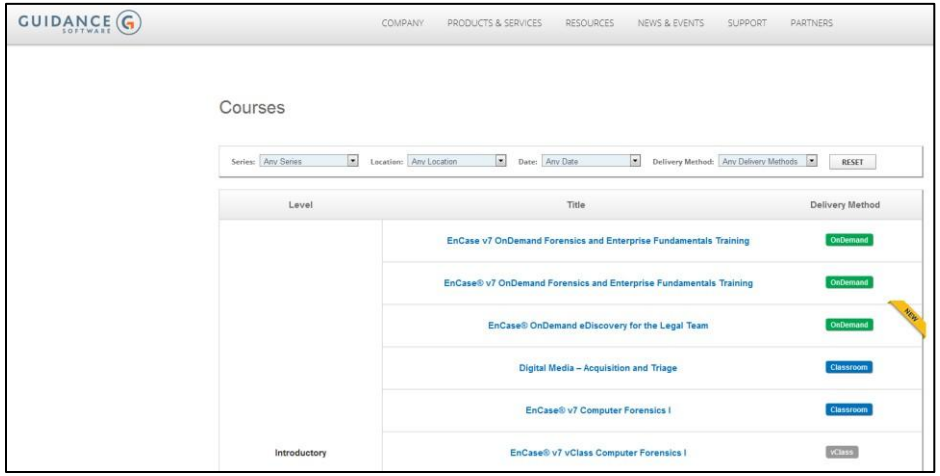
BitCurator User Group
https://groups.google.com/forum/#!forum/bitcurator-users



Digital Curation List
https://groups.google.com/forum/#!forum/digital-curation

Further Education





Thank you!

Go forth and curate the bits!

Most resources from the BitCuratorEdu project are intentionally left with basic formatting and without project branding. We encourage educators, practitioners, and students to adapt these materials as much as needed and share them widely.

The BitCuratorEdu project is a three-year effort funded by the Institute of Museum and Library Services (IMLS) to study and advance the adoption of digital forensics tools and methods in libraries and archives through professional education efforts. This project is a partnership between Educopia Institute and the School of Information and Library Science at the University of North Carolina at Chapel Hill, along with the Council of State Archivists (CoSA) and several Masters-level programs in library and information science.