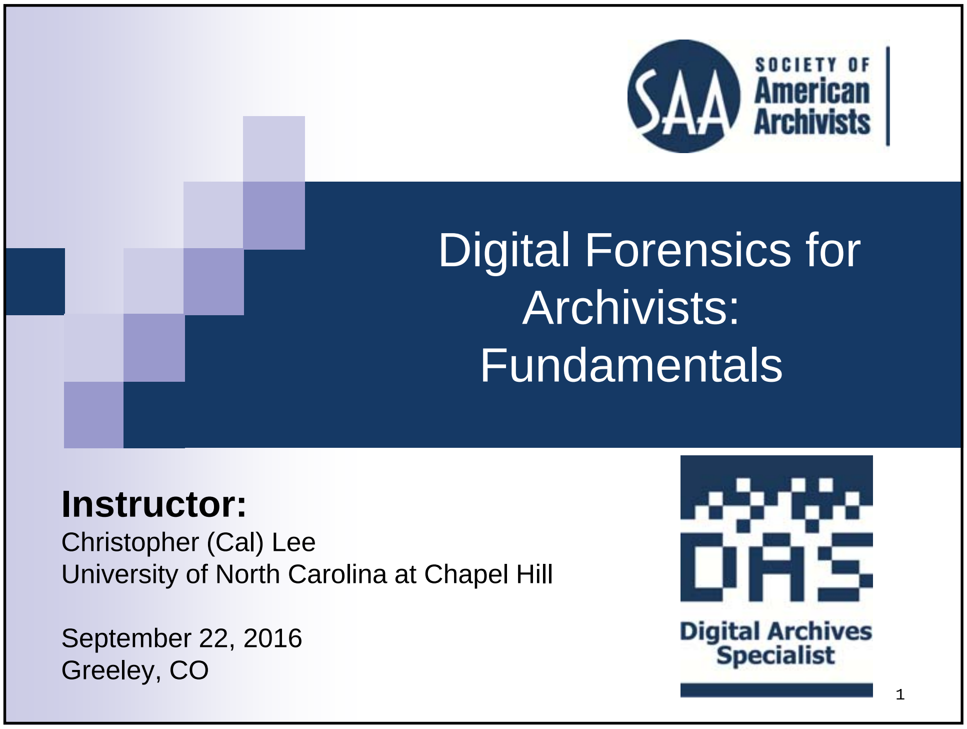


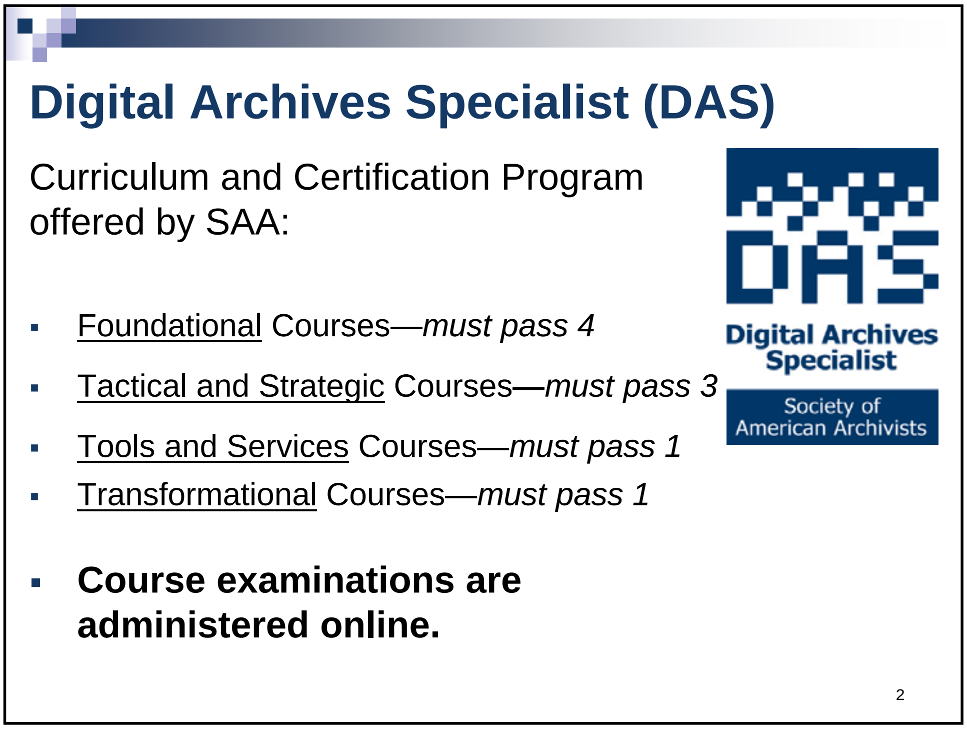
Society of American Archivists

Digital Forensics for Archivists: Fundamentals

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

September 22, 2016
Greeley, CO

Digital Archives Specialist



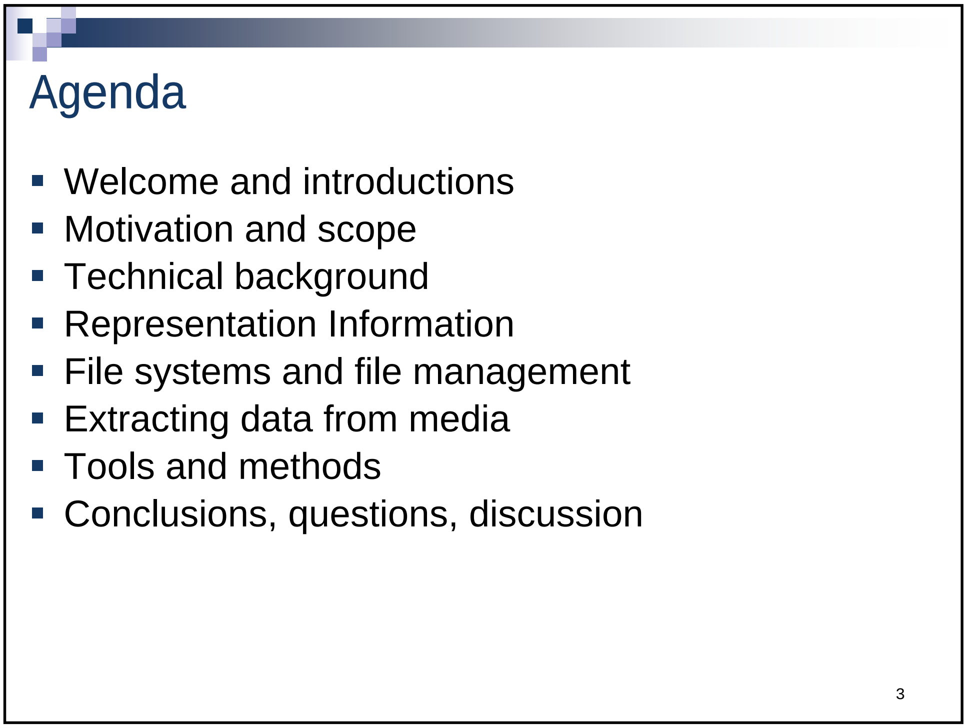
Digital Archives Specialist (DAS)

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.**

Digital Archives Specialist
Society of American Archivists



Agenda

- Welcome and introductions
- Motivation and scope
- Technical background
- Representation Information
- File systems and file management
- Extracting data from media
- Tools and methods
- Conclusions, questions, discussion

Welcome and Introductions

Motivation and Scope

Applying Digital Forensics to Archival Work

Many archivists know how to process this stuff:

Source: The Processing Table: Reflections on a manuscripts internship at the Lilly Library.
https://processingtable.wordpress.com/tag/archival-processing/

How about processing this stuff?

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

Same Goals as When Acquiring Analog Materials

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

Same Fundamental Archival Principles Apply

Provenance	• 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.

But you might need some of this stuff:

Luckily, there are a lot of people with expertise in using such tools in places like this:

El Paso County Sheriff’s Office (Colorado)

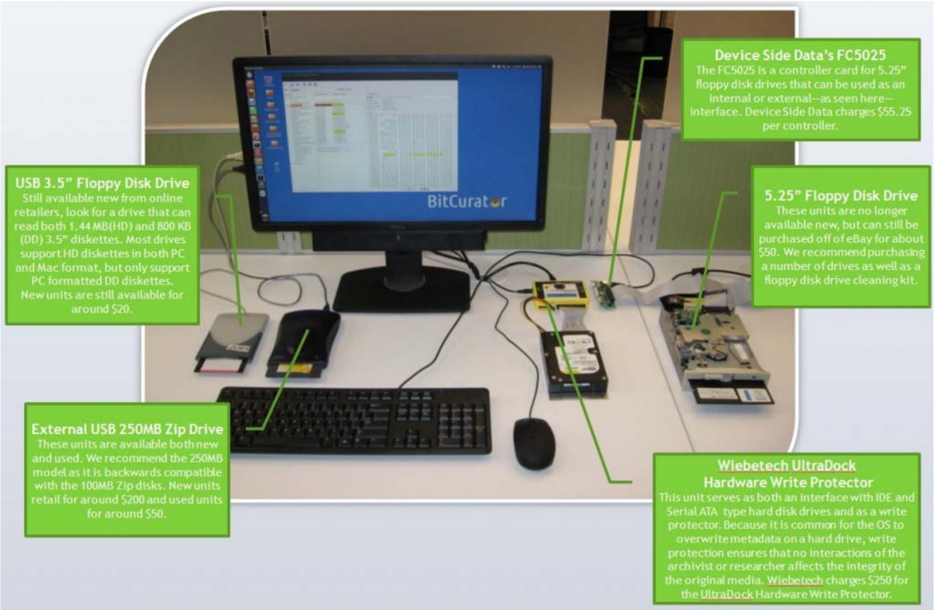
http://shr.elpasoco.com/Law+Enforcement+Bureau/Investigations+Division/Computer+Crime+Lab.htm

Here’s what it looks like in libraries and archives:

Stanford University Libraries and Academic Information Resources (SULAIR)

British Library, London

UNC School of Information and Library Science



Device Side Data's FC5025
The FC5025 is a controller card for 5.25" floppy disk drives that can be used as an internal or external—as seen here—interface. Device Side Data charges $55.25 per controller.

USB 3.5" Floppy Disk Drive
Still available new from online retailers, look for a drive that can read both 1.44 MB(HD) and 800 KB (DD) 3.5" diskettes. Most drives support HD diskettes in both PC and Mac format, but only support PC formatted DD diskettes. New units are still available for around $20.

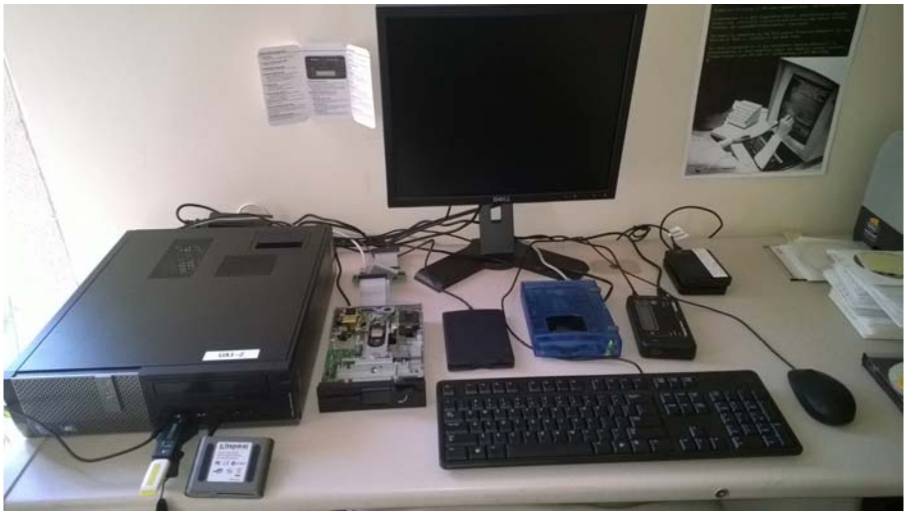
5.25" Floppy Disk Drive
These units are no longer available new, but can still be purchased off of eBay for about $50. We recommend purchasing a number of drives as well as a floppy disk drive cleaning kit.

External USB 250MB Zip Drive
These units are available both new and used. We recommend the 250MB model as it is backwards compatible with the 100MB Zip disks. New units retail for around $200 and used units for around $50.

Wiebetech UltraDock Hardware Write Protector
This unit serves as both an interface with IDE and Serial ATA type hard disk drives and as a write protector. Because it is common for the OS to overwrite metadata on a hard drive, write protection ensures that no interactions of the archivist or researcher affects the integrity of the original media. Wiebetech charges $250 for the UltraDock Hardware Write Protector.

http://www.bitcurator.net/building-a-digital-curation-workstation-with-bitcurator-update/

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



http://practicaltechnologyforarchives.org/issue2_goldman/

Motivation

- Archivists are often responsible for acquiring or helping others access materials on removable storage media
- Information is often not packaged nor described as one would hope
- Information professionals must extract whatever useful information resides on the medium, while avoiding the accidental alteration of data or metadata

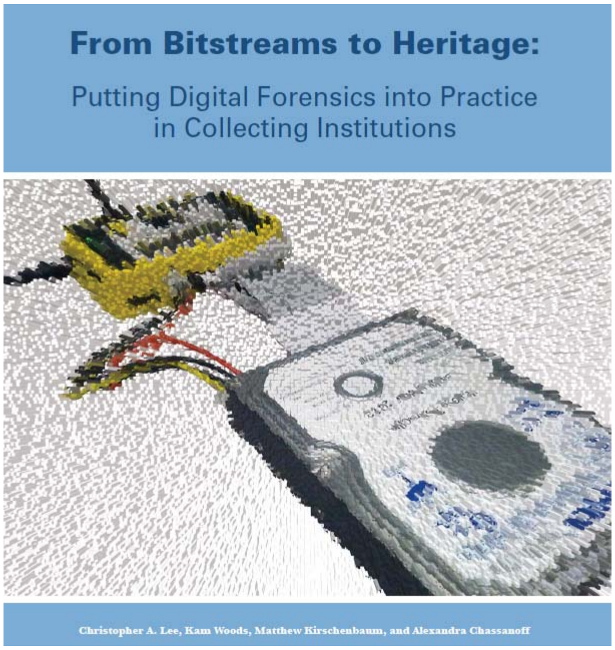
Digital Forensics Can Help Archivists to Fulfill their Principles

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

Applying Digital Forensics to Digital Collections – Previous Work*

- Ross and Gow (1999) - potential relevance of advances in data recovery and digital forensics to collecting institutions
- More recently - active stream of literature related to use of forensic tools and methods for digital collections, including activities at the British Library, National Library of Australia and Indiana University
- PERPOS (Georgia Tech) – has applied data capture and extraction to US presidential materials
- "Computer Forensics and Born-Digital Content in Cultural Heritage Collections" - symposium and report (2010)
- Born Digital Collections: An Inter-Institutional Model for Stewardship (AIMS) - framework for the stewardship of born-digital materials, including digital forensics methods
- Digital Records Forensics project - has articulated connections between the concepts of digital forensics and archival science

*See citations in:
http://ica2012.ica.org/files/pdf/Full%20papers%20upload/ica12Final00290.pdf



From Bitstreams to Heritage:

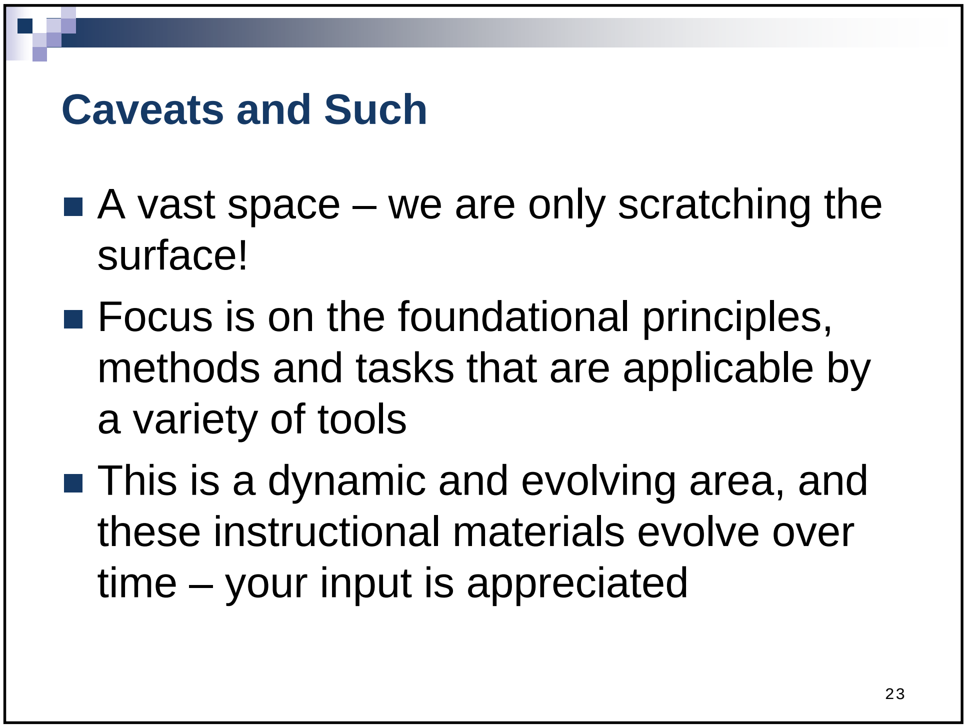
Putting Digital Forensics into Practice in Collecting Institutions

Christopher A. Lee, Kam Woods, Matthew Kirschenbaum, and Alexandra Chassanoff

http://www.bitcurator.net/docs/bitstreams-to-heritage.pdf

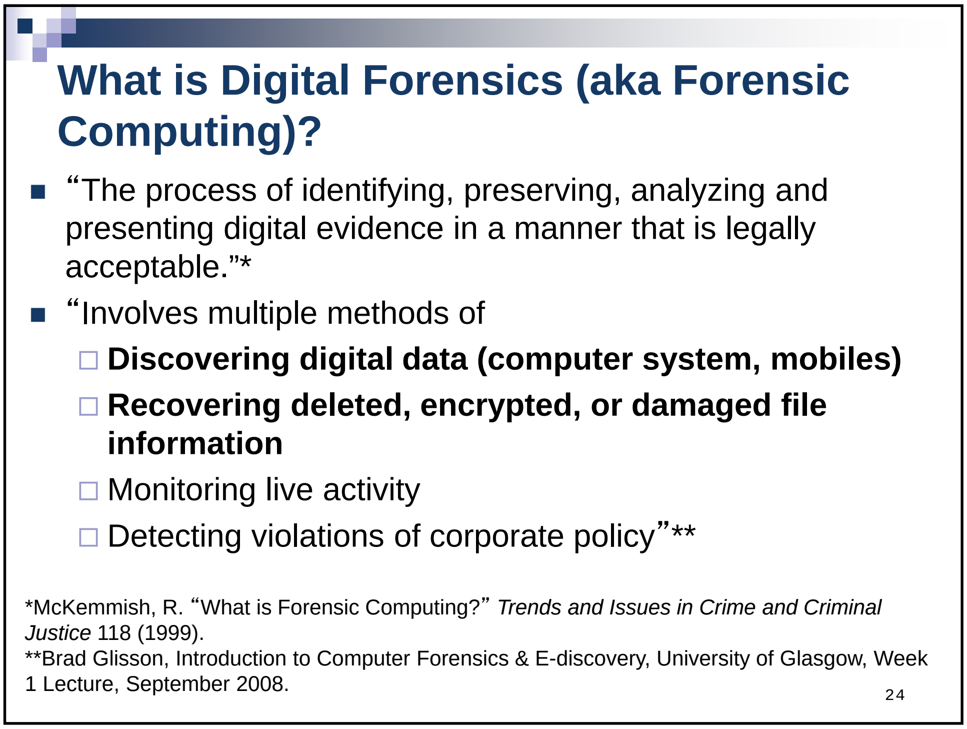
What this Course Covers

- Computational operations
- Layers of hardware and software that allow bitstreams on digital media to be read as files
- Roles and relationships of these layers
- Tools and techniques for ensuring completeness and evidential value of data



Caveats and Such

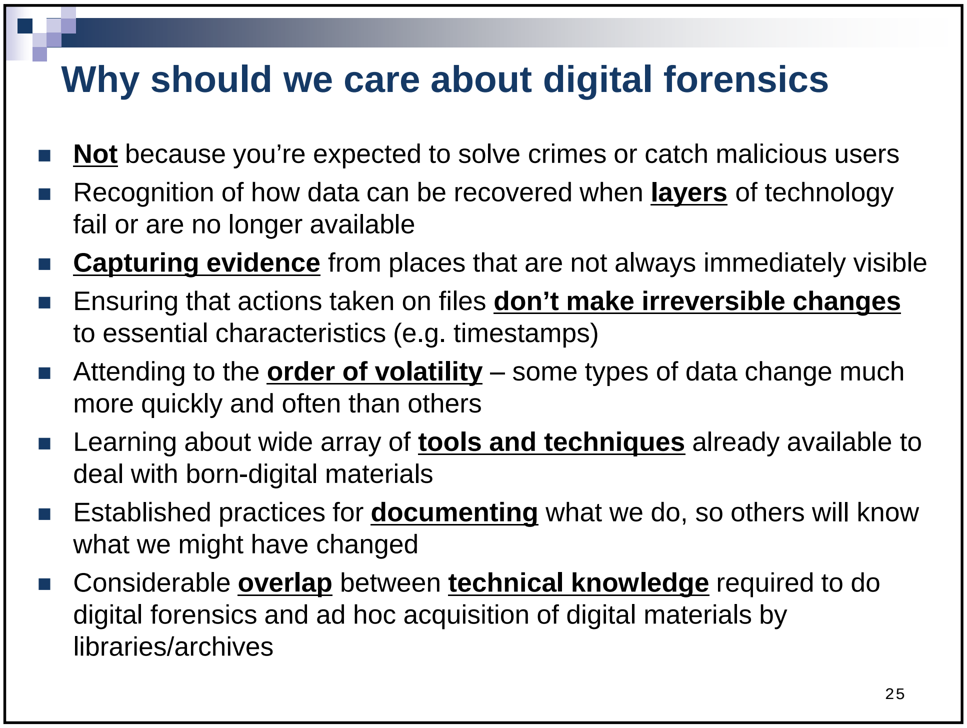
- A vast space – we are only scratching the surface!
- Focus is on the foundational principles, methods and tasks that are applicable by a variety of tools
- This is a dynamic and evolving area, and these instructional materials evolve over time – your input is appreciated



What is Digital Forensics (aka Forensic Computing)?

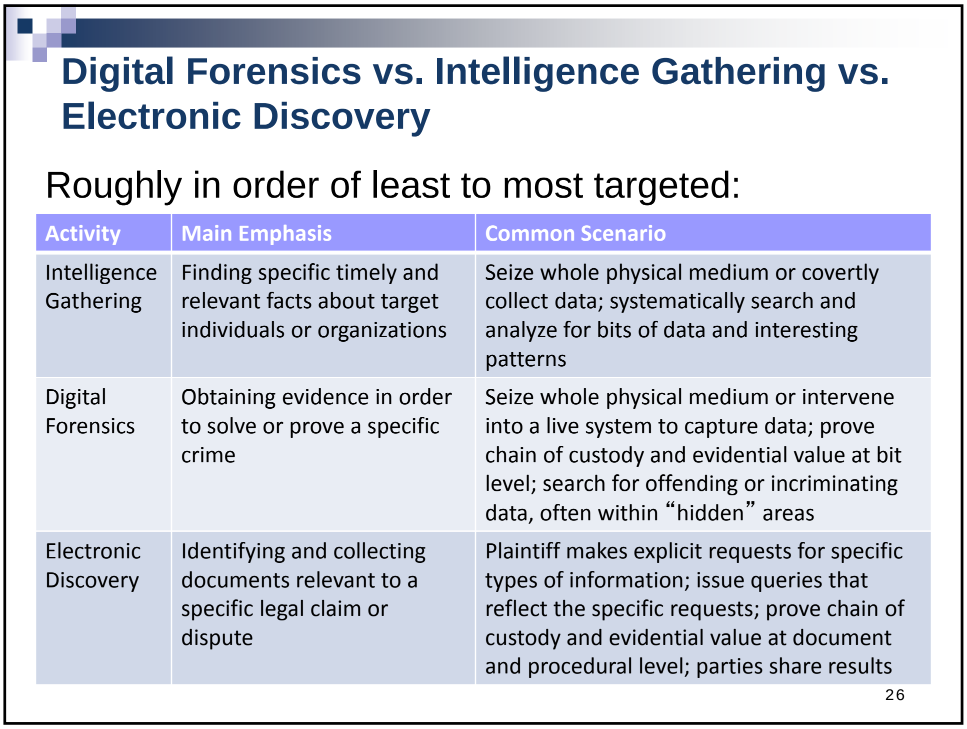
- “The process of identifying, preserving, analyzing and presenting digital evidence in a manner that is legally acceptable.”*
- “Involves multiple methods of
  - □ **Discovering digital data (computer system, mobiles)**
  - □ **Recovering deleted, encrypted, or damaged file information**
  - □ Monitoring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

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



Why should we care about digital forensics

- **Not** because you're expected to solve crimes or catch malicious users
- Recognition of how data can be recovered when **layers** of technology fail or are no longer available
- **Capturing evidence** from places that are not always immediately visible
- Ensuring that actions taken on files **don't make irreversible changes** to essential characteristics (e.g. timestamps)
- Attending to the **order of volatility** – some types of data change much more quickly and often than others
- Learning about wide array of **tools and techniques** already available to deal with born-digital materials
- Established practices for **documenting** what we do, so others will know what we might have changed
- Considerable **overlap** between **technical knowledge** required to do digital forensics and ad hoc acquisition of digital materials by libraries/archives



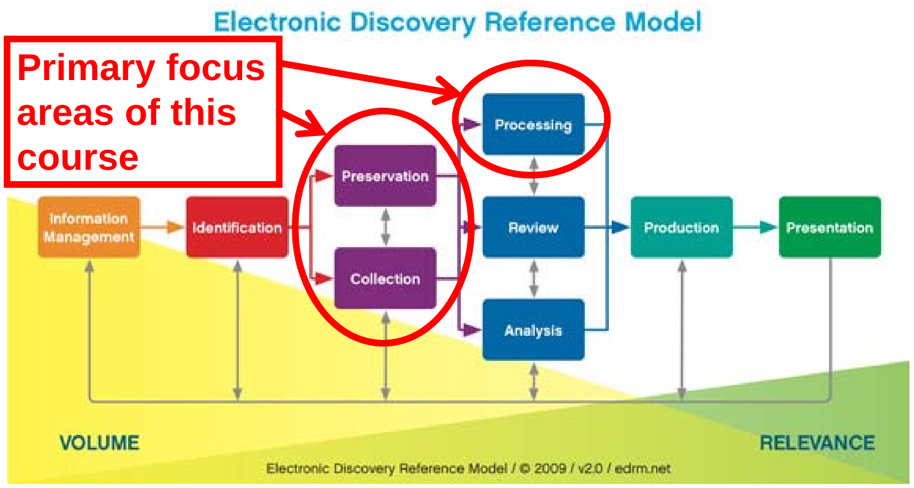
Digital Forensics vs. Intelligence Gathering vs. Electronic Discovery

Roughly in order of least to most targeted:

Activity	Main Emphasis	Common Scenario
Intelligence Gathering	Finding specific timely and relevant facts about target individuals or organizations	Seize whole physical medium or covertly collect data; systematically search and analyze for bits of data and interesting patterns
Digital Forensics	Obtaining evidence in order to solve or prove a specific crime	Seize whole physical medium or intervene into a live system to capture data; prove chain of custody and evidential value at bit level; search for offending or incriminating data, often within "hidden" areas
Electronic Discovery	Identifying and collecting documents relevant to a specific legal claim or dispute	Plaintiff makes explicit requests for specific types of information; issue queries that reflect the specific requests; prove chain of custody and evidential value at document and procedural level; parties share results

Common Digital Forensics Scenarios

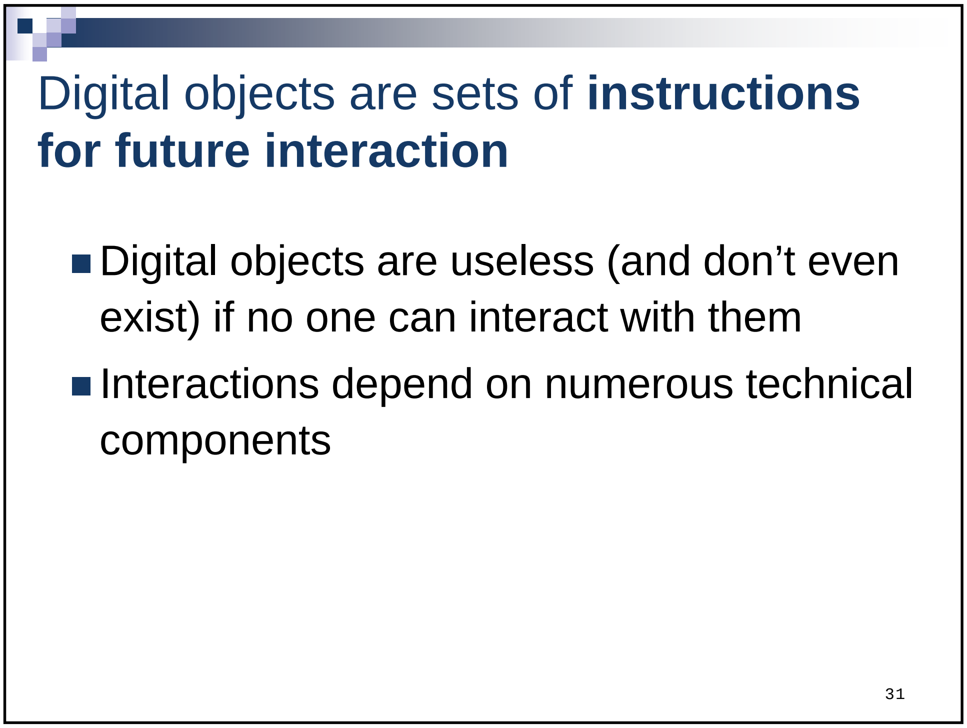
- Evidence seized from home/office of “person of interest” in a criminal investigation (dead forensics)
- Response to system security breach, to determine what was done, by whom and how (live forensics)



Primary focus areas of this course

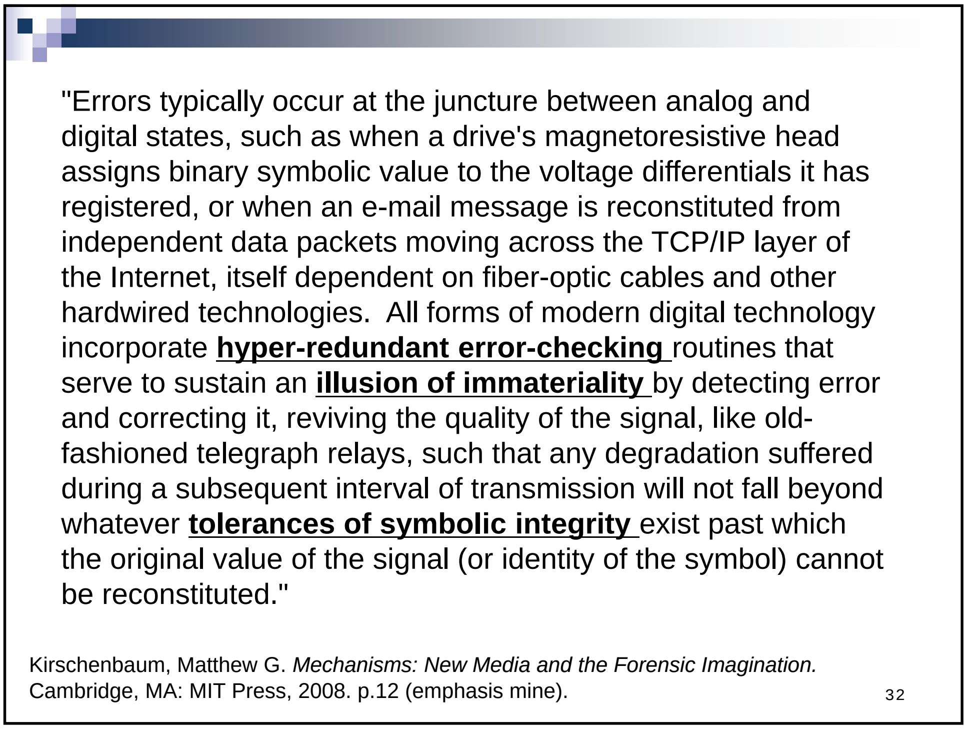
Technical Background

Nature of Digital Materials



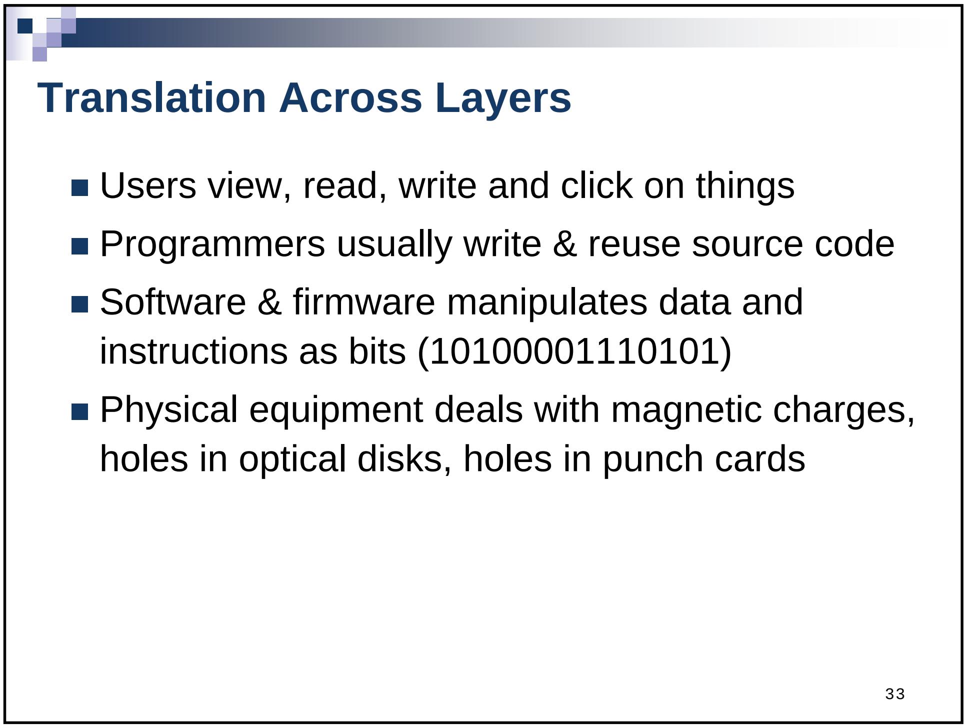
Digital objects are sets of **instructions for future interaction**

- Digital objects are useless (and don’t even exist) if no one can interact with them
- Interactions depend on numerous technical components



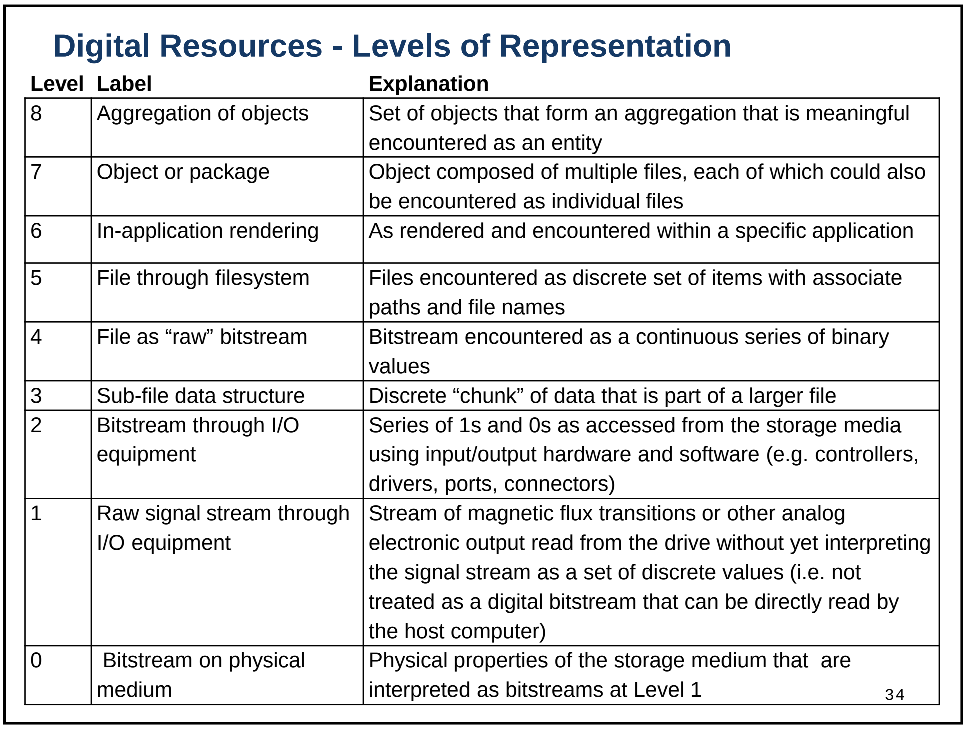
"Errors typically occur at the juncture between analog and digital states, such as when a drive's magnetoresistive head assigns binary symbolic value to the voltage differentials it has registered, or when an e-mail message is reconstituted from independent data packets moving across the TCP/IP layer of the Internet, itself dependent on fiber-optic cables and other hardwired technologies. All forms of modern digital technology incorporate **hyper-redundant error-checking** routines that serve to sustain an **illusion of immateriality** by detecting error and correcting it, reviving the quality of the signal, like old-fashioned telegraph relays, such that any degradation suffered during a subsequent interval of transmission will not fall beyond whatever **tolerances of symbolic integrity** exist past which the original value of the signal (or identity of the symbol) cannot be reconstituted."

Kirschenbaum, Matthew G. *Mechanisms: New Media and the Forensic Imagination.* Cambridge, MA: MIT Press, 2008. p.12 (emphasis mine).



Translation Across Layers

- Users view, read, write and click on things
- Programmers usually write & reuse source code
- Software & firmware manipulates data and instructions as bits (10100001110101)
- Physical equipment deals with magnetic charges, holes in optical disks, holes in punch cards



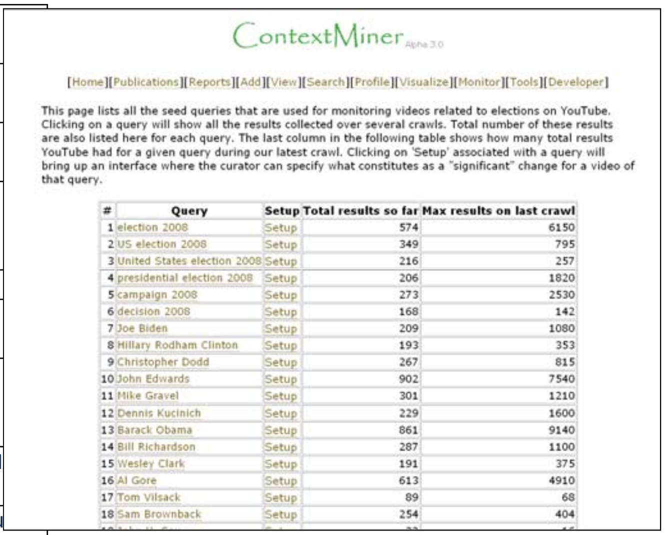
Digital Resources - Levels of Representation

Level	Label	Explanation
8	Aggregation of objects	Set of objects that form an aggregation that is meaningful encountered as an entity
7	Object or package	Object composed of multiple files, each of which could also be encountered as individual files
6	In-application rendering	As rendered and encountered within a specific application
5	File through filesystem	Files encountered as discrete set of items with associate paths and file names
4	File as "raw" bitstream	Bitstream encountered as a continuous series of binary values
3	Sub-file data structure	Discrete "chunk" of data that is part of a larger file
2	Bitstream through I/O equipment	Series of 1s and 0s as accessed from the storage media using input/output hardware and software (e.g. controllers, drivers, ports, connectors)
1	Raw signal stream through I/O equipment	Stream of magnetic flux transitions or other analog electronic output read from the drive without yet interpreting the signal stream as a set of discrete values (i.e. not treated as a digital bitstream that can be directly read by the host computer)
0	Bitstream on physical medium	Physical properties of the storage medium that are interpreted as bitstreams at Level 1

Interaction Examples

Level

Aggregation of objects
Object or package
In-application rendering
File through filesystem
File as “raw” bitstream
Sub-file data structure
Bitstream through I/O equipment
Raw signal stream through I equipment
Bitstream on physical mediu



ContextMiner Alpha 3.0

[Home][Publications][Reports][Add][View][Search][Profile][Visualize][Monitor][Tools][Developer]

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

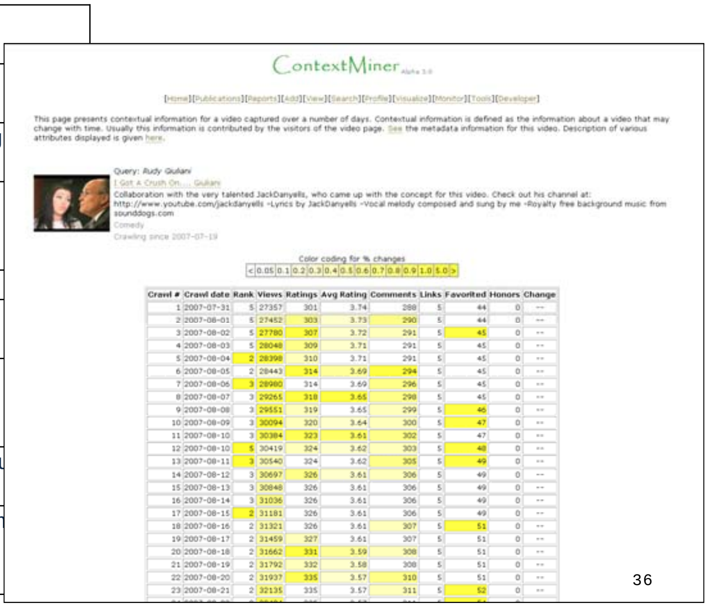
#	Query	Setup	Total results so far	Max results on last crawl
1	election 2008	Setup	574	6150
2	US election 2008	Setup	349	795
3	United States election 2008	Setup	216	257
4	presidential election 2008	Setup	206	1820
5	campaign 2008	Setup	273	2530
6	decision 2008	Setup	168	142
7	Joe Biden	Setup	209	1080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

35

Interaction Examples

Level

Aggregation of objects
Object or package
In-application rendering
File through filesystem
File as “raw” bitstream
Sub-file data structure
Bitstream through I/O equipment
Raw signal stream throu equipment
Bitstream on physical m



ContextMiner Alpha 3.0

[Home][Publications][Reports][Add][View][Search][Profile][Visualize][Monitor][Tools][Developer]

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

Query: *Rudy Giuliani*

I Got A Crush On.... Giuliani

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

Comedy

Crawling since 2007-07-19

Color coding for % changes

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

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

36

Interaction Examples

Level

Aggregation of objects

Object or package

In-application rendering

File through filesystem

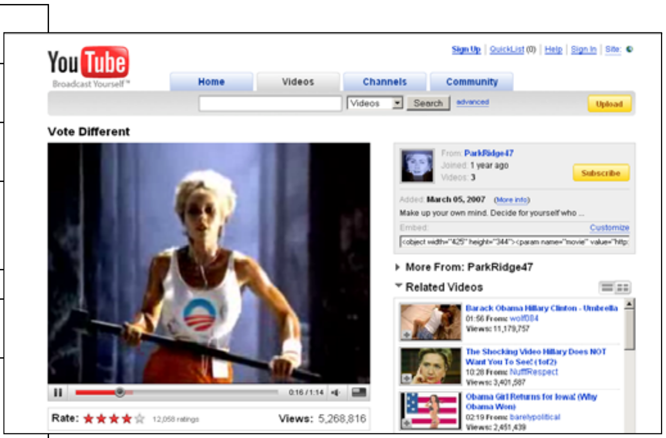
File as “raw” bitstream

Sub-file data structure

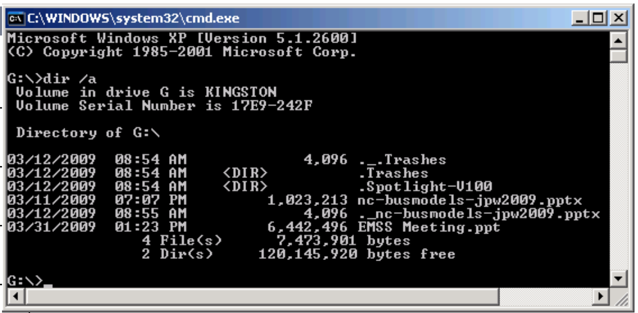
Bitstream through I/O equipment

Raw signal stream through I/O equipment

Bitstream on physical medium



You Tube
Broadcast Yourself™
Sign Up | QuickList (0) | Help | Sign In | Site:
Home
Videos
Channels
Community
Videos
Search
advanced
Upload
Vote Different
From: ParkRidge47
Joined: 1 year ago
Videos: 3
Subscribe
Added: March 05, 2007 (More info)
Make up your own mind. Decide for yourself who ...
Embed:
Customize
<object width="425" height="344"><param name="movie" value="http:
More From: ParkRidge47
Related Videos
Barack Obama Hillary Clinton - Umbrella
01:56 From: wolf084
Views: 11,179,757
The Shocking Video Hillary Does NOT Want You To See! (1of2)
10:28 From: NuffRespect
Views: 3,401,587
Obama Girl Returns for Iowa! (Why Obama Won)
02:19 From: barelypolitical
Views: 2,451,439
0:16 / 1:14
Rate: ★★★★☆ 12,058 ratings
Views: 5,268,816



```
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

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

 Directory of G:\

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

G:\>_
```

Level

Aggregation of objects

Object or package

In-application rendering

File through filesystem

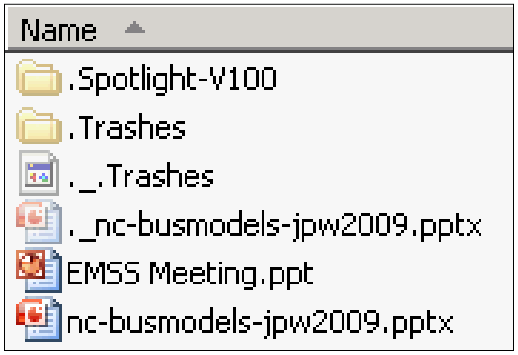
File as “raw” bitstream

Sub-file data structure

Bitstream through I/O equipment

Raw signal stream through I/O equipment

Bitstream on physical medium



Name

.Spotlight-V100

.Trashes

._.Trashes

._nc-busmodels-jpw2009.pptx

EMSS Meeting.ppt

nc-busmodels-jpw2009.pptx

Interaction Examples

Level

Aggregation of objects

Object or package

In-application rendering

File through filesystem

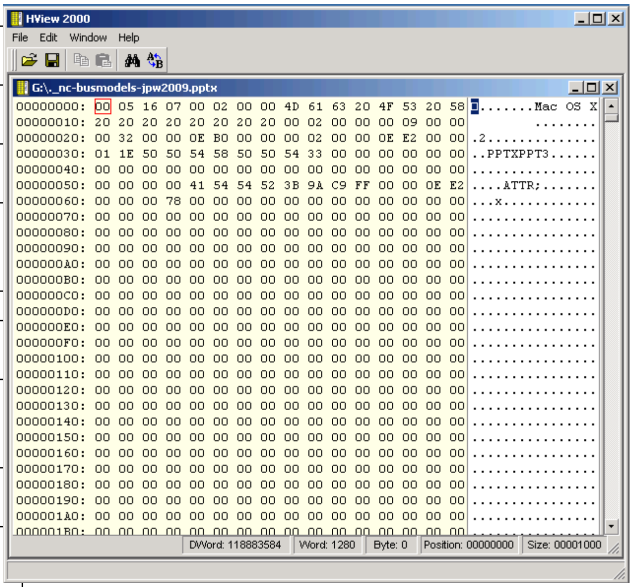
File as “raw” bitstream

Sub-file data structure

Bitstream through I/O equipment

Raw signal stream through I/O equipment

Bitstream on physical medium



```
00000000: 00 05 16 07 00 02 00 00 4D 61 63 20 4F 53 20 58 ........Mac OS X
00000010: 20 20 20 20 20 20 20 20 00 02 00 00 00 09 00 00         ........
00000020: 00 32 00 00 0E B0 00 00 00 02 00 00 0E E2 00 00 .2..............
00000030: 01 1E 50 50 54 58 50 50 54 33 00 00 00 00 00 00 ..PPTXPPT3......
00000040: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000050: 00 00 00 00 41 54 54 52 3B 9A C9 FF 00 00 0E E2 ....ATTR;.......
00000060: 00 00 00 78 00 00 00 00 00 00 00 00 00 00 00 00 ...x............
00000070: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000080: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000090: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
000000A0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
000000B0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
000000C0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
000000D0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
000000E0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
000000F0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000100: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000110: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000120: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000130: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000140: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000150: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000160: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000170: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000180: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000190: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
000001A0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
000001B0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
```

DWord: 118883584 | Word: 1280 | Byte: 0 | Position: 00000000 | Size: 00001000

Interaction Examples

Level

Aggregation of objects

Object or package

In-application rendering

File through filesystem

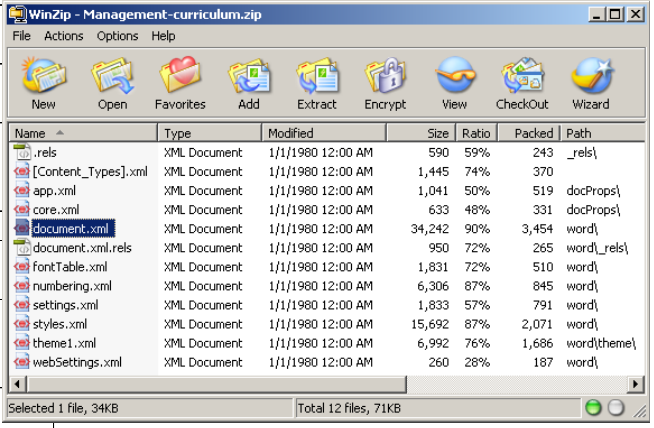
File as “raw” bitstream

Sub-file data structure

Bitstream through I/O equipment

Raw signal stream through equipment

Bitstream on physical medium



Name	Type	Modified	Size	Ratio	Packed	Path
.rels	XML Document	1/1/1980 12:00 AM	590	59%	243	_rels\
[Content_Types].xml	XML Document	1/1/1980 12:00 AM	1,445	74%	370	
app.xml	XML Document	1/1/1980 12:00 AM	1,041	50%	519	docProps\
core.xml	XML Document	1/1/1980 12:00 AM	633	48%	331	docProps\
document.xml	XML Document	1/1/1980 12:00 AM	34,242	90%	3,454	word\
document.xml.rels	XML Document	1/1/1980 12:00 AM	950	72%	265	word_rels\
fontTable.xml	XML Document	1/1/1980 12:00 AM	1,831	72%	510	word\
numbering.xml	XML Document	1/1/1980 12:00 AM	6,306	87%	845	word\
settings.xml	XML Document	1/1/1980 12:00 AM	1,833	57%	791	word\
styles.xml	XML Document	1/1/1980 12:00 AM	15,692	87%	2,071	word\
theme1.xml	XML Document	1/1/1980 12:00 AM	6,992	76%	1,686	word\theme\
webSettings.xml	XML Document	1/1/1980 12:00 AM	260	28%	187	word\

Selected 1 file, 34KB | Total 12 files, 71KB

Interaction Examples

Level

Aggregation of objec

Object or package

In-application render

File through filesyste

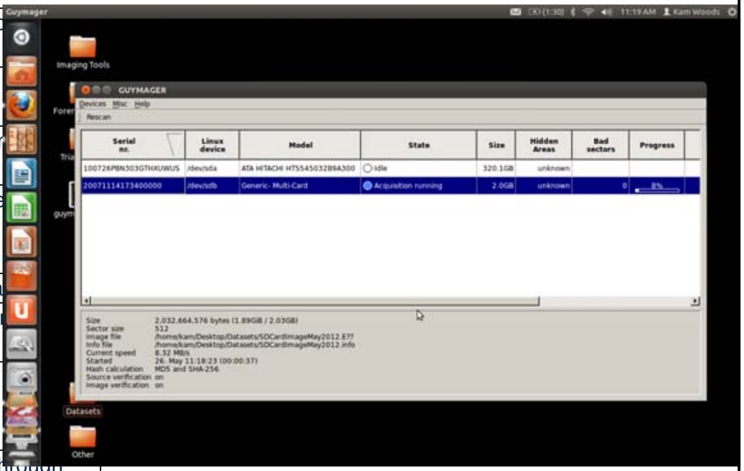
File as “raw” bitstrea

Sub-file data structu

Bitstream through I/O equipment

Raw signal stream through I/O equipment

Bitstream on physical medium



Serial nr.	Linux device	Model	State	Size	Hidden Areas	Bad sectors	Progress
100726PBN303GTHXUWUS	/dev/sda	ATA HITACHI HTS545032B9A300	○ Idle	320.1GB	unknown		
20071114173400000	/dev/sdb	Generic- Multi-Card	● Acquisition running	2.0GB	unknown	0	8%

```
Size                2,032,664,576 bytes (1.89GiB / 2.03GB)
Sector size         512
Image file          /home/kam/Desktop/Datasets/SDCardImageMay2012.E??
Info file           /home/kam/Desktop/Datasets/SDCardImageMay2012.info
Current speed       8.32 MB/s
Started             26. May 11:18:23 (00:00:37)
Hash calculation    MD5 and SHA-256
Source verification on
Image verification  on
```

Level

Aggregation of

Object or pack

In-application r

File through file

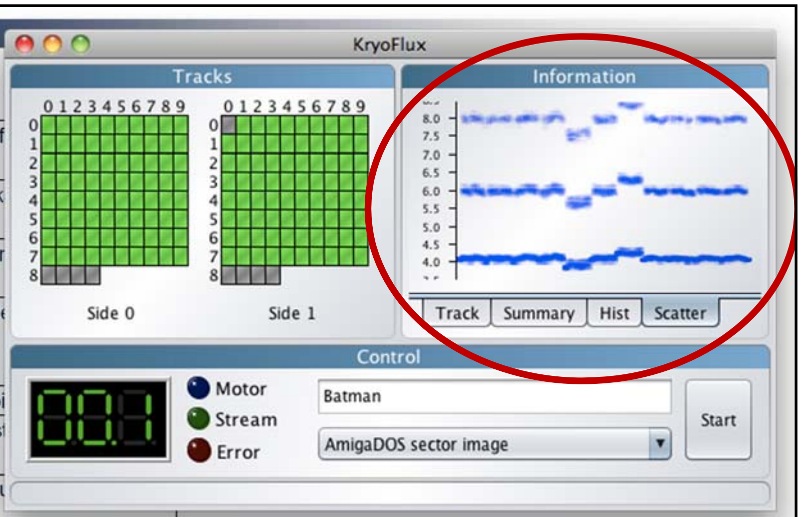
File as “raw” bi

Sub-file data st

Bitstream throu equipment

Raw signal stream through I/O equipment

Bitstream on physical medium

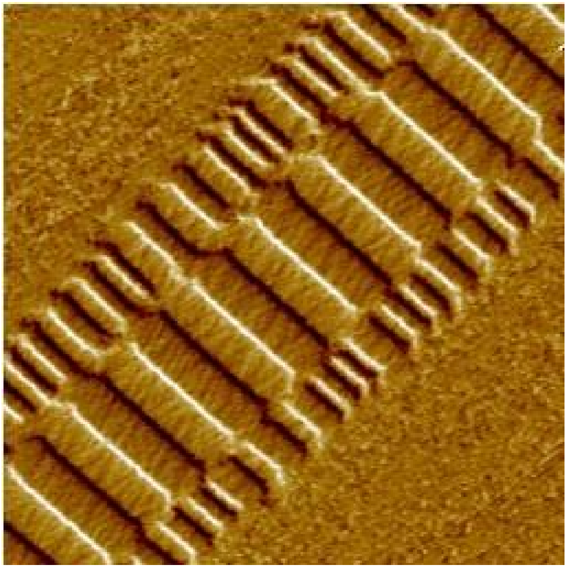


Batman

AmigaDOS sector image ▼

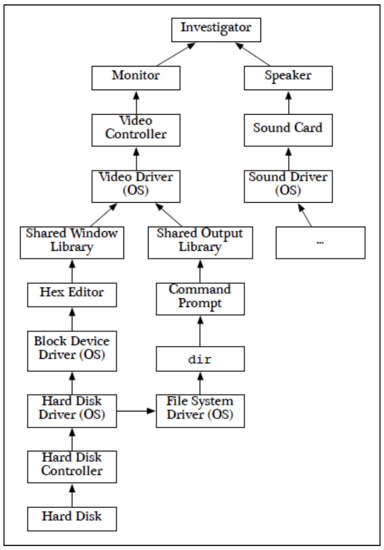
Interaction Examples

Level
Aggregation of objects
Object or package
In-application rendering
File through filesystem
File as “raw” bitstream
Sub-file data structure
Bitstream through I/O equipment
Raw signal stream through I/O equipment
Bitstream on physical medium

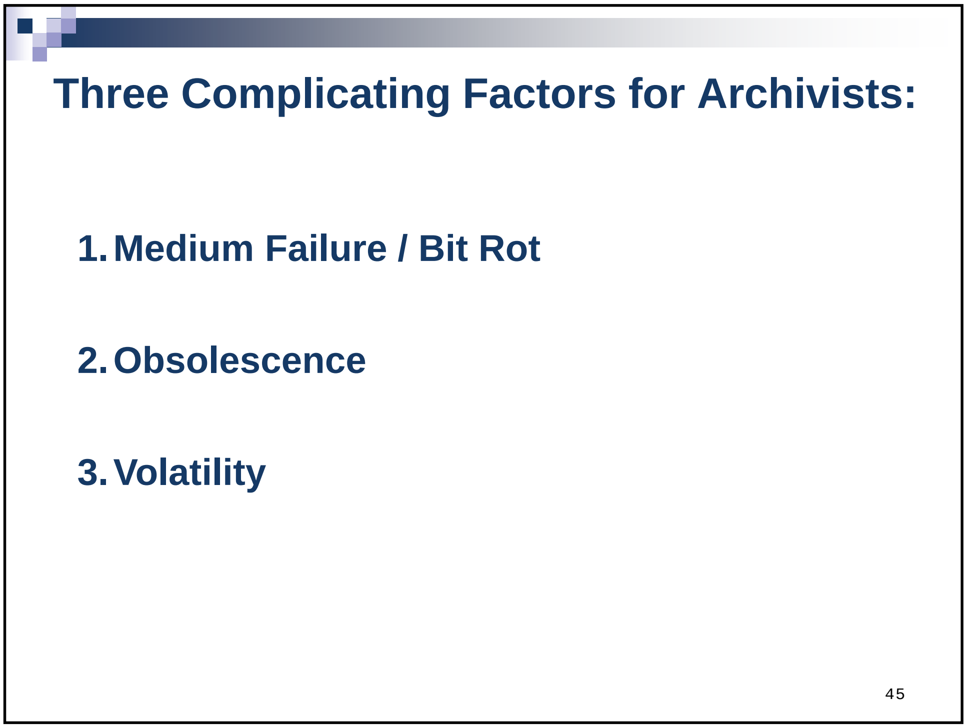


Veeco Instruments. http://www.veeco.com/library/nanotheater_detail.php?type=application&id=78&app_id=34

Multiple Paths for Viewing Bits

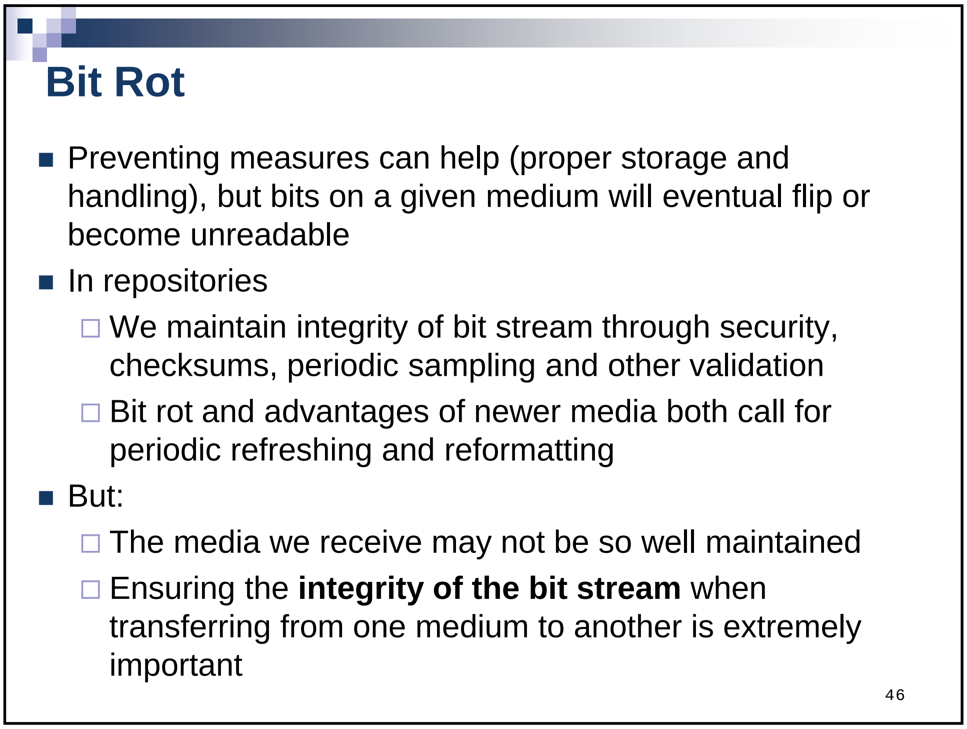


Carrier, Brian D. "A Hypothesis-Based Approach to Digital Forensic Investigations." Doctoral Dissertation, Purdue University, 2006. Figure 3-3 (p.60)



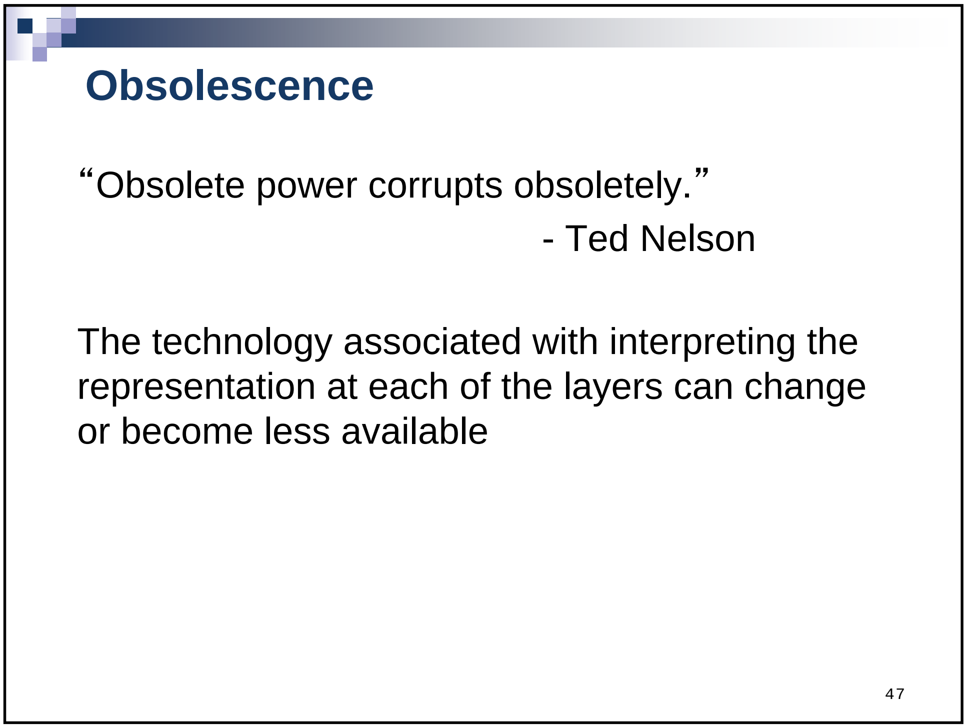
Three Complicating Factors for Archivists:

1. Medium Failure / Bit Rot
2. Obsolescence
3. Volatility



Bit Rot

- Preventing measures can help (proper storage and handling), but bits on a given medium will eventual flip or become unreadable
- In repositories
  - We maintain integrity of bit stream through security, checksums, periodic sampling and other validation
  - Bit rot and advantages of newer media both call for periodic refreshing and reformatting
- But:
  - The media we receive may not be so well maintained
  - Ensuring the **integrity of the bit stream** when transferring from one medium to another is extremely important

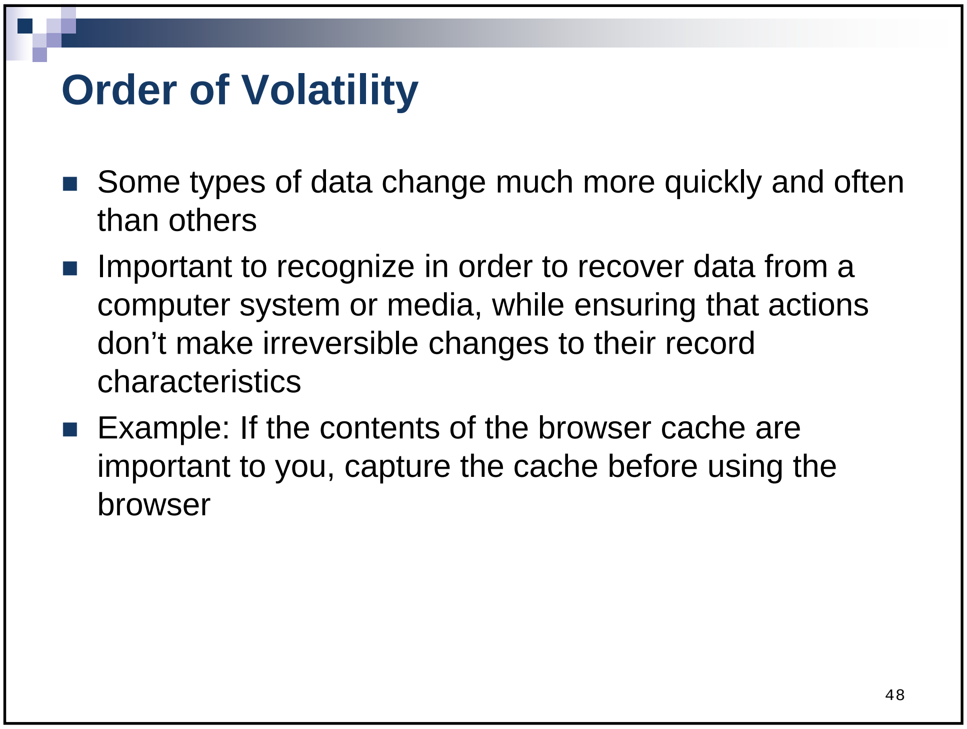


Obsolescence

"Obsolete power corrupts obsoletely."

- Ted Nelson

The technology associated with interpreting the representation at each of the layers can change or become less available



Order of Volatility

- Some types of data change much more quickly and often than others
- Important to recognize in order to recover data from a computer system or media, while ensuring that actions don't make irreversible changes to their record characteristics
- Example: If the contents of the browser cache are important to you, capture the cache before using the browser

How and where does a computer store information?

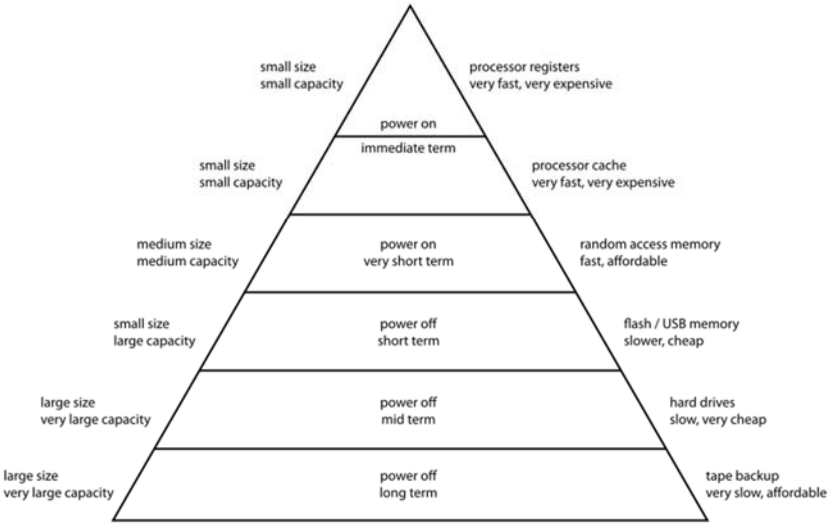
Bits – How Data are Conveyed in Computers

- Variable voltage electrical signals or pulses of light
- Bit represents a tiny “switch” with two possible states – on/off, true/false, 1/0
- Bit string or bitstream: a consecutive sequence of bits (e.g. 101000111010101)
- Rarely meaningful to humans – when looking at bitstream, usually use a hex editor (discussed later)

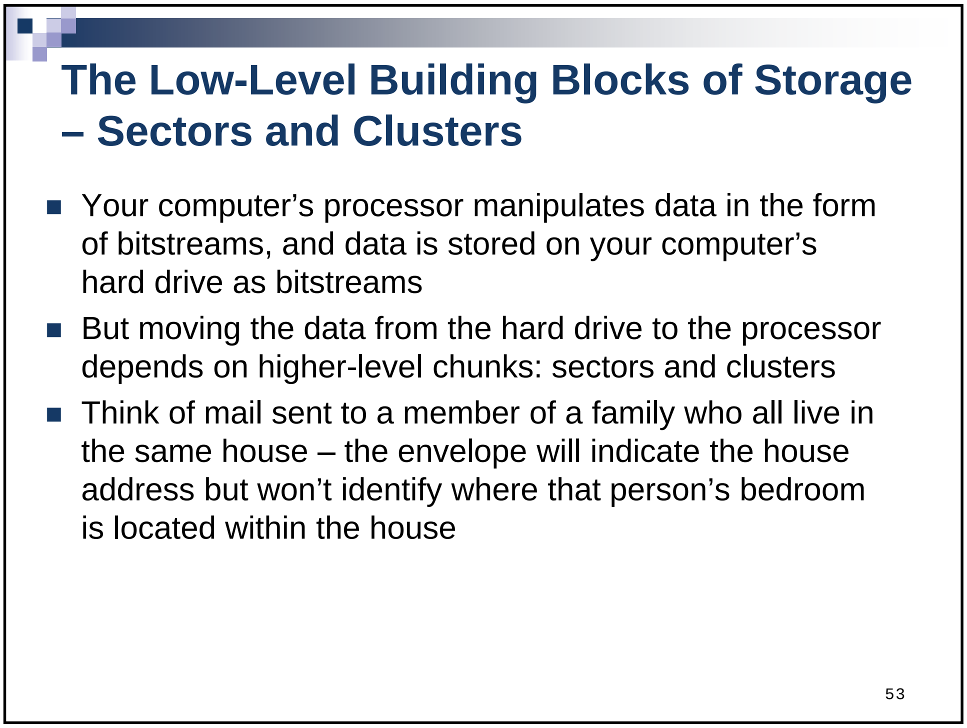
Motivations for Storage Hierarchy

- Different forms of memory/storage have significantly different costs and performance
- Store recent data close by, in fast, expensive, volatile storage
- Store data that has not been used recently and is rarely used in slower, cheaper, less volatile storage

Computer Memory Hierarchy

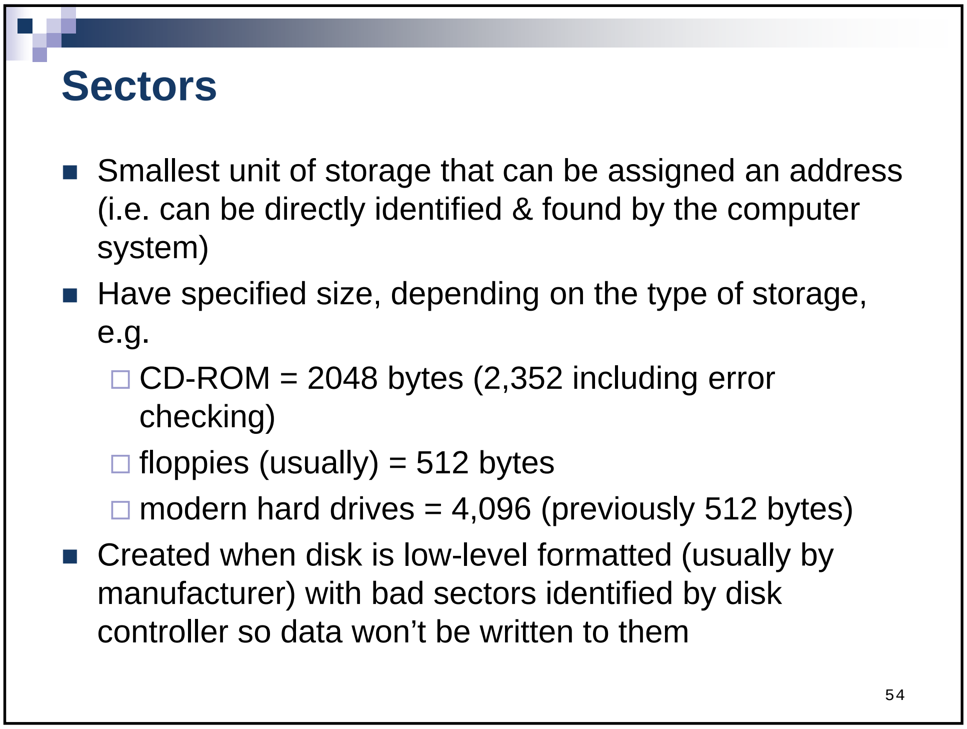


Source: http://en.wikipedia.org/wiki/File:ComputerMemoryHierarchy.svg



The Low-Level Building Blocks of Storage – Sectors and Clusters

- Your computer’s processor manipulates data in the form of bitstreams, and data is stored on your computer’s hard drive as bitstreams
- But moving the data from the hard drive to the processor depends on higher-level chunks: sectors and clusters
- Think of mail sent to a member of a family who all live in the same house – the envelope will indicate the house address but won’t identify where that person’s bedroom is located within the house



Sectors

- Smallest unit of storage that can be assigned an address (i.e. can be directly identified & found by the computer system)
- Have specified size, depending on the type of storage, e.g.
  - CD-ROM = 2048 bytes (2,352 including error checking)
  - floppies (usually) = 512 bytes
  - modern hard drives = 4,096 (previously 512 bytes)
- Created when disk is low-level formatted (usually by manufacturer) with bad sectors identified by disk controller so data won’t be written to them

Clusters

- Groups of sectors
- Smallest unit of storage that can be tracked by the operating system
- Sizes depends on operating system, type & size of storage device – examples are 2048 bytes (4 sectors of 512 bytes) or 4096 bytes
- Defined during high-level formatting performed by operating system

Magnetic Disk (e.g. Hard Drive or Floppy)

- Bits stored as magnetic fields of different polarity
- Magnetized surface of disk rotates under a read/write head
- Divided into tracks (like rings of a tree)
- Tracks divided into sectors and clusters
- Windows: File Allocation Table (FAT) or Master File Table (for NTFS) indicates, for given file, what clusters contain its content

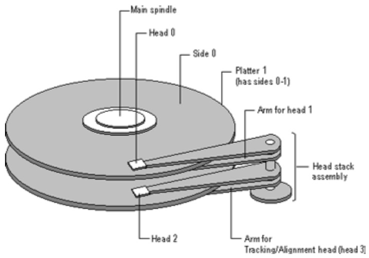
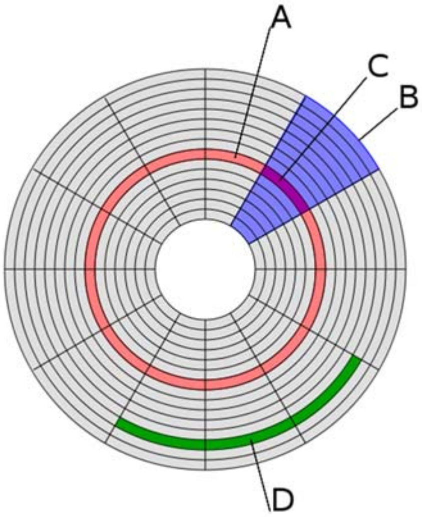


Image from : “Concepts.” In Active UNDELETE v2.0 Documentation. Active Data Recovery Software. www.active-undelete.com/3tracks.htm

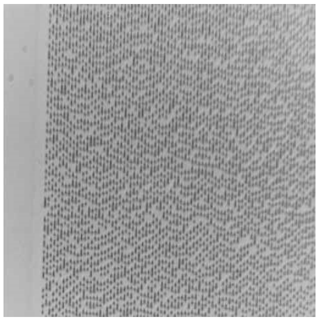


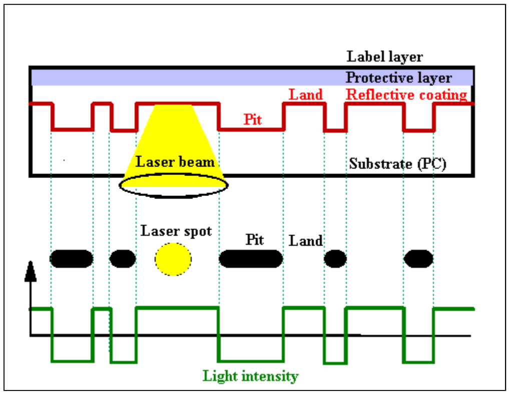
Hard Drive Structure:

A = track
B = sector
C = sector of a track
D = cluster

Source: http://en.wikipedia.org/wiki/File:Disk-structure2.svg

Optical Media – CD-ROM as Example

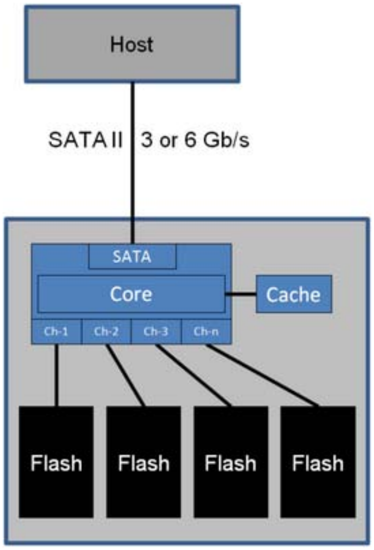




Source of Images: Compact Disk (CD). USByte.
http://www.usbyte.com/common/compact_disk_3.htm

Solid-State Drives (SSDs)

Source:
http://www.tomshardware.com/gallery/Samsung-SSD-256-ToggleDDR,0101-260898-0-0-0-0-jpg-.html



- Uses integrated circuits to store data
- No moving parts
- Can be read using same I/O equipment as used for hard drives
- Increasingly common in laptops

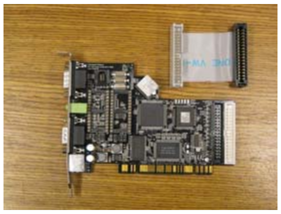
Source:
http://www.tomshardware.com/gallery/ssd-controller-external-cache,0101-260900-0-0-0-0-jpg-.html

Floppy Disks

- Physical storage is similar to hard drives described above (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…

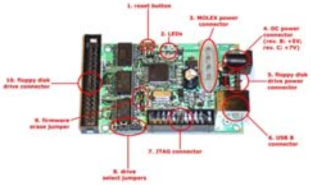
Floppy Controller Hardware

CatWeasel[1] (no longer available)



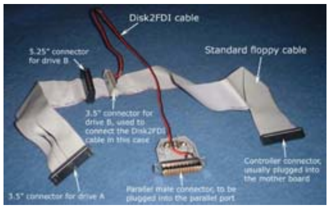
Disc Ferret[2]

Kryoflux[3]

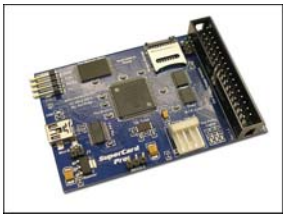


FC 5025[4]

Disk2FDI[5]


Disk2FDI cable
5.25" connector for drive B
Standard floppy cable
3.5" connector for drive B, used to connect the Disk2FDI cable in this case
Controller connector, usually plugged into the mother board
3.5" connector for drive A
Parallel male connector, to be plugged into the parallel port

SuperCard Pro[6]

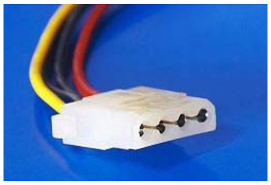


1. http://lib.stanford.edu/digitial-forensics-stanford-university-libraries/catweasel-universal-floppy-drive-controller
2. http://discferret.com/wiki/DiscFerret
3. http://www.kryoflux.com/
4. http://www.deviceside.com/fc5025.html
5. http://disk2fdi.joguin.com/D2FCABLE.htm
6. http://www.cbmstuff.com/proddetail.php?prod=SCP

61

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

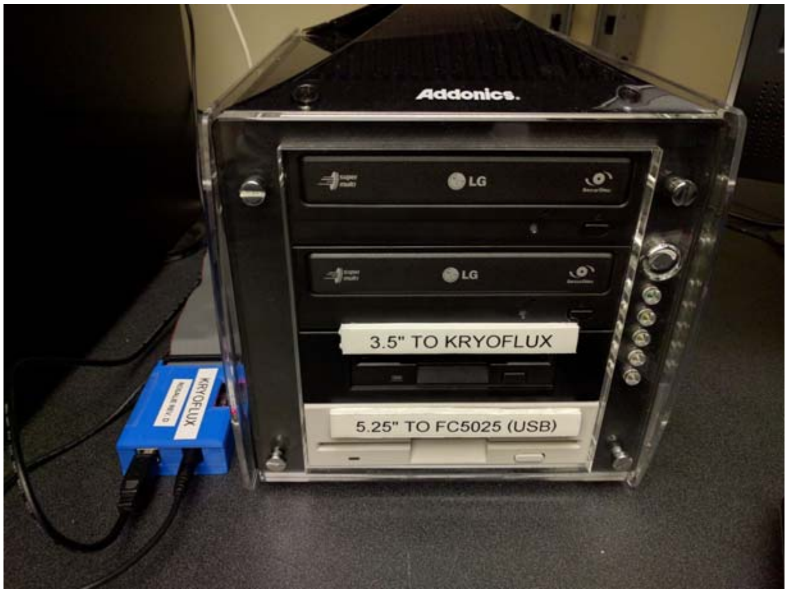


http://en.wikipedia.org/wiki/File:Molex_female_connector.jpg

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

62

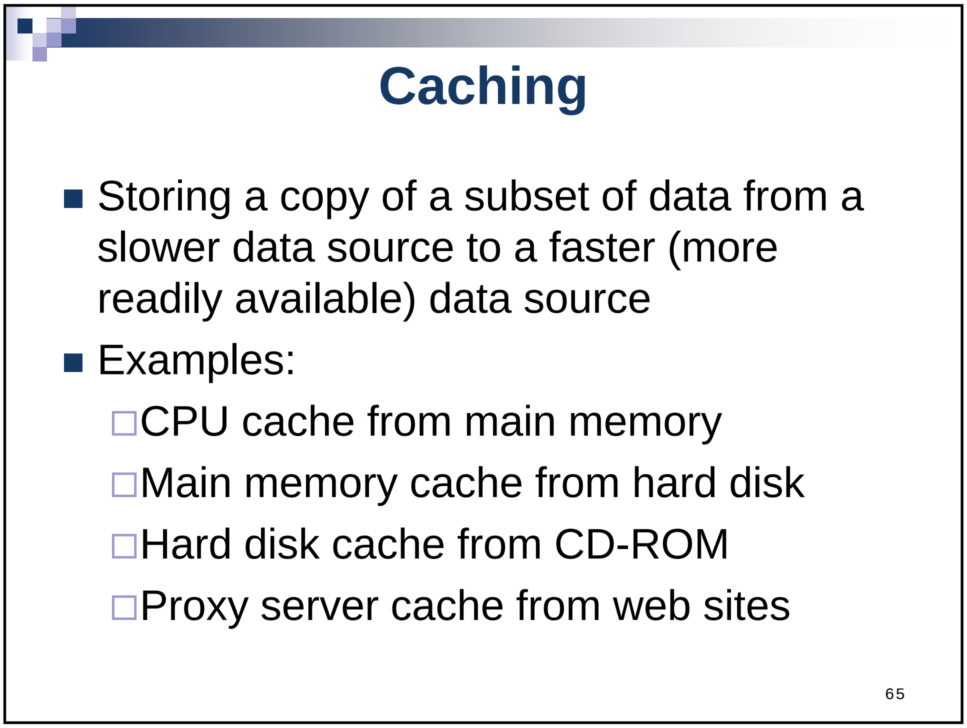
Kryoflux Running on a “Mini JukeBox”*



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

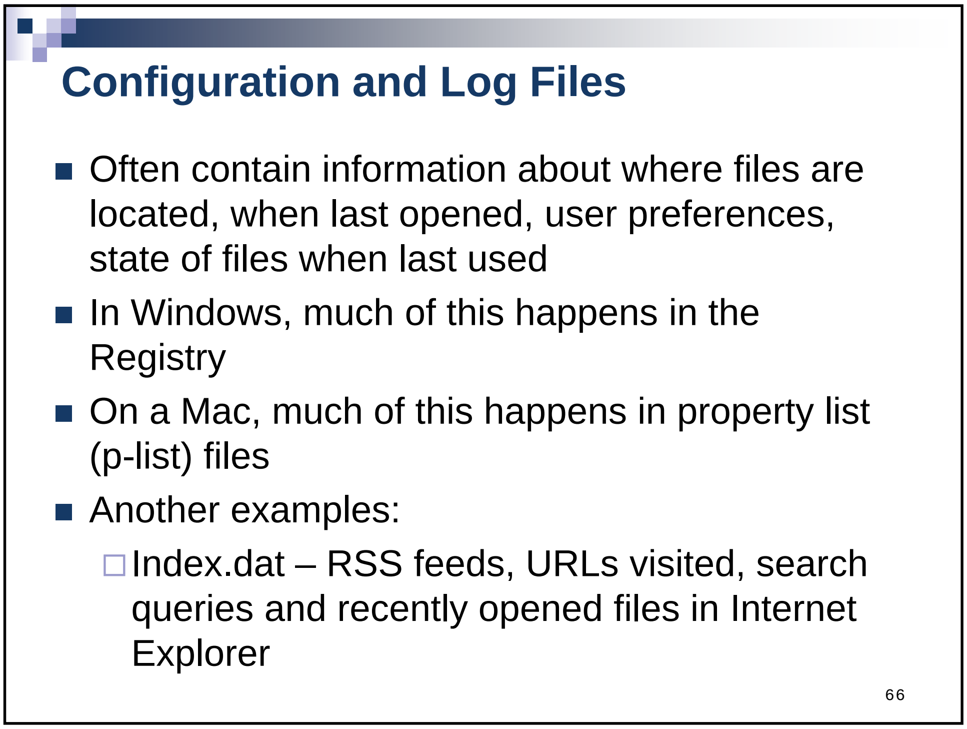
Areas Designed to Store Temporary Data

- Files on disk used for virtual memory management – e.g. “swap files” in Windows 95/98, “page files” in Windows NT/2000/XP
- Temp files
- Various caches - e.g. browser cache, which includes copies of recently downloaded files
- “Recent documents” in Windows
- Cookies – “expires” attribute can indicate quick deletion or long-term retention
- History files – e.g. browsing & download history



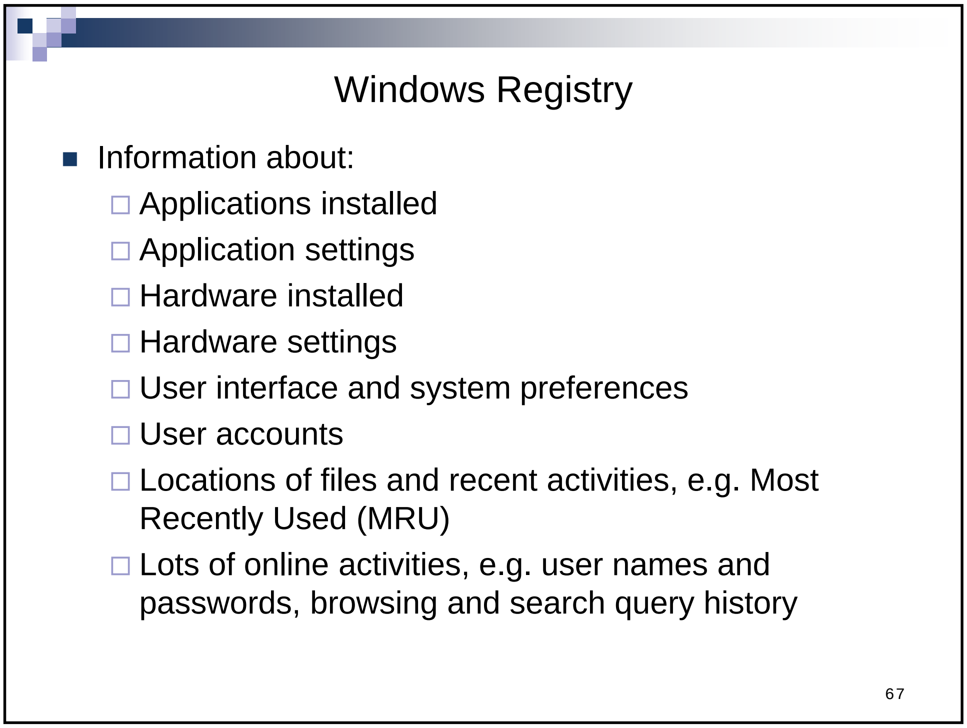
Caching

- Storing a copy of a subset of data from a slower data source to a faster (more readily available) data source
- Examples:
  - □CPU cache from main memory
  - □Main memory cache from hard disk
  - □Hard disk cache from CD-ROM
  - □Proxy server cache from web sites



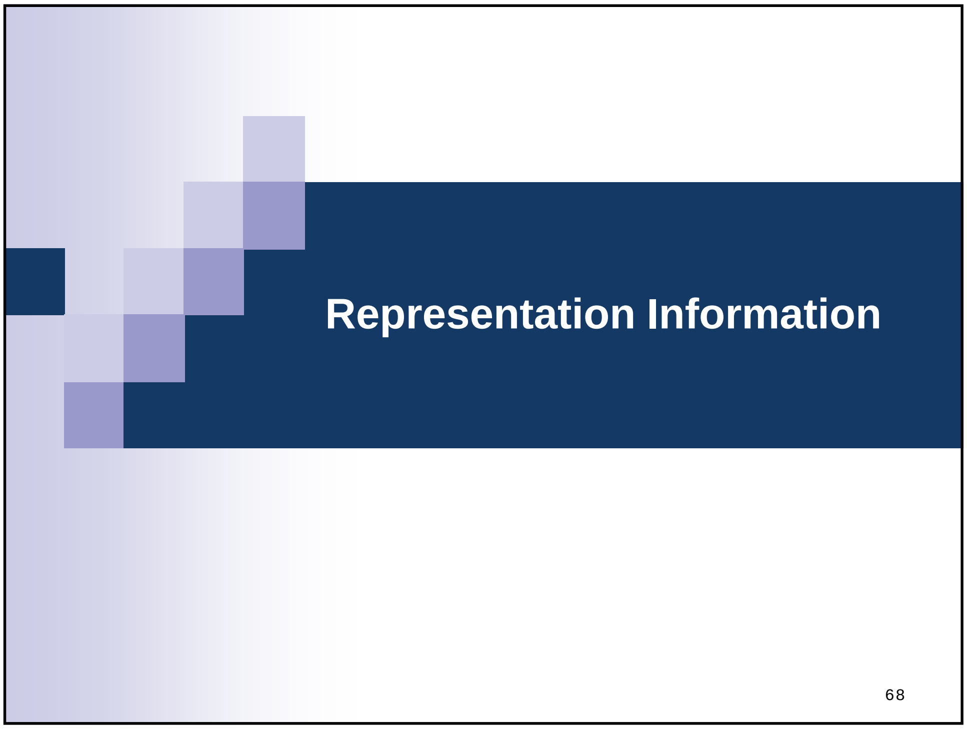
Configuration and Log Files

- Often contain information about where files are located, when last opened, user preferences, state of files when last used
- In Windows, much of this happens in the Registry
- On a Mac, much of this happens in property list (p-list) files
- Another examples:
  - □Index.dat – RSS feeds, URLs visited, search queries and recently opened files in Internet Explorer



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. user names and passwords, browsing and search query history



Representation Information

“No computation without representation”

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

69

The 8 bits highlighted in the bit stream shown below can be interpreted in many ways, e.g., as an integer, a simple character code, a sound, a floating point number, an image, a logical bitmap, etc.

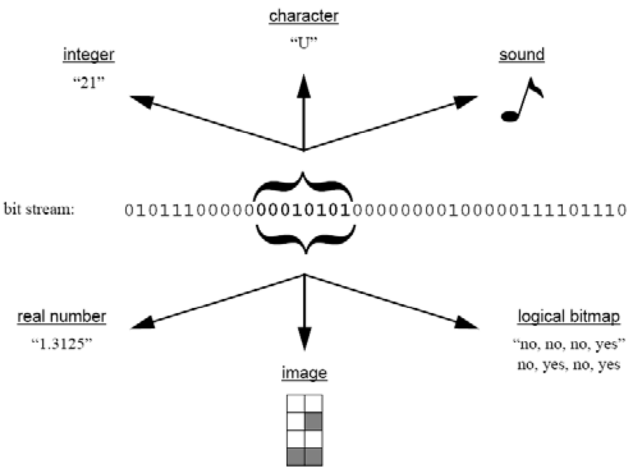


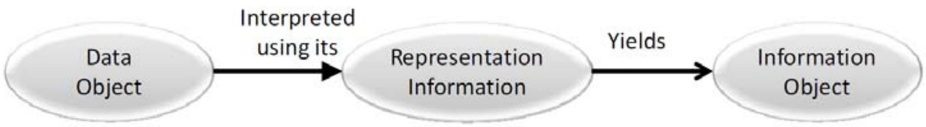
Figure 4: A bit stream can represent anything at all

Rothenberg, Jeff. "Ensuring the Longevity of Digital Information." Washington, DC: Council on Library and Information Resources, 1999.

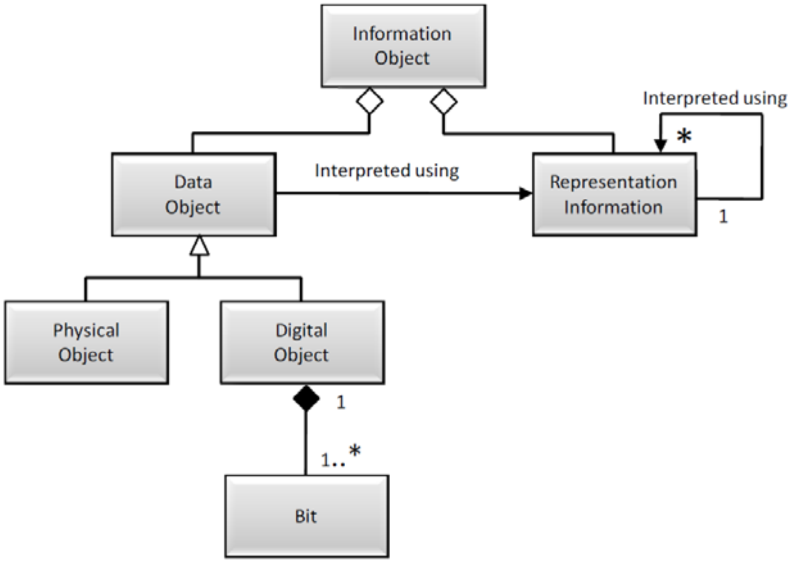
70

Representation Information

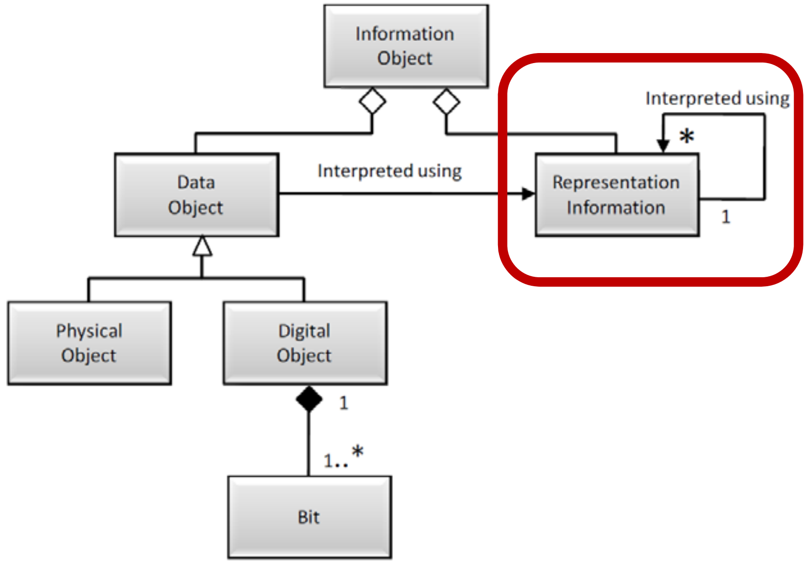
- "Information that maps a Data Object into more meaningful concepts" (OAIS) - makes humanly-perceptible properties happen
- Examples: file format, encoding scheme, data type



Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-M-2 (Magenta Book). Consultative Committee for Space Data Systems, 2012. [ISO 14721:2012]. Figure 2-2.



Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-M-2 (Magenta Book). Consultative Committee for Space Data Systems, 2012. [ISO 14721:2012]. Figure 4-10.



Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-M-2 (Magenta Book). Consultative Committee for Space Data Systems, 2012. [ISO 14721:2012]. Figure 4-10.

Representation Information can Reside in Many Places

- Within digital object itself
- Stored separately as metadata
- Encoded within software required to read and parse the digital object

Finding Representation Information Within a File

- Keys fields
- Headers
- Manifests

The 4 bits at the start of this bit stream are intended to be read as the "key" integer 7, meaning that the remaining bytes in the bit stream are each 7 bits long. However, there is no way to tell from the bit stream itself how long the key integer is; if we were to erroneously read the first 5 bits of the bit stream as the key (instead of the first 4), we would erroneously conclude that the remaining bytes were each 15 bits long.

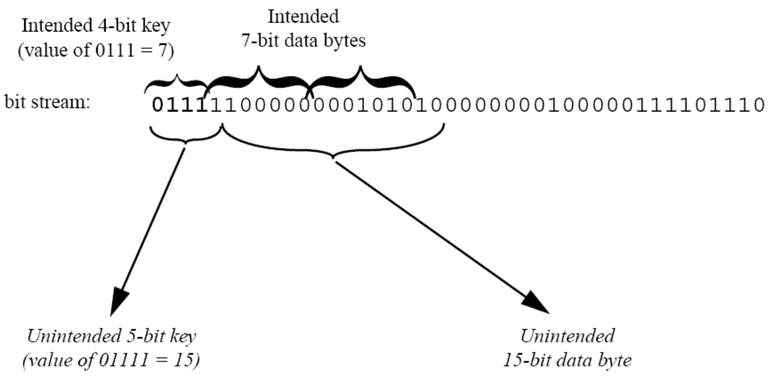


Figure 6: Bit streams cannot be made self-explanatory

Rothenberg, Jeff. "Ensuring the Longevity of Digital Information." Washington, DC: Council on Library and Information Resources, 1999.

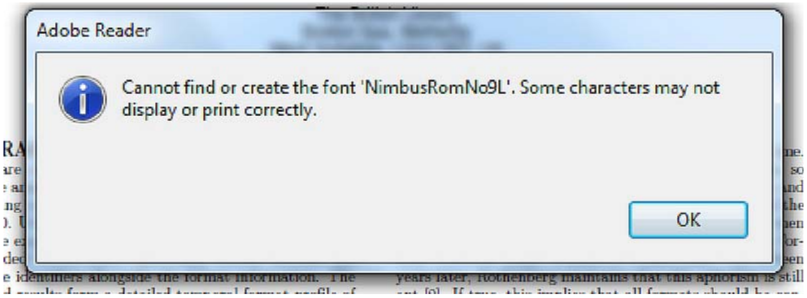
Not Just a Series of Bytes – Pointers and Offsets

- Pointer – reference within a file or programming code that leads from one place to another
  - □ Causes the data to be read out of serial order (i.e. a jump from one place to another place that does not immediately follow it within the data stream)
  - □ Ability to resolve the pointer is essential

- Offset – location that's some given distance from a starting point
  - □ Location calculated by adding offset to a base address (location)
  - □ Again, ability to resolve the offset to the precise location within a bitstream is essential

Fonts and Character Encoding

Font

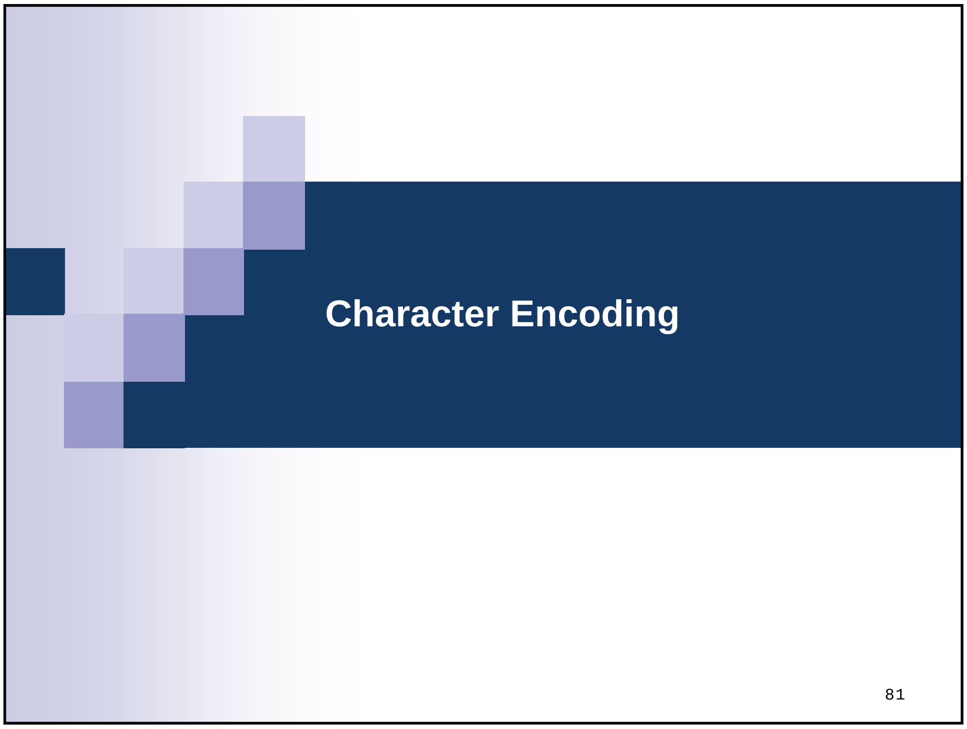


Cannot find or create the font 'NimbusRomNo9L'. Some characters may not display or print correctly.

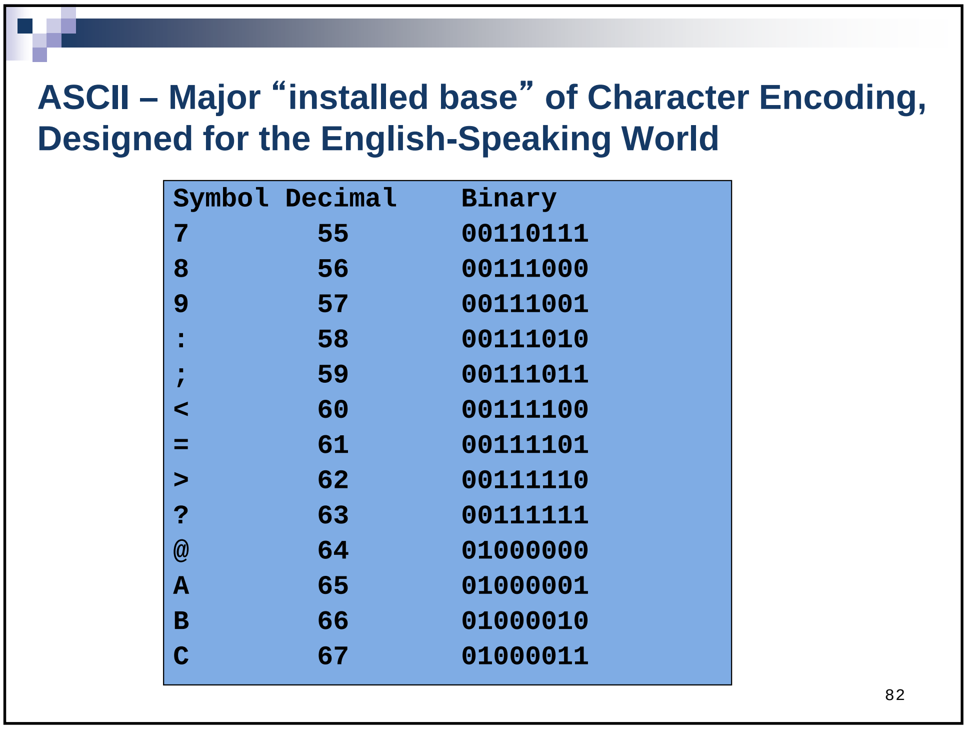
PDF with missing fonts by prwheatly, on Flickr.

Font

- Determines how characters will appear on screen (generation of glyphs)
- Same character can appear completely different in two different fonts
- Can be a major issue in digital preservation, when consistent rendering is important
- Not usually a focus of digital forensics or data recovery efforts, which focus on simply making sense of the characters within a bitstream
- However, changes of fonts within a document can provide some hints to versioning and authorship



Character Encoding



ASCII – Major “installed base” of Character Encoding, Designed for the English-Speaking World

Symbol	Decimal	Binary
7	55	00110111
8	56	00111000
9	57	00111001
:	58	00111010
;	59	00111011
<	60	00111100
=	61	00111101
>	62	00111110
?	63	00111111
@	64	01000000
A	65	01000001
B	66	01000010
C	67	01000011

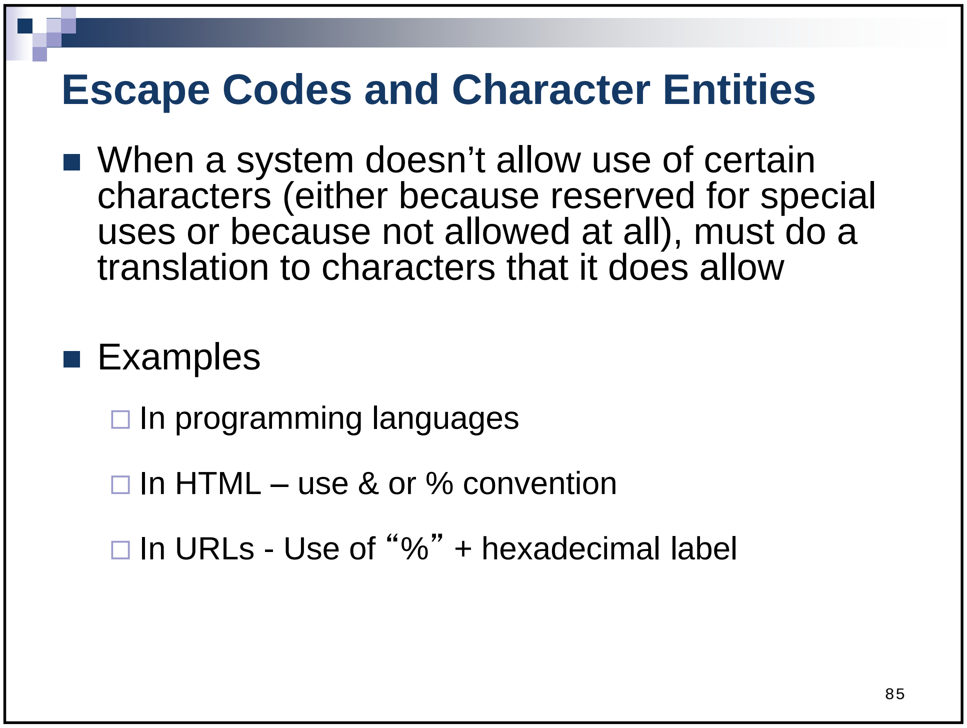
Unicode

- **<u>Huge</u>** number of possible characters – not limited to 8-bits for each
- Mapped to unique codes (numbers)
- Standard first published in 1991
- Current version is 8.0 (June 2015)

UTF-8

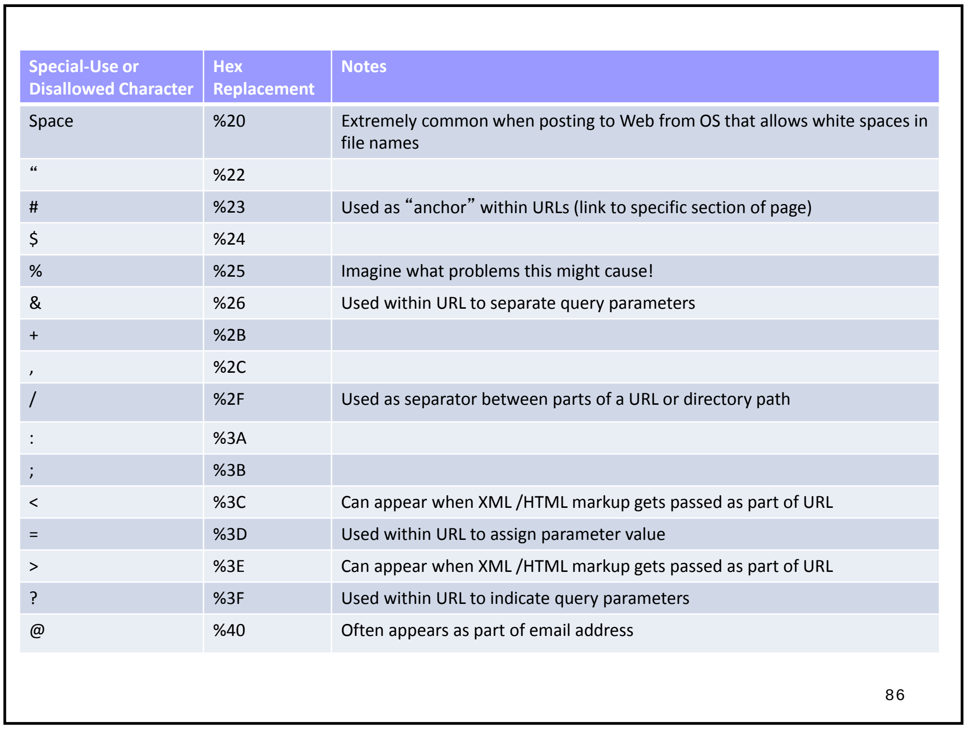
- *Unicode Transformation Format (UTF)* = set of conventions for how specific Unicode code points are represented as unique byte sequences
- UTF 8 is widely used – including in email and web pages
- Codes 0 to 127 are backward compatible with ASCII

See: Frequently Asked Questions: UTF-8, UTF-16, UTF-32 & BOM. Unicode, Inc.
http://www.unicode.org/faq/utf_bom.html



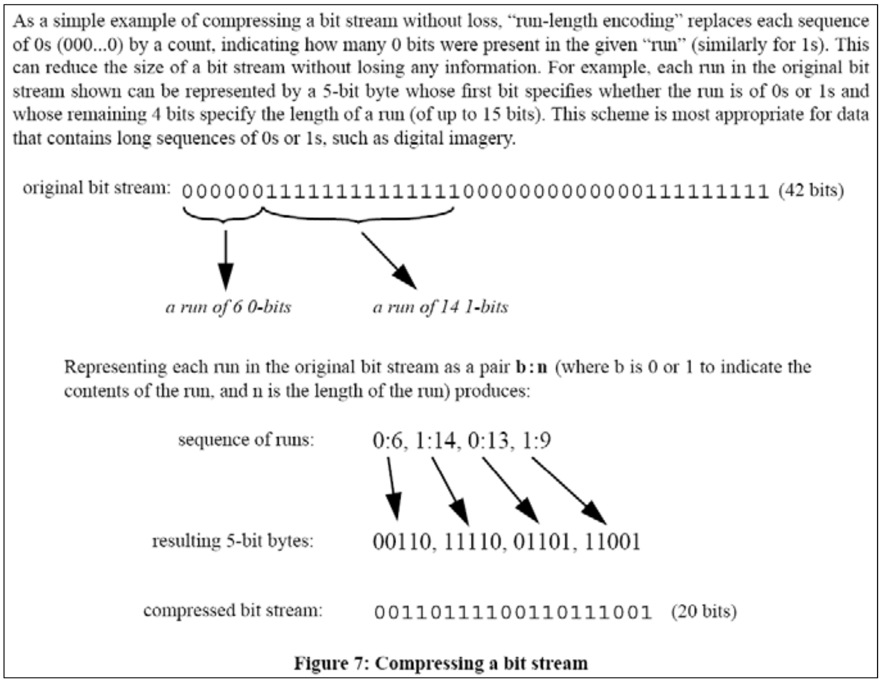
Escape Codes and Character Entities

- When a system doesn’t allow use of certain characters (either because reserved for special uses or because not allowed at all), must do a translation to characters that it does allow
- Examples
  - □ In programming languages
  - □ In HTML – use & or % convention
  - □ In URLs - Use of “%” + hexadecimal label



Special-Use or Disallowed Character	Hex Replacement	Notes
Space	%20	Extremely common when posting to Web from OS that allows white spaces in file names
“	%22	
#	%23	Used as “anchor” within URLs (link to specific section of page)
$	%24	
%	%25	Imagine what problems this might cause!
&	%26	Used within URL to separate query parameters
+	%2B	
,	%2C	
/	%2F	Used as separator between parts of a URL or directory path
:	%3A	
;	%3B	
<	%3C	Can appear when XML /HTML markup gets passed as part of URL
=	%3D	Used within URL to assign parameter value
>	%3E	Can appear when XML /HTML markup gets passed as part of URL
?	%3F	Used within URL to indicate query parameters
@	%40	Often appears as part of email address

Compression



As a simple example of compressing a bit stream without loss, "run-length encoding" replaces each sequence of 0s (000...0) by a count, indicating how many 0 bits were present in the given "run" (similarly for 1s). This can reduce the size of a bit stream without losing any information. For example, each run in the original bit stream shown can be represented by a 5-bit byte whose first bit specifies whether the run is of 0s or 1s and whose remaining 4 bits specify the length of a run (of up to 15 bits). This scheme is most appropriate for data that contains long sequences of 0s or 1s, such as digital imagery.

original bit stream: 000000111111111111110000000000000111111111 (42 bits)

a run of 6 0-bits   *a run of 14 1-bits*

Representing each run in the original bit stream as a pair **b:n** (where b is 0 or 1 to indicate the contents of the run, and n is the length of the run) produces:

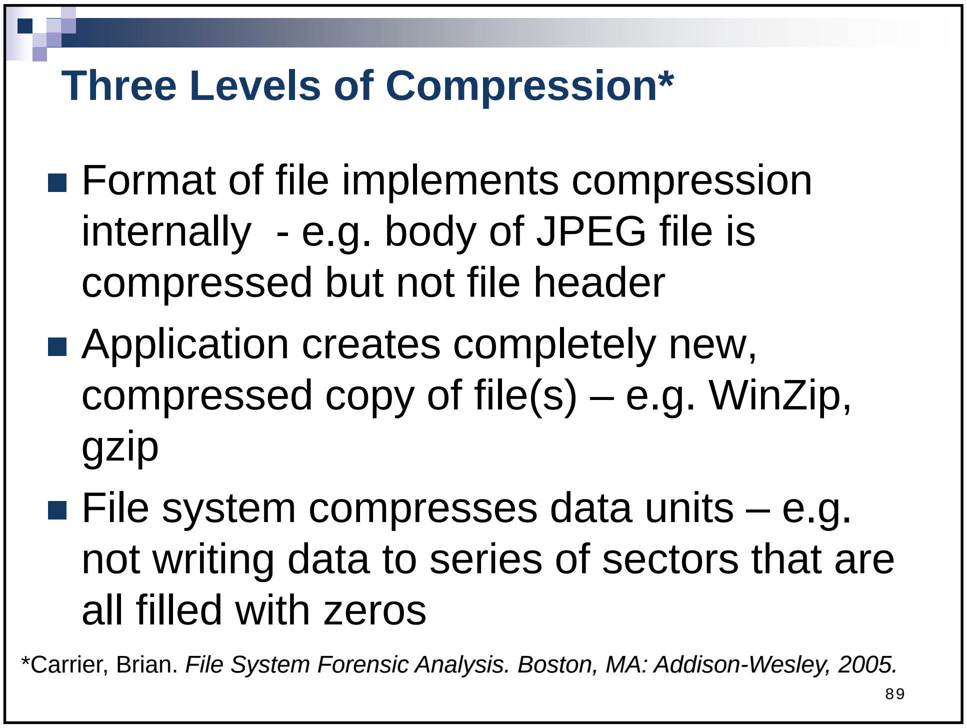
sequence of runs: 0:6, 1:14, 0:13, 1:9

resulting 5-bit bytes: 00110, 11110, 01101, 11001

compressed bit stream: 00110111100110111001 (20 bits)

Figure 7: Compressing a bit stream

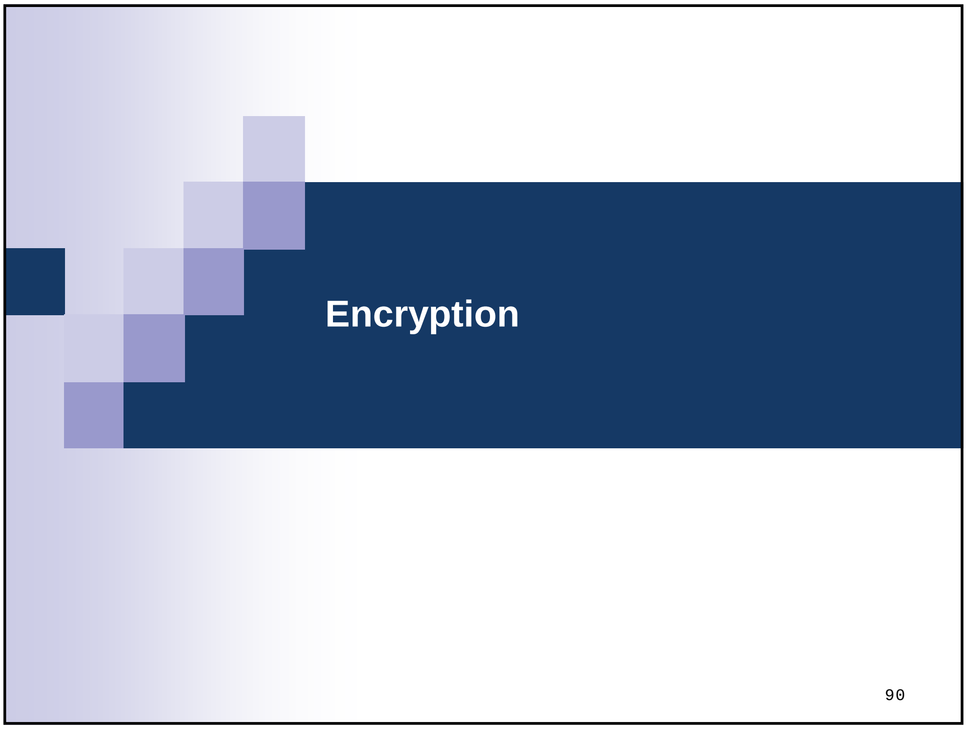
Rothenberg, Jeff. "Ensuring the Longevity of Digital Information." Washington, DC: Council on Library and Information Resources, 1999



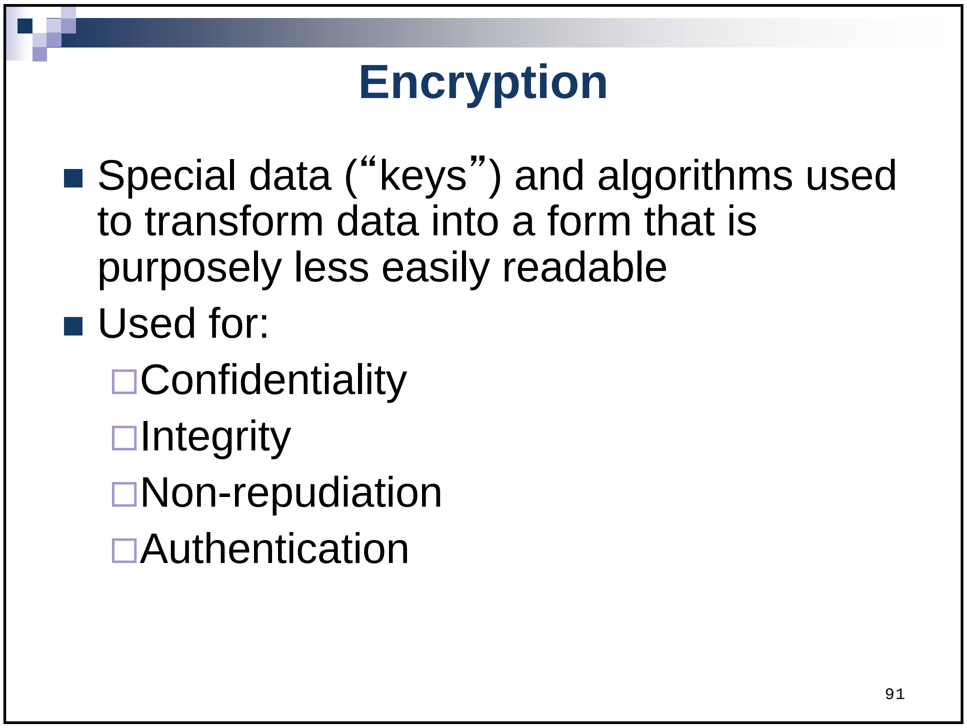
Three Levels of Compression*

- Format of file implements compression internally - e.g. body of JPEG file is compressed but not file header
- Application creates completely new, compressed copy of file(s) – e.g. WinZip, gzip
- File system compresses data units – e.g. not writing data to series of sectors that are all filled with zeros

*Carrier, Brian. *File System Forensic Analysis. Boston, MA: Addison-Wesley, 2005.*

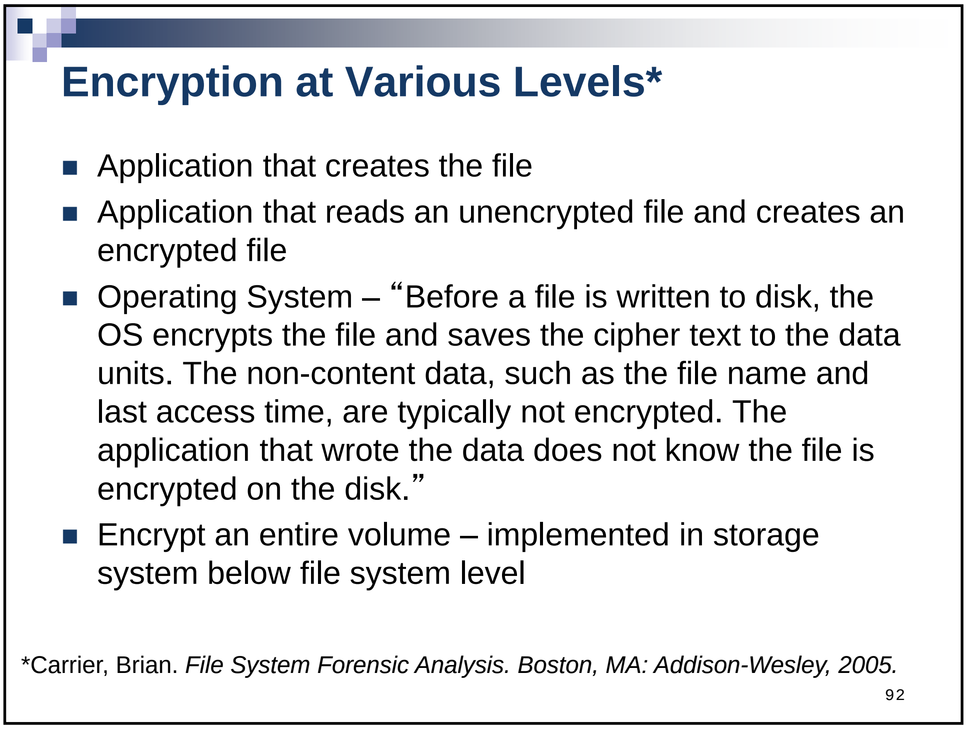


Encryption



Encryption

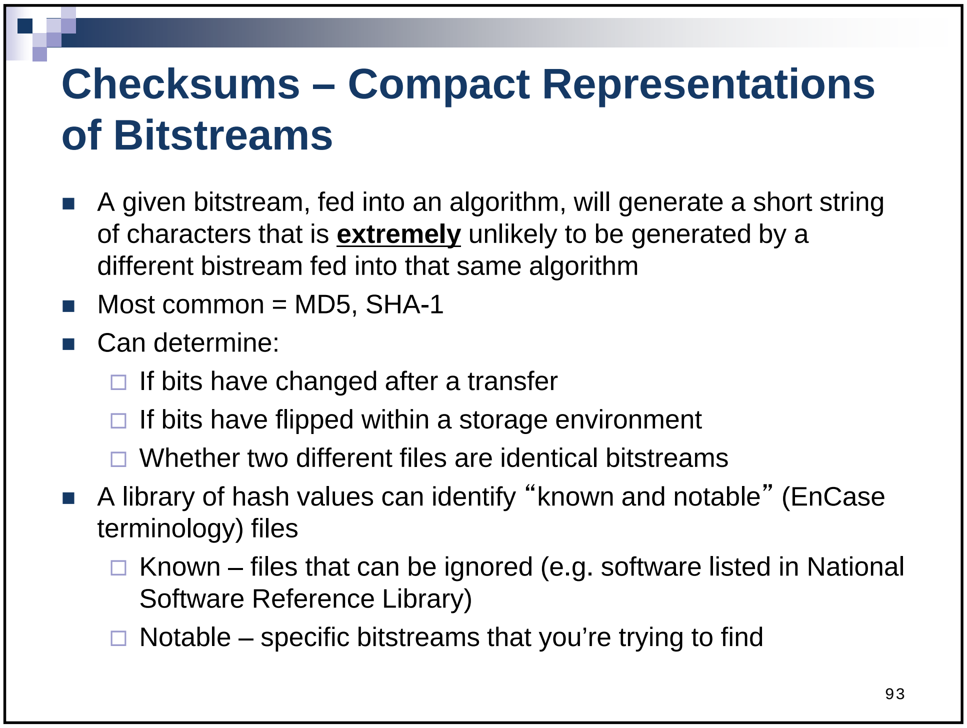
- Special data (“keys”) and algorithms used to transform data into a form that is purposely less easily readable
- Used for:
  - □Confidentiality
  - □Integrity
  - □Non-repudiation
  - □Authentication



Encryption at Various Levels*

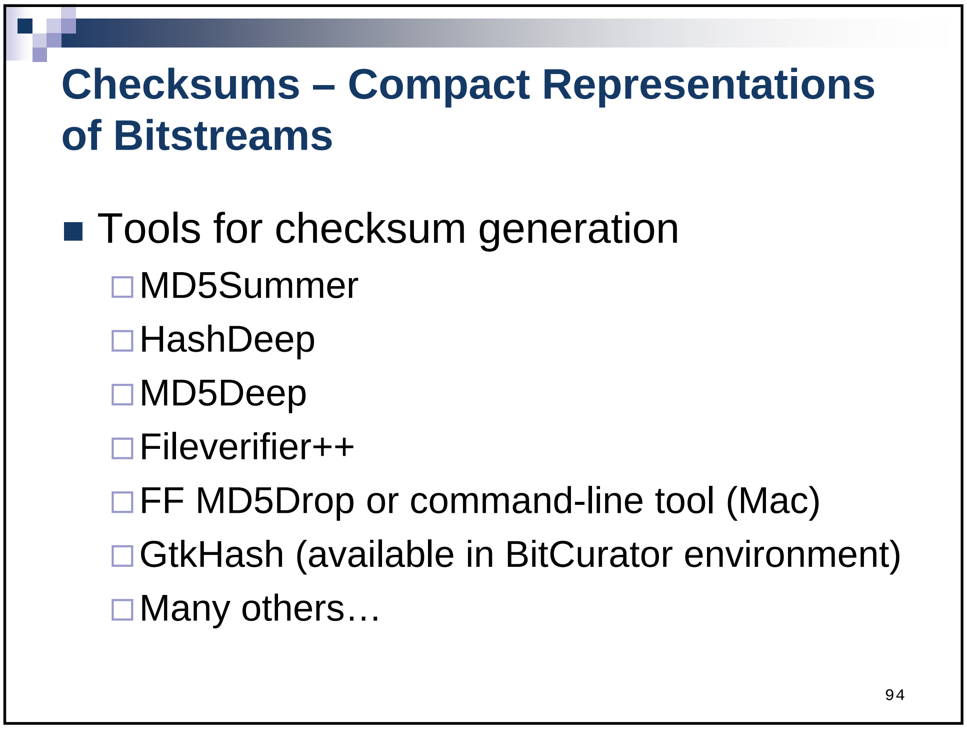
- Application that creates the file
- Application that reads an unencrypted file and creates an encrypted file
- Operating System – “Before a file is written to disk, the OS encrypts the file and saves the cipher text to the data units. The non-content data, such as the file name and last access time, are typically not encrypted. The application that wrote the data does not know the file is encrypted on the disk.”
- Encrypt an entire volume – implemented in storage system below file system level

*Carrier, Brian. *File System Forensic Analysis. Boston, MA: Addison-Wesley, 2005.*



Checksums – Compact Representations of Bitstreams

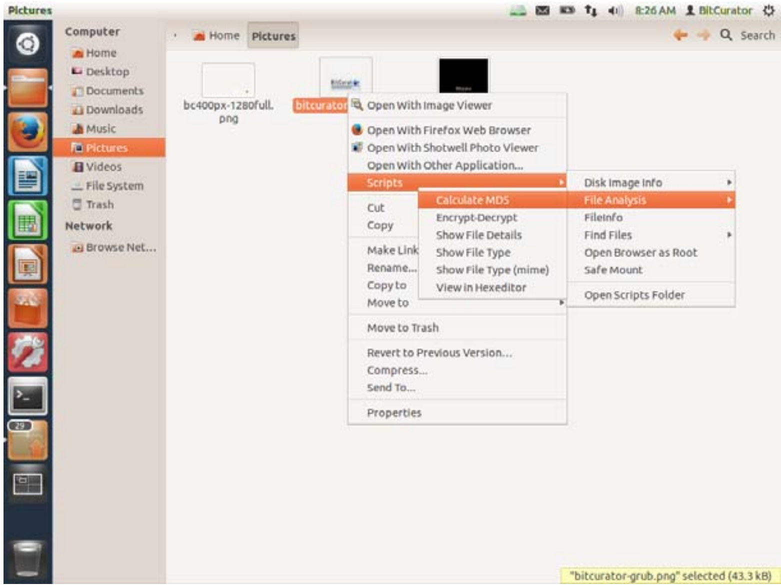
- A given bitstream, fed into an algorithm, will generate a short string of characters that is **extremely** unlikely to be generated by a different bistream fed into that same algorithm
- Most common = MD5, SHA-1
- Can determine:
  - □ If bits have changed after a transfer
  - □ If bits have flipped 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



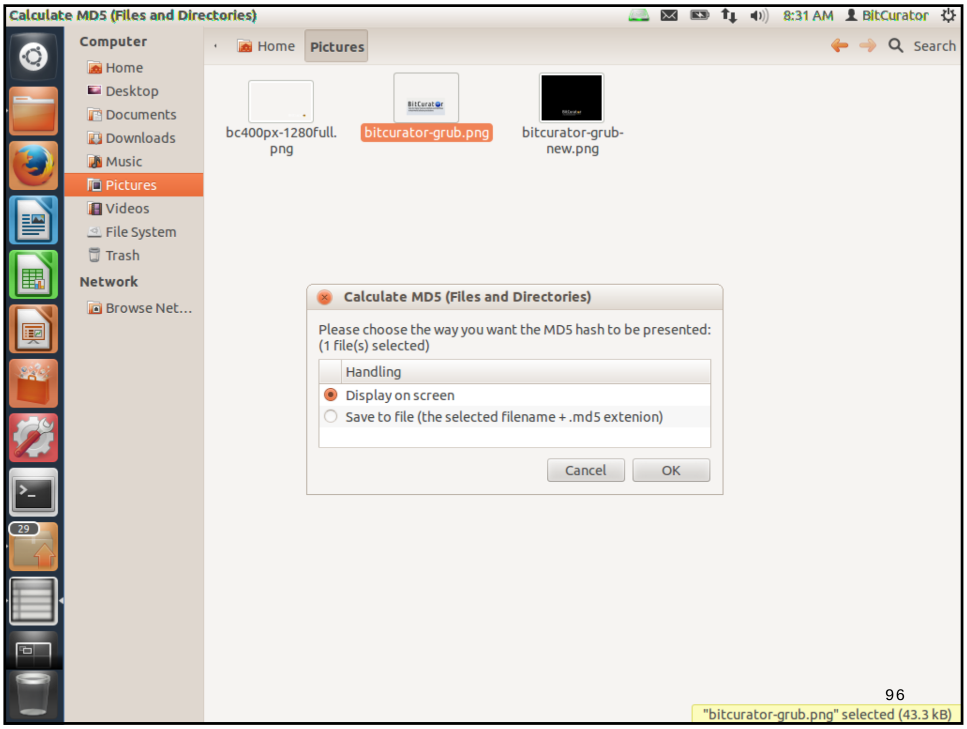
Checksums – Compact Representations of Bitstreams

- Tools for checksum generation
  - □ MD5Summer
  - □ HashDeep
  - □ MD5Deep
  - □ Fileverifier++
  - □ FF MD5Drop or command-line tool (Mac)
  - □ GtkHash (available in BitCurator environment)
  - □ Many others…

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



95



Calculate MD5 (Files and Directories)

Please choose the way you want the MD5 hash to be presented:
(1 file(s) selected)

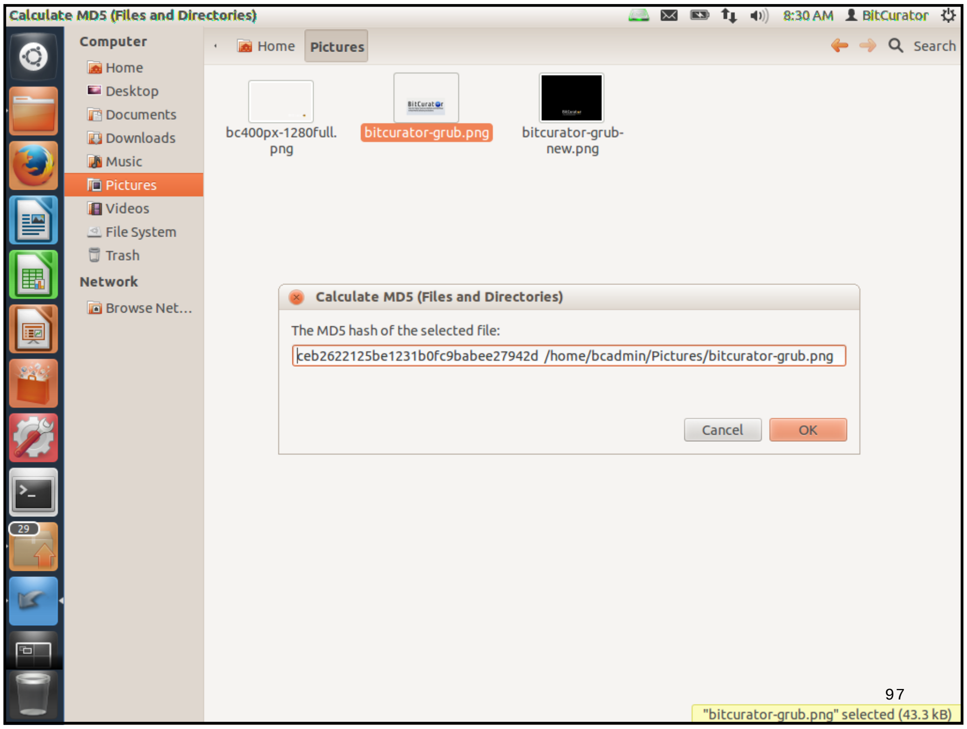
Handling

● Display on screen

○ Save to file (the selected filename + .md5 extenion)

Cancel OK

96



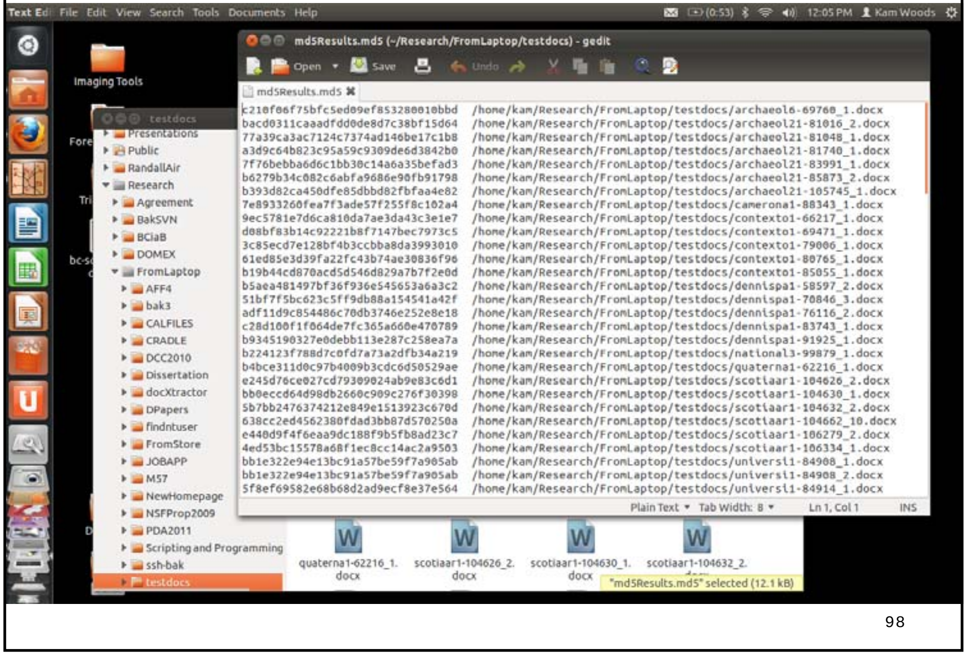
Calculate MD5 (Files and Directories)

The MD5 hash of the selected file:

```
ceb2622125be1231b0fc9babee27942d  /home/bcadmin/Pictures/bitcurator-grub.png
```

"bitcurator-grub.png" selected (43.3 kB)

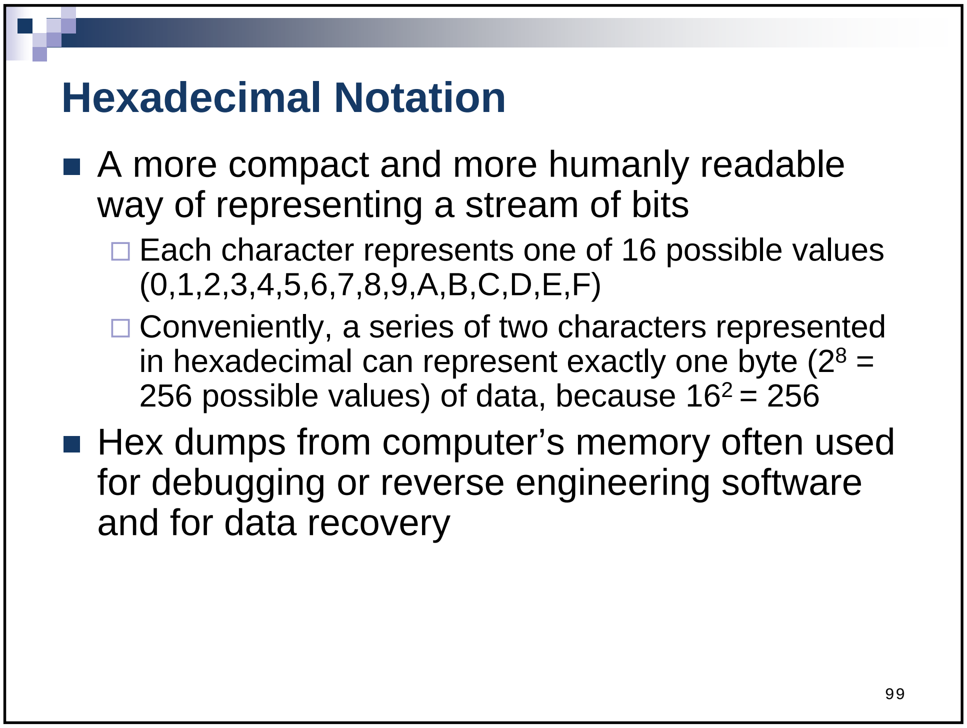
MD5 Hashes of an Entire Directory of Files



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

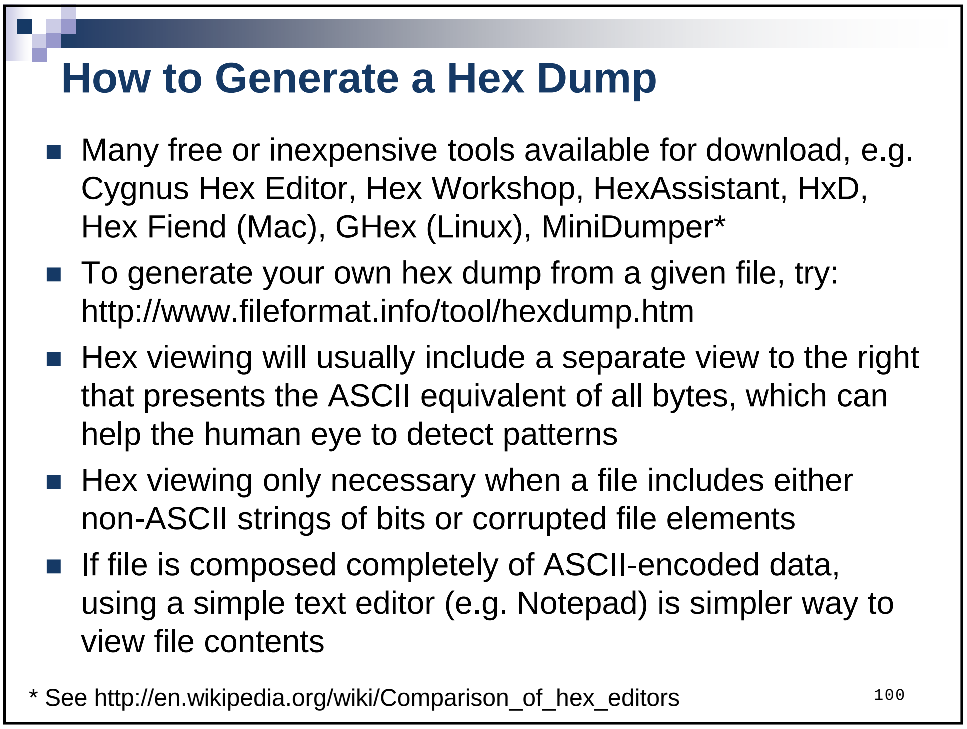
```
c210f06f75bfc5ed09ef853280010bbd  /home/kam/Research/FromLaptop/testdocs/archaeol6-69760_1.docx
bacd0311caaadfdd0de8d7c38bf15d64  /home/kam/Research/FromLaptop/testdocs/archaeol21-81016_2.docx
77a39ca3ac7124c7374ad146be17c1b8  /home/kam/Research/FromLaptop/testdocs/archaeol21-81048_1.docx
a3d9c64b823c95a59c9309de6d3842b0  /home/kam/Research/FromLaptop/testdocs/archaeol21-81740_1.docx
7f76bebba6d6c1bb30c14a6a35befad3  /home/kam/Research/FromLaptop/testdocs/archaeol21-83991_1.docx
b6279b34c082c6abfa9686e90fb91798  /home/kam/Research/FromLaptop/testdocs/archaeol21-85873_2.docx
b393d82ca450dfe85dbbd82fbfaa4e82  /home/kam/Research/FromLaptop/testdocs/archaeol21-105745_1.docx
7e8933260fea7f3ade57f255f8c102a4  /home/kam/Research/FromLaptop/testdocs/camerona1-88343_1.docx
9ec5781e7d6ca810da7ae3da43c3e1e7  /home/kam/Research/FromLaptop/testdocs/contexto1-66217_1.docx
d08bf83b14c92221b8f7147bec7973c5  /home/kam/Research/FromLaptop/testdocs/contexto1-69471_1.docx
3c85ecd7e128bf4b3ccbba8da3993010  /home/kam/Research/FromLaptop/testdocs/contexto1-79006_1.docx
61ed85e3d39fa22fc43b74ae30836f96  /home/kam/Research/FromLaptop/testdocs/contexto1-80765_1.docx
b19b44cd870acd5d546d829a7b7f2e0d  /home/kam/Research/FromLaptop/testdocs/contexto1-85055_1.docx
b5aea481497bf36f936e545653a6a3c2  /home/kam/Research/FromLaptop/testdocs/dennispa1-58597_2.docx
51bf7f5bc623c5ff9db88a154541a42f  /home/kam/Research/FromLaptop/testdocs/dennispa1-70846_3.docx
adf11d9c854486c70db3746e252e8e18  /home/kam/Research/FromLaptop/testdocs/dennispa1-76116_2.docx
c28d100f1f064de7fc365a660e470789  /home/kam/Research/FromLaptop/testdocs/dennispa1-83743_1.docx
b9345190327e0debb113e287c258ea7a  /home/kam/Research/FromLaptop/testdocs/dennispa1-91925_1.docx
b224123f788d7c0fd7a73a2dfb34a219  /home/kam/Research/FromLaptop/testdocs/national3-99879_1.docx
b4bce311d0c97b4009b3cdc6d50529ae  /home/kam/Research/FromLaptop/testdocs/quaterna1-62216_1.docx
e245d76ce027cd79309024ab9e83c6d1  /home/kam/Research/FromLaptop/testdocs/scotiaar1-104626_2.docx
bb0eccd64d98db2660c909c276f30398  /home/kam/Research/FromLaptop/testdocs/scotiaar1-104630_1.docx
5b7bb2476374212e849e1513923c670d  /home/kam/Research/FromLaptop/testdocs/scotiaar1-104632_2.docx
638cc2ed4562380fdad3bb87d570250a  /home/kam/Research/FromLaptop/testdocs/scotiaar1-104662_10.docx
e440d9f4f6eaa9dc188f9b5fb8ad23c7  /home/kam/Research/FromLaptop/testdocs/scotiaar1-106279_2.docx
4ed53bc15578a68f1ec8cc14ac2a9503  /home/kam/Research/FromLaptop/testdocs/scotiaar1-106334_1.docx
bb1e322e94e13bc91a57be59f7a905ab  /home/kam/Research/FromLaptop/testdocs/universi1-84908_1.docx
bb1e322e94e13bc91a57be59f7a905ab  /home/kam/Research/FromLaptop/testdocs/universi1-84908_2.docx
5f8ef69582e68b68d2ad9ecf8e37e564  /home/kam/Research/FromLaptop/testdocs/universi1-84914_1.docx
```

"md5Results.md5" selected (12.1 kB)



Hexadecimal Notation

- A more compact and more humanly readable way of representing a stream of bits
  - □ 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

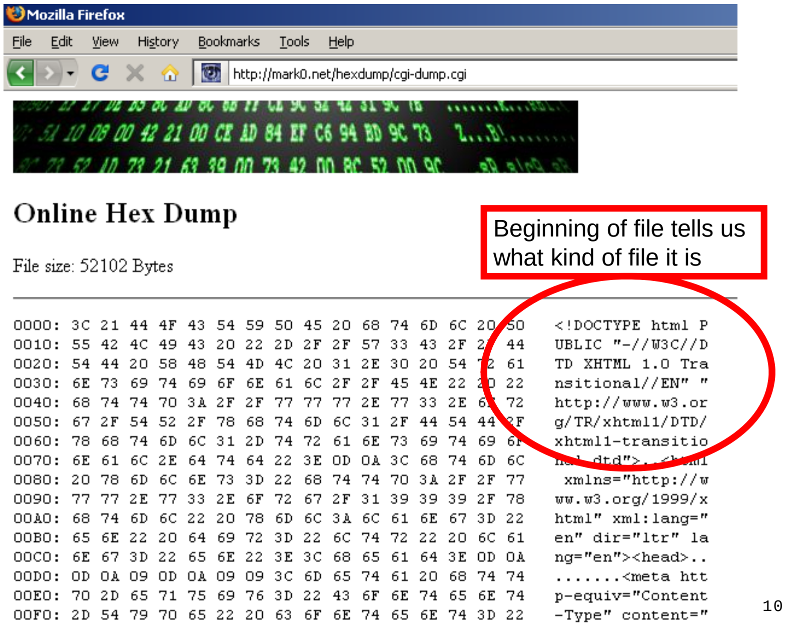


How to Generate a Hex Dump

- Many free or inexpensive tools available for download, e.g. Cygnus Hex Editor, Hex Workshop, HexAssistant, HxD, Hex Fiend (Mac), GHex (Linux), MiniDumper*
- To generate your own hex dump from a given file, try: http://www.fileformat.info/tool/hexdump.htm
- Hex viewing will usually include a separate view to the right that presents the ASCII equivalent of all bytes, which can help the human eye to detect patterns
- Hex viewing only necessary when a file includes either non-ASCII strings of bits or corrupted file elements
- If file is composed completely of ASCII-encoded data, using a simple text editor (e.g. Notepad) is simpler way to view file contents

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

Syllabus for a class (HTML File)


Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://mark0.net/hexdump/cgi-dump.cgi

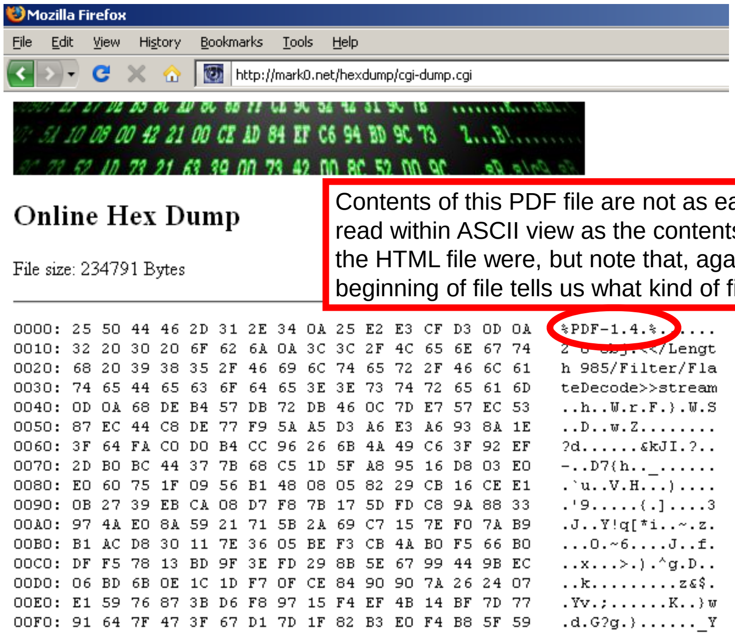
Online Hex Dump

File size: 52102 Bytes

Beginning of file tells us what kind of file it is

```
0000: 3C 21 44 4F 43 54 59 50 45 20 68 74 6D 6C 20 50   <!DOCTYPE html P
0010: 55 42 4C 49 43 20 22 2D 2F 2F 57 33 43 2F 2F 44   UBLIC "-//W3C//D
0020: 54 44 20 58 48 54 4D 4C 20 31 2E 30 20 54 72 61   TD XHTML 1.0 Tra
0030: 6E 73 69 74 69 6F 6E 61 6C 2F 2F 45 4E 22 20 22   nsitional//EN" "
0040: 68 74 74 70 3A 2F 2F 77 77 77 2E 77 33 2E 6F 72   http://www.w3.or
0050: 67 2F 54 52 2F 78 68 74 6D 6C 31 2F 44 54 44 2F   g/TR/xhtml1/DTD/
0060: 78 68 74 6D 6C 31 2D 74 72 61 6E 73 69 74 69 6F   xhtml1-transitio
0070: 6E 61 6C 2E 64 74 64 22 3E 0D 0A 3C 68 74 6D 6C   nal.dtd">..<html
0080: 20 78 6D 6C 6E 73 3D 22 68 74 74 70 3A 2F 2F 77    xmlns="http://w
0090: 77 77 2E 77 33 2E 6F 72 67 2F 31 39 39 39 2F 78   ww.w3.org/1999/x
00A0: 68 74 6D 6C 22 20 78 6D 6C 3A 6C 61 6E 67 3D 22   html" xml:lang="
00B0: 65 6E 22 20 64 69 72 3D 22 6C 74 72 22 20 6C 61   en" dir="ltr" la
00C0: 6E 67 3D 22 65 6E 22 3E 3C 68 65 61 64 3E 0D 0A   ng="en"><head>..
00D0: 0D 0A 09 0D 0A 09 09 3C 6D 65 74 61 20 68 74 74   .......<meta htt
00E0: 70 2D 65 71 75 69 76 3D 22 43 6F 6E 74 65 6E 74   p-equiv="Content
00F0: 2D 54 79 70 65 22 20 63 6F 6E 74 65 6E 74 3D 22   -Type" content="
```

Slides from a lecture (PDF/A file)


Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://mark0.net/hexdump/cgi-dump.cgi

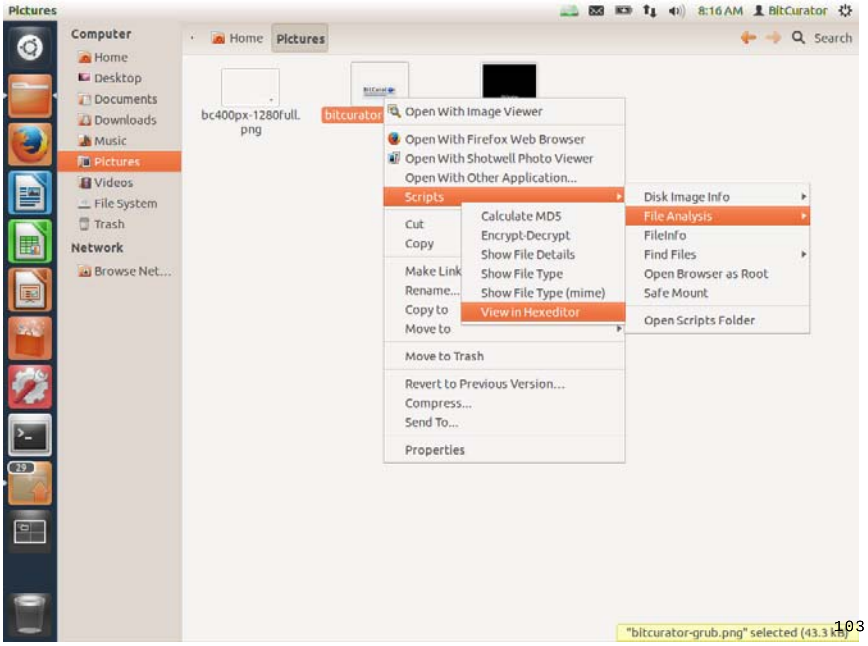
Online Hex Dump

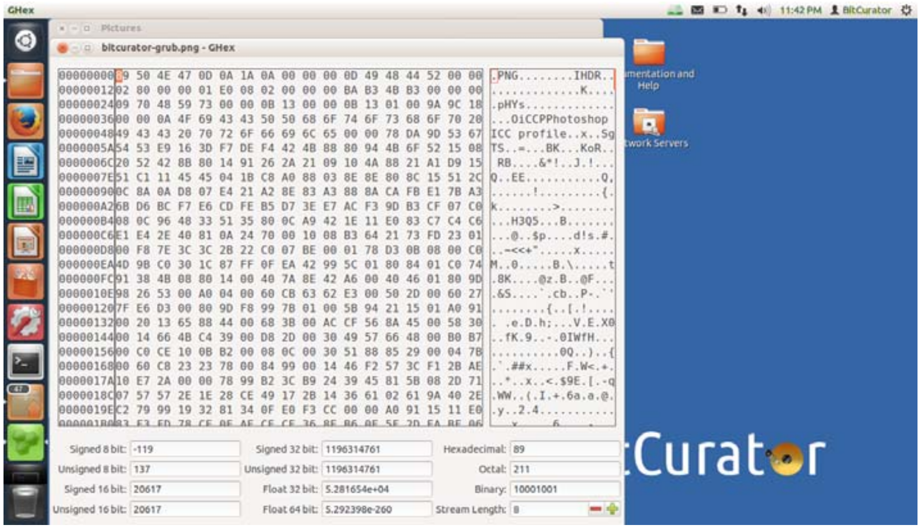
File size: 234791 Bytes

Contents of this PDF file are not as easy to read within ASCII view as the contents of the HTML file were, but note that, again, beginning of file tells us what kind of file it is

```
0000: 25 50 44 46 2D 31 2E 34 0A 25 E2 E3 CF D3 0D 0A   %PDF-1.4.%......
0010: 32 20 30 20 6F 62 6A 0A 3C 3C 2F 4C 65 6E 67 74   2 0 obj.<</Lengt
0020: 68 20 39 38 35 2F 46 69 6C 74 65 72 2F 46 6C 61   h 985/Filter/Fla
0030: 74 65 44 65 63 6F 64 65 3E 3E 73 74 72 65 61 6D   teDecode>>stream
0040: 0D 0A 68 DE B4 57 DB 72 DB 46 0C 7D E7 57 EC 53   ..h..W.r.F.}.W.S
0050: 87 EC 44 C8 DE 77 F9 5A A5 D3 A6 E3 A6 93 8A 1E   ..D..w.Z........
0060: 3F 64 FA C0 D0 B4 CC 96 26 6B 4A 49 C6 3F 92 EF   ?d......&kJI.?..
0070: 2D B0 BC 44 37 7B 68 C5 1D 5F A8 95 16 D8 03 E0   -..D7{h.._......
0080: E0 60 75 1F 09 56 B1 48 08 05 82 29 CB 16 CE E1   .`u..V.H...)....
0090: 0B 27 39 EB CA 08 D7 F8 7B 17 5D FD C8 9A 88 33   .'9.....{.]....3
00A0: 97 4A E0 8A 59 21 71 5B 2A 69 C7 15 7E F0 7A B9   .J..Y!q[*i..~.z.
00B0: B1 AC D8 30 11 7E 36 05 BE F3 CB 4A B0 F5 66 B0   ...0.~6....J..f.
00C0: DF F5 78 13 BD 9F 3E FD 29 8B 5E 67 99 44 9B EC   ..x...>.).^g.D..
00D0: 06 BD 6B 0E 1C 1D F7 0F CE 84 90 90 7A 26 24 07   ..k.........z&$.
00E0: E1 59 76 87 3B D6 F8 97 15 F4 EF 4B 14 BF 7D 77   .Yv.;......K..}w
00F0: 91 64 7F 47 3F 67 D1 7D 1F 82 B3 E0 F4 B8 5F 59   .d.G?g.}......_Y
```

In the BitCurator environment:

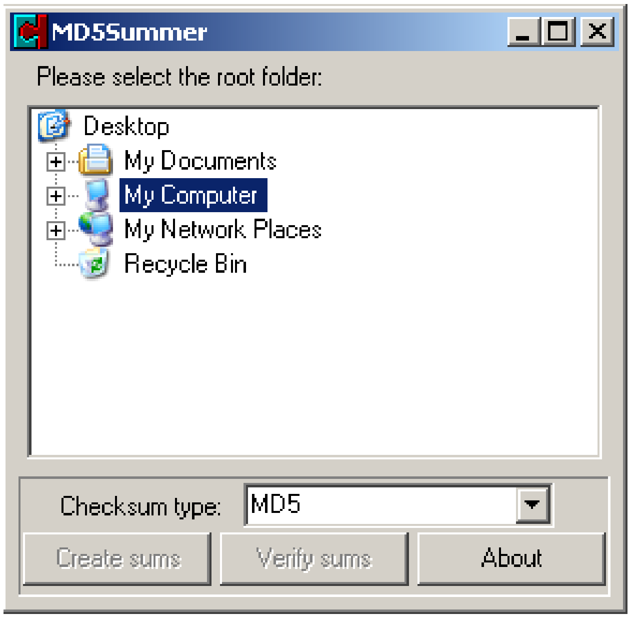




```
0000000089 50 4E 47 0D 0A 1A 0A 00 00 00 0D 49 48 44 52 00 00 .PNG........IHDR..
0000001202 80 00 00 01 E0 08 02 00 00 00 BA B3 4B B3 00 00 00 ..............K....
0000002409 70 48 59 73 00 00 0B 13 00 00 0B 13 01 00 9A 9C 18 .pHYs.............
0000003600 00 0A 4F 69 43 43 50 50 68 6F 74 6F 73 68 6F 70 20 ...OiCCPPhotoshop 
0000004849 43 43 20 70 72 6F 66 69 6C 65 00 00 78 DA 9D 53 67 ICC profile..x..Sg
0000005A54 53 E9 16 3D F7 DE F4 42 4B 88 80 94 4B 6F 52 15 08 TS..=...BK...KoR..
0000006C20 52 42 8B 80 14 91 26 2A 21 09 10 4A 88 21 A1 D9 15  RB....&*!..J.!...
0000007E51 C1 11 45 45 04 1B C8 A0 88 03 8E 8E 80 8C 15 51 2C Q..EE...........Q,
000000900C 8A 0A D8 07 E4 21 A2 8E 83 A3 88 8A CA FB E1 7B A3 .......!.........{.
000000A26B D6 BC F7 E6 CD FE B5 D7 3E E7 AC F3 9D B3 CF 07 C0 k........>........
000000B408 0C 96 48 33 51 35 80 0C A9 42 1E 11 E0 83 C7 C4 C6 ...H3Q5...B.......
000000C6E1 E4 2E 40 81 0A 24 70 00 10 08 B3 64 21 73 FD 23 01 ...@..$p....d!s.#.
000000D800 F8 7E 3C 3C 2B 22 C0 07 BE 00 01 78 D3 0B 08 00 C0 ..~<<+".....x.....
000000EA4D 9B C0 30 1C 87 FF 0F EA 42 99 5C 01 80 84 01 C0 74 M..0.....B.\.....t
000000FC91 38 4B 08 80 14 00 40 7A 8E 42 A6 00 40 46 01 80 9D .8K....@z.B..@F...
0000010E98 26 53 00 A0 04 00 60 CB 63 62 E3 00 50 2D 00 60 27 .&S....`.cb..P-.`'
000001207F E6 D3 00 80 9D F8 99 7B 01 00 5B 94 21 15 01 A0 91 ........{..[.!....
0000013200 20 13 65 88 44 00 68 3B 00 AC CF 56 8A 45 00 58 30 . .e.D.h;...V.E.X0
0000014400 14 66 4B C4 39 00 D8 2D 00 30 49 57 66 48 00 B0 B7 ..fK.9..-.0IWfH...
0000015600 C0 CE 10 0B B2 00 08 0C 00 30 51 88 85 29 00 04 7B ..........0Q..)..{
0000016800 60 C8 23 23 78 00 84 99 00 14 46 F2 57 3C F1 2B AE .`.##x.....F.W<.+.
0000017A10 E7 2A 00 00 78 99 B2 3C B9 24 39 45 81 5B 08 2D 71 ..*..x..<.$9E.[.-q
0000018C07 57 57 2E 1E 28 CE 49 17 2B 14 36 61 02 61 9A 40 2E .WW..(.I.+.6a.a.@.
0000019EC2 79 99 19 32 81 34 0F E0 F3 CC 00 00 A0 91 15 11 E0 .y..2.4...........
```

Signed 8 bit: -119
Unsigned 8 bit: 137
Signed 16 bit: 20617
Unsigned 16 bit: 20617
Signed 32 bit: 1196314761
Unsigned 32 bit: 1196314761
Float 32 bit: 5.281654e+04
Float 64 bit: 5.292398e-260
Hexadecimal: 89
Octal: 211
Binary: 10001001
Stream Length: 8

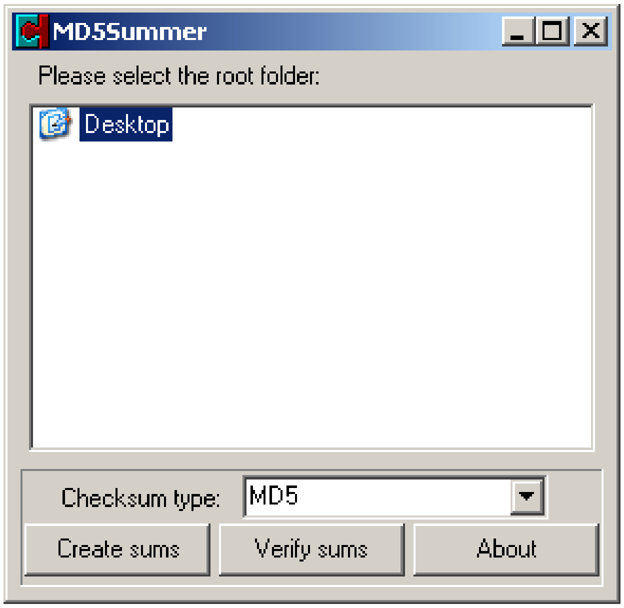
Let’s corrupt a bitstream.



MD5Summer

Please select the root folder:

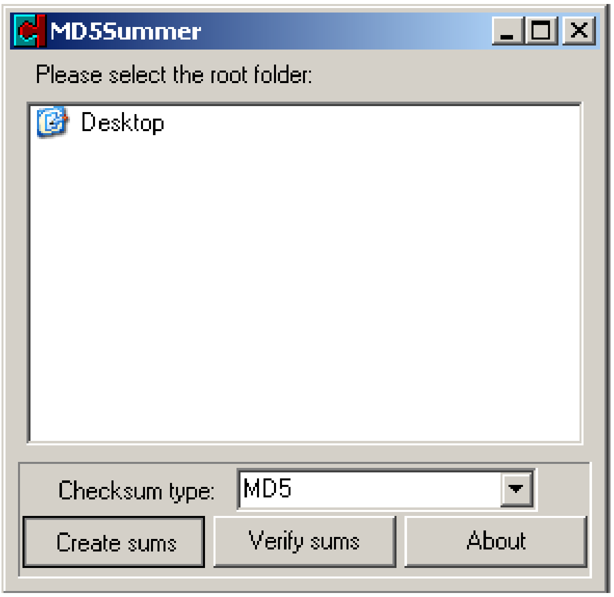
Checksum type: MD5



MD5Summer

Please select the root folder:

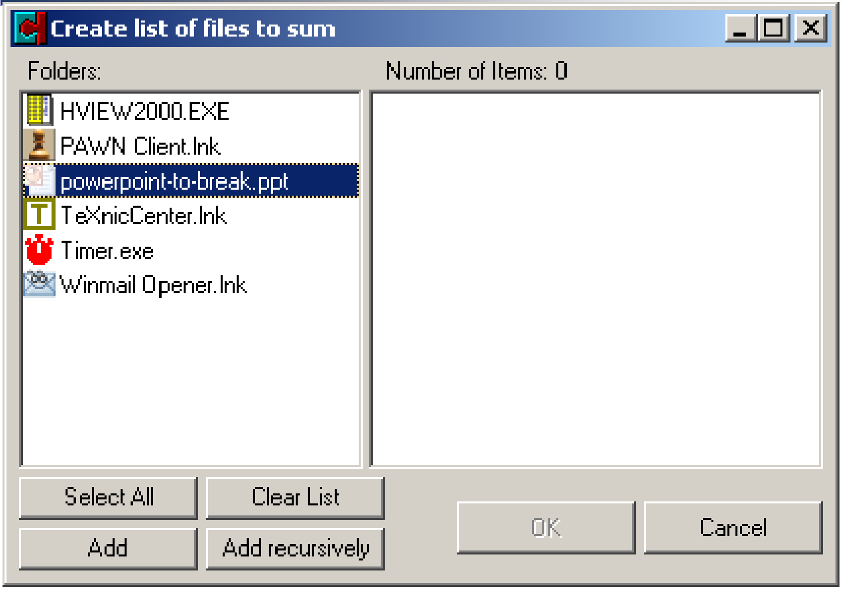
Checksum type: MD5



MD5Summer

Please select the root folder:

Checksum type: MD5

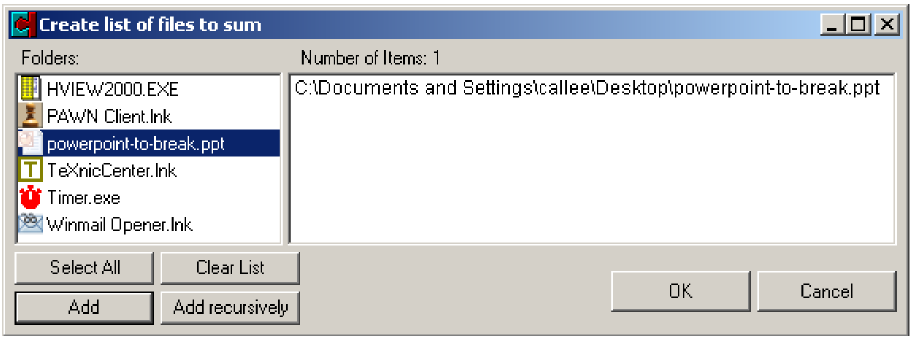


Create list of files to sum

Folders:

- HVIEW2000.EXE
- PAWN Client.lnk
- powerpoint-to-break.ppt
- TeXnicCenter.lnk
- Timer.exe
- Winmail Opener.lnk

Number of Items: 0



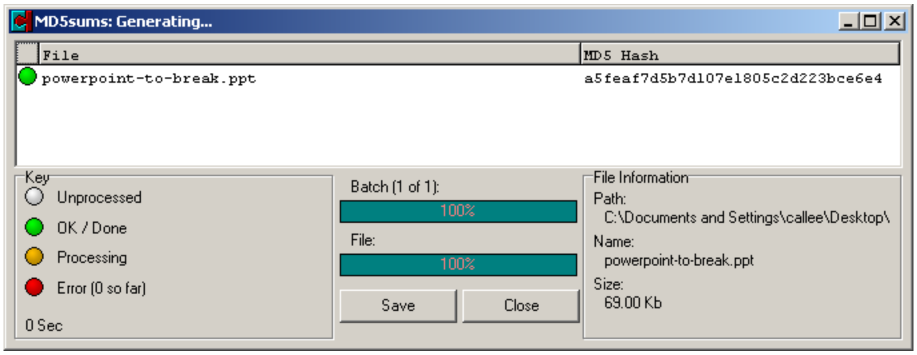
Create list of files to sum

Folders:

- HVIEW2000.EXE
- PAWN Client.lnk
- powerpoint-to-break.ppt
- TeXnicCenter.lnk
- Timer.exe
- Winmail Opener.lnk

Number of Items: 1

C:\Documents and Settings\callee\Desktop\powerpoint-to-break.ppt

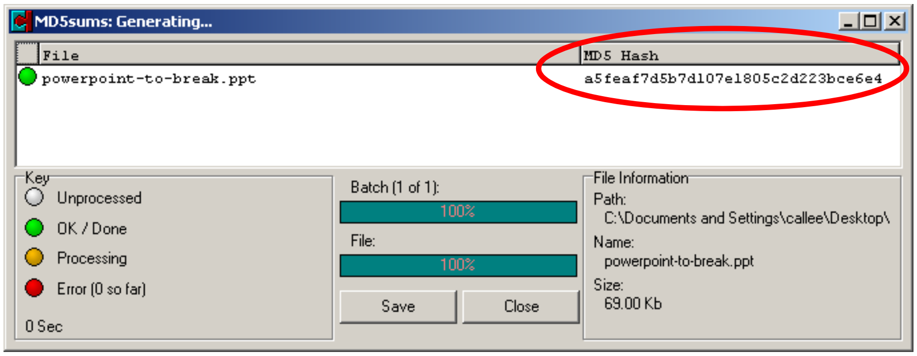


File	MD5 Hash
● powerpoint-to-break.ppt	a5feaf7d5b7d107e1805c2d223bce6e4

Key
○ Unprocessed
● OK / Done
● Processing
● Error (0 so far)
0 Sec

Batch (1 of 1): 100%
File: 100%

File Information
Path:
C:\Documents and Settings\callee\Desktop\
Name:
powerpoint-to-break.ppt
Size:
69.00 Kb

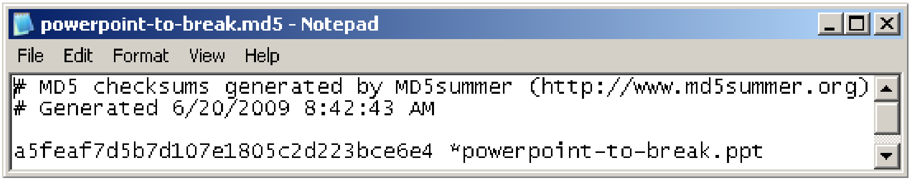


File	MD5 Hash
● powerpoint-to-break.ppt	a5feaf7d5b7d107e1805c2d223bce6e4

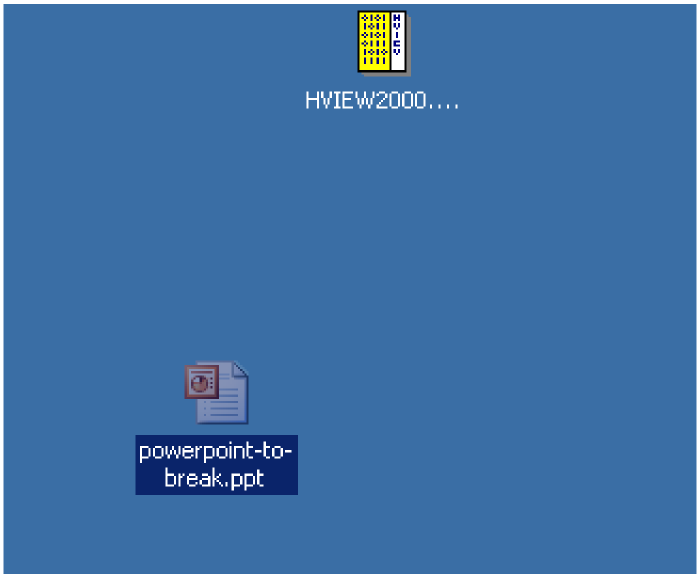
Key
○ Unprocessed
● OK / Done
● Processing
● Error (0 so far)
0 Sec

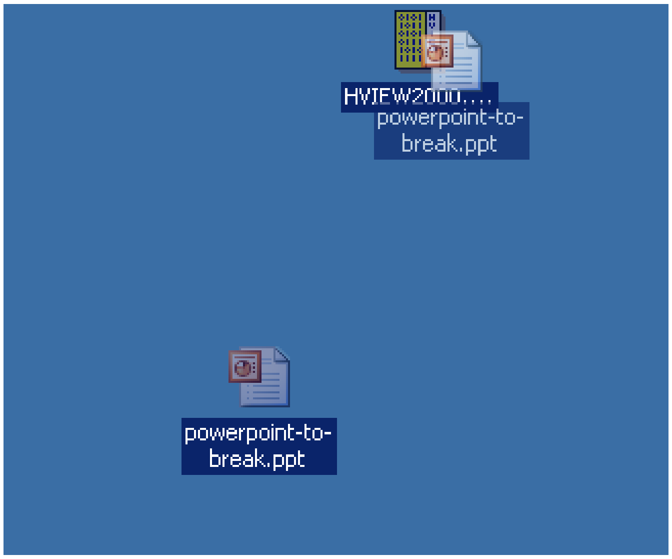
Batch (1 of 1): 100%
File: 100%

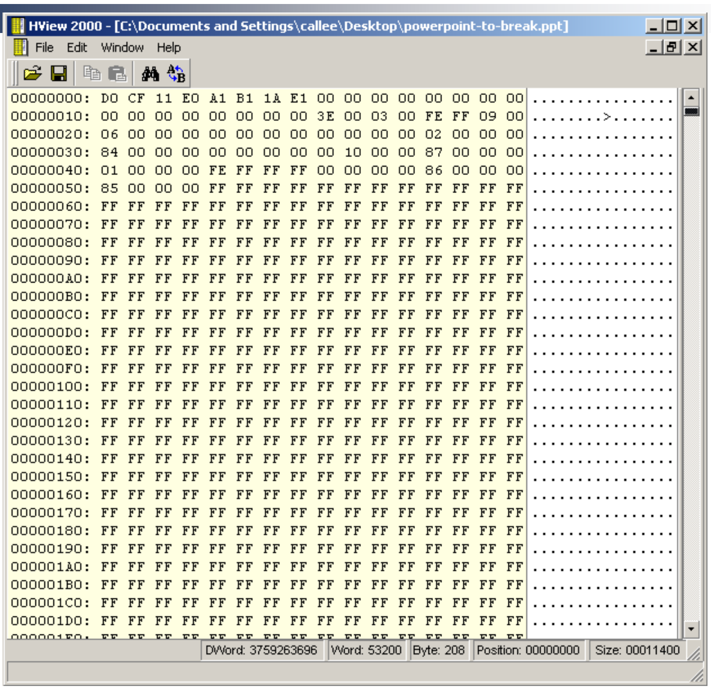
File Information
Path:
C:\Documents and Settings\callee\Desktop\
Name:
powerpoint-to-break.ppt
Size:
69.00 Kb



```
# MD5 checksums generated by MD5summer (http://www.md5summer.org)
# Generated 6/20/2009 8:42:43 AM

a5feaf7d5b7d107e1805c2d223bce6e4 *powerpoint-to-break.ppt
```

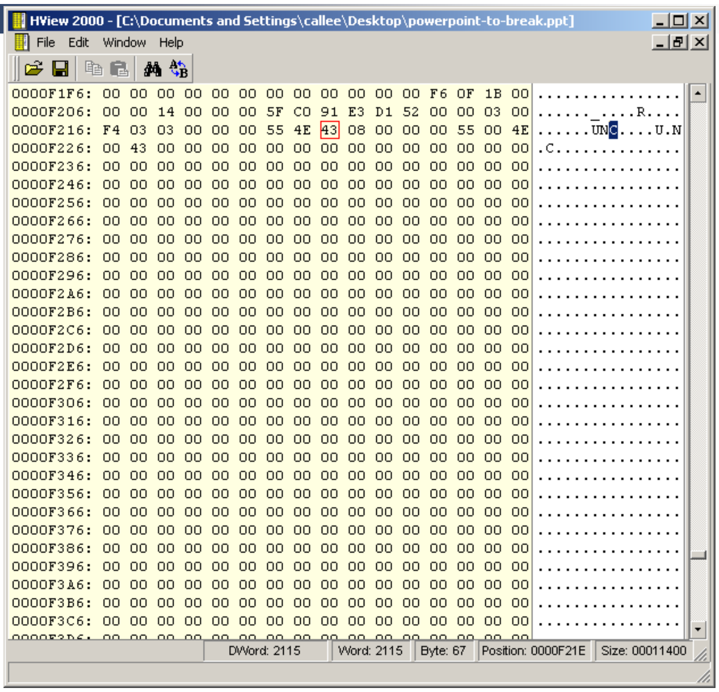





HView 2000 - [C:\Documents and Settings\callee\Desktop\powerpoint-to-break.ppt]
File
Edit
Window
Help

```
00000000: D0 CF 11 E0 A1 B1 1A E1 00 00 00 00 00 00 00 00  ................
00000010: 00 00 00 00 00 00 00 00 3E 00 03 00 FE FF 09 00  ........>.......
00000020: 06 00 00 00 00 00 00 00 00 00 00 00 02 00 00 00  ................
00000030: 84 00 00 00 00 00 00 00 00 10 00 00 87 00 00 00  ................
00000040: 01 00 00 00 FE FF FF FF 00 00 00 00 86 00 00 00  ................
00000050: 85 00 00 00 FF FF FF FF FF FF FF FF FF FF FF FF  ................
00000060: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
00000070: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
00000080: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
00000090: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
000000A0: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
000000B0: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
000000C0: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
000000D0: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
000000E0: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
000000F0: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
00000100: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
00000110: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
00000120: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
00000130: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
00000140: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
00000150: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
00000160: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
00000170: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
00000180: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
00000190: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
000001A0: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
000001B0: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
000001C0: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
000001D0: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF  ................
000001E0: FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
```

DWord: 3759263696 | Word: 53200 | Byte: 208 | Position: 00000000 | Size: 00011400

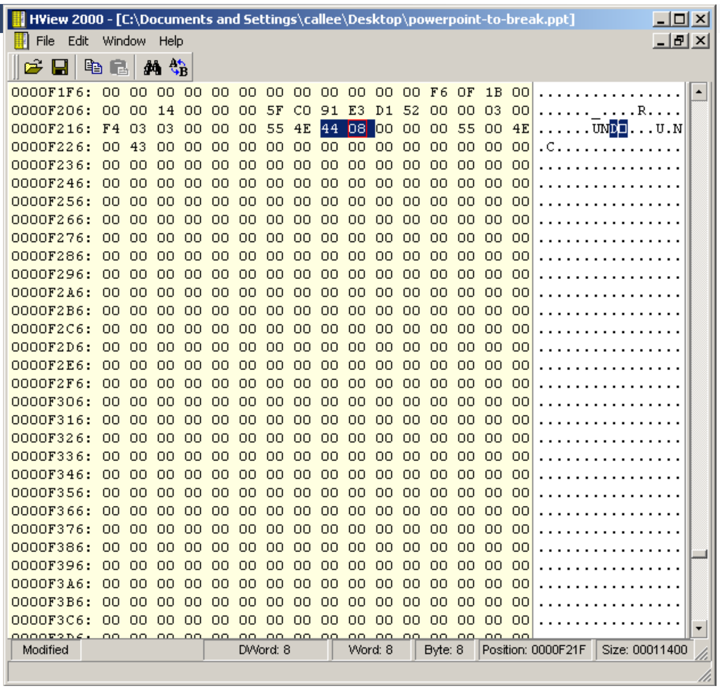


HView 2000 - [C:\Documents and Settings\callee\Desktop\powerpoint-to-break.ppt]
File
Edit
Window
Help

```
0000F1F6: 00 00 00 00 00 00 00 00 00 00 00 00 F6 0F 1B 00 ................
0000F206: 00 00 14 00 00 00 5F C0 91 E3 D1 52 00 00 03 00 ......_....R....
0000F216: F4 03 03 00 00 00 55 4E 43 08 00 00 00 55 00 4E ......UNC....U.N
0000F226: 00 43 00 00 00 00 00 00 00 00 00 00 00 00 00 00 .C..............
0000F236: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F246: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F256: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F266: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F276: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F286: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F296: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F2A6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F2B6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F2C6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F2D6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F2E6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F2F6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F306: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F316: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F326: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F336: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F346: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F356: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F366: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F376: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F386: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F396: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F3A6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F3B6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F3C6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
```

DWord: 2115 | Word: 2115 | Byte: 67 | Position: 0000F21E | Size: 00011400

117

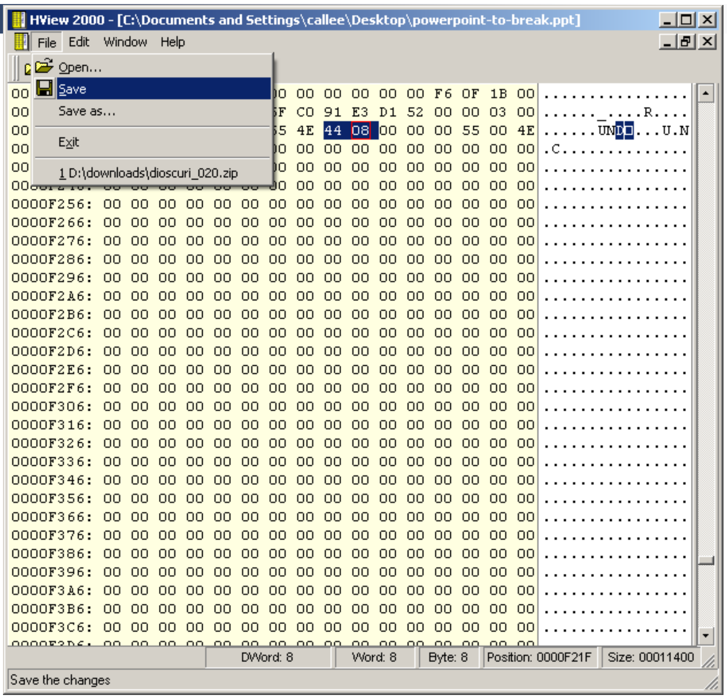


HView 2000 - [C:\Documents and Settings\callee\Desktop\powerpoint-to-break.ppt]
File
Edit
Window
Help

```
0000F1F6: 00 00 00 00 00 00 00 00 00 00 00 00 F6 0F 1B 00 ................
0000F206: 00 00 14 00 00 00 5F C0 91 E3 D1 52 00 00 03 00 ......_....R....
0000F216: F4 03 03 00 00 00 55 4E 44 08 00 00 00 55 00 4E ......UND....U.N
0000F226: 00 43 00 00 00 00 00 00 00 00 00 00 00 00 00 00 .C..............
0000F236: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F246: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F256: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F266: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F276: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F286: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F296: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F2A6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F2B6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F2C6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F2D6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F2E6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F2F6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F306: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F316: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F326: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F336: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F346: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F356: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F366: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F376: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F386: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F396: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F3A6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F3B6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
0000F3C6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
```

Modified | DWord: 8 | Word: 8 | Byte: 8 | Position: 0000F21F | Size: 00011400

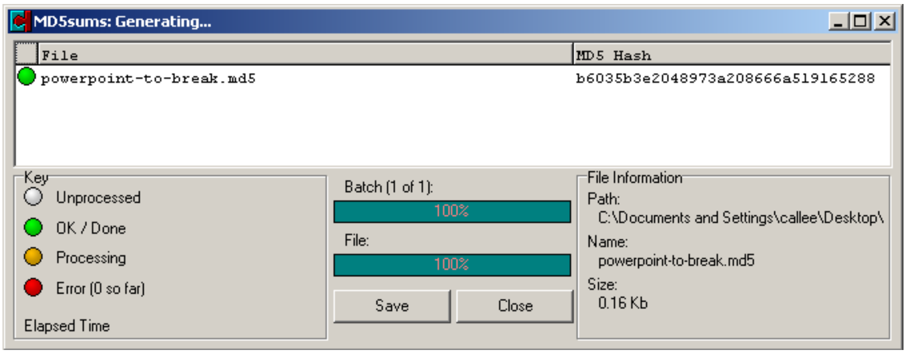
118



HView 2000 - [C:\Documents and Settings\callee\Desktop\powerpoint-to-break.ppt]
File
Edit
Window
Help
Open...
Save
Save as...
Exit
1 D:\downloads\dioscuri_020.zip

```
00                           00 00 00 00 00 F6 0F 1B 00  ................
00                           C0 91 E3 D1 52 00 00 03 00  ......_....R....
00                           4E 44 08 00 00 00 55 00 4E  ......UND□...U.N
00                           00 00 00 00 00 00 00 00 00  .C..............
00                           00 00 00 00 00 00 00 00 00  ................
00                           00 00 00 00 00 00 00 00 00  ................
0000F256: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F266: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F276: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F286: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F296: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F2A6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F2B6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F2C6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F2D6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F2E6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F2F6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F306: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F316: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F326: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F336: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F346: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F356: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F366: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F376: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F386: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F396: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F3A6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F3B6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
0000F3C6: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
```

```
DWord: 8   Word: 8   Byte: 8   Position: 0000F21F   Size: 00011400
Save the changes
```

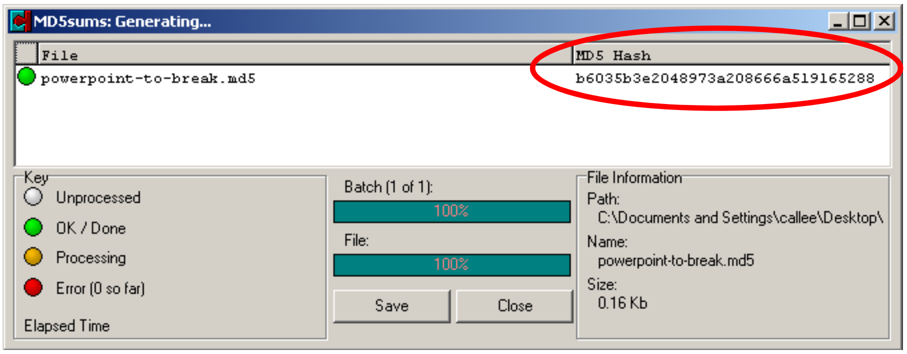


File	MD5 Hash
● powerpoint-to-break.md5	b6035b3e2048973a208666a519165288

Key
○ Unprocessed
● OK / Done
● Processing
● Error (0 so far)
Elapsed Time

Batch (1 of 1): 100%
File: 100%

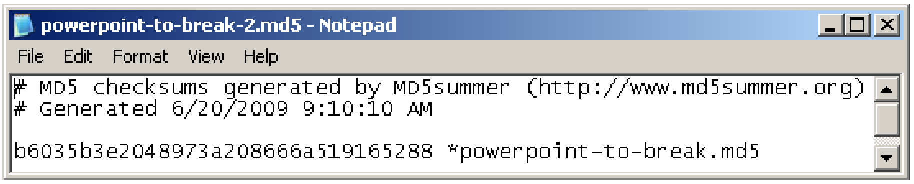
File Information
Path:
C:\Documents and Settings\callee\Desktop\
Name:
powerpoint-to-break.md5
Size:
0.16 Kb



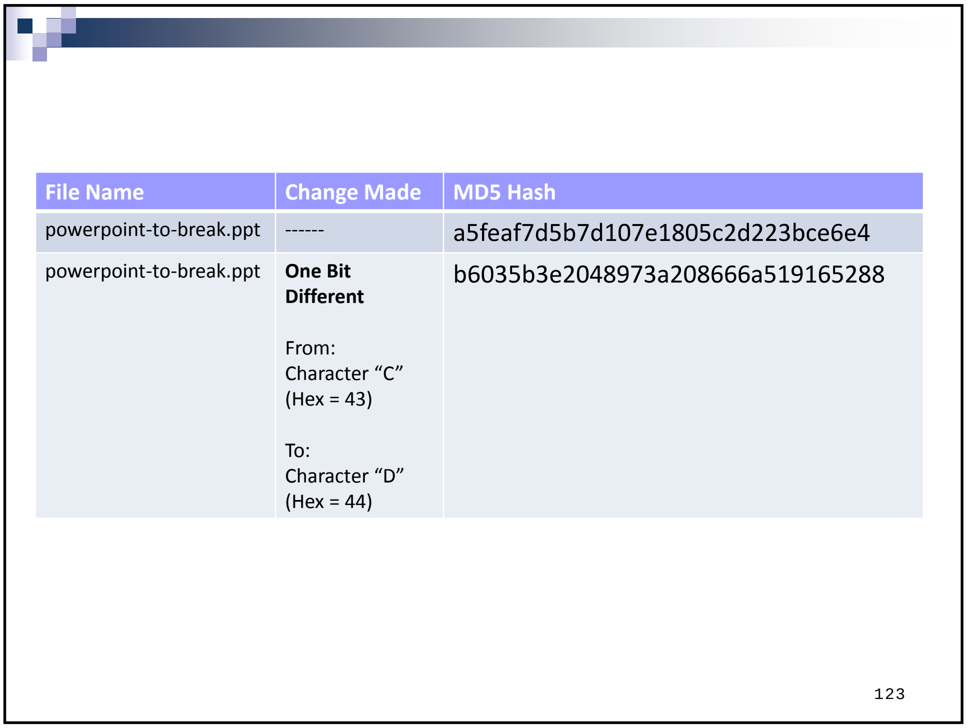
File	MD5 Hash
powerpoint-to-break.md5	b6035b3e2048973a208666a519165288

Path:
C:\Documents and Settings\callee\Desktop\
Name:
powerpoint-to-break.md5
Size:
0.16 Kb

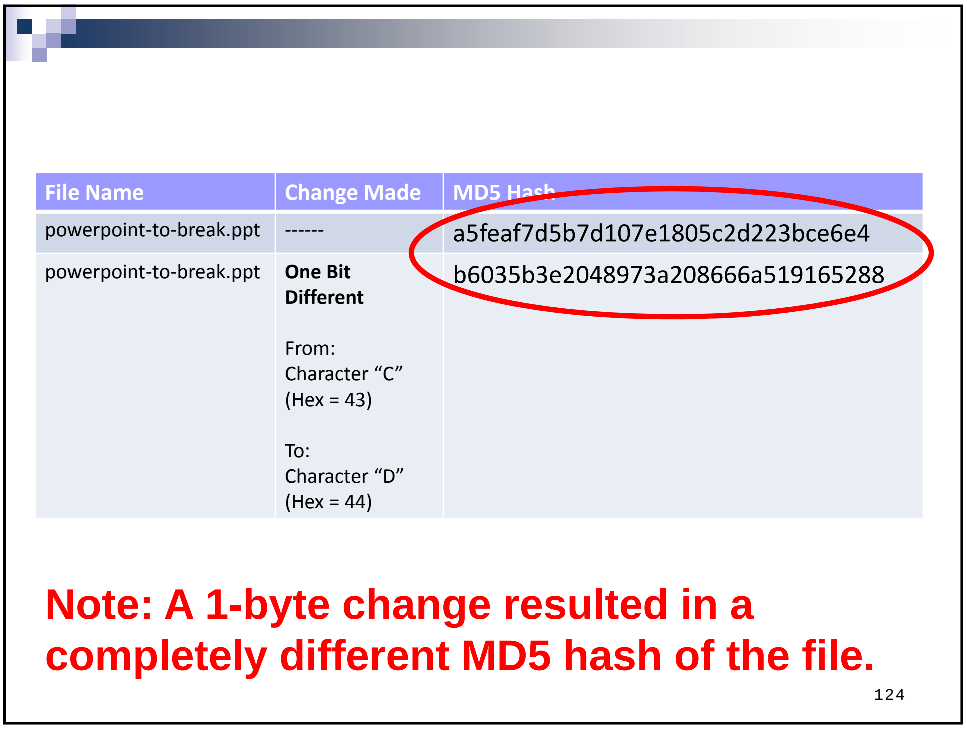
That doesn't look right.
Let's compare it to our previous MD5 hash...


powerpoint-to-break-2.md5 - Notepad
File Edit Format View Help

```
# MD5 checksums generated by MD5summer (http://www.md5summer.org)
# Generated 6/20/2009 9:10:10 AM

b6035b3e2048973a208666a519165288 *powerpoint-to-break.md5
```



File Name	Change Made	MD5 Hash
powerpoint-to-break.ppt	------	a5feaf7d5b7d107e1805c2d223bce6e4
powerpoint-to-break.ppt	**One Bit Different** From: Character “C” (Hex = 43) To: Character “D” (Hex = 44)	b6035b3e2048973a208666a519165288



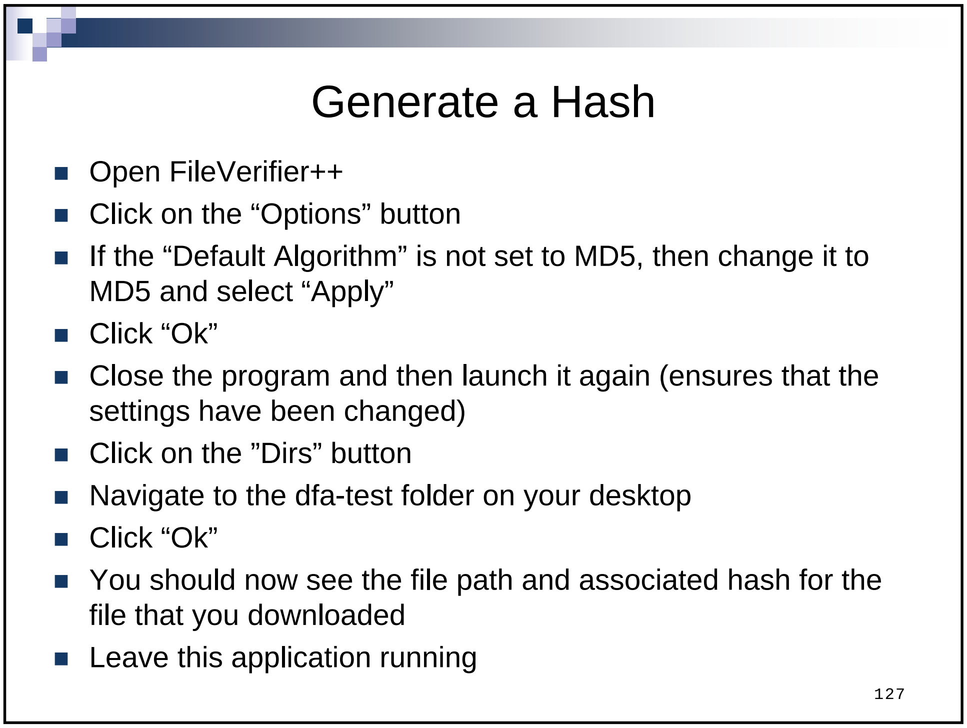
File Name	Change Made	MD5 Hash
powerpoint-to-break.ppt	------	a5feaf7d5b7d107e1805c2d223bce6e4
powerpoint-to-break.ppt	**One Bit Different** From: Character “C” (Hex = 43) To: Character “D” (Hex = 44)	b6035b3e2048973a208666a519165288

Note: A 1-byte change resulted in a completely different MD5 hash of the file.

Now do it yourself!

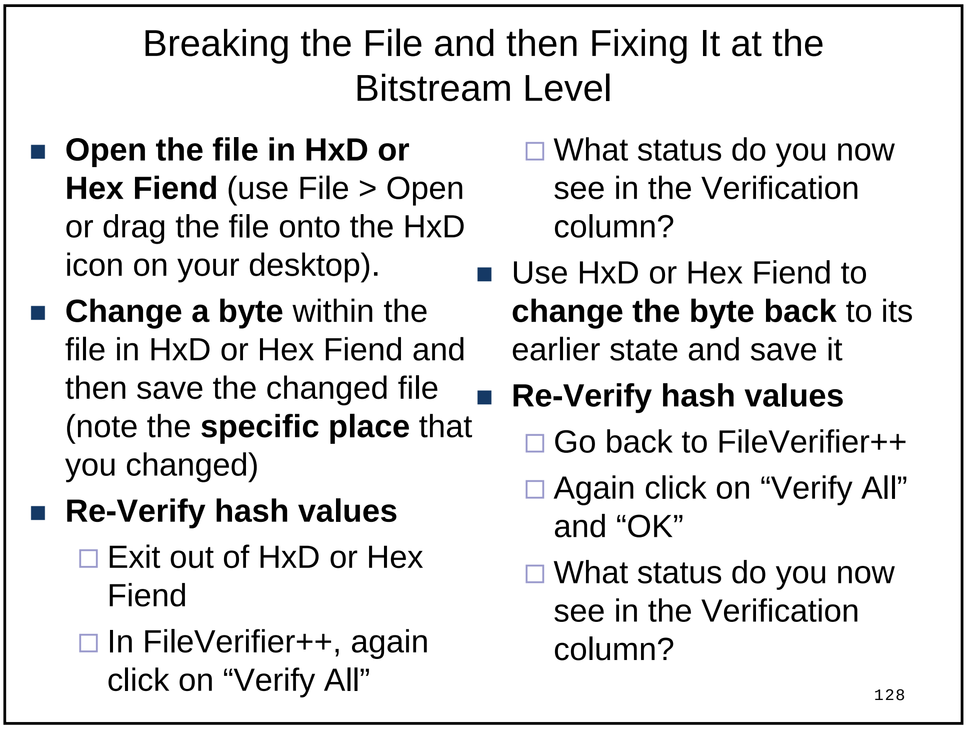
Get yourself a Microsoft Word File:

- Create a new folder on your desktop called dfa-test
- Find a file by searching in Google on a topic of interest along with “filetype:doc” or “filetype:docx”
- Save the file to your dfa-test folder



Generate a Hash

- Open FileVerifier++
- Click on the “Options” button
- If the “Default Algorithm” is not set to MD5, then change it to MD5 and select “Apply”
- Click “Ok”
- Close the program and then launch it again (ensures that the settings have been changed)
- Click on the ”Dirs” button
- Navigate to the dfa-test folder on your desktop
- Click “Ok”
- You should now see the file path and associated hash for the file that you downloaded
- Leave this application running



Breaking the File and then Fixing It at the Bitstream Level

- **Open the file in HxD or Hex Fiend** (use File > Open or drag the file onto the HxD icon on your desktop).
- **Change a byte** within the file in HxD or Hex Fiend and then save the changed file (note the **specific place** that you changed)
- **Re-Verify hash values**
  - □ Exit out of HxD or Hex Fiend
  - □ In FileVerifier++, again click on “Verify All”
  - □ What status do you now see in the Verification column?
- Use HxD or Hex Fiend to **change the byte back** to its earlier state and save it
- **Re-Verify hash values**
  - □ Go back to FileVerifier++
  - □ Again click on “Verify All” and “OK”
  - □ What status do you now see in the Verification column?

Breaking the File and then “Fixing” It Using MS Word

- **Open the file in MS Word**
- **Change one word** within the file’s content (note the **specific place** that you changed) and save it
- **Re-Verify hash values**
  - □ Exit out of MS Office
  - □ In FileVerifier++, again click on “Verify All”
  - □ What status do you now see in the Verification column?
- Use PowerPoint/Word to **change the text back** to its earlier state and save it
- **Re-Verify hash values**
  - □ Go back to FileVerifier++
  - □ Again click on “Verify All” and “OK”
  - □ What status do you now see in the Verification column?

Note on MD5 – Potential Collisions

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

Implications of MD5 Being "Broken"

- Rarely a concern when MD5 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 hashes to address this (SHA-2)
- MD5 is still widely used, because it is fast to calculate and widely supported

What have we covered so far?

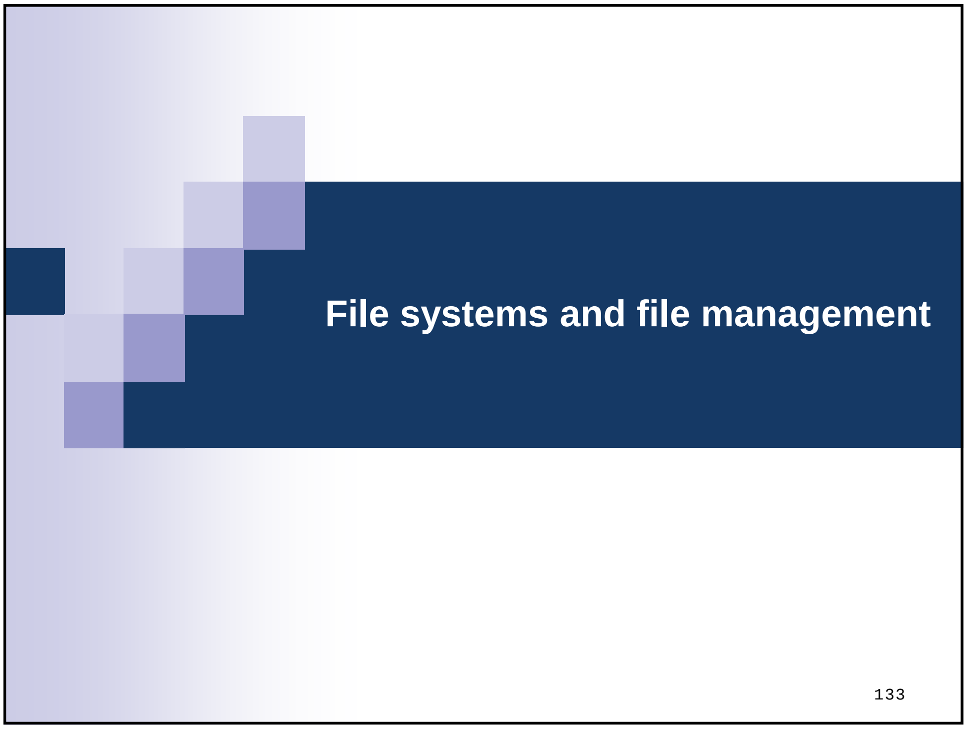
✓ **Welcome and introductions**

✓ **Motivation and scope**

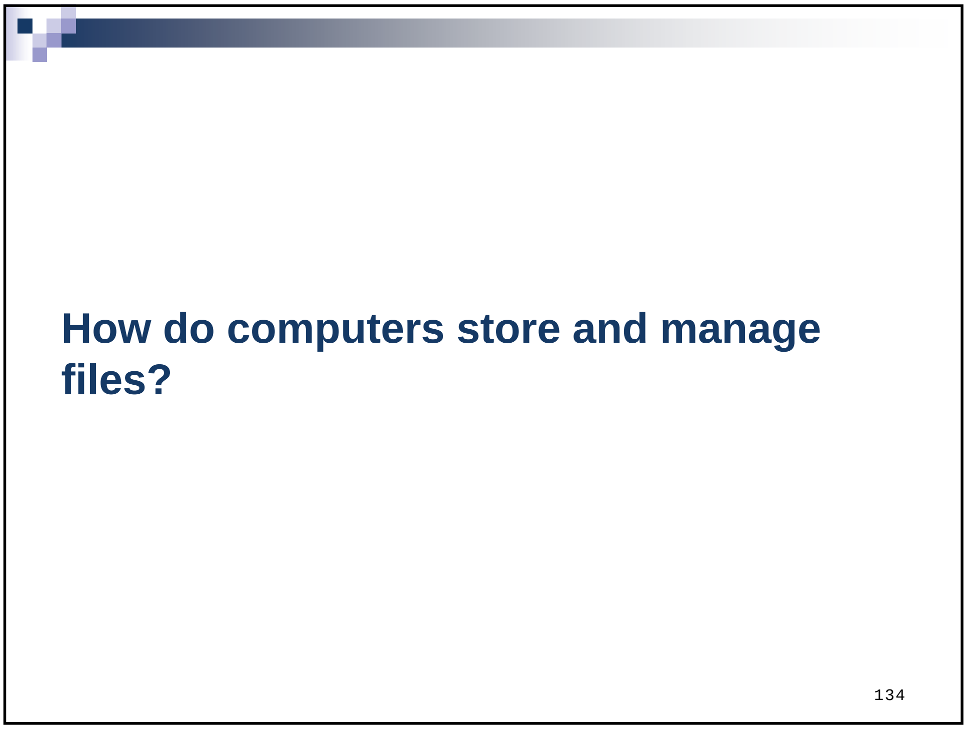
✓ **Technical background**

✓ **Representation Information**

4. File systems and file management
5. Extracting data from media
6. Tools and methods
7. Conclusions, questions, discussion



File systems and file management



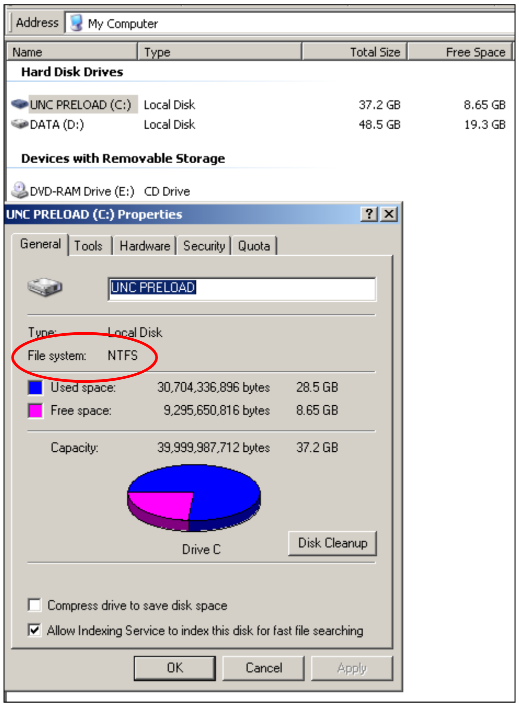
How do computers store and manage files?

Volumes and Partitions

- **Volume**
  - □ Storage area defined at the logical OS level, which has a single filesystem & usually resides on one disk partition
- **Partition**
  - □ Exists at physical, media-specific level
  - □ May be used to set up multiple operating systems on same computer

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)



Address: My Computer

Name	Type	Total Size	Free Space
Hard Disk Drives			
UNC PRELOAD (C:)	Local Disk	37.2 GB	8.65 GB
DATA (D:)	Local Disk	48.5 GB	19.3 GB
Devices with Removable Storage			
DVD-RAM Drive (E:)	CD Drive		

UNC PRELOAD (C:) Properties

UNC PRELOAD

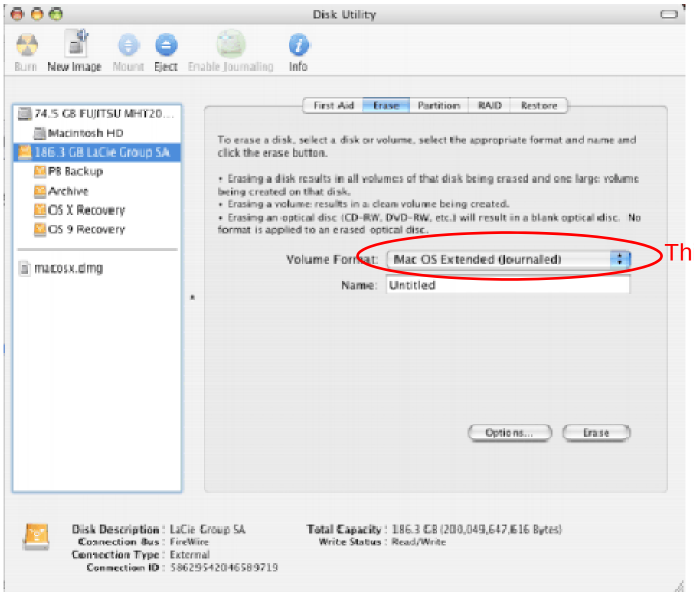
Type: Local Disk

File system: NTFS

■	Used space:	30,704,336,896 bytes	28.5 GB
■	Free space:	9,295,650,816 bytes	8.65 GB
	Capacity:	39,999,987,712 bytes	37.2 GB

☐ Compress drive to save disk space

☑ Allow Indexing Service to index this disk for fast file searching



To erase a disk, select a disk or volume, select the appropriate format and name and click the erase button.

- Erasing a disk results in all volumes of that disk being erased and one large volume being created on that disk.
- Erasing a volume results in a clean volume being created.
- Erasing an optical disc (CD-RW, DVD-RW, etc.) will result in a blank optical disc. No format is applied to an erased optical disc.

Volume Format: Mac OS Extended (Journaled)

Name: Untitled

Disk Description : LaCie Group SA
Connection Bus : FireWire
Connection Type : External
Connection ID : 58629542046589719

Total Capacity : 186.3 GB (200,049,647,616 Bytes)
Write Status : Read/Write

This is HFS+

Source for underlying screenshot: “How to Use 1 External Drive between Mac and PC,” nicknack, February 5, 2007, http://www.gigacrate.com/Articles/?p=53

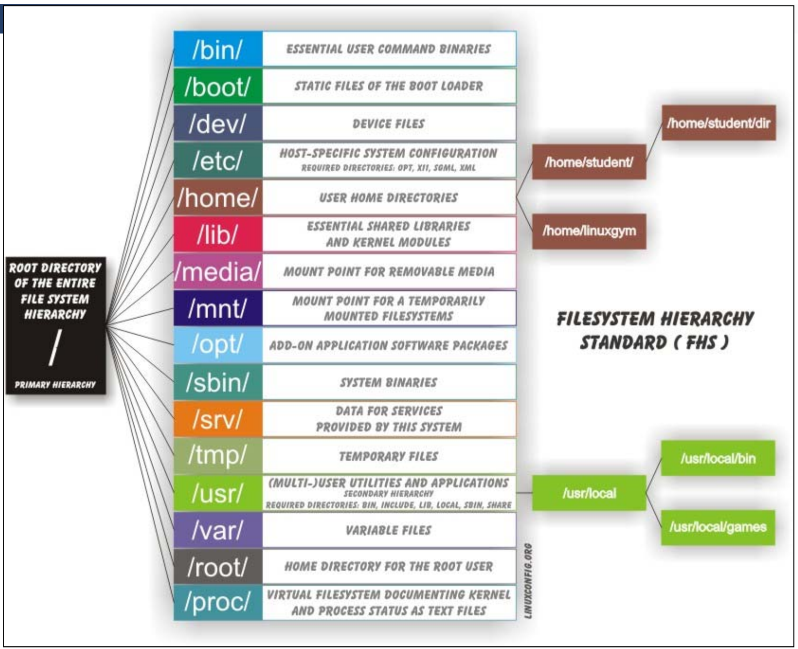
File System Examples

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

File System Examples

Name	Operating System(s) Using it as Native File System [often other OSs can also recognize it]
FAT12, FAT16	
FAT32 (VFAT)	
exFAT	s, SD cards)
NTFS	Vista
MFS	
HFS (Hierarchical File System)	
HFS+	
APFS	
ext, ext2, ext3, ext4 (Extended File System)	
XFS	
HPFS (High Performance File System)	
ISOFS (ISO 9660)	
JFS1 (Journaled File System)	
ReiserFS	
UFS (Unix File System) aka FFS (Fast File System)	

The filesystems you’ re most likely to encounter within archival collections



/bin/	ESSENTIAL USER COMMAND BINARIES
/boot/	STATIC FILES OF THE BOOT LOADER
/dev/	DEVICE FILES
/etc/	HOST-SPECIFIC SYSTEM CONFIGURATION REQUIRED DIRECTORIES: OPT, X11, SGML, XML
/home/	USER HOME DIRECTORIES
/lib/	ESSENTIAL SHARED LIBRARIES AND KERNEL MODULES
/media/	MOUNT POINT FOR REMOVABLE MEDIA
/mnt/	MOUNT POINT FOR A TEMPORARILY MOUNTED FILESYSTEMS
/opt/	ADD-ON APPLICATION SOFTWARE PACKAGES
/sbin/	SYSTEM BINARIES
/srv/	DATA FOR SERVICES PROVIDED BY THIS SYSTEM
/tmp/	TEMPORARY FILES
/usr/	(MULTI-)USER UTILITIES AND APPLICATIONS SECONDARY HIERARCHY REQUIRED DIRECTORIES: BIN, INCLUDE, LIB, LOCAL, SBIN, SHARE
/var/	VARIABLE FILES
/root/	HOME DIRECTORY FOR THE ROOT USER
/proc/	VIRTUAL FILESYSTEM DOCUMENTING KERNEL AND PROCESS STATUS AS TEXT FILES

Source: Linuxconfig.org – Filesytem Basics. http://www.linuxconfig.org/Filesystem_Basics

Microsoft File Systems:

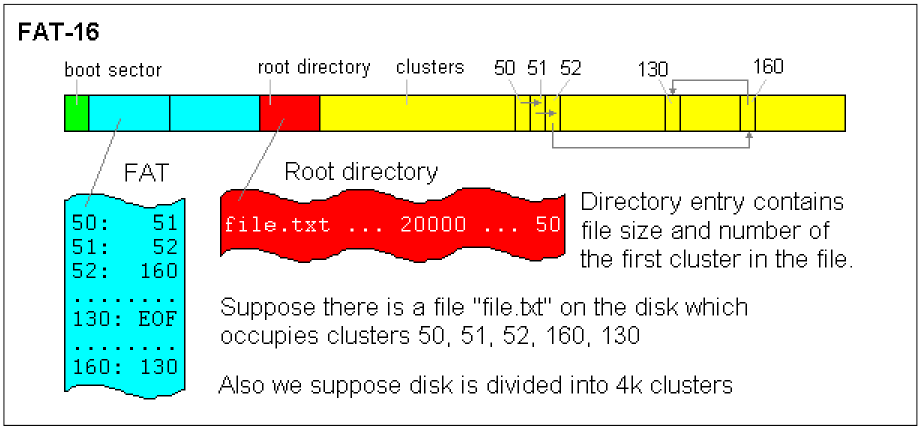
FAT and NTFS

FAT

- Supported by all versions of Windows and most versions of UNIX
- Common in many removable media, e.g. thumb drives, memory cards

http://mono-for-android.1047100.n5.nabble.com/detect-SD-Card-path-td5710218.html

http://www.ubergizmo.com/2008/09/usb-thumb-drive/



```
50:    51
51:    52
52:   160
........
130:  EOF
........
160:  130
```

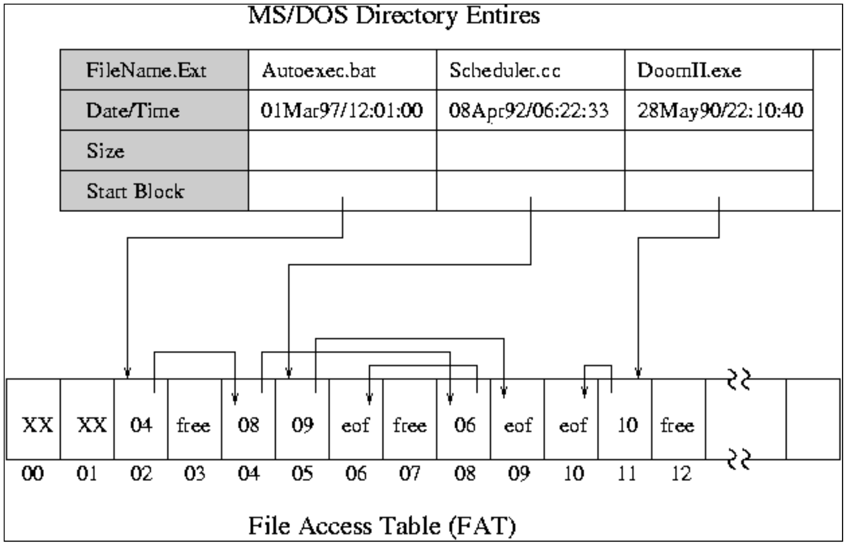
file.txt ... 20000 ... 50

Directory entry contains file size and number of the first cluster in the file.

Suppose there is a file "file.txt" on the disk which occupies clusters 50, 51, 52, 160, 130

Also we suppose disk is divided into 4k clusters

Source: Mikhail, Ranish. “Partitioning Primer.” August 5, 1998. http://www.ranish.com/part/primer.htm



MS/DOS Directory Entires

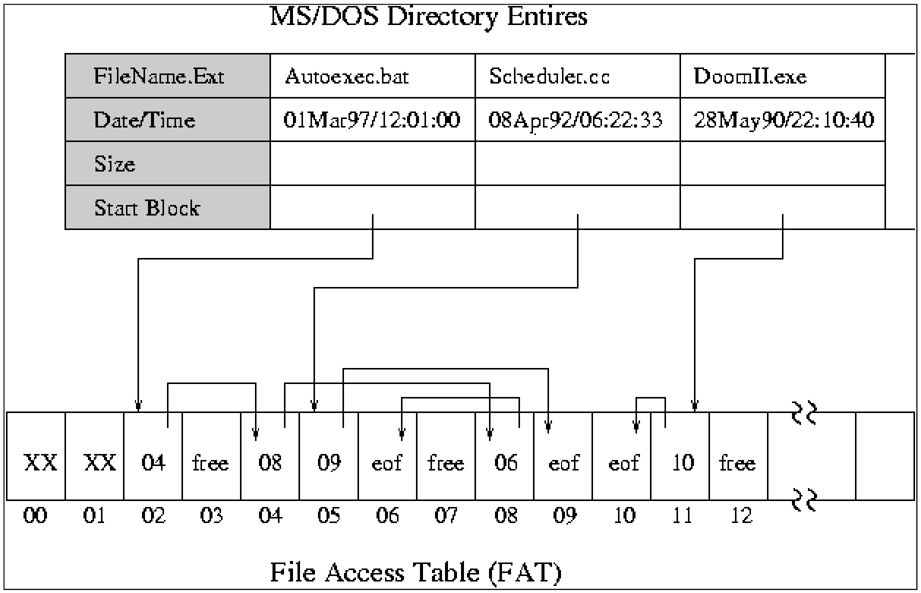
FileName.Ext	Autoexec.bat	Scheduler.cc	DoomII.exe
Date/Time	01Mar97/12:01:00	08Apr92/06:22:33	28May90/22:10:40
Size			
Start Block			

XX	XX	04	free	08	09	eof	free	06	eof	eof	10	free
00	01	02	03	04	05	06	07	08	09	10	11	12

File Access Table (FAT)

Source: Duong, Duc. "I/O devices and File systems." Vietnam OpenCourseWare. November 18, 2008. http://www.vocw.edu.vn/content/m10107/1.3/.

What Happens When a File is Deleted from a Directory

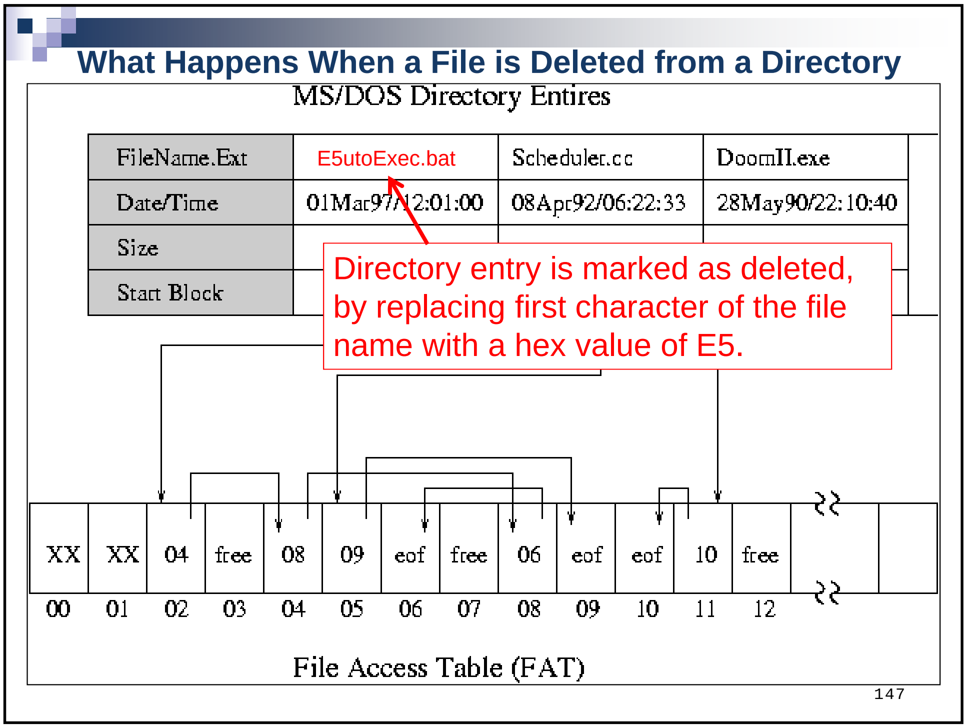


MS/DOS Directory Entires

FileName.Ext	Autoexec.bat	Scheduler.cc	DoomII.exe
Date/Time	01Mar97/12:01:00	08Apr92/06:22:33	28May90/22:10:40
Size			
Start Block			

XX	XX	04	free	08	09	eof	free	06	eof	eof	10	free
00	01	02	03	04	05	06	07	08	09	10	11	12

File Access Table (FAT)



What Happens When a File is Deleted from a Directory

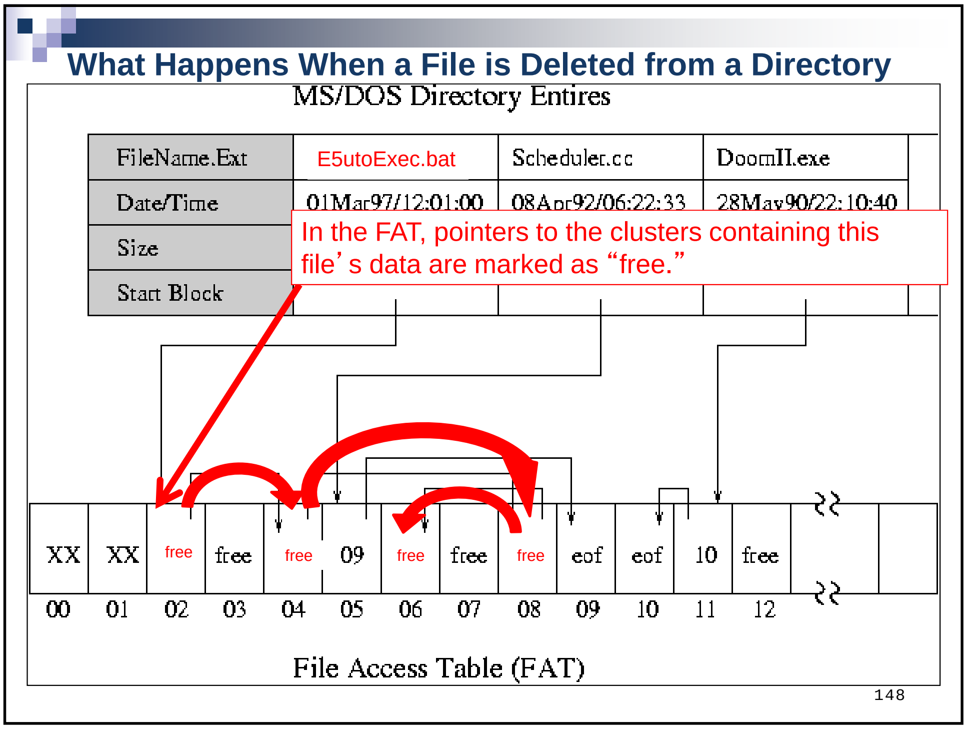
MS/DOS Directory Entires

FileName.Ext	E5utoExec.bat	Scheduler.cc	DoomII.exe
Date/Time	01Mar97/12:01:00	08Apr92/06:22:33	28May90/22:10:40
Size			
Start Block			

Directory entry is marked as deleted, by replacing first character of the file name with a hex value of E5.

XX	XX	04	free	08	09	eof	free	06	eof	eof	10	free
00	01	02	03	04	05	06	07	08	09	10	11	12

File Access Table (FAT)



What Happens When a File is Deleted from a Directory

MS/DOS Directory Entires

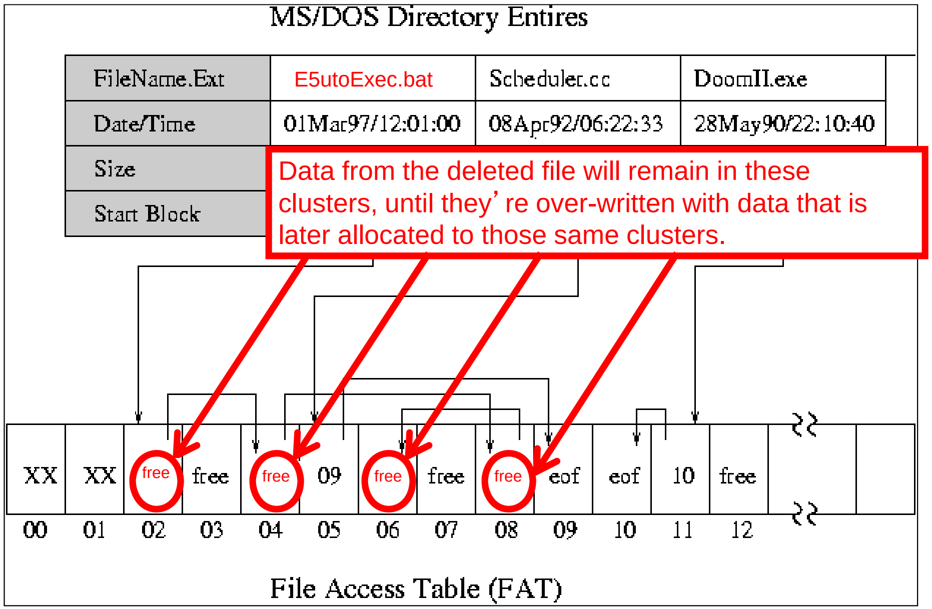
FileName.Ext	E5utoExec.bat	Scheduler.cc	DoomII.exe
Date/Time	01Mar97/12:01:00	08Apr92/06:22:33	28May90/22:10:40
Size			
Start Block			

In the FAT, pointers to the clusters containing this file’s data are marked as “free.”

XX	XX	free	free	free	09	free	free	free	eof	eof	10	free
00	01	02	03	04	05	06	07	08	09	10	11	12

File Access Table (FAT)

What Happens When a File is Deleted from a Directory



MS/DOS Directory Entires

FileName.Ext	E5utoExec.bat	Scheduler.cc	DoomII.exe
Date/Time	01Mar97/12:01:00	08Apr92/06:22:33	28May90/22:10:40
Size			
Start Block			

Data from the deleted file will remain in these clusters, until they’re over-written with data that is later allocated to those same clusters.

XX	XX	free	free	free	09	free	free	free	eof	eof	10	free
00	01	02	03	04	05	06	07	08	09	10	11	12

File Access Table (FAT)

NTFS

- Directory and FAT functions are combined in the Master File Table ($MFT)
- Each MFT record is assigned a unique number
- Good for forensic discovery:
  - For small files (< about 600 bytes), content is stored directly in the MFT itself & remains until overwritten by another MFT record
- Not so good for forensic discovery:
  - After deletion of a file, NTFS replaces (overwrites) the MFT record the next time a new file is created

“Archive” Formats as Portable File Systems

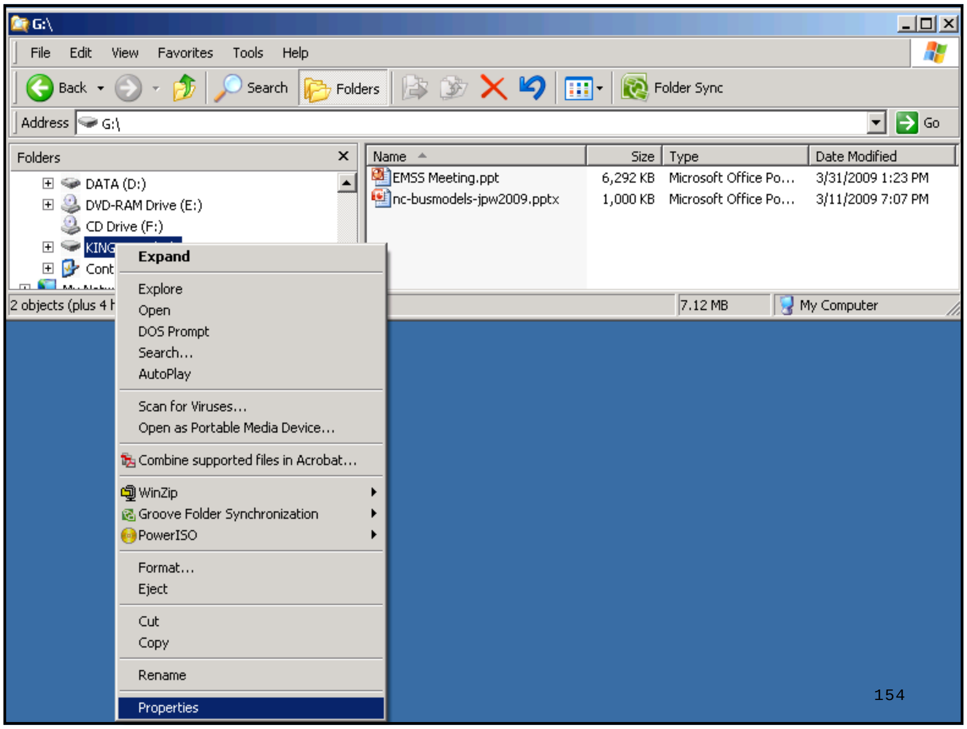
- Most popular: zip and tar
- Retain important metadata that was in original file system, but does add a layer of representation information (packaging and possibly compression) that software needs to understand
- Compression also reduces robustness in the face of bit loss (any given bit flip is more likely to prevent recovery/rendering of content)
- Note that both zip and tar can be stored uncompressed (more common with tar)

Bit-Level Treatment of Individual Files

Example scenario:
A professor lets a couple of speakers use a USB thumb drive for their presentation slides.

Let's see what's on it...

153



G:\

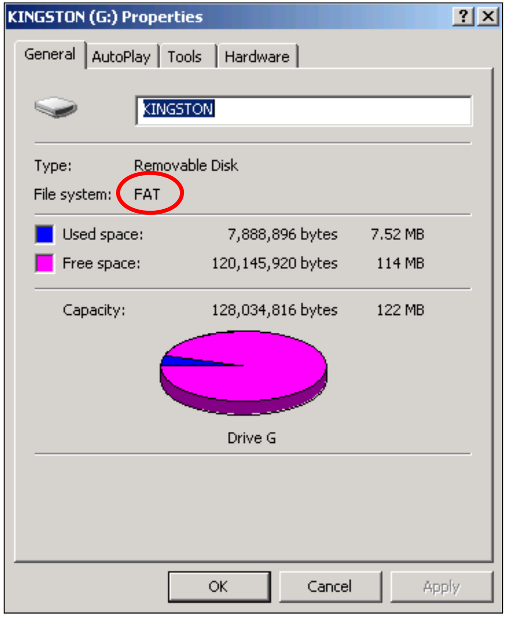
Name	Size	Type	Date Modified
EMSS Meeting.ppt	6,292 KB	Microsoft Office Po...	3/31/2009 1:23 PM
nc-busmodels-jpw2009.pptx	1,000 KB	Microsoft Office Po...	3/11/2009 7:07 PM

2 objects (plus 4 h

7.12 MB

My Computer

154



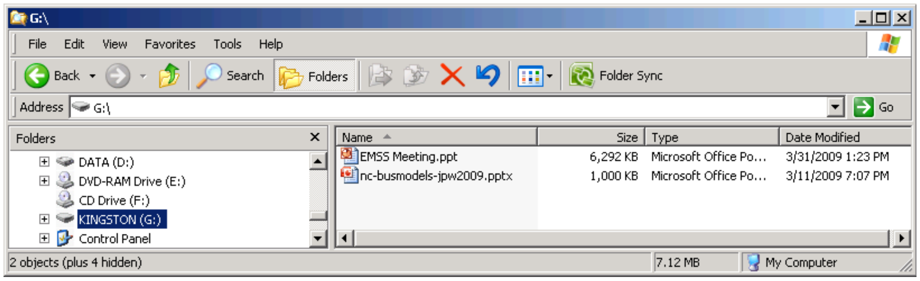
KINGSTON (G:) Properties

KINGSTON

Type: Removable Disk

File system: FAT

🟦	Used space:	7,888,896 bytes	7.52 MB
🟪	Free space:	120,145,920 bytes	114 MB
	Capacity:	128,034,816 bytes	122 MB



G:\

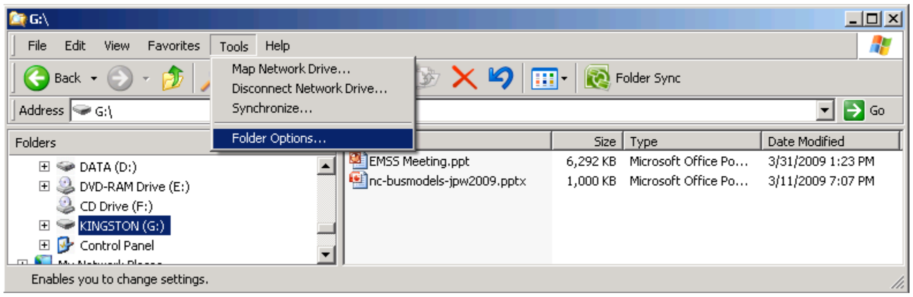
Address G:\

Folders

- DATA (D:)
- DVD-RAM Drive (E:)
- CD Drive (F:)
- KINGSTON (G:)
- Control Panel

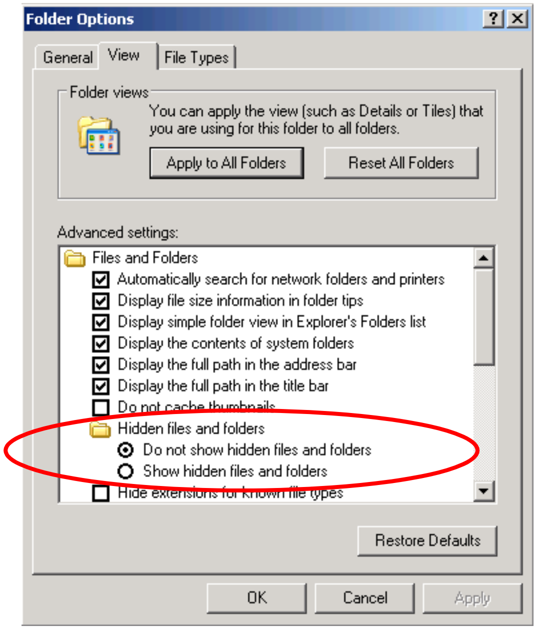
Name	Size	Type	Date Modified
EMSS Meeting.ppt	6,292 KB	Microsoft Office Po...	3/31/2009 1:23 PM
nc-busmodels-jpw2009.pptx	1,000 KB	Microsoft Office Po...	3/11/2009 7:07 PM

2 objects (plus 4 hidden) 7.12 MB My Computer



Name	Size	Type	Date Modified
EMSS Meeting.ppt	6,292 KB	Microsoft Office Po…	3/31/2009 1:23 PM
nc-busmodels-jpw2009.pptx	1,000 KB	Microsoft Office Po…	3/11/2009 7:07 PM

Enables you to change settings.



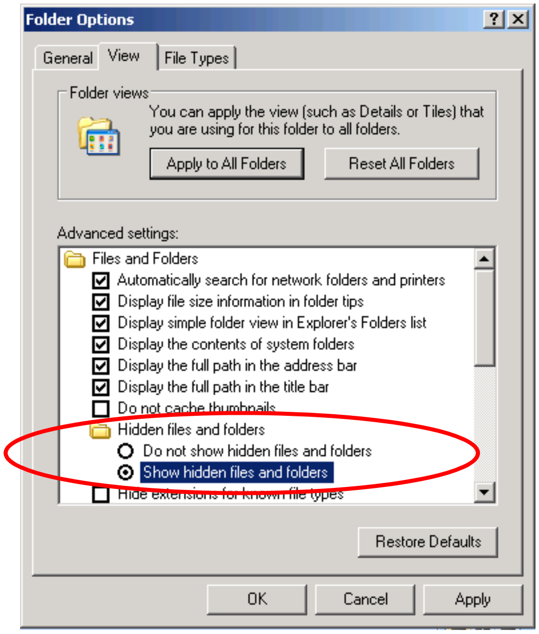
You can apply the view (such as Details or Tiles) that you are using for this folder to all folders.

Advanced settings:

Files and Folders
- ☑ Automatically search for network folders and printers
- ☑ Display file size information in folder tips
- ☑ Display simple folder view in Explorer's Folders list
- ☑ Display the contents of system folders
- ☑ Display the full path in the address bar
- ☑ Display the full path in the title bar
- □ Do not cache thumbnails

Hidden files and folders
- ● Do not show hidden files and folders
- ○ Show hidden files and folders

- □ Hide extensions for known file types



Folder Options

General | View | File Types

Folder views

You can apply the view (such as Details or Tiles) that you are using for this folder to all folders.

Apply to All Folders | Reset All Folders

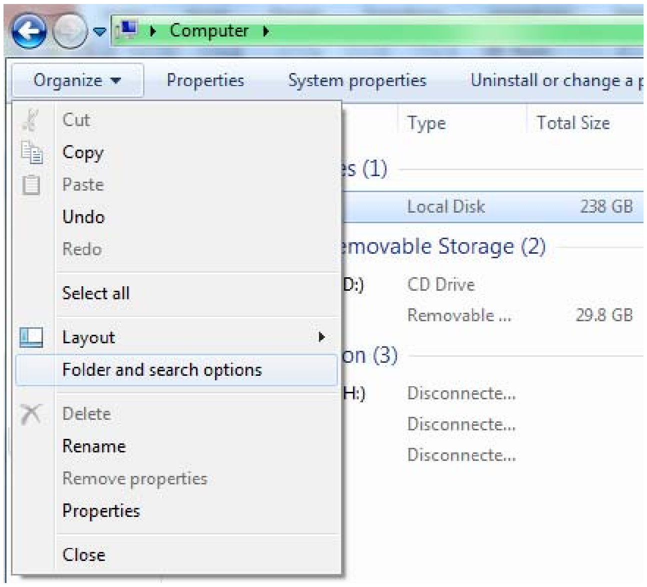
Advanced settings:

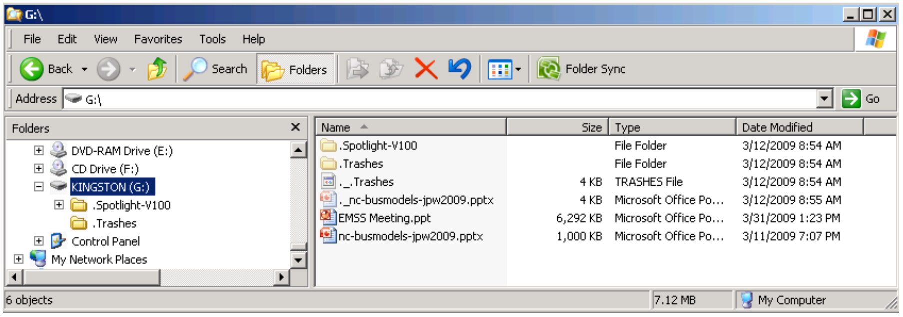
- Files and Folders
  - ☑ Automatically search for network folders and printers
  - ☑ Display file size information in folder tips
  - ☑ Display simple folder view in Explorer's Folders list
  - ☑ Display the contents of system folders
  - ☑ Display the full path in the address bar
  - ☑ Display the full path in the title bar
  - □ Do not cache thumbnails
  - Hidden files and folders
    - ○ Do not show hidden files and folders
    - ● Show hidden files and folders
  - □ Hide extensions for known file types

Restore Defaults

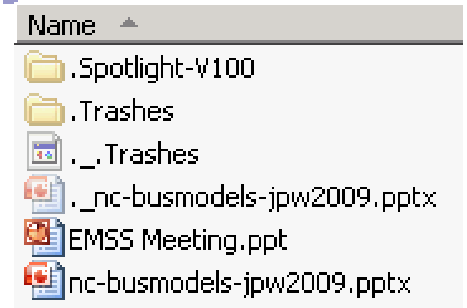
OK | Cancel | Apply

Note: The previous screenshots were from an older version of Windows. On a later version, you'll probably see something like this instead:

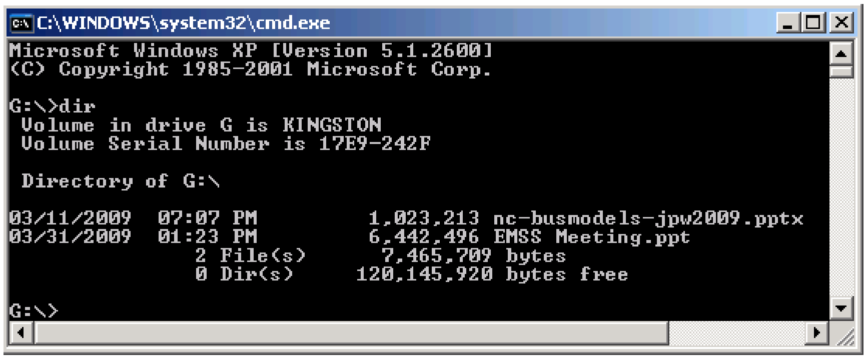




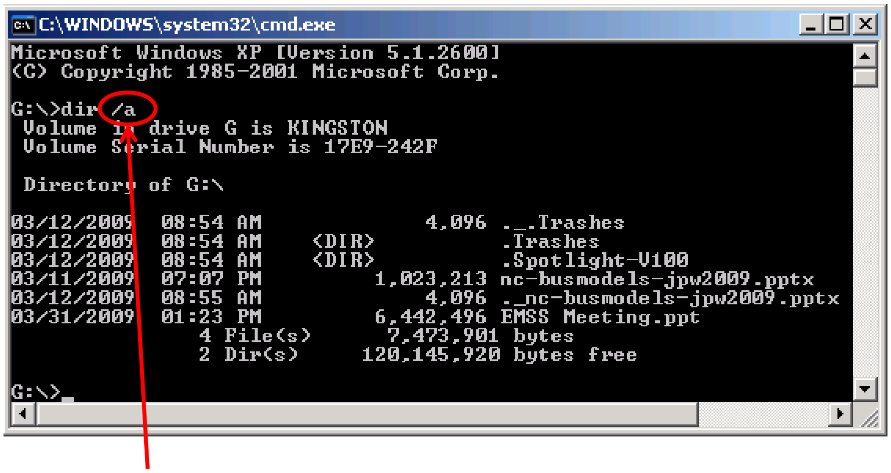
Name	Size	Type	Date Modified
.Spotlight-V100		File Folder	3/12/2009 8:54 AM
.Trashes		File Folder	3/12/2009 8:54 AM
._.Trashes	4 KB	TRASHES File	3/12/2009 8:54 AM
._nc-busmodels-jpw2009.pptx	4 KB	Microsoft Office Po...	3/12/2009 8:55 AM
EMSS Meeting.ppt	6,292 KB	Microsoft Office Po...	3/31/2009 1:23 PM
nc-busmodels-jpw2009.pptx	1,000 KB	Microsoft Office Po...	3/11/2009 7:07 PM



Name
.Spotlight-V100
.Trashes
._.Trashes
._nc-busmodels-jpw2009.pptx
EMSS Meeting.ppt
nc-busmodels-jpw2009.pptx

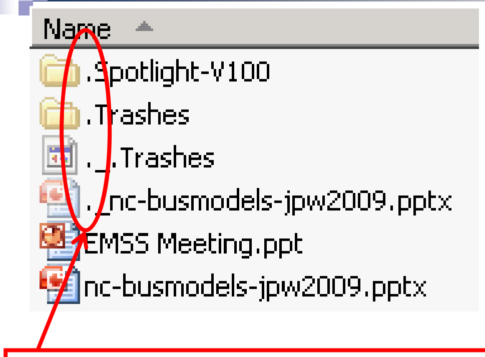


```
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

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

 Directory of G:\

03/11/2009  07:07 PM         1,023,213 nc-busmodels-jpw2009.pptx
03/31/2009  01:23 PM         6,442,496 EMSS Meeting.ppt
               2 File(s)      7,465,709 bytes
               0 Dir(s)     120,145,920 bytes free

G:\>
```



```
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

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

 Directory of G:\

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

G:\>_
```

/a switch shows directories, read-only files, hidden files, system files



Name

.Spotlight-V100

.Trashes

._.Trashes

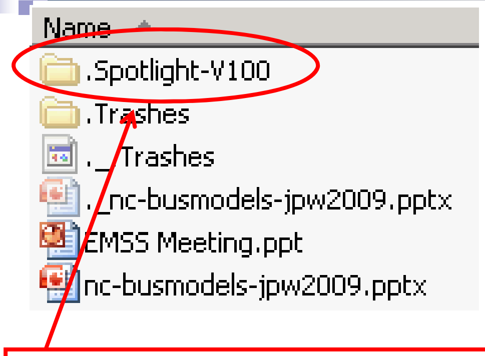
._nc-busmodels-jpw2009.pptx

EMSS Meeting.ppt

nc-busmodels-jpw2009.pptx

Period at beginning of file names indicates “hidden” files

- Not visible by default in Macintosh Finder
- Visible through Windows Explorer (with proper settings), DOS (using /a switch: dir /a) or Unix (-a switch: ls –a)



Name

.Spotlight-V100

.Trashes

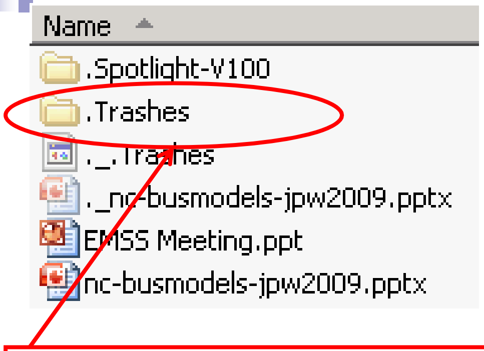
._.Trashes

._nc-busmodels-jpw2009.pptx

EMSS Meeting.ppt

nc-busmodels-jpw2009.pptx

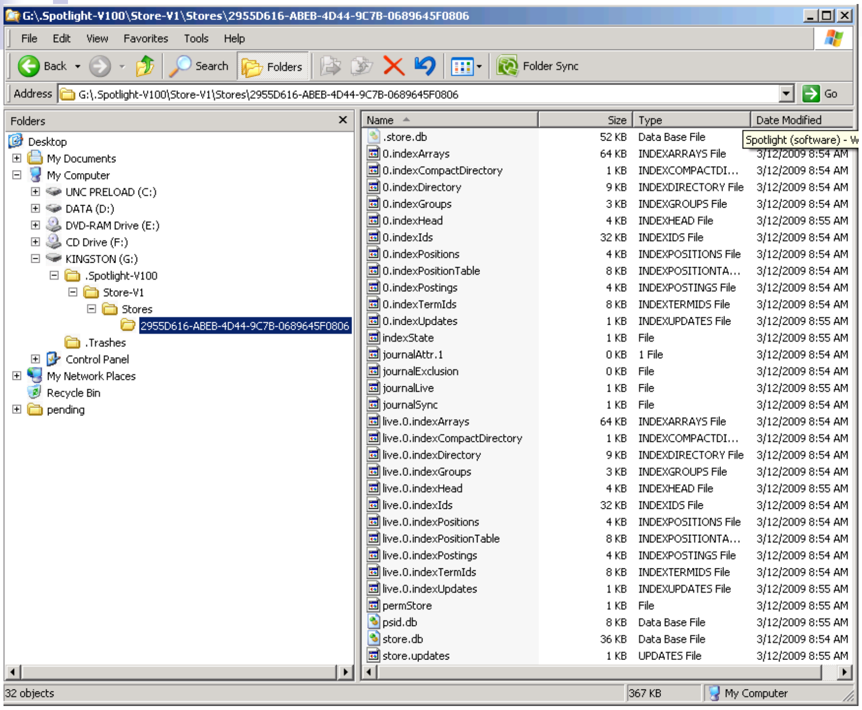
Spotlight is the desktop search utility on the Macintosh OS X. The contents of this folder serve as an index of all files that were on the thumb drive that last time was last used on a Mac OS X computer.



Name

.Spotlight-V100
.Trashes
._.Trashes
._nc-busmodels-jpw2009.pptx
EMSS Meeting.ppt
nc-busmodels-jpw2009.pptx

On a Mac, each user's home directory has a .Trash folder, and each volume has a .Trashes folder. This is similar to the Windows Recycling Bin .Trashes for each volume contains a separate sub-folder for the trash files contributed by each user (named by UID).



G:\.Spotlight-V100\Store-V1\Stores\2955D616-ABEB-4D44-9C7B-0689645F0806

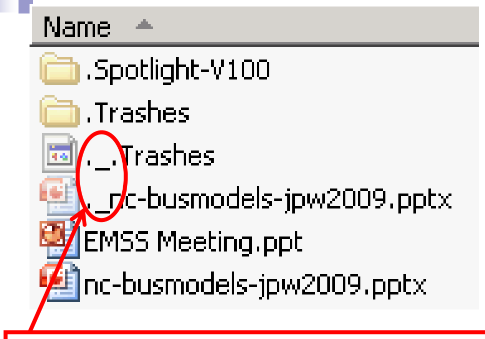
Address G:\.Spotlight-V100\Store-V1\Stores\2955D616-ABEB-4D44-9C7B-0689645F0806

Folders

Desktop
My Documents
My Computer
UNC PRELOAD (C:)
DATA (D:)
DVD-RAM Drive (E:)
CD Drive (F:)
KINGSTON (G:)
.Spotlight-V100
Store-V1
Stores
2955D616-ABEB-4D44-9C7B-0689645F0806
.Trashes
Control Panel
My Network Places
Recycle Bin
pending

Name	Size	Type	Date Modified
.store.db	52 KB	Data Base File	Spotlight (software) - W
0.indexArrays	64 KB	INDEXARRAYS File	3/12/2009 8:54 AM
0.indexCompactDirectory	1 KB	INDEXCOMPACTDI...	3/12/2009 8:54 AM
0.indexDirectory	9 KB	INDEXDIRECTORY File	3/12/2009 8:54 AM
0.indexGroups	3 KB	INDEXGROUPS File	3/12/2009 8:54 AM
0.indexHead	4 KB	INDEXHEAD File	3/12/2009 8:55 AM
0.indexIds	32 KB	INDEXIDS File	3/12/2009 8:54 AM
0.indexPositions	4 KB	INDEXPOSITIONS File	3/12/2009 8:54 AM
0.indexPositionTable	8 KB	INDEXPOSITIONTA...	3/12/2009 8:54 AM
0.indexPostings	4 KB	INDEXPOSTINGS File	3/12/2009 8:54 AM
0.indexTermIds	8 KB	INDEXTERMIDS File	3/12/2009 8:54 AM
0.indexUpdates	1 KB	INDEXUPDATES File	3/12/2009 8:55 AM
indexState	1 KB	File	3/12/2009 8:55 AM
journalAttr.1	0 KB	1 File	3/12/2009 8:54 AM
journalExclusion	0 KB	File	3/12/2009 8:54 AM
journalLive	1 KB	File	3/12/2009 8:55 AM
journalSync	1 KB	File	3/12/2009 8:54 AM
live.0.indexArrays	64 KB	INDEXARRAYS File	3/12/2009 8:54 AM
live.0.indexCompactDirectory	1 KB	INDEXCOMPACTDI...	3/12/2009 8:54 AM
live.0.indexDirectory	9 KB	INDEXDIRECTORY File	3/12/2009 8:54 AM
live.0.indexGroups	3 KB	INDEXGROUPS File	3/12/2009 8:54 AM
live.0.indexHead	4 KB	INDEXHEAD File	3/12/2009 8:55 AM
live.0.indexIds	32 KB	INDEXIDS File	3/12/2009 8:54 AM
live.0.indexPositions	4 KB	INDEXPOSITIONS File	3/12/2009 8:54 AM
live.0.indexPositionTable	8 KB	INDEXPOSITIONTA...	3/12/2009 8:54 AM
live.0.indexPostings	4 KB	INDEXPOSTINGS File	3/12/2009 8:54 AM
live.0.indexTermIds	8 KB	INDEXTERMIDS File	3/12/2009 8:54 AM
live.0.indexUpdates	1 KB	INDEXUPDATES File	3/12/2009 8:55 AM
permStore	1 KB	File	3/12/2009 8:55 AM
psid.db	8 KB	Data Base File	3/12/2009 8:55 AM
store.db	36 KB	Data Base File	3/12/2009 8:54 AM
store.updates	1 KB	UPDATES File	3/12/2009 8:55 AM

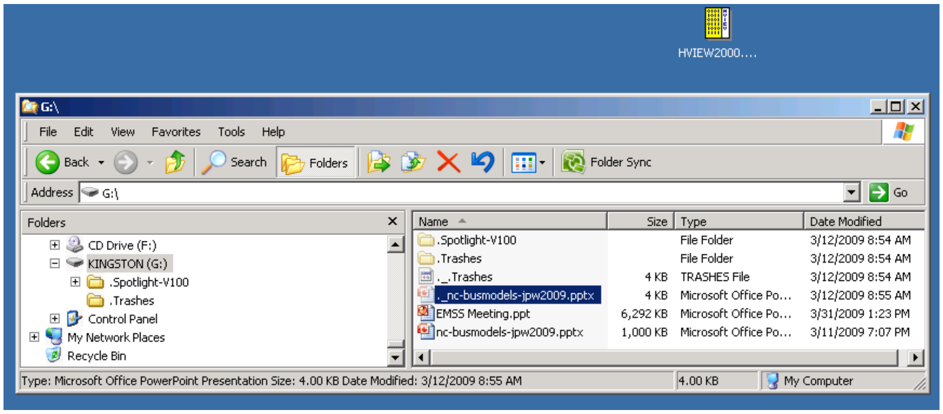
32 objects | 367 KB | My Computer



Name
.Spotlight-V100
.Trashes
._.Trashes
._nc-busmodels-jpw2009.pptx
EMSS Meeting.ppt
nc-busmodels-jpw2009.pptx

Files beginning with ._ are resource fork files, which the Mac file system uses to store icon, data type, and some other metadata about the file (content being stored in the data fork)

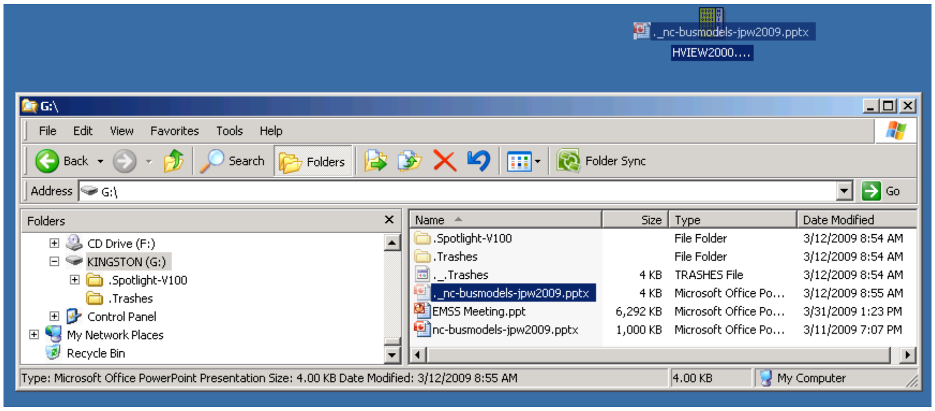
Maybe I should look at a hex dump of one of them to be sure...



Address G:\

Name	Size	Type	Date Modified
.Spotlight-V100		File Folder	3/12/2009 8:54 AM
.Trashes		File Folder	3/12/2009 8:54 AM
._.Trashes	4 KB	TRASHES File	3/12/2009 8:54 AM
._nc-busmodels-jpw2009.pptx	4 KB	Microsoft Office Po...	3/12/2009 8:55 AM
EMSS Meeting.ppt	6,292 KB	Microsoft Office Po...	3/31/2009 1:23 PM
nc-busmodels-jpw2009.pptx	1,000 KB	Microsoft Office Po...	3/11/2009 7:07 PM

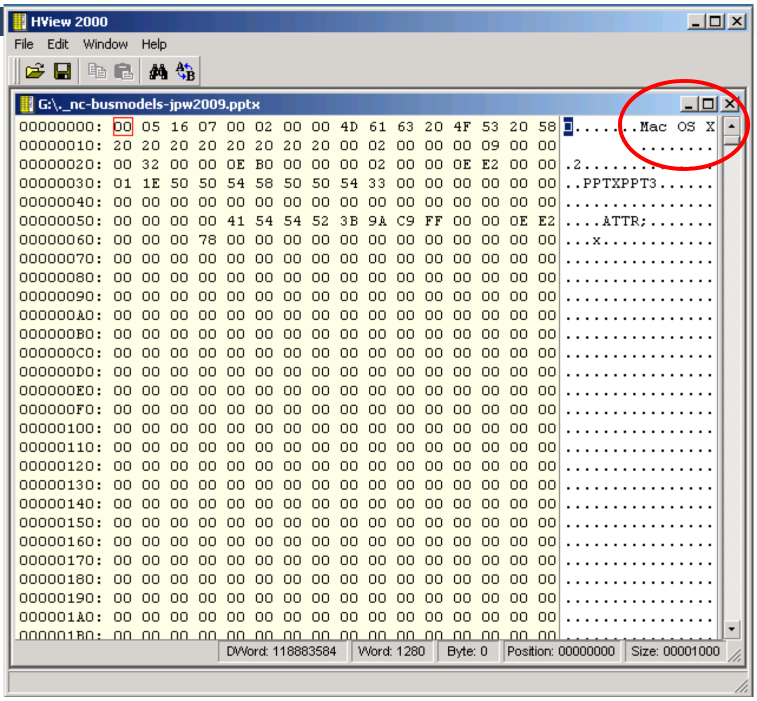
Type: Microsoft Office PowerPoint Presentation Size: 4.00 KB Date Modified: 3/12/2009 8:55 AM 4.00 KB My Computer



Name	Size	Type	Date Modified
.Spotlight-V100		File Folder	3/12/2009 8:54 AM
.Trashes		File Folder	3/12/2009 8:54 AM
._.Trashes	4 KB	TRASHES File	3/12/2009 8:54 AM
._nc-busmodels-jpw2009.pptx	4 KB	Microsoft Office Po...	3/12/2009 8:55 AM
EMSS Meeting.ppt	6,292 KB	Microsoft Office Po...	3/31/2009 1:23 PM
nc-busmodels-jpw2009.pptx	1,000 KB	Microsoft Office Po...	3/11/2009 7:07 PM

Type: Microsoft Office PowerPoint Presentation Size: 4.00 KB Date Modified: 3/12/2009 8:55 AM | 4.00 KB | My Computer

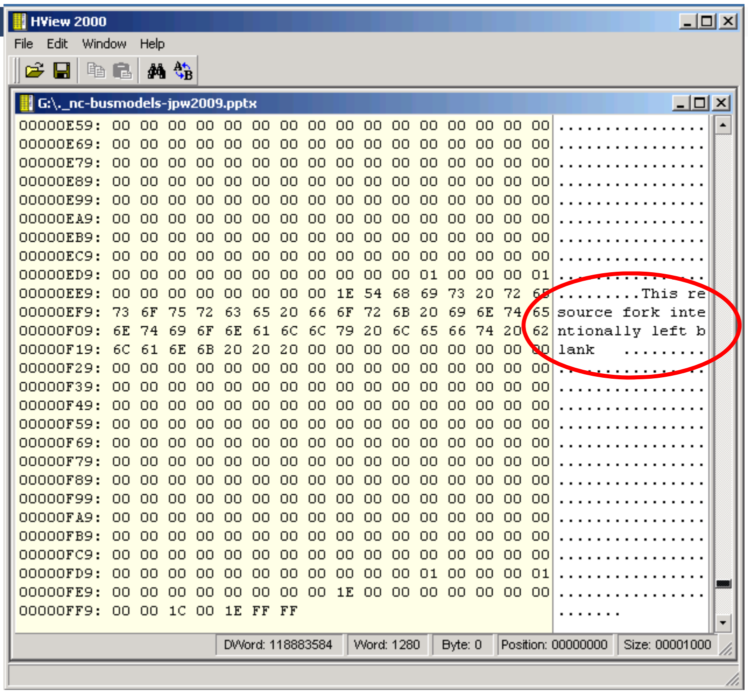
171



```
00000000: 00 05 16 07 00 02 00 00 4D 61 63 20 4F 53 20 58 ........Mac OS X
00000010: 20 20 20 20 20 20 20 20 00 02 00 00 00 09 00 00         ........
00000020: 00 32 00 00 0E B0 00 00 00 02 00 00 0E E2 00 00 .2..............
00000030: 01 1E 50 50 54 58 50 50 54 33 00 00 00 00 00 00 ..PPTXPPT3......
00000040: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000050: 00 00 00 00 41 54 54 52 3B 9A C9 FF 00 00 0E E2 ....ATTR;.......
00000060: 00 00 00 78 00 00 00 00 00 00 00 00 00 00 00 00 ...x............
00000070: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000080: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000090: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
000000A0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
000000B0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
000000C0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
000000D0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
000000E0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
000000F0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000100: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000110: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000120: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000130: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000140: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000150: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000160: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000170: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000180: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000190: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
000001A0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
000001B0: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
```

DWord: 118883584 | Word: 1280 | Byte: 0 | Position: 00000000 | Size: 00001000

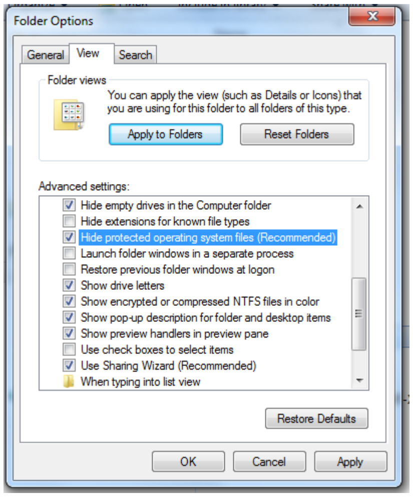
172



```
00000E59: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000E69: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000E79: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000E89: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000E99: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000EA9: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000EB9: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000EC9: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000ED9: 00 00 00 00 00 00 00 00 00 00 00 01 00 00 00 01 ................
00000EE9: 00 00 00 00 00 00 00 00 1E 54 68 69 73 20 72 65 .........This re
00000EF9: 73 6F 75 72 63 65 20 66 6F 72 6B 20 69 6E 74 65 source fork inte
00000F09: 6E 74 69 6F 6E 61 6C 6C 79 20 6C 65 66 74 20 62 ntionally left b
00000F19: 6C 61 6E 6B 20 20 20 00 00 00 00 00 00 00 00 00 lank   .........
00000F29: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000F39: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000F49: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000F59: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000F69: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000F79: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000F89: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000F99: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000FA9: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000FB9: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000FC9: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ................
00000FD9: 00 00 00 00 00 00 00 00 00 00 00 01 00 00 00 01 ................
00000FE9: 00 00 00 00 00 00 00 00 1E 00 00 00 00 00 00 00 ................
00000FF9: 00 00 1C 00 1E FF FF                            .......
```

DWord: 118883584 | Word: 1280 | Byte: 0 | Position: 00000000 | Size: 00001000

This wasn't necessary for seeing the hidden files (resource forks) in the previous example, but another system change you might want to make on your processing/forensics workstation to show system files that are normally hidden from the user:

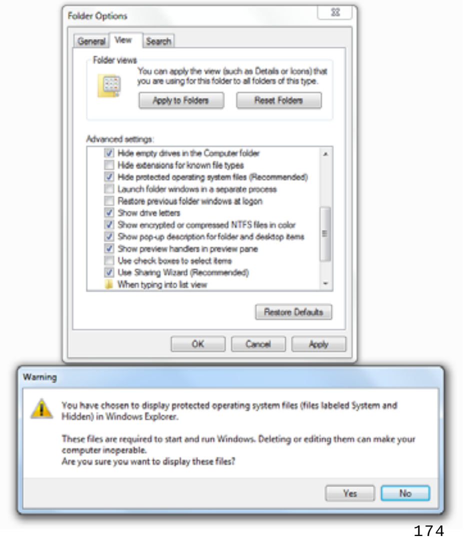


Folder views

You can apply the view (such as Details or Icons) that you are using for this folder to all folders of this type.

Advanced settings:

- ☑ Hide empty drives in the Computer folder
- □ Hide extensions for known file types
- ☑ Hide protected operating system files (Recommended)
- □ Launch folder windows in a separate process
- □ 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
- □ Use check boxes to select items
- ☑ Use Sharing Wizard (Recommended)
- When typing into list view



Folder views

You can apply the view (such as Details or Icons) that you are using for this folder to all folders of this type.

Advanced settings:

- ☑ Hide empty drives in the Computer folder
- □ Hide extensions for known file types
- ☑ Hide protected operating system files (Recommended)
- □ Launch folder windows in a separate process
- □ 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
- □ Use check boxes to select items
- ☑ Use Sharing Wizard (Recommended)
- When typing into list view

You have chosen to display protected operating system files (files labeled System and Hidden) in Windows 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?

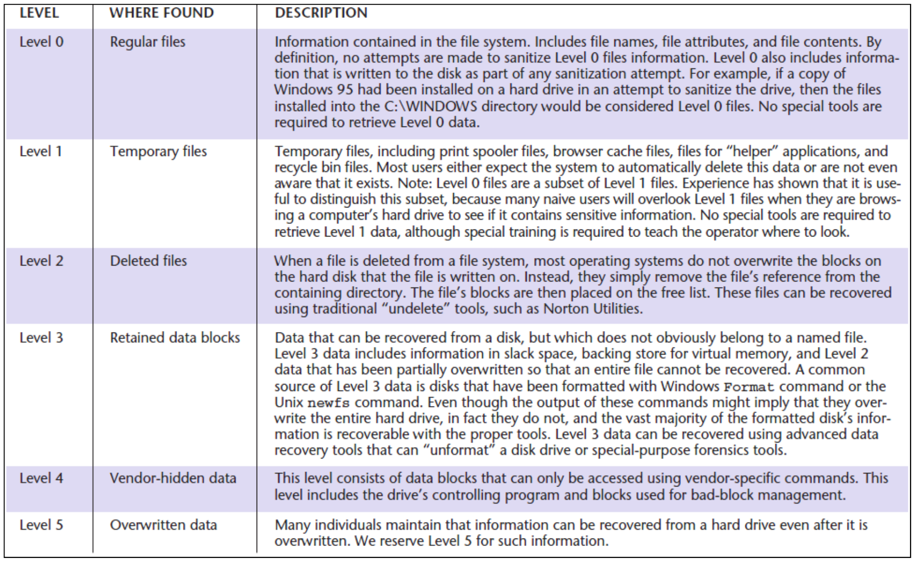
Forms of “Hidden Data”

<u>Not</u> just what you see when you open a file in its native application

Listed roughly in order of difficulty for identification and retrieval

175

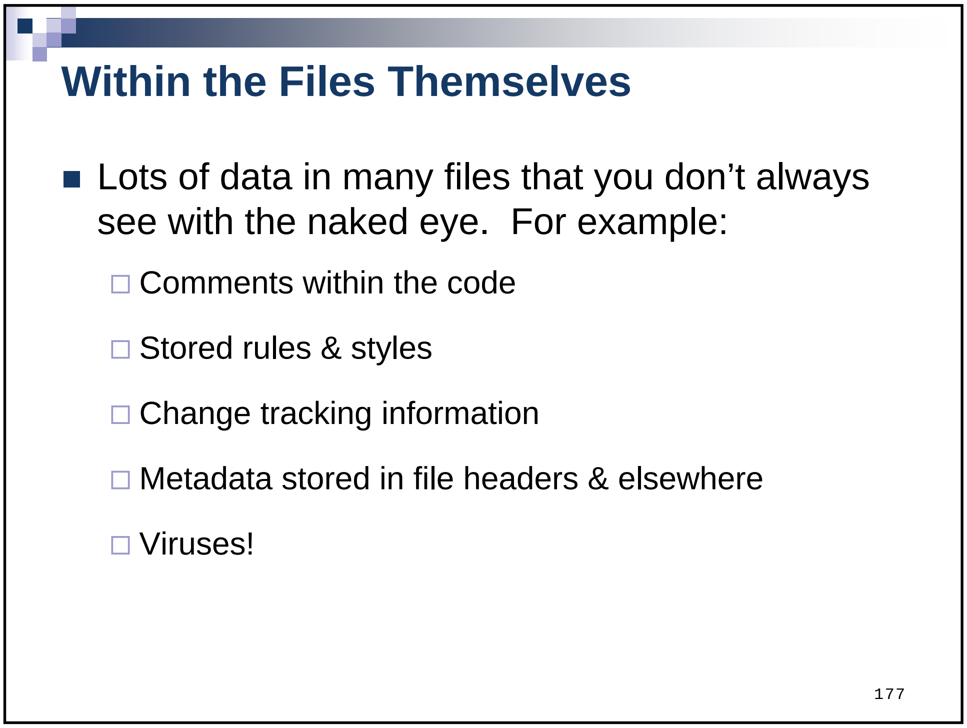
Sanitization Taxonomy



LEVEL	WHERE FOUND	DESCRIPTION
Level 0	Regular files	Information contained in the file system. Includes file names, file attributes, and file contents. By definition, no attempts are made to sanitize Level 0 files information. Level 0 also includes information that is written to the disk as part of any sanitization attempt. For example, if a copy of Windows 95 had been installed on a hard drive in an attempt to sanitize the drive, then the files installed into the C:\WINDOWS directory would be considered Level 0 files. No special tools are required to retrieve Level 0 data.
Level 1	Temporary files	Temporary files, including print spooler files, browser cache files, files for “helper” applications, and recycle bin files. Most users either expect the system to automatically delete this data or are not even aware that it exists. Note: Level 0 files are a subset of Level 1 files. Experience has shown that it is useful to distinguish this subset, because many naive users will overlook Level 1 files when they are browsing a computer’s hard drive to see if it contains sensitive information. No special tools are required to retrieve Level 1 data, although special training is required to teach the operator where to look.
Level 2	Deleted files	When a file is deleted from a file system, most operating systems do not overwrite the blocks on the hard disk that the file is written on. Instead, they simply remove the file’s reference from the containing directory. The file’s blocks are then placed on the free list. These files can be recovered using traditional “undelete” tools, such as Norton Utilities.
Level 3	Retained data blocks	Data that can be recovered from a disk, but which does not obviously belong to a named file. Level 3 data includes information in slack space, backing store for virtual memory, and Level 2 data that has been partially overwritten so that an entire file cannot be recovered. A common source of Level 3 data is disks that have been formatted with Windows `Format` command or the Unix `newfs` command. Even though the output of these commands might imply that they overwrite the entire hard drive, in fact they do not, and the vast majority of the formatted disk’s information is recoverable with the proper tools. Level 3 data can be recovered using advanced data recovery tools that can “unformat” a disk drive or special-purpose forensics tools.
Level 4	Vendor-hidden data	This level consists of data blocks that can only be accessed using vendor-specific commands. This level includes the drive’s controlling program and blocks used for bad-block management.
Level 5	Overwritten data	Many individuals maintain that information can be recovered from a hard drive even after it is overwritten. We reserve Level 5 for such information.

Garfinkel, Simson L., and Abhi Shelat. "Remembrance of Data Passed: A Study of Disk Sanitization Practices." *IEEE Security and Privacy* 1 (2003): 17-27.

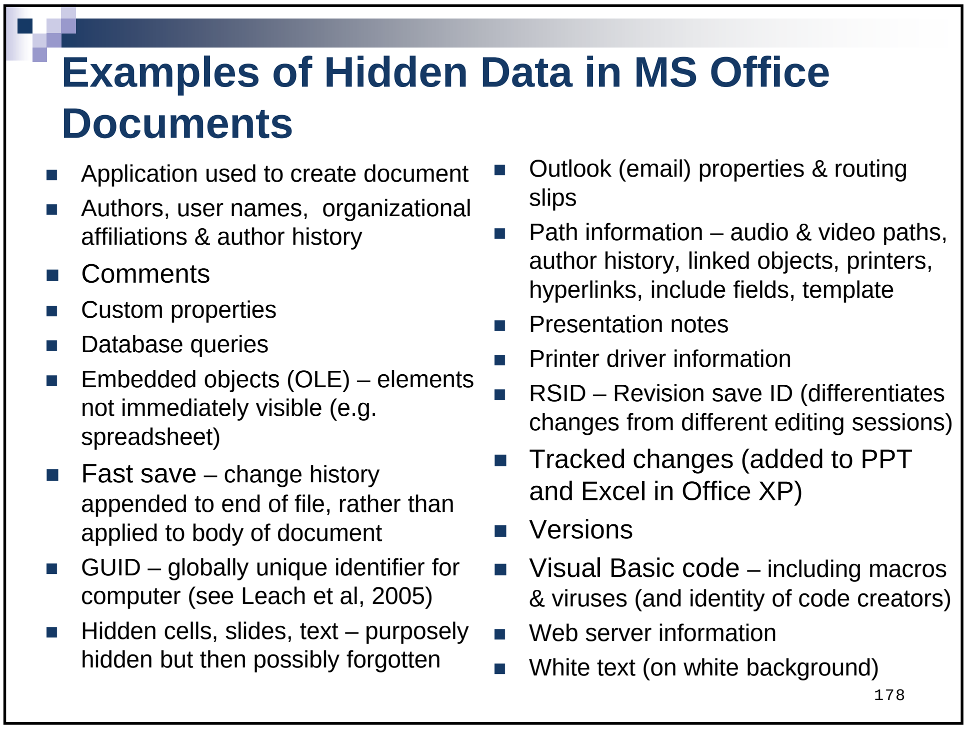
176



Within the Files Themselves

- Lots of data in many files that you don't always see with the naked eye. For example:
  - □ Comments within the code
  - □ Stored rules & styles
  - □ Change tracking information
  - □ Metadata stored in file headers & elsewhere
  - □ Viruses!

177

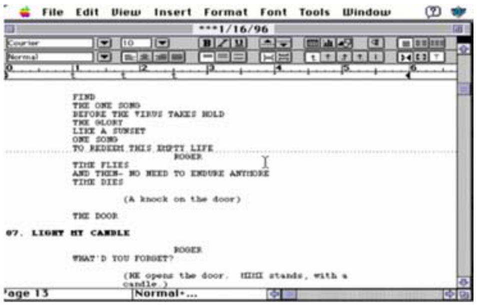


Examples of Hidden Data in MS Office Documents

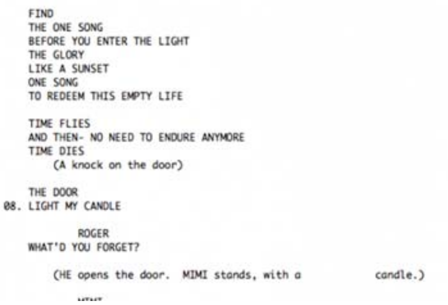
- Application used to create document
- Authors, user names, organizational affiliations & author history
- Comments
- Custom properties
- Database queries
- Embedded objects (OLE) – elements not immediately visible (e.g. spreadsheet)
- Fast save – change history appended to end of file, rather than applied to body of document
- GUID – globally unique identifier for computer (see Leach et al, 2005)
- Hidden cells, slides, text – purposely hidden but then possibly forgotten
- Outlook (email) properties & routing slips
- Path information – audio & video paths, author history, linked objects, printers, hyperlinks, include fields, template
- Presentation notes
- Printer driver information
- RSID – Revision save ID (differentiates changes from different editing sessions)
- Tracked changes (added to PPT and Excel in Office XP)
- Versions
- Visual Basic code – including macros & viruses (and identity of code creators)
- Web server information
- White text (on white background)

178

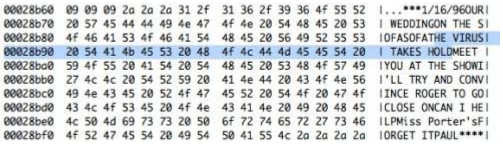
Jonathan Larson Fast Save Example



```
          FIND
          THE ONE SONG
          BEFORE THE VIRUS TAKES HOLD
          THE GLORY
          LIKE A SUNSET
          ONE SONG
          TO REDEEM THIS EMPTY LIFE
                         ROGER
          TIME FLIES
          AND THEN- NO NEED TO ENDURE ANYMORE
          TIME DIES

                 (A knock on the door)

          THE DOOR

07. LIGHT MY CANDLE

                         ROGER
          WHAT'D YOU FORGET?

                 (HE opens the door.  MIMI stands, with a
                 candle.)
```



```
    FIND
    THE ONE SONG
    BEFORE YOU ENTER THE LIGHT
    THE GLORY
    LIKE A SUNSET
    ONE SONG
    TO REDEEM THIS EMPTY LIFE

    TIME FLIES
    AND THEN- NO NEED TO ENDURE ANYMORE
    TIME DIES
       (A knock on the door)

    THE DOOR
08. LIGHT MY CANDLE

          ROGER
    WHAT'D YOU FORGET?

       (HE opens the door.  MIMI stands, with a          candle.)
```



```
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****|
```

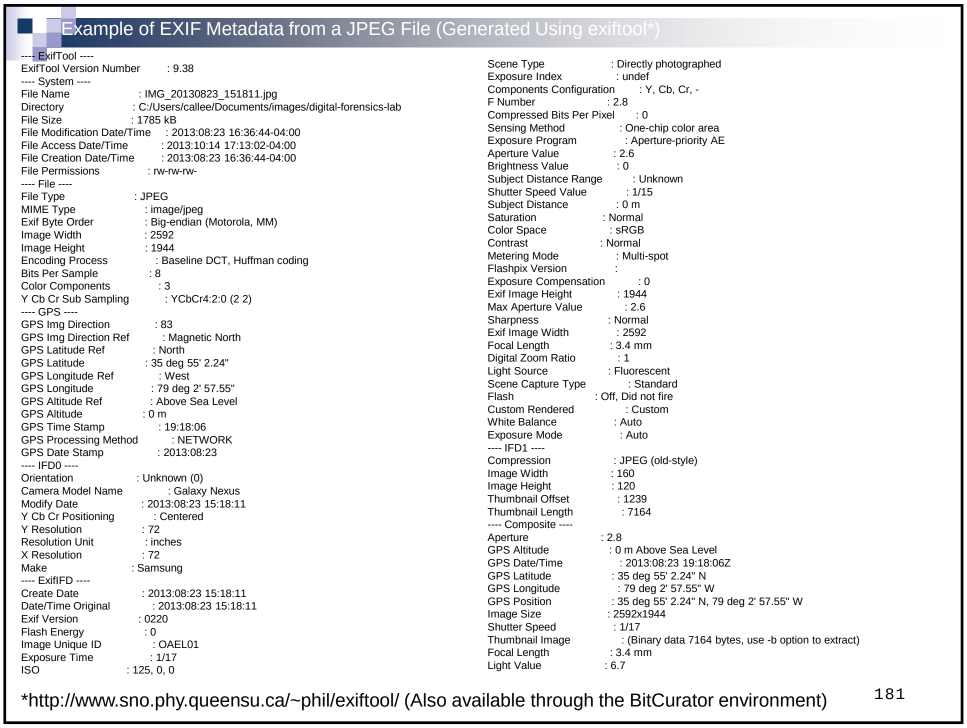
http://www.nypl.org/blog/2011/04/22/no-day-today-look-jonathan-larsons-word-files

179

Hidden Image Data

- Content outside crop area
- Layered objects (hidden from view)
- Pixel information in resized and embedded image
- Metadata:
  - □GIF – comment extensions and application extensions
  - □JPEG - camera use, date/time, distance settings, location, thumbnail image

180

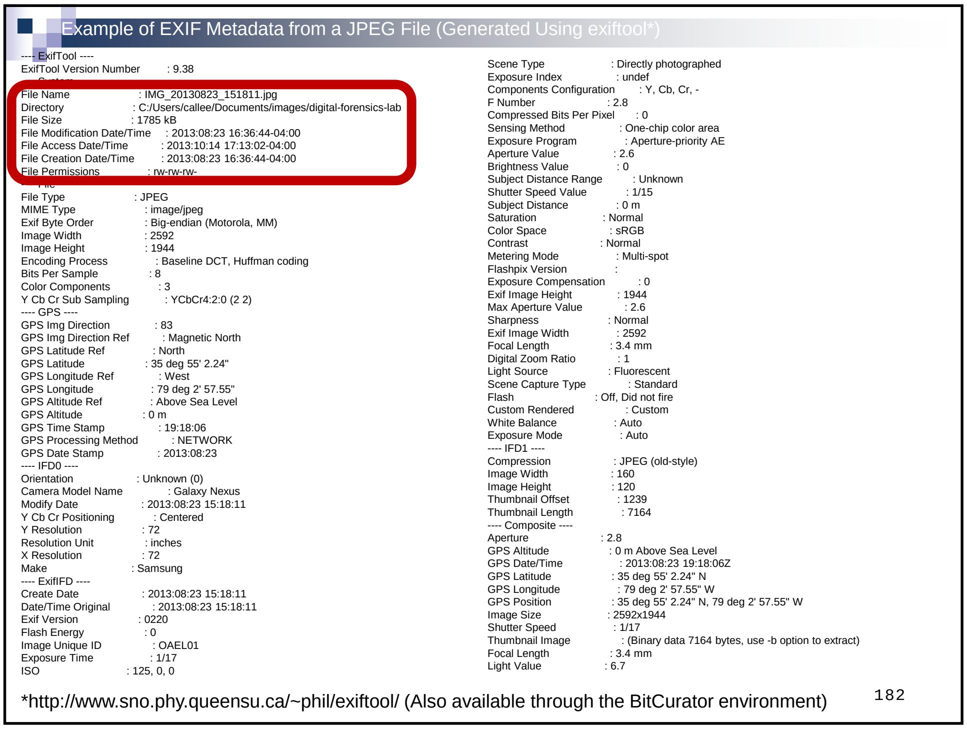


Example of EXIF Metadata from a JPEG File (Generated Using exiftool*)

```
---- 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
```

*http://www.sno.phy.queensu.ca/~phil/exiftool/ (Also available through the BitCurator environment)

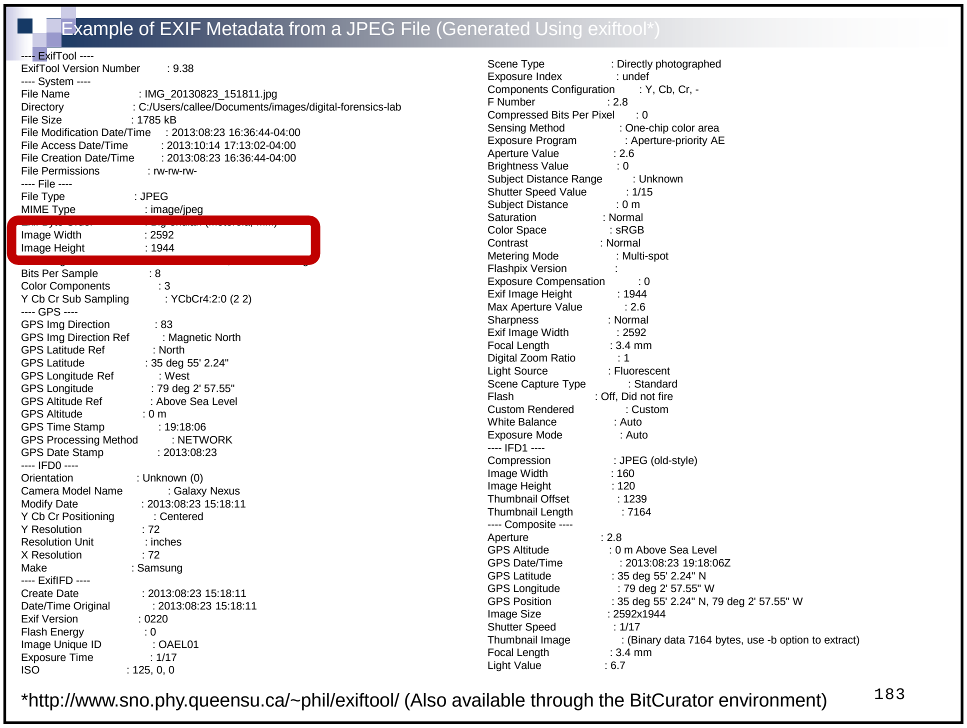


Example of EXIF Metadata from a JPEG File (Generated Using exiftool*)

```
---- ExifTool ----
ExifTool Version Number        : 9.38
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 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
```

*http://www.sno.phy.queensu.ca/~phil/exiftool/ (Also available through the BitCurator environment)



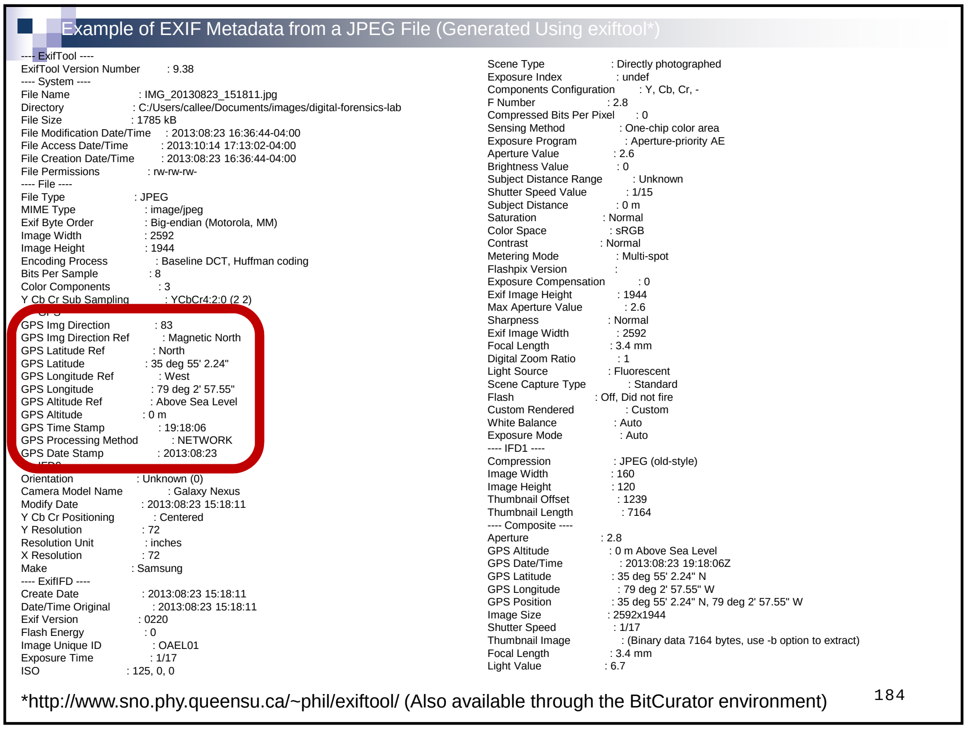
Example of EXIF Metadata from a JPEG File (Generated Using exiftool*)

```
---- 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
```

*http://www.sno.phy.queensu.ca/~phil/exiftool/ (Also available through the BitCurator environment)

183



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
```

*http://www.sno.phy.queensu.ca/~phil/exiftool/ (Also available through the BitCurator environment)

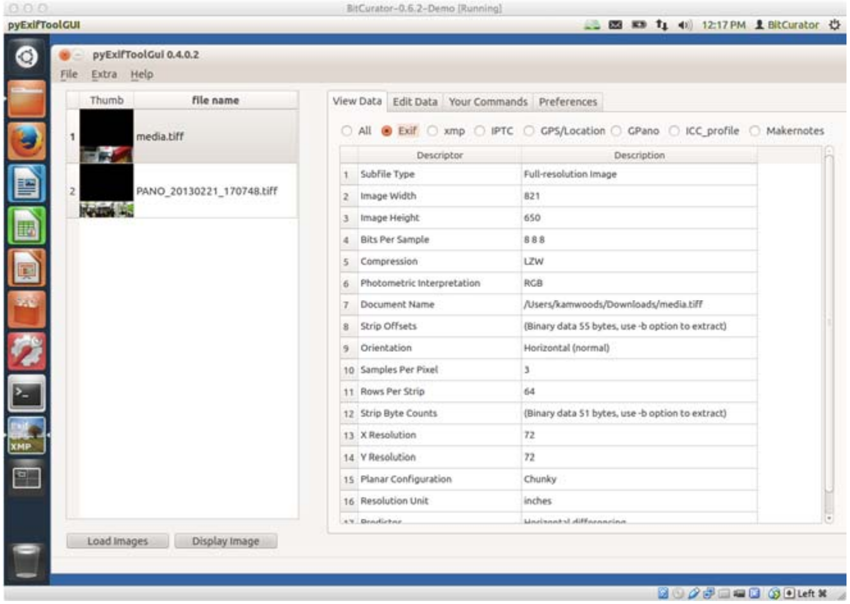
184

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
```

*http://www.sno.phy.queensu.ca/~phil/exiftool/ (Also available through the BitCurator environment)

EXIF Metadata From Header of a TIFF File*

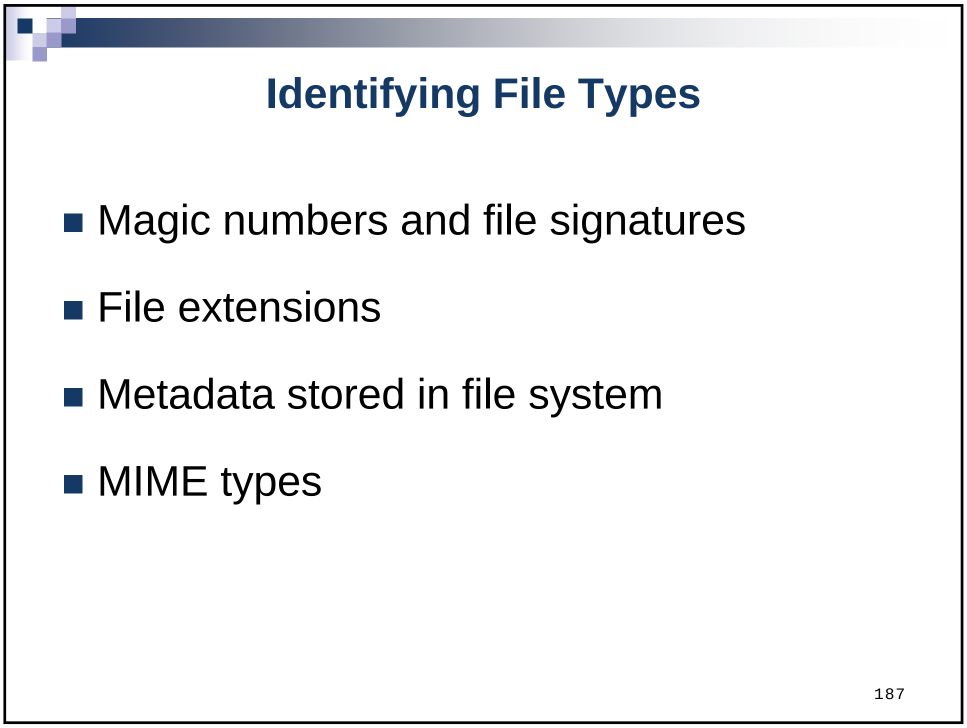


	Thumb	File name
1		media.tiff
2		PANO_20130221_170748.tiff

○ All ● Exif ○ xmp ○ IPTC ○ GPS/Location ○ GPano ○ ICC_profile ○ Makernotes

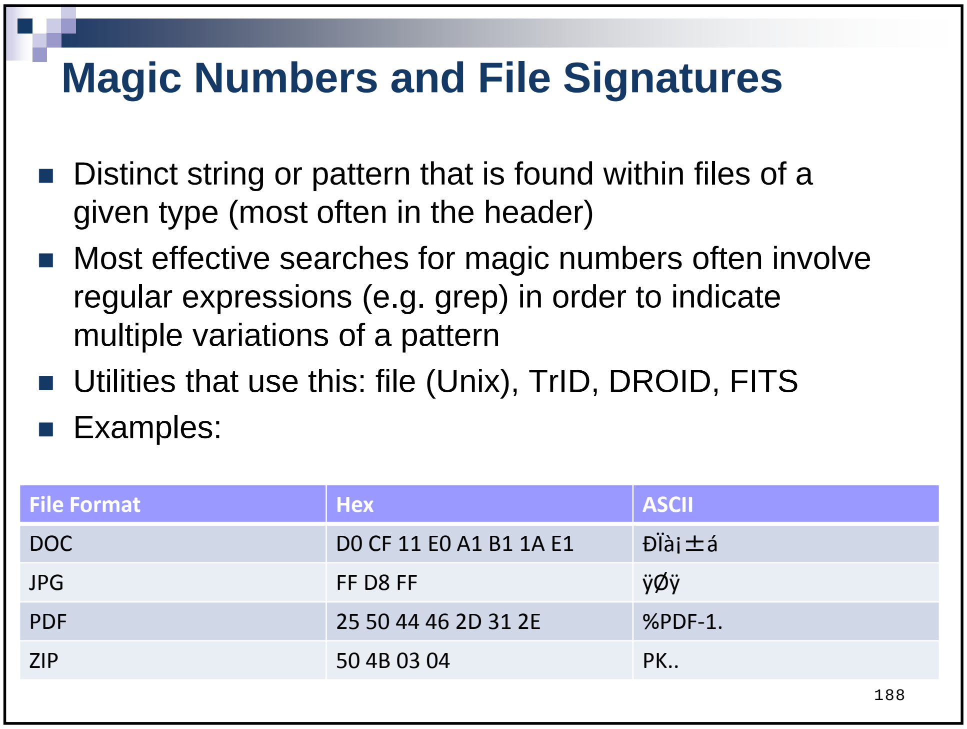
	Descriptor	Description
1	SubFile Type	Full-resolution Image
2	Image Width	821
3	Image Height	650
4	Bits Per Sample	8 8 8
5	Compression	LZW
6	Photometric Interpretation	RGB
7	Document Name	/Users/kamwoods/Downloads/media.tiff
8	Strip Offsets	(Binary data 55 bytes, use -b option to extract)
9	Orientation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

*Using pyExifToolGUI in the BitCurator environment



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..

Try it Yourself!

- Go back to the file that you downloaded earlier and open it again in HxD or Hex Fiend
- What do you see at the beginning?
- Is there a pattern that matches one of these?

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..

189

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

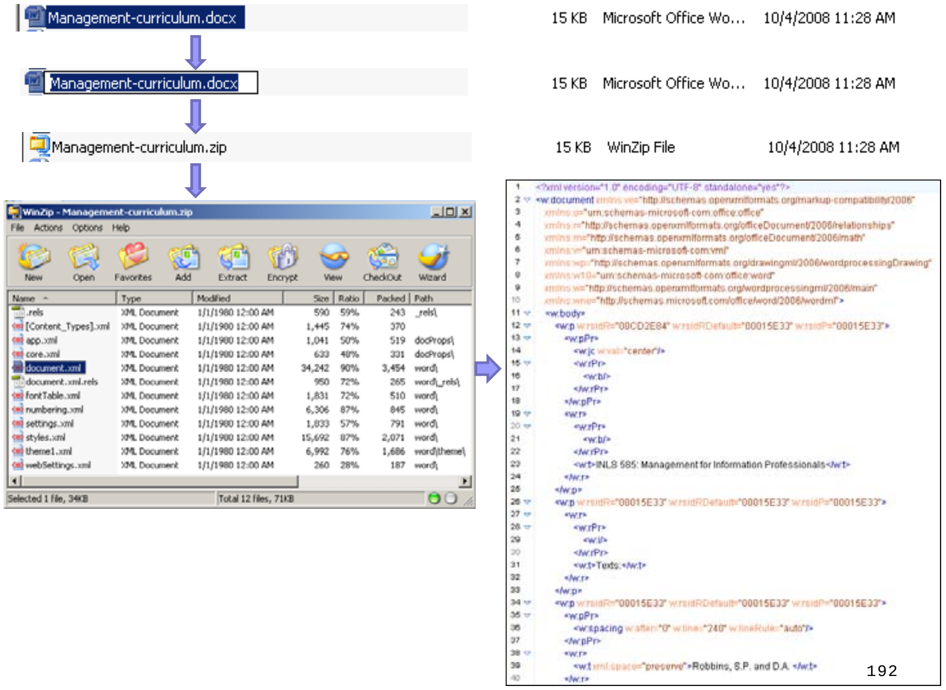
- FITS "identifies, validates, and extracts technical metadata for various file formats. It wraps several third-party open source tools, normalizes and consolidates their output, and reports any errors. FITS was created by the Harvard University Library Office for Information Systems for use in its Digital Repository Service (DRS)."
- Tools currently bundled into it:
  - □ Jhove
  - □ Exiftool
  - □ National Library of New Zealand Metadata Extractor
  - □ DROID
  - □ FFIdent
  - □ File Utility (windows)
- Note: you can find and run FITS from the command line in the BitCurator environment (found in "Additional Tools")

190

File Extensions

- Changing file extension usually changes default application that 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 portion of 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)

191



Management-curriculum.docx 15 KB Microsoft Office Wo... 10/4/2008 11:28 AM

Management-curriculum.docx 15 KB Microsoft Office Wo... 10/4/2008 11:28 AM

Management-curriculum.zip 15 KB WinZip File 10/4/2008 11:28 AM

WinZip - Management-curriculum.zip

Name	Type	Modified	Size	Ratio	Packed	Path
.rels	XML Document	1/1/1980 12:00 AM	590	59%	243	_rels\
[Content_Types].xml	XML Document	1/1/1980 12:00 AM	1,445	74%	370	
app.xml	XML Document	1/1/1980 12:00 AM	1,041	50%	519	docProps\
core.xml	XML Document	1/1/1980 12:00 AM	633	48%	331	docProps\
document.xml	XML Document	1/1/1980 12:00 AM	34,242	90%	3,454	word\
document.xml.rels	XML Document	1/1/1980 12:00 AM	950	72%	265	word_rels\
fontTable.xml	XML Document	1/1/1980 12:00 AM	1,831	72%	510	word\
numbering.xml	XML Document	1/1/1980 12:00 AM	6,306	87%	845	word\
settings.xml	XML Document	1/1/1980 12:00 AM	1,833	57%	791	word\
styles.xml	XML Document	1/1/1980 12:00 AM	15,692	87%	2,071	word\
theme1.xml	XML Document	1/1/1980 12:00 AM	6,992	76%	1,686	word\theme\
webSettings.xml	XML Document	1/1/1980 12:00 AM	260	28%	187	word\

Selected 1 file, 34KB Total 12 files, 71KB

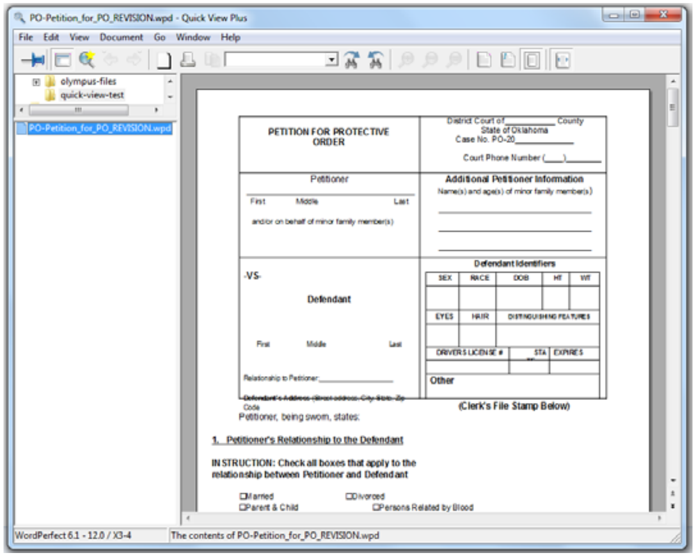
```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<w:document xmlns:ve="http://schemas.openxmlformats.org/markup-compatibility/2006"
  xmlns:o="urn:schemas-microsoft-com:office:office"
  xmlns:r="http://schemas.openxmlformats.org/officeDocument/2006/relationships"
  xmlns:m="http://schemas.openxmlformats.org/officeDocument/2006/math"
  xmlns:v="urn:schemas-microsoft-com:vml"
  xmlns:wp="http://schemas.openxmlformats.org/drawingml/2006/wordprocessingDrawing"
  xmlns:w10="urn:schemas-microsoft-com:office:word"
  xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main"
  xmlns:wne="http://schemas.microsoft.com/office/word/2006/wordml">
  <w:body>
    <w:p w:rsidR="00CD2E84" w:rsidRDefault="00015E33" w:rsidP="00015E33">
      <w:pPr>
        <w:jc w:val="center"/>
        <w:rPr>
          <w:b/>
        </w:rPr>
      </w:pPr>
      <w:r>
        <w:rPr>
          <w:b/>
        </w:rPr>
        <w:t>INLS 585: Management for Information Professionals</w:t>
      </w:r>
    </w:p>
    <w:p w:rsidR="00015E33" w:rsidRDefault="00015E33" w:rsidP="00015E33">
      <w:r>
        <w:rPr>
          <w:i/>
        </w:rPr>
        <w:t>Texts:</w:t>
      </w:r>
    </w:p>
    <w:p w:rsidR="00015E33" w:rsidRDefault="00015E33" w:rsidP="00015E33">
      <w:pPr>
        <w:spacing w:after="0" w:line="240" w:lineRule="auto"/>
      </w:pPr>
      <w:r>
        <w:t xml:space="preserve">Robbins, S.P. and D.A. </w:t>
      </w:r>
```

192

MIME types

- 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

Reproducing In-Application Rendering (Sort of) for Archival Processing


PO-Petition_for_PO_REVISION.wpd - Quick View Plus
File
Edit
View
Document
Go
Window
Help
olympus-files
quick-view-test
PO-Petition_for_PO_REVISION.wpd
WordPerfect 6.1 - 12.0 / X3-4
The contents of PO-Petition_for_PO_REVISION.wpd

PETITION FOR PROTECTIVE ORDER	District Court of____________ County State of Oklahoma Case No. PO-20____________ Court Phone Number (___)________
Petitioner First Middle Last and/or on behalf of minor family member(s)	**Additional Petitioner Information** Name(s) and age(s) of minor family member(s)
-VS- **Defendant** First Middle Last Relationship to Petitioner:________________ Defendant's Address (Street address, City State, Zip Code	**Defendant Identifiers**

SEX	RACE	DOB	HT	WT
EYES	HAIR	DISTINGUISHING FEATURES		
DRIVERS LICENSE #	STA	EXPIRES		
Other				

(Clerk's File Stamp Below)

Petitioner, being sworn, states:

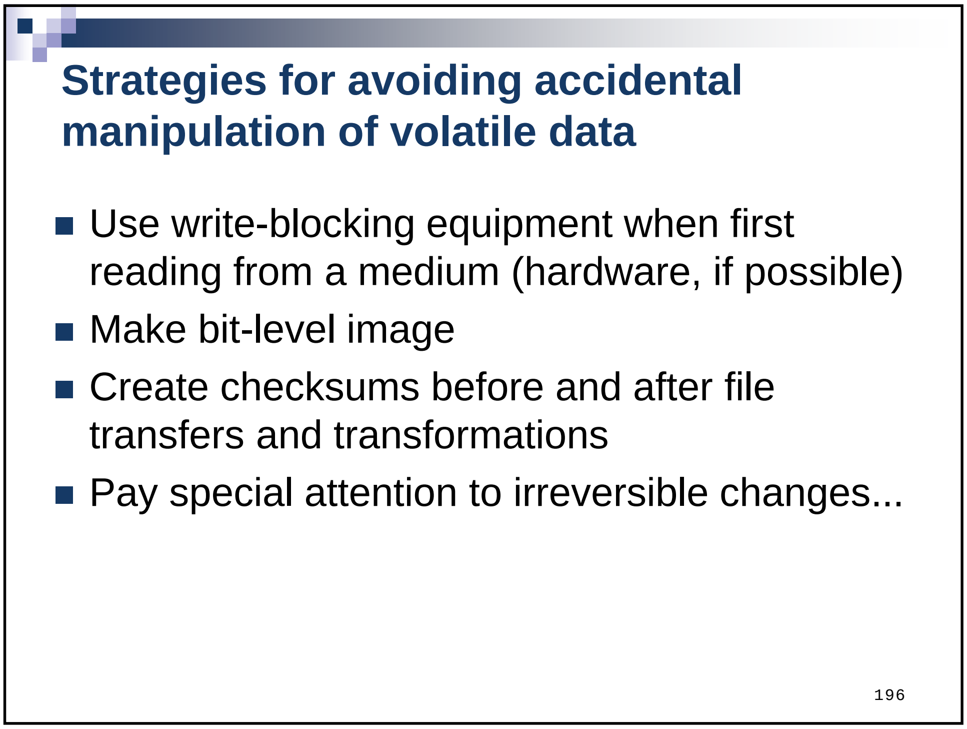
1. Petitioner's Relationship to the Defendant

INSTRUCTION: Check all boxes that apply to the relationship between Petitioner and Defendant

□Married □Divorced
□Parent & Child □Persons Related by Blood

Viewing a WordPerfect document on a computer that doesn't have WordPerfect installed, using Quick View Plus

Properly Extracting Data from Media



Strategies for avoiding accidental manipulation of volatile data

- Use write-blocking equipment when first reading from a medium (hardware, if possible)
- Make bit-level image
- Create checksums before and after file transfers and transformations
- Pay special attention to irreversible changes...

Examples of Irreversible Changes

- Lossy compression (e.g. JPEG)
- Lower-quality surrogate (e.g. thumbnail image, access copy of video)
- Format conversion (e.g. Word to PDF/A, Excel to CSV)
- Character encoding (e.g. EBCDIC to ASCII)
- Normalization of data values (e.g. date values in a database to a common date encoding)
- Rewriting pointers (e.g. links in a web site from absolute to relative or vice versa)
- Overwriting older versions files or values with newer versions
- Pulling files out of their native file system

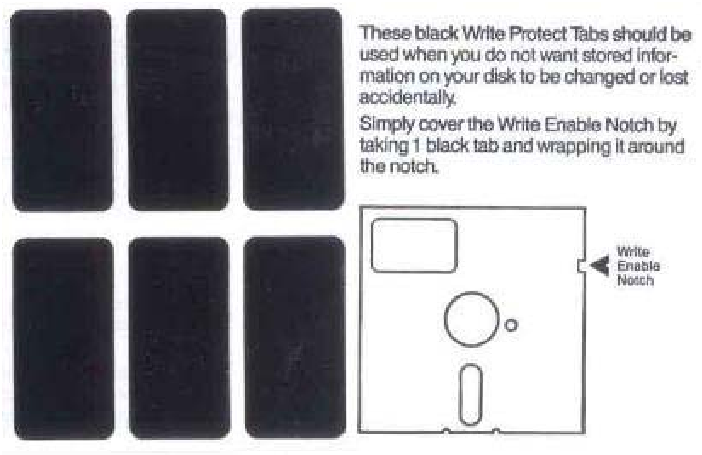
Write Blocking – One-Way Streets for Data

- 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

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



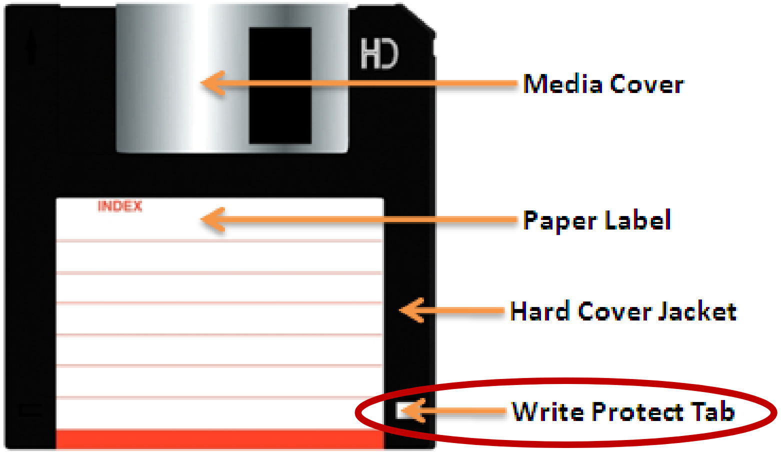
These black Write Protect Tabs should be used when you do not want stored information on your disk to be changed or lost accidentally.

Simply cover the Write Enable Notch by taking 1 black tab and wrapping it around the notch.

Write
Enable
Notch

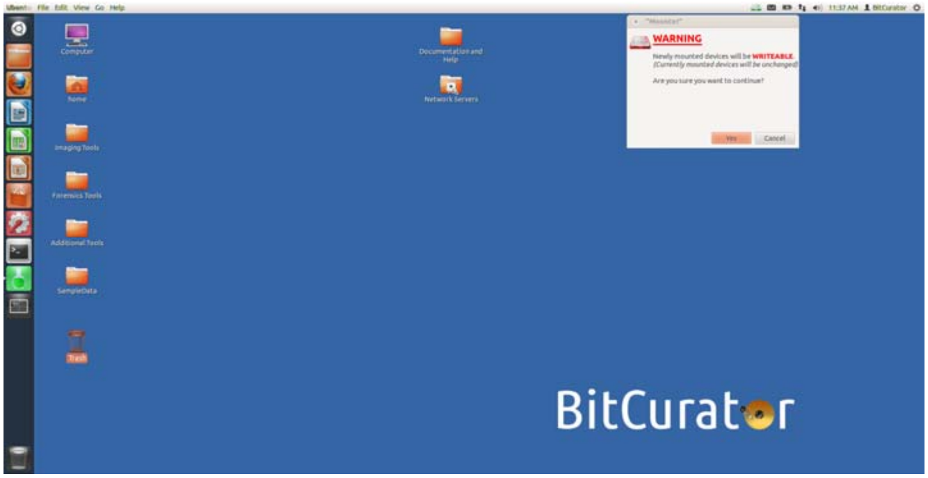
http://en.wikipedia.org/wiki/File:Floppy_tabs_3x2.jpg

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



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



WARNING

Newly mounted devices will be WRITEABLE.
(Currently mounted devices will be unchanged)

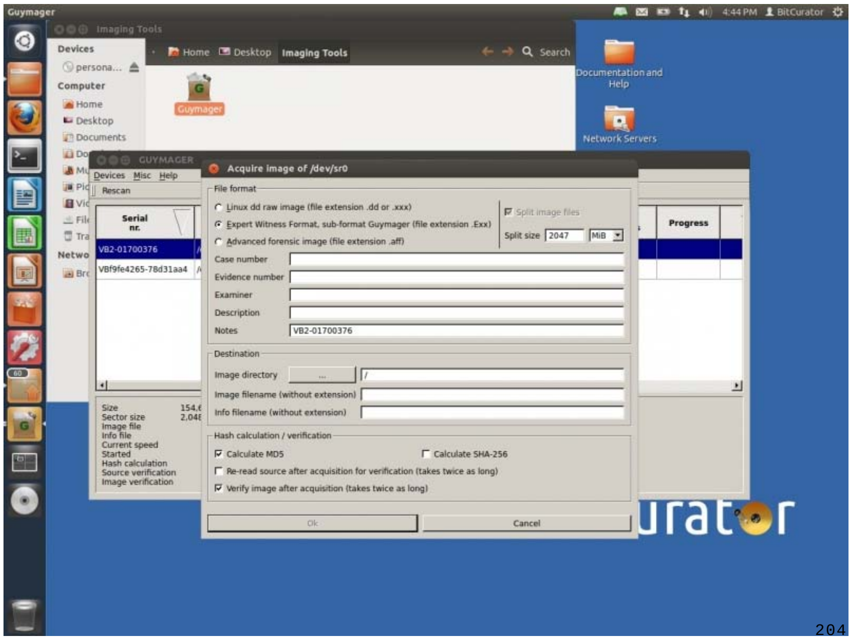
Are you sure you want to continue?

Getting below the File System – Low-Level Copying

- 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
- Most commonly used tool is dd (or variant) - UNIX program for low-level copying and conversion of data from a storage device
- More specialized tools for creating forensic images include:
  - FTK Imager
  - Guymager
  - Imaging utilities in commercial applications (e.g. EnCase)

203

Main Acquisition Interface for Guymager



VB2-01700376
VBf9fe4265-78d31aa4

Acquire image of /dev/sr0

File format

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

☑ Split image files
Split size 2047 MiB

Case number
Evidence number
Examiner
Description
Notes VB2-01700376

Destination

Image directory ... /
Image filename (without extension)
Info filename (without extension)

Hash calculation / verification

☑ Calculate MD5 □ Calculate SHA-256
□ Re-read source after acquisition for verification (takes twice as long)
☑ Verify image after acquisition (takes twice as long)

Ok Cancel

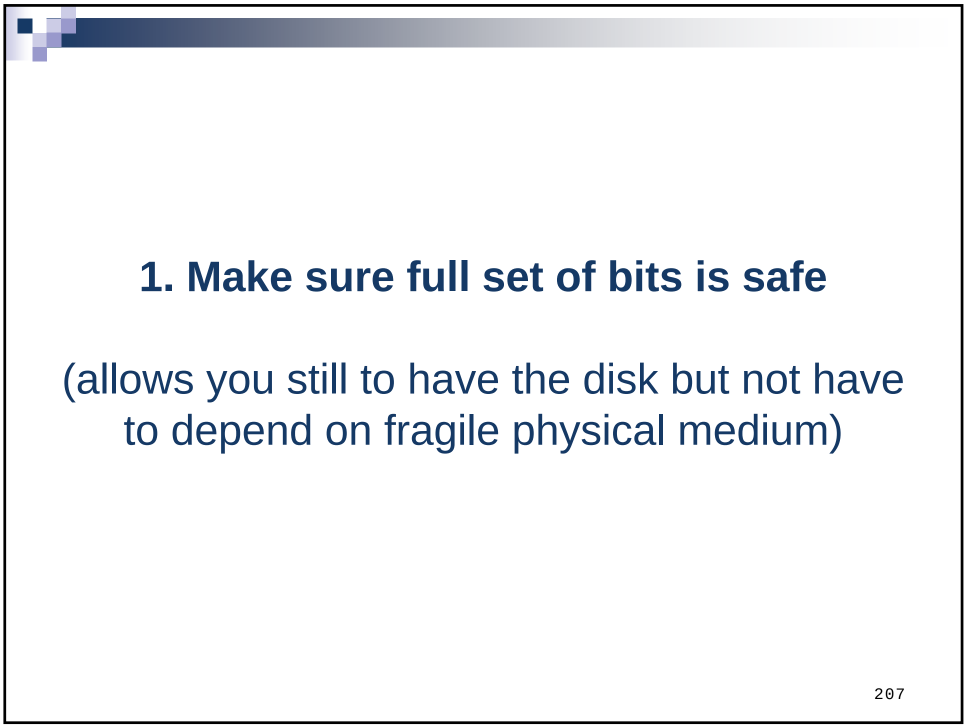
204

“The file system provides little if any help when you want to know about details of where and how it stores information; indeed, the entire **purpose of file systems is to hide such detail**. In order to look under the file system, you have to **bypass the file system code** and use tools that duplicate some of the file system's functionality.”*

*Farmer, Dan, and Wietse Venema. *Forensic Discovery*. Upper Saddle River, NJ: Addison-Wesley, 2005. (emphasis added)

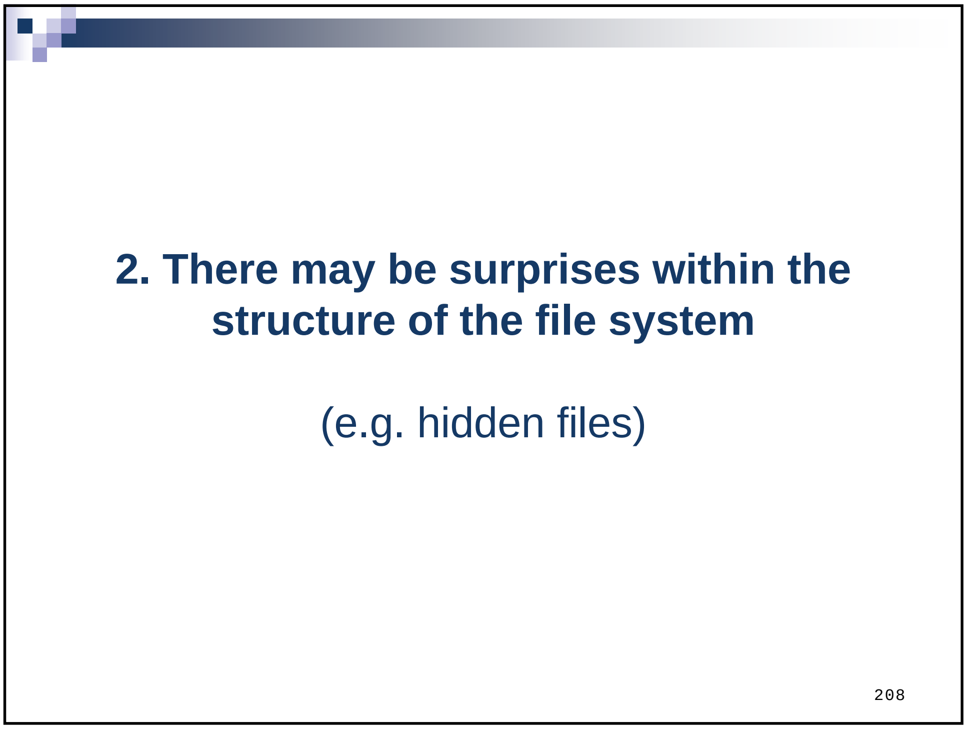
Why Make a Bit-Level Image?

7 Reasons



1. Make sure full set of bits is safe

(allows you still to have the disk but not have to depend on fragile physical medium)



2. There may be surprises within the structure of the file system

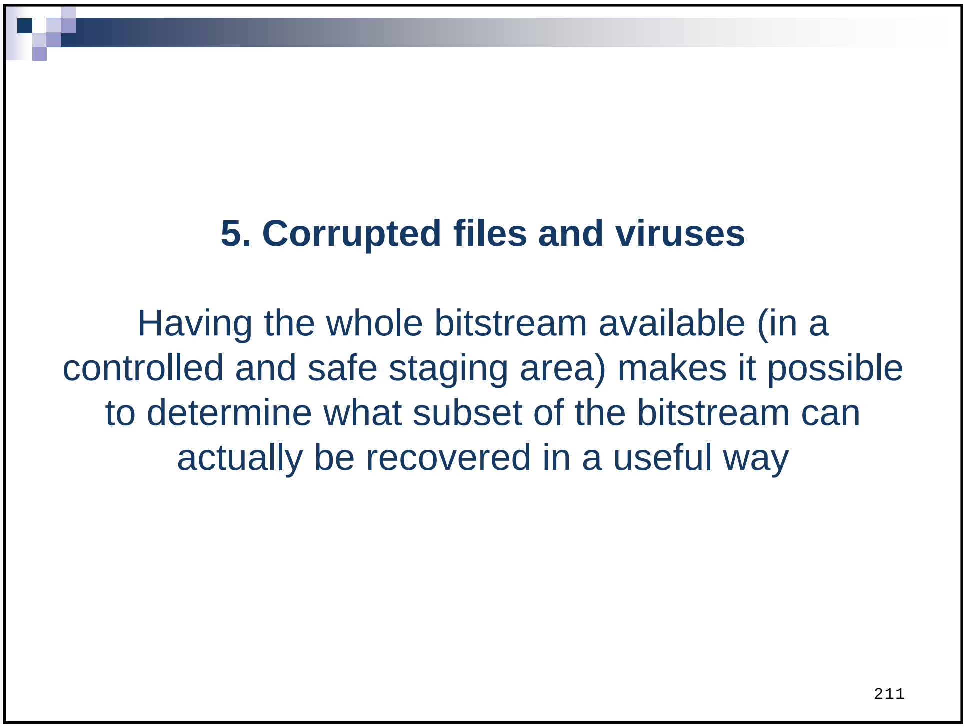
(e.g. hidden files)

3. You could inadvertently change something in the act of examining or dealing with the files

- Byte order
- Character encoding
- File system information
  - □ MAC values
  - □ Access permissions
  - □ File typing
  - □ File sizes
- For example, when using a Windows machine to accession an HFS (Mac) disk, good idea to image the disk right away, so errors in translation across file systems can be noticed and corrected

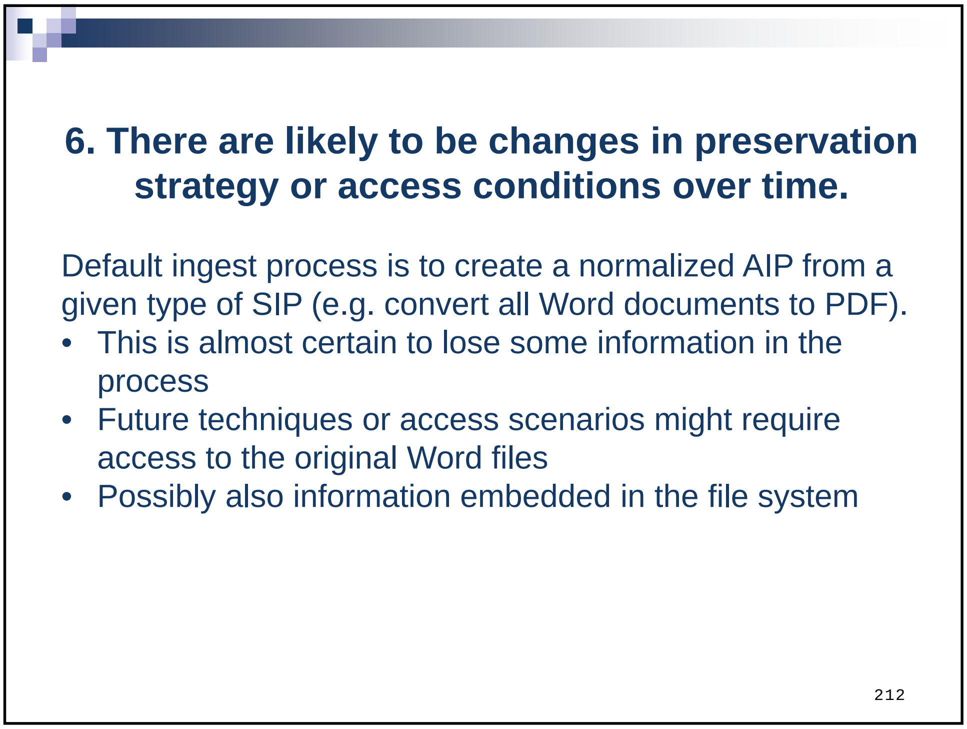
4. Proof of file integrity and chain of custody

If there are questions about whether a given source was the basis for a given set of digital objects, one can go back to the original bits and compare hash values



5. Corrupted files and viruses

Having the whole bitstream available (in a controlled and safe staging area) makes it possible to determine what subset of the bitstream can actually be recovered in a useful way



6. There are likely to be changes in preservation strategy or access conditions over time.

Default ingest process is to create a normalized AIP from a given type of SIP (e.g. convert all Word documents to PDF).

- This is almost certain to lose some information in the process
- Future techniques or access scenarios might require access to the original Word files
- Possibly also information embedded in the file system

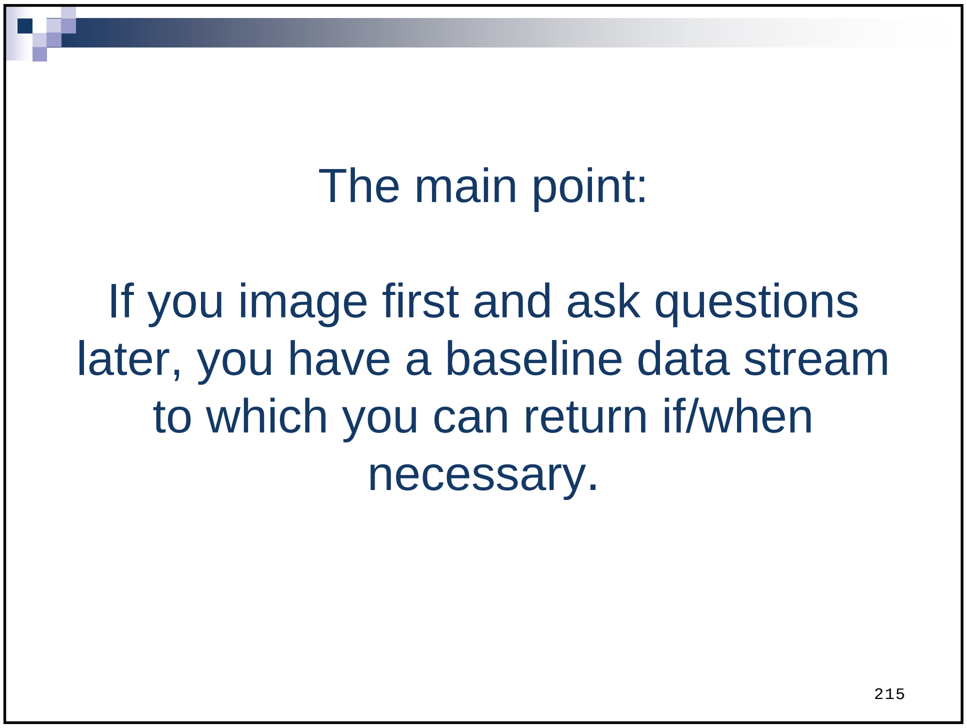
7. Embedded Contextual Information and User Artifacts

Depending on understanding of arrangement with the Producer, hidden data can also serve as important evidence for the curation of a collection, e.g. traces of data that indicate what application created the files, login or password information that's necessary for accessing various data sources*

*For further discussion of possibilities, see: Garfinkel, Simson, and David Cox. "Finding and Archiving the Internet Footprint." Paper presented at the First Digital Lives Research Conference: Personal Digital Archives for the 21st Century, London, UK, February 9-11, 2009.

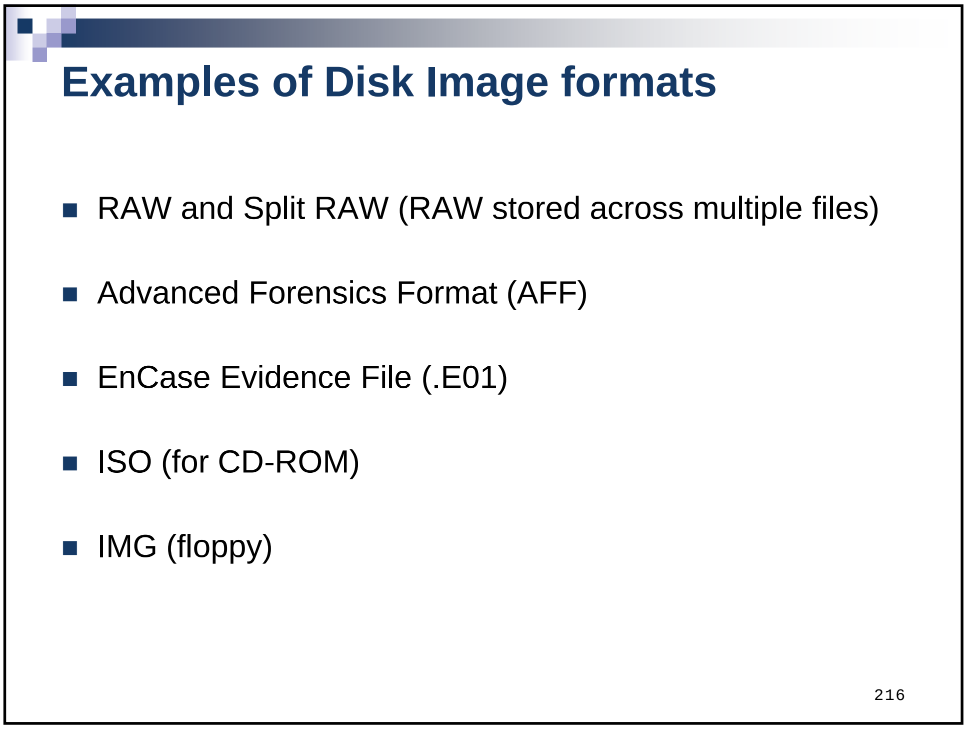
A major rule of digital curation is to minimize irreversible transformations. Copying files off of the original bitstream and then discarding that bistream is just such an irreversible change (no way to then derive the original bits from the files you have).

It's pretty easy to screw up files or file attributes in numerous ways, and the imaging step can dramatically increase the possibility that those screw ups are reversible.



The main point:

If you image first and ask questions later, you have a baseline data stream to which you can return if/when necessary.



Examples of Disk Image formats

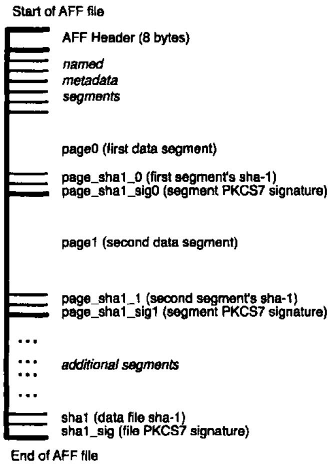
- RAW and Split RAW (RAW stored across multiple files)
- Advanced Forensics Format (AFF)
- EnCase Evidence File (.E01)
- ISO (for CD-ROM)
- IMG (floppy)

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).

AFF

- Original AFF format: single file containing segments with drive data and metadata. Can be compressed.
- Large AFF files can be broken into multiple AFD format files. The smaller AFD files can be readily moved around a FAT32 file system which limits files to 2GB or stored on DVDs, which have similar size restrictions.
- AFM format stores the metadata in an AFF file, and the disk data in a separate raw file. This format allows analysis tools that support the raw format to access the data, but without losing the metadata.

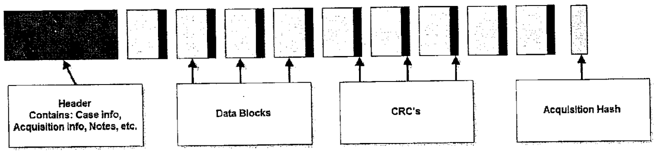

Start of AFF file
AFF Header (8 bytes)
named metadata segments
page0 (first data segment)
page_sha1_0 (first segment's sha-1)
page_sha1_sig0 (segment PKCS7 signature)
page1 (second data segment)
page_sha1_1 (second segment's sha-1)
page_sha1_sig1 (segment PKCS7 signature)
additional segments
sha1 (data file sha-1)
sha1_sig (file PKCS7 signature)
End of AFF file

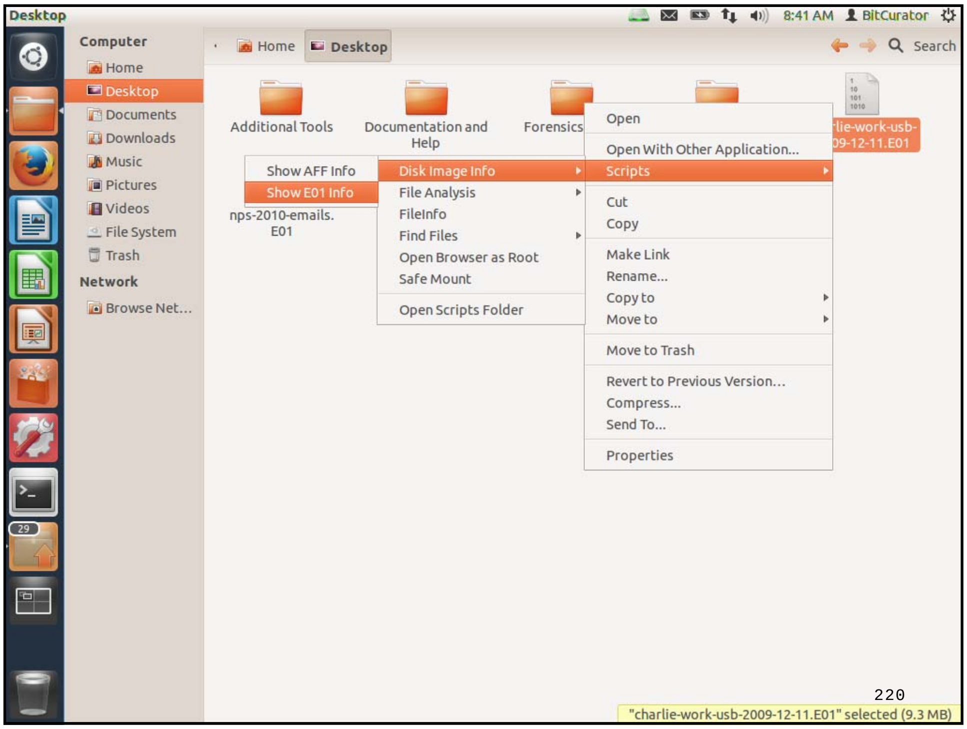
AFF File
(Not to Scale)

Image source: Simson Garfinkel, David Malan, Karl-Alexander Dubec, Christopher Stevens, and Cecile Pham, "AFF An Open Extensible Format for Disk Imaging," *Advances in Digital Forensics II*, edited by Martin S. Olivier and Sujeet Shenoi (New York: Springer, 2006), 13-28.

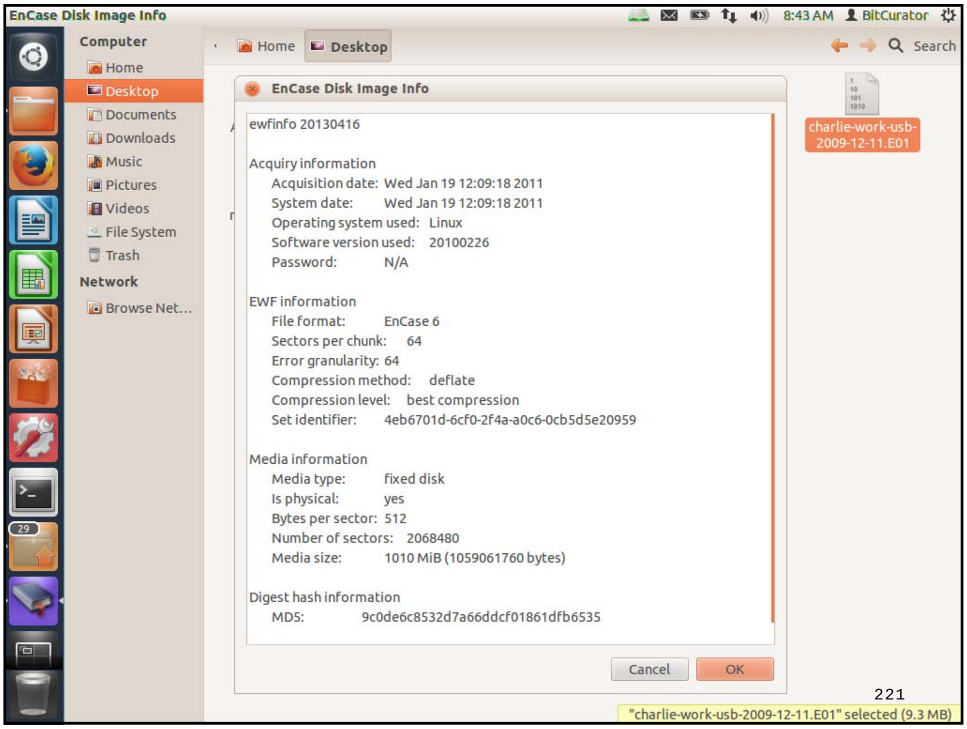
Expert Witness Format – EWF (EnCase)

- 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) - http://sourceforge.net/projects/libewf/files/





"charlie-work-usb-2009-12-11.E01" selected (9.3 MB)



EnCase Disk Image Info

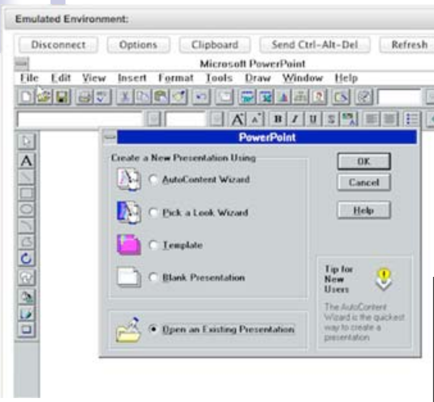
```
ewfinfo 20130416

Acquiry information
    Acquisition date: Wed Jan 19 12:09:18 2011
    System date:      Wed Jan 19 12:09:18 2011
    Operating system used:  Linux
    Software version used:  20100226
    Password:         N/A

EWF information
    File format:      EnCase 6
    Sectors per chunk:    64
    Error granularity: 64
    Compression method:   deflate
    Compression level:    best compression
    Set identifier:   4eb6701d-6cf0-2f4a-a0c6-0cb5d5e20959

Media information
    Media type:       fixed disk
    Is physical:      yes
    Bytes per sector: 512
    Number of sectors:    2068480
    Media size:       1010 MiB (1059061760 bytes)

Digest hash information
    MD5:          9c0de6c8532d7a66ddcf01861dfb6535
```

"charlie-work-usb-2009-12-11.E01" selected (9.3 MB)

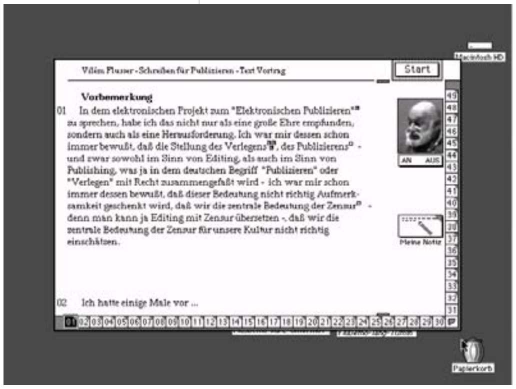
Four Ways to Interact with Disk Images

- Emulation
- Mount them like regular drives:
  - For ISO images - disk utilities in Mac OS or Windows 8/10
  - For forensically packaged disk images: ewfmount, OSFMount, BitCurator (mounting scripts built into the environment)
- Inspect them as forensic objects
  - FTK Imager
  - Autopsy
  - BitCurator (Disk Image Access tool)
- Dynamically navigate them from within a web browser (BCA Webtools)

Emulation as a Service



The AutoContent Wizard is the quickest way to create a presentation



Vilém Flusser - Schreiben für Publizieren - Text Vortrag

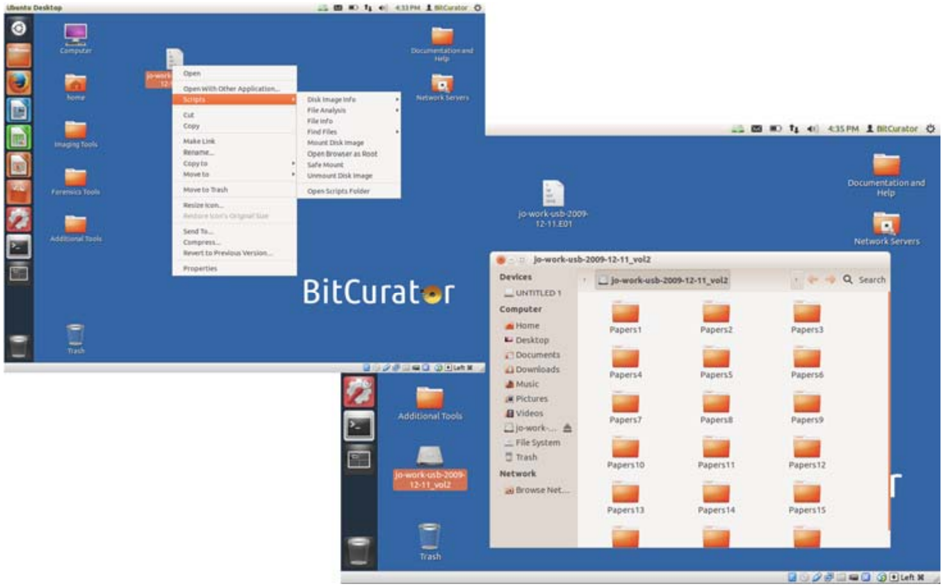
Vorbemerkung

01 In dem elektronischen Projekt zum "Elektronischen Publizieren" zu sprechen, habe ich das nicht nur als eine große Ehre empfunden, sondern auch als eine Herausforderung. Ich war mir dessen schon immer bewußt, daß die Stellung des Verlegens, des Publizierens - und zwar sowohl im Sinn von Editing, als auch im Sinn von Publishing, was ja in dem deutschen Begriff "Publizieren" oder "Verlegen" mit Recht zusammengefaßt wird - ich war mir schon immer dessen bewußt, daß dieser Bedeutung nicht richtig Aufmerksamkeit geschenkt wird, daß wir die zentrale Bedeutung der Zensur - denn man kann ja Editing mit Zensur übersetzen -, daß wir die zentrale Bedeutung der Zensur für unsere Kultur nicht richtig einschätzen.

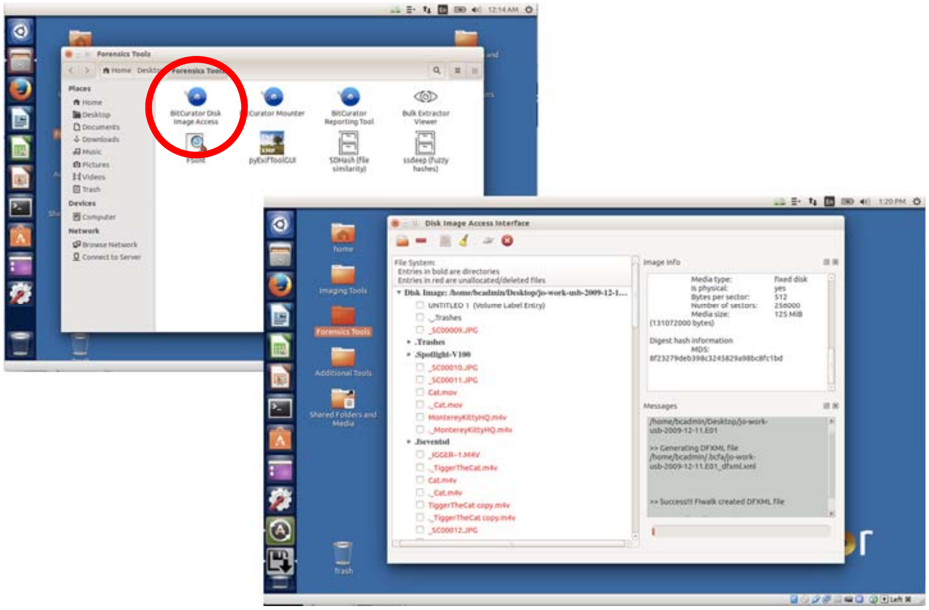
02 Ich hatte einige Male vor ...

http://bw-fla.uni-freiburg.de/demos.html

Mounting a Forensically Packaged Disk Image in the BitCurator Environment



Exporting Files from a Disk Image



File System:
Entries in bold are directories
Entries in red are unallocated/deleted files

```
▾ Disk Image: /home/bcadmin/Desktop/jo-work-usb-2009-12-1...
    □ UNTITLED 1 (Volume Label Entry)
    □ ._Trashes
    □ _SC00009.JPG
  ▸ .Trashes
  ▸ .Spotlight-V100
    □ _SC00010.JPG
    □ _SC00011.JPG
    □ Cat.mov
    □ ._Cat.mov
    □ MontereyKittyHQ.m4v
    □ ._MontereyKittyHQ.m4v
  ▸ .fseventsd
    □ _IGGER~1.M4V
    □ ._TiggerTheCat.m4v
    □ Cat.m4v
    □ ._Cat.m4v
    □ TiggerTheCat copy.m4v
    □ ._TiggerTheCat copy.m4v
    □ _SC00012.JPG
```

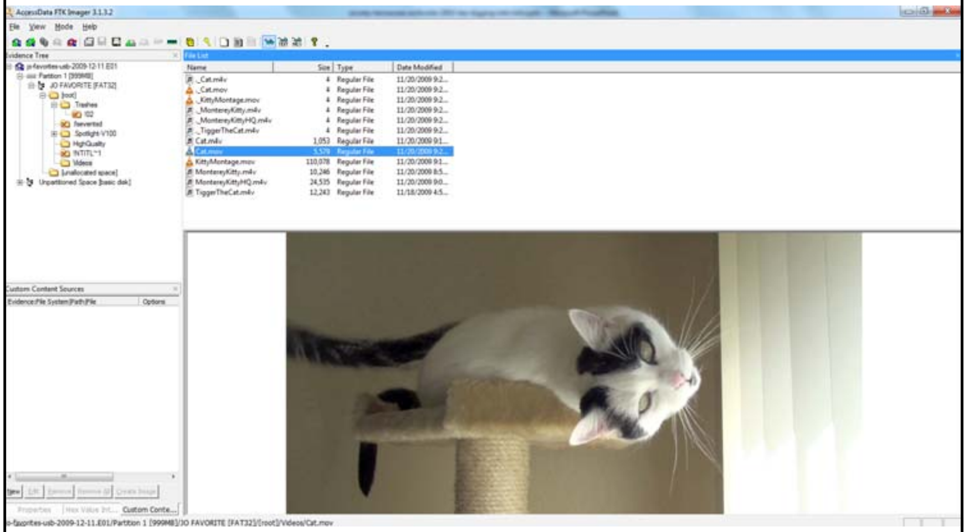
Image Info

```
            Media type:          Fixed disk
            Is physical:         yes
            Bytes per sector:    512
            Number of sectors:   256000
            Media size:          125 MiB
(131072000 bytes)

Digest hash information
            MD5:
8f23279deb398c3245829a98bc8fc1bd
```

Messages

```
/home/bcadmin/Desktop/jo-work-
usb-2009-12-11.E01

>> Generating DFXML file
/home/bcadmin/.bcfa/jo-work-
usb-2009-12-11.E01_dfxml.xml

>> Success!!! Fiwalk created DFXML file
```

Viewing Contents of a Disk Image (.E01 format) in FTK Imager

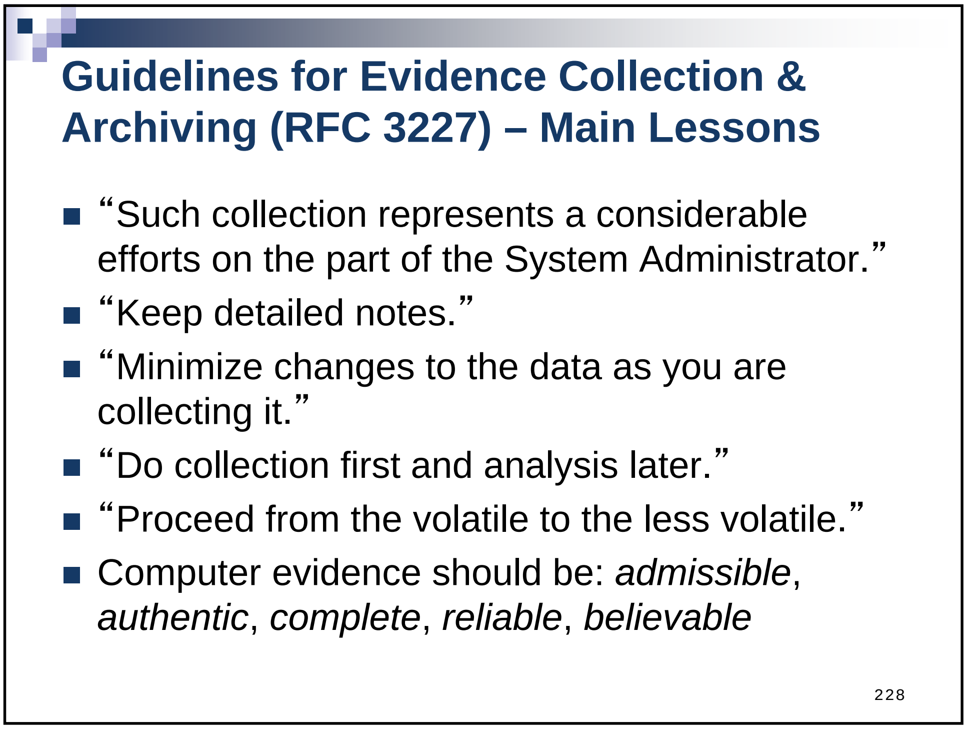


```
⊟ jo-favorites-usb-2009-12-11.E01
  ⊟ Partition 1 [999MB]
    ⊟ JO FAVORITE [FAT32]
      ⊟ [root]
        ⊟ .Trashes
            502
          .fseventsd
        ⊞ .Spotlight-V100
          HighQuality
          INTITL~1
          Videos
        [unallocated space]
  ⊞ Unpartitioned Space [basic disk]
```

Name	Size	Type	Date Modified
._Cat.m4v	4	Regular File	11/20/2009 9:2...
._Cat.mov	4	Regular File	11/20/2009 9:2...
._KittyMontage.mov	4	Regular File	11/20/2009 9:2...
._MontereyKitty.m4v	4	Regular File	11/20/2009 9:2...
._MontereyKittyHQ.m4v	4	Regular File	11/20/2009 9:2...
._TiggerTheCat.m4v	4	Regular File	11/20/2009 9:2...
Cat.m4v	1,053	Regular File	11/20/2009 9:1...
Cat.mov	5,578	Regular File	11/20/2009 9:2...
KittyMontage.mov	110,078	Regular File	11/20/2009 9:1...
MontereyKitty.m4v	10,246	Regular File	11/20/2009 8:5...
MontereyKittyHQ.m4v	24,535	Regular File	11/20/2009 9:0...
TiggerTheCat.m4v	12,243	Regular File	11/18/2009 4:5...

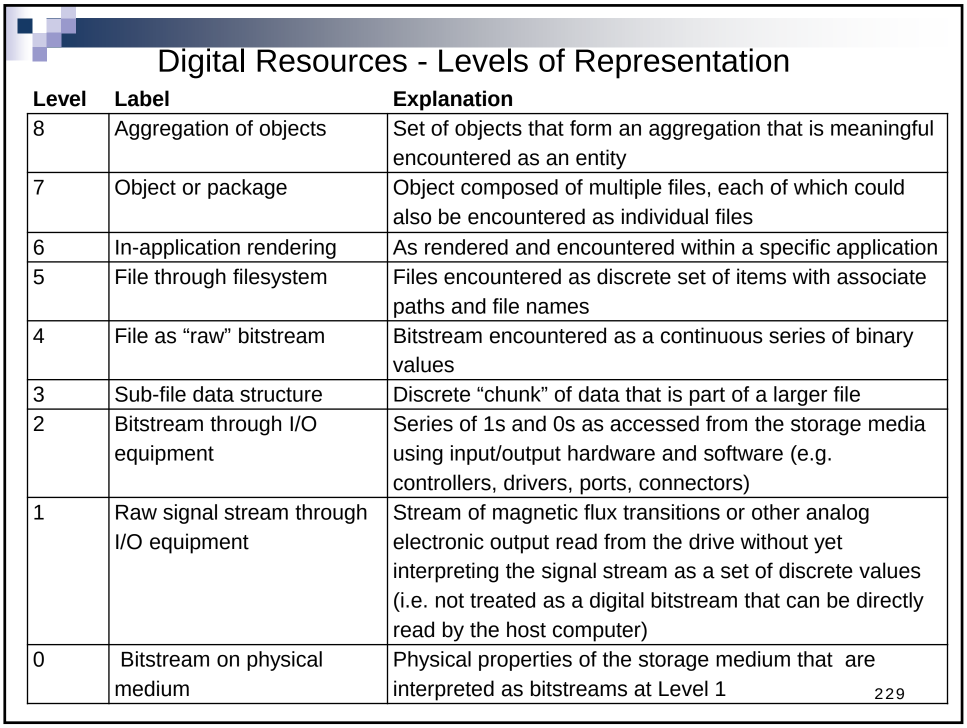
o-favorites-usb-2009-12-11.E01/Partition 1 [999MB]/JO FAVORITE [FAT32]/[root]/Videos/Cat.mov

Forensic analysis meets the archives



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

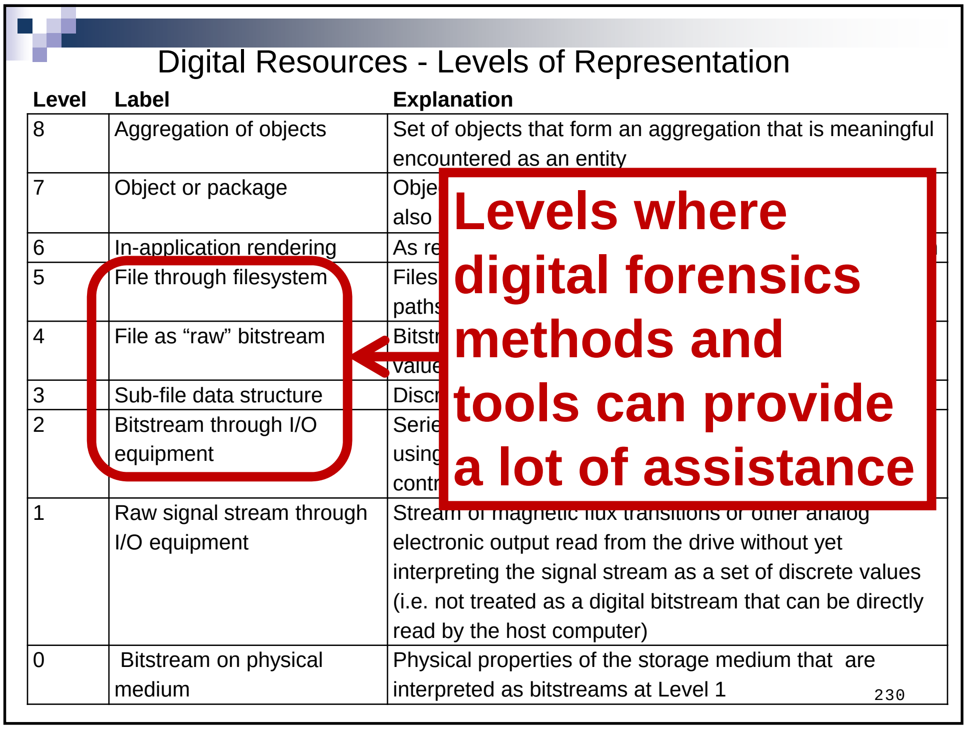
- "Such collection represents a considerable efforts on the part of the System Administrator."
- "Keep detailed notes."
- "Minimize changes to the data as you are collecting it."
- "Do collection first and analysis later."
- "Proceed from the volatile to the less volatile."
- Computer evidence should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



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

229



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	Obje also
6	In-application rendering	As re
5	File through filesystem	Files paths
4	File as "raw" bitstream	Bitstr value
3	Sub-file data structure	Discr
2	Bitstream through I/O equipment	Serie using contr
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

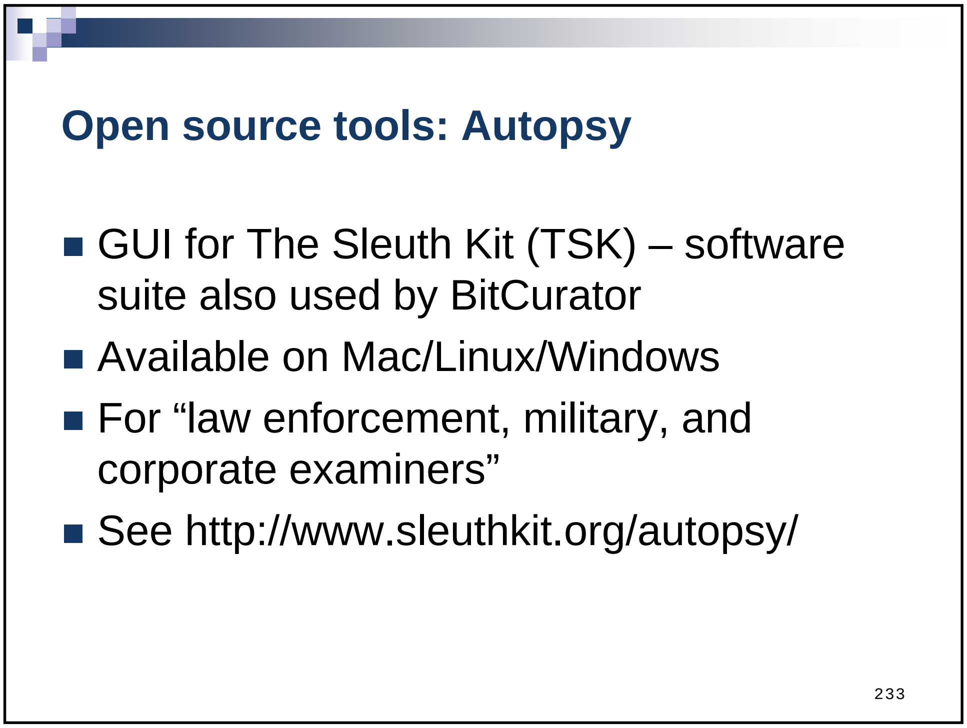
230

Digital Forensics Industry and Tools

- Commercial products:
  - □EnCase (Guidance Software)
  - □FTK (AccessData)
- Open source tools – see:
  - □http://www.sleuthkit.org/autopsy/
  - □http://bitcurator.net
  - □http://www.forensicswiki.org

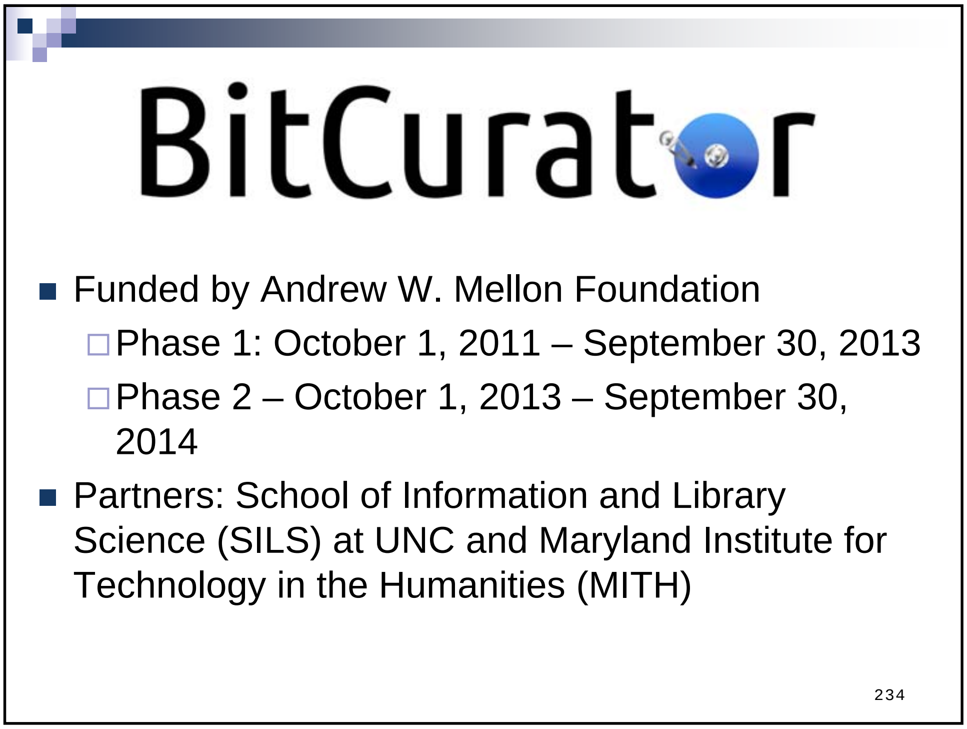
Commercial products

- Pros
  - □GUI for non-specialists
  - □Powerful search and indexing
  - □Bookmarking, annotation, reporting
- Cons
  - □Designed for law enforcement
  - □High cost, low uptake in archival community



Open source tools: Autopsy

- GUI for The Sleuth Kit (TSK) – software suite also used by BitCurator
- Available on Mac/Linux/Windows
- For “law enforcement, military, and corporate examiners”
- See http://www.sleuthkit.org/autopsy/



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

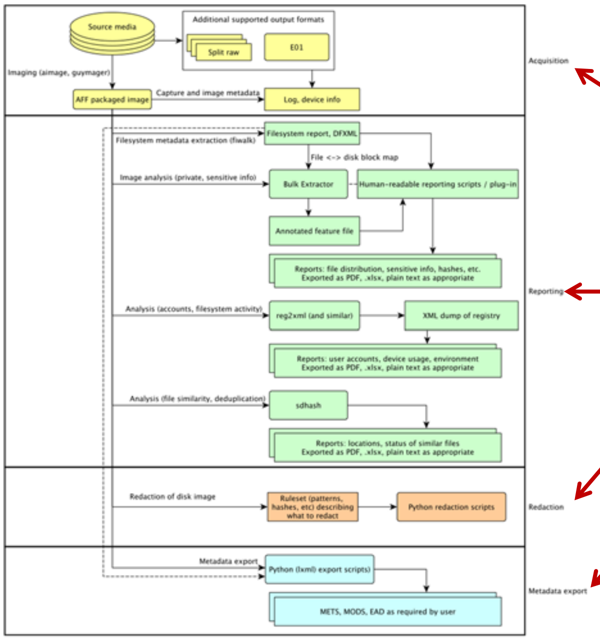
- Develop a system for professionals in libraries, archives and museums that incorporates the functionality of open-source digital forensics tools
- Address two fundamental needs not usually addressed by the digital forensics industry:
  - □ incorporation into the workflow of archives/library ingest 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)
  - □ Coming very soon: installation scripts to turn any Ubuntu Linux machine into a BitCurator Environment
  - □ Self-contained Linux environment in a virtual machine using e.g. Virtual Box or VMWare
  - □ As individual components run directly in your own Linux environment or (whenever possible) Windows environment

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

BitCurator-Supported Workflow


Source media
Additional supported output formats
Split raw
E01
Imaging (aimage, guymager)
AFF packaged image
Capture and image metadata
Log, device info
Acquisition
Filesystem metadata extraction (fiwalk)
Filesystem report, DFXML
File <-> disk block map
Image analysis (private, sensitive info)
Bulk Extractor
Human-readable reporting scripts / plug-in
Annotated feature file
Reports: file distribution, sensitive info, hashes, etc. Exported as PDF, .xlsx, plain text as appropriate
Reporting
Analysis (accounts, filesystem activity)
reg2xml (and similar)
XML dump of registry
Reports: user accounts, device usage, environment Exported as PDF, .xlsx, plain text as appropriate
Analysis (file similarity, deduplication)
sdhash
Reports: locations, status of similar files Exported as PDF, .xlsx, plain text as appropriate
Redaction of disk image
Ruleset (patterns, hashes, etc) describing what to redact
Python redaction scripts
Redaction
Metadata export
Python (lxml) export scripts)
METS, MODS, EAD as required by user
Metadata export

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

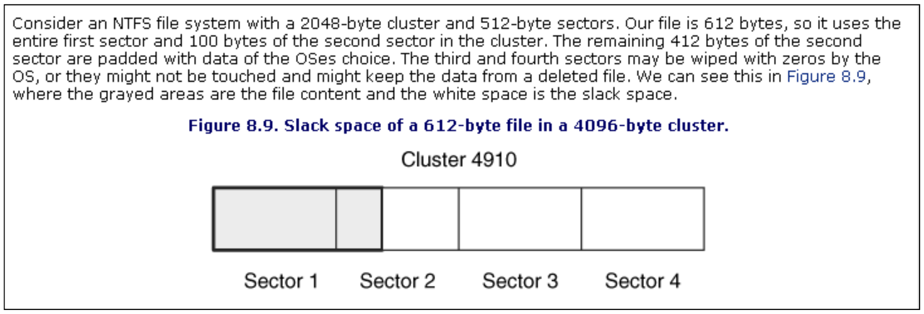
See: http://bitcurator.net

Digging Deeper – Unallocated Space and Data Carving

Unallocated Data Segments in Storage and File Slack

- Recall that computers store data within storage segments of given size (clusters)
- Unallocated storage
  - When file is deleted, pointer to data in each segment is removed, but data will sit in that segment until overwritten by something else (i.e. clusters are freed up for further use, but data within them can persist for a while)
  - Special case is when file is only partially written to a medium (e.g. failed copy due to insufficient space) and then “deleted” but data remains on medium until overwritten
- Overwriting of storage across a computer system is not uniform – “most activity accesses the same data, programs, and other resources over and over again”*

*Farmer, Dan, and Wietse Venema. *Forensic Discovery*. Upper Saddle River, NJ: Addison-Wesley, 2005.



Consider an NTFS file system with a 2048-byte cluster and 512-byte sectors. Our file is 612 bytes, so it uses the entire first sector and 100 bytes of the second sector in the cluster. The remaining 412 bytes of the second sector are padded with data of the OSes choice. The third and fourth sectors may be wiped with zeros by the OS, or they might not be touched and might keep the data from a deleted file. We can see this in Figure 8.9, where the grayed areas are the file content and the white space is the slack space.

Figure 8.9. Slack space of a 612-byte file in a 4096-byte cluster.

Source: Carrier, Brian. *File System Forensic Analysis*. Boston, MA: Addison-Wesley, 2005.

Data in “Slack” Space

- Storage space from the end of the file to the end of the last cluster used to store the file
  - □ **RAM slack** (in last sector of file) - If there is not enough data in file to fill the last sector, OS fills the space with whatever data is in memory at the moment (can be anything since computer was last booted) – no longer happens in recent versions of Windows (overwrites with zeros)
  - □ **Drive slack** (in last cluster of file) - If data from a file does not completely fill the cluster, then data remaining from previous files will remain at the end of the cluster

Data Carving

- Extracting data from raw data blocks, rather than examining the contents of files through the file system.
- Can be done to reconstruct files, identify deleted information, and find data that was purposely hidden in various ways

Carving Taxonomy

Carving: Extract data (files) from raw data (blocks).

Block-Based Carving: Analyze input on block-by-block basis to determine if a block is part of a possible output file.

Statistical Carving: Algorithm that analyzes the input on characteristic or statistic for example, entropy) to determine if the input is part of a possible output file.

Header/Footer Carving: A method for carving files out of raw data using a distinct header (start of file marker) and footer (end of file marker).

Header/Maximum (file) size Carving: A method for carving files out of raw data using a distinct header (start of file marker) and a maximum (file) size.

Header/Embedded Length Carving: A method for carving files out of raw data using a distinct header and a file length (size) which is embedded in the file format

File structure based Carving: A method for carving files out of raw data using a certain level of knowledge of the internal structure of file types.

Semantic Carving: A method for carving files based on a linguistic analysis of the file's content.

Fragment Recovery Carving (split carving): A carving method in which two or more fragments are reassembled to form the original file or object.

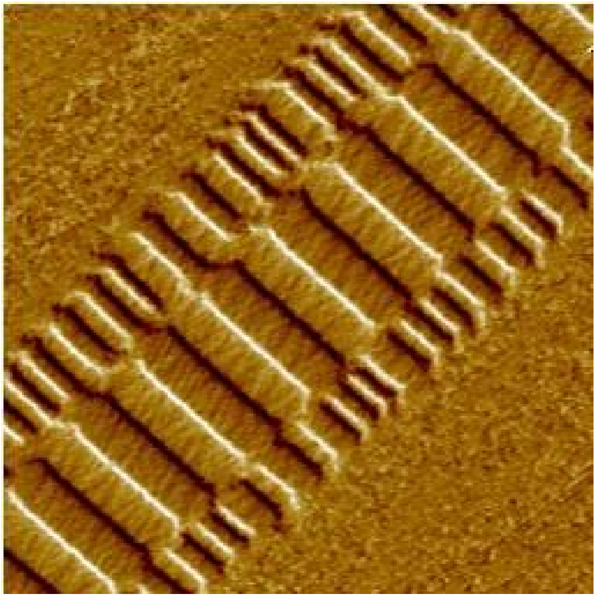
Repackaging Carving: A carving method that modifies the extracted data by adding new headers, footers, or other information so that it can be viewed with standard utilities.

Source: http://www.forensicswiki.org/wiki/File_Carving

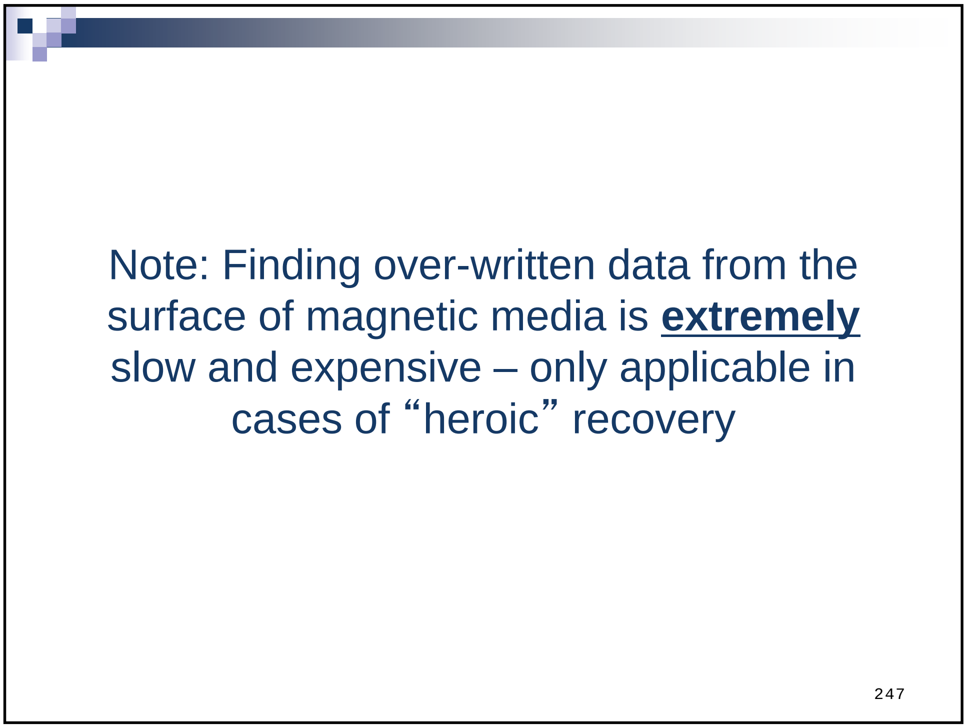
Digging as Deep as Possible – Physical Medium

“Because of the way magnetic media are written it is very difficult to lose everything. With sufficient resources much material that most of us would expect to be lost can be recovered.”*

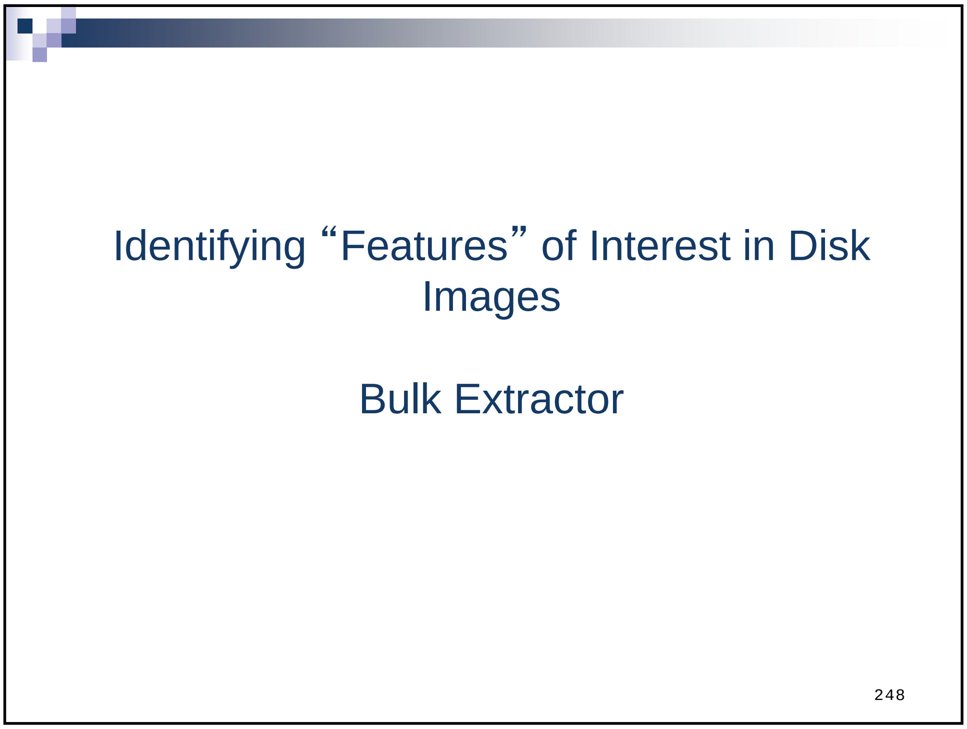
*Ross, Seamus, and Ann Gow. "Digital Archaeology: Rescuing Neglected and Damaged Data Resources.“ London: British Library, 1999.



Veeco Instruments. http://www.veeco.com/library/nanotheater_detail.php?type=application&id=78&app_id=34



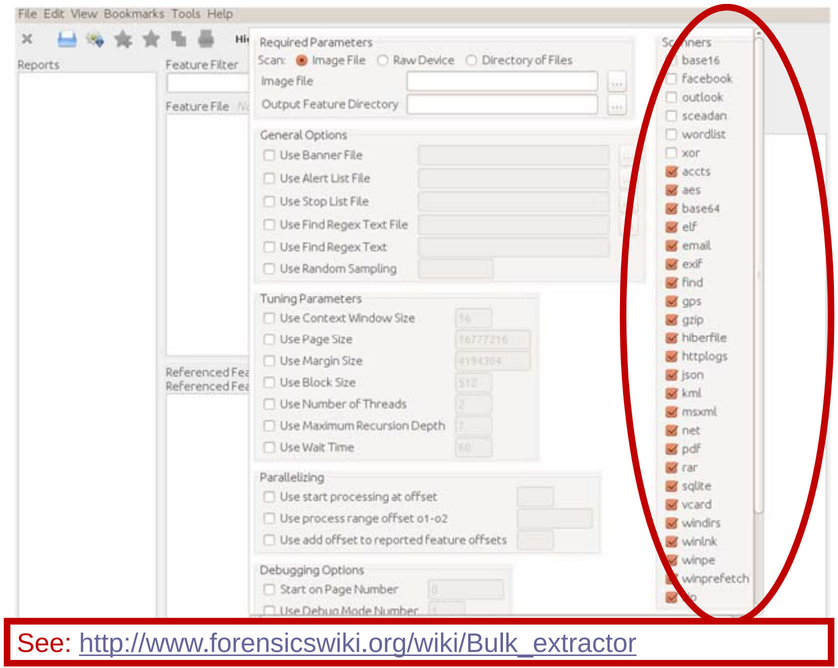
Note: Finding over-written data from the surface of magnetic media is **extremely** slow and expensive – only applicable in cases of “heroic” recovery



Identifying “Features” of Interest in Disk Images

Bulk Extractor

Identifying Potentially Sensitive Data using Bulk Extractor - Scanning Options



Required Parameters
Scan: ● Image File ○ Raw Device ○ Directory of Files
Image File
Output Feature Directory

General Options
- □ Use Banner File
- □ Use Alert List File
- □ Use Stop List File
- □ Use Find Regex Text File
- □ Use Find Regex Text
- □ Use Random Sampling

Tuning Parameters
- □ Use Context Window Size 16
- □ Use Page Size 16777216
- □ Use Margin Size 4194304
- □ Use Block Size 512
- □ Use Number of Threads 2
- □ Use Maximum Recursion Depth 7
- □ Use Wait Time 60

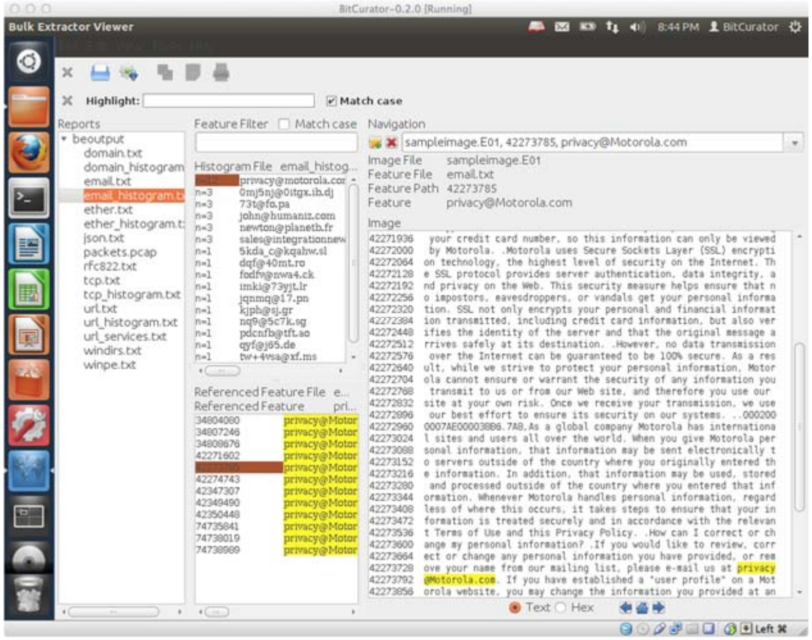
Parallelizing
- □ Use start processing at offset
- □ Use process range offset o1-o2
- □ Use add offset to reported feature offsets

Debugging Options
- □ Start on Page Number 0
- □ Use Debug Mode Number 1

Scanners
- □ base16
- □ facebook
- □ outlook
- □ sceadan
- □ wordlist
- □ xor
- ☑ accts
- ☑ aes
- ☑ base64
- ☑ elf
- ☑ email
- ☑ exif
- ☑ find
- ☑ gps
- ☑ gzip
- ☑ hiberfile
- ☑ httplogs
- ☑ json
- ☑ kml
- ☑ msxml
- ☑ net
- ☑ pdf
- ☑ rar
- ☑ sqlite
- ☑ vcard
- ☑ windirs
- ☑ winlnk
- ☑ winpe
- ☑ winprefetch
- ☑ zip

See: http://www.forensicswiki.org/wiki/Bulk_extractor

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



Reports
- beoutput
  - domain.txt
  - domain_histogram
  - email.txt
  - email_histogram.t
  - ether.txt
  - ether_histogram.t
  - json.txt
  - packets.pcap
  - rfc822.txt
  - tcp.txt
  - tcp_histogram.txt
  - url.txt
  - url_histogram.txt
  - url_services.txt
  - windirs.txt
  - winpe.txt

Histogram File email_histog...

	privacy@motorola.cor
n=3	0mj5nj@0itgx.ib.dj
n=3	73t@fo.pa
n=3	john@humaniz.com
n=3	newton@planetb.fr
n=3	sales@integrationnew
n=1	5kda_c@kqahw.sl
n=1	dqf@40mt.ro
n=1	fodfv@nwa4.ck
n=1	imki@73yjt.lr
n=1	jqnmq@17.pn
n=1	kjph@sj.gr
n=1	nq9@5c7k.sg
n=1	pdcnfb@tft.ao
n=1	qyf@j65.de
n=1	tw+4vsa@xf.ms

Referenced Feature File e...

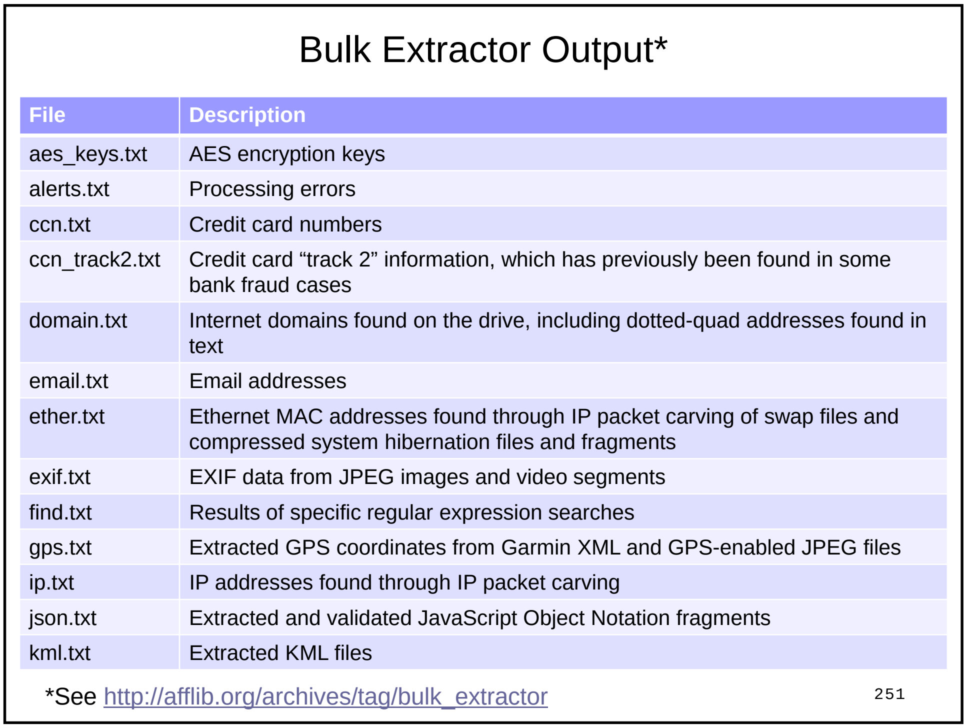
Referenced Feature	pri...
34804080	privacy@Motor
34807246	privacy@Motor
34808676	privacy@Motor
42271602	privacy@Motor
	privacy@Motor
42274743	privacy@Motor
42347307	privacy@Motor
42349490	privacy@Motor
42350448	privacy@Motor
74735841	privacy@Motor
74738019	privacy@Motor
74738989	privacy@Motor

Navigation: sampleimage.E01, 42273785, privacy@Motorola.com

Image File sampleimage.E01
Feature File email.txt
Feature Path 42273785
Feature privacy@Motorola.com

Image

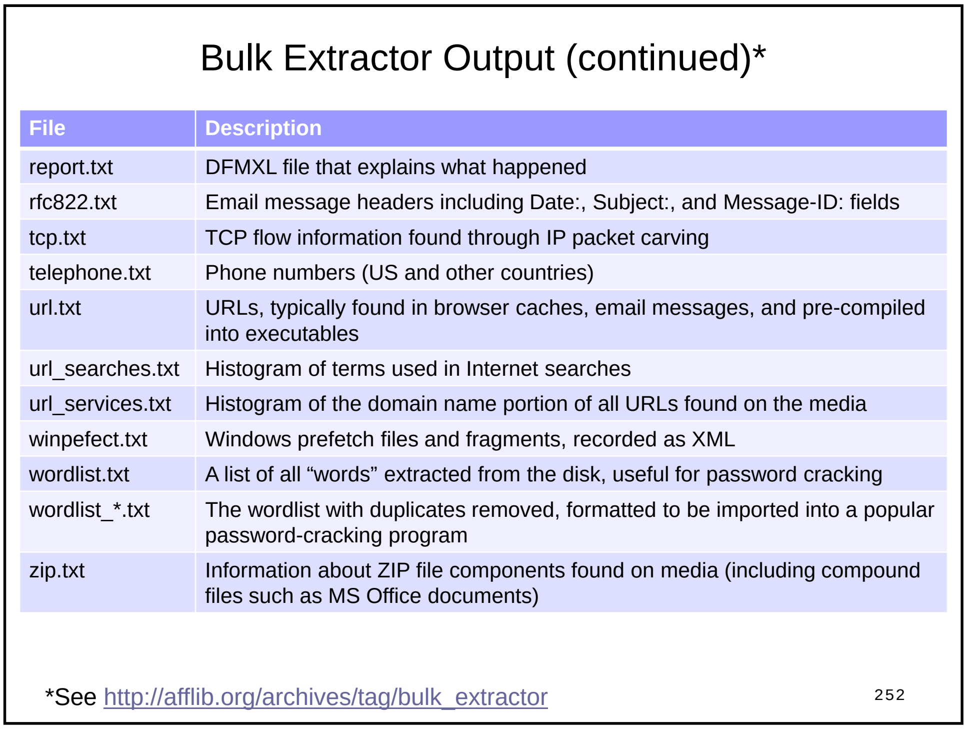
```
42271936  your credit card number, so this information can only be viewed
42272000  by Motorola. .Motorola uses Secure Sockets Layer (SSL) encrypti
42272064 on technology, the highest level of security on the Internet. Th
42272128 e SSL protocol provides server authentication, data integrity, a
42272192 nd privacy on the Web. This security measure helps ensure that n
42272256 o impostors, eavesdroppers, or vandals get your personal informa
42272320 tion. SSL not only encrypts your personal and financial informat
42272384 ion transmitted, including credit card information, but also ver
42272448 ifies the identity of the server and that the original message a
42272512 rrives safely at its destination. .However, no data transmission
42272576  over the Internet can be guaranteed to be 100% secure. As a res
42272640 ult, while we strive to protect your personal information, Motor
42272704 ola cannot ensure or warrant the security of any information you
42272768  transmit to us or from our Web site, and therefore you use our
42272832 site at your own risk. Once we receive your transmission, we use
42272896  our best effort to ensure its security on our systems. ..000200
42272960 0007AE00003BB6.7A8.As a global company Motorola has internationa
42273024 l sites and users all over the world. When you give Motorola per
42273088 sonal information, that information may be sent electronically t
42273152 o servers outside of the country where you originally entered th
42273216 e information. In addition, that information may be used, stored
42273280  and processed outside of the country where you entered that inf
42273344 ormation. Whenever Motorola handles personal information, regard
42273408 less of where this occurs, it takes steps to ensure that your in
42273472 formation is treated securely and in accordance with the relevan
42273536 t Terms of Use and this Privacy Policy. .How can I correct or ch
42273600 ange my personal information? .If you would like to review, corr
42273664 ect or change any personal information you have provided, or rem
42273728 ove your name from our mailing list, please e-mail us at privacy
42273792 @Motorola.com. If you have established a "user profile" on a Mot
42273856 orola website, you may change the information you provided at an
```



Bulk Extractor Output*

File	Description
aes_keys.txt	AES encryption keys
alerts.txt	Processing errors
ccn.txt	Credit card numbers
ccn_track2.txt	Credit card “track 2” information, which has previously been found in some bank fraud cases
domain.txt	Internet domains found on the drive, including dotted-quad addresses found in text
email.txt	Email addresses
ether.txt	Ethernet MAC addresses found through IP packet carving of swap files and compressed system hibernation files and fragments
exif.txt	EXIF data from JPEG images and video segments
find.txt	Results of specific regular expression searches
gps.txt	Extracted GPS coordinates from Garmin XML and GPS-enabled JPEG files
ip.txt	IP addresses found through IP packet carving
json.txt	Extracted and validated JavaScript Object Notation fragments
kml.txt	Extracted KML files

*See http://afflib.org/archives/tag/bulk_extractor

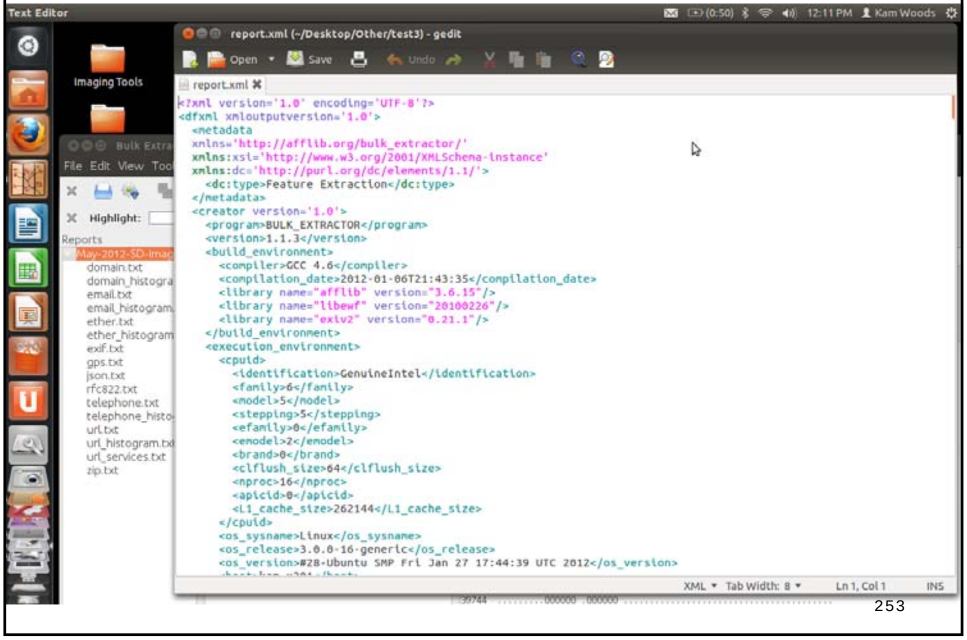


Bulk Extractor Output (continued)*

File	Description
report.txt	DFMXL file that explains what happened
rfc822.txt	Email message headers including Date:, Subject:, and Message-ID: fields
tcp.txt	TCP flow information found through IP packet carving
telephone.txt	Phone numbers (US and other countries)
url.txt	URLs, typically found in browser caches, email messages, and pre-compiled into executables
url_searches.txt	Histogram of terms used in Internet searches
url_services.txt	Histogram of the domain name portion of all URLs found on the media
winpefect.txt	Windows prefetch files and fragments, recorded as XML
wordlist.txt	A list of all “words” extracted from the disk, useful for password cracking
wordlist_*.txt	The wordlist with duplicates removed, formatted to be imported into a popular password-cracking program
zip.txt	Information about ZIP file components found on media (including compound files such as MS Office documents)

*See http://afflib.org/archives/tag/bulk_extractor

Technical Metadata (about the System Used to do the Capture) in a Bulk Extractor Report



```
<?xml version='1.0' encoding='UTF-8'?>
<dfxml xmloutputversion='1.0'>
  <metadata
  xmlns='http://afflib.org/bulk_extractor/'
  xmlns:xsi='http://www.w3.org/2001/XMLSchema-instance'
  xmlns:dc='http://purl.org/dc/elements/1.1/'>
    <dc:type>Feature Extraction</dc:type>
  </metadata>
  <creator version='1.0'>
    <program>BULK_EXTRACTOR</program>
    <version>1.1.3</version>
    <build_environment>
      <compiler>GCC 4.6</compiler>
      <compilation_date>2012-01-06T21:43:35</compilation_date>
      <library name="afflib" version="3.6.15"/>
      <library name="libewf" version="20100226"/>
      <library name="exiv2" version="0.21.1"/>
    </build_environment>
    <execution_environment>
      <cpuid>
        <identification>GenuineIntel</identification>
        <family>6</family>
        <model>5</model>
        <stepping>5</stepping>
        <efamily>0</efamily>
        <emodel>2</emodel>
        <brand>0</brand>
        <clflush_size>64</clflush_size>
        <nproc>16</nproc>
        <apicid>0</apicid>
        <L1_cache_size>262144</L1_cache_size>
      </cpuid>
      <os_sysname>Linux</os_sysname>
      <os_release>3.0.0-16-generic</os_release>
      <os_version>#28-Ubuntu SMP Fri Jan 27 17:44:39 UTC 2012</os_version>
```

253

A Real World Example of Forensic Feature Extraction:

Jeb Bush dumps emails including social security numbers of Florida residents online

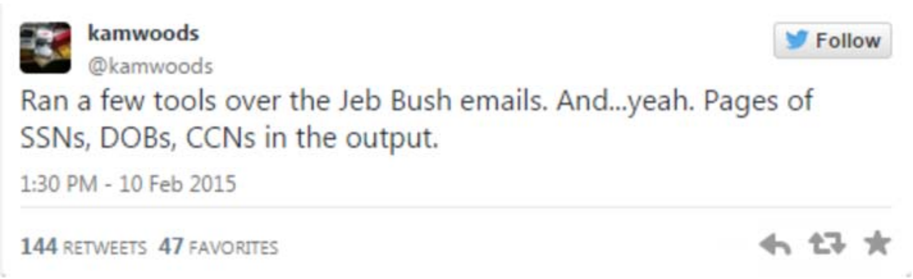
Florida man strikes again

By T.C. Sottek on February 10, 2015 01:37 pm Email @chillmage

https://www.theverge.com/2015/2/10/8013531/jeb-bush-florida-email-dump-privacy

254

A Real World Example of Forensic Feature Extraction:



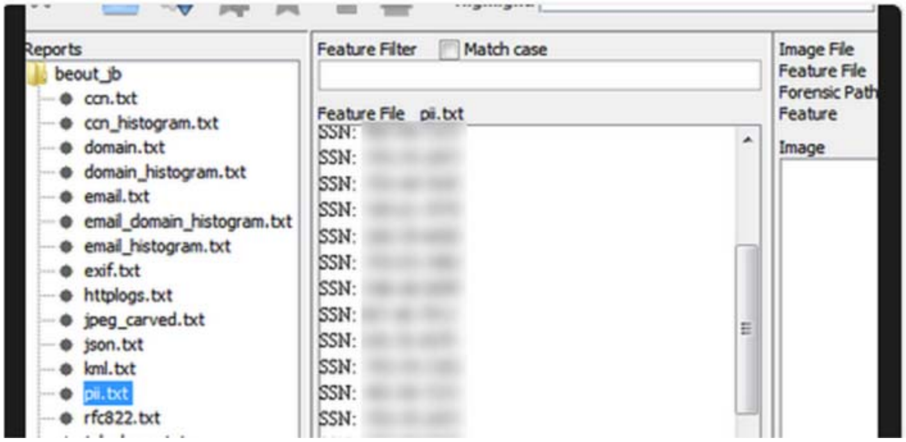
kamwoods
@kamwoods

Ran a few tools over the Jeb Bush emails. And...yeah. Pages of SSNs, DOBs, CCNs in the output.

1:30 PM - 10 Feb 2015

144 RETWEETS 47 FAVORITES

A Real World Example of Forensic Feature Extraction:



Reports
- beout_jb
  - ccn.txt
  - ccn_histogram.txt
  - domain.txt
  - domain_histogram.txt
  - email.txt
  - email_domain_histogram.txt
  - email_histogram.txt
  - exif.txt
  - httplogs.txt
  - jpeg_carved.txt
  - json.txt
  - kml.txt
  - pii.txt
  - rfc822.txt

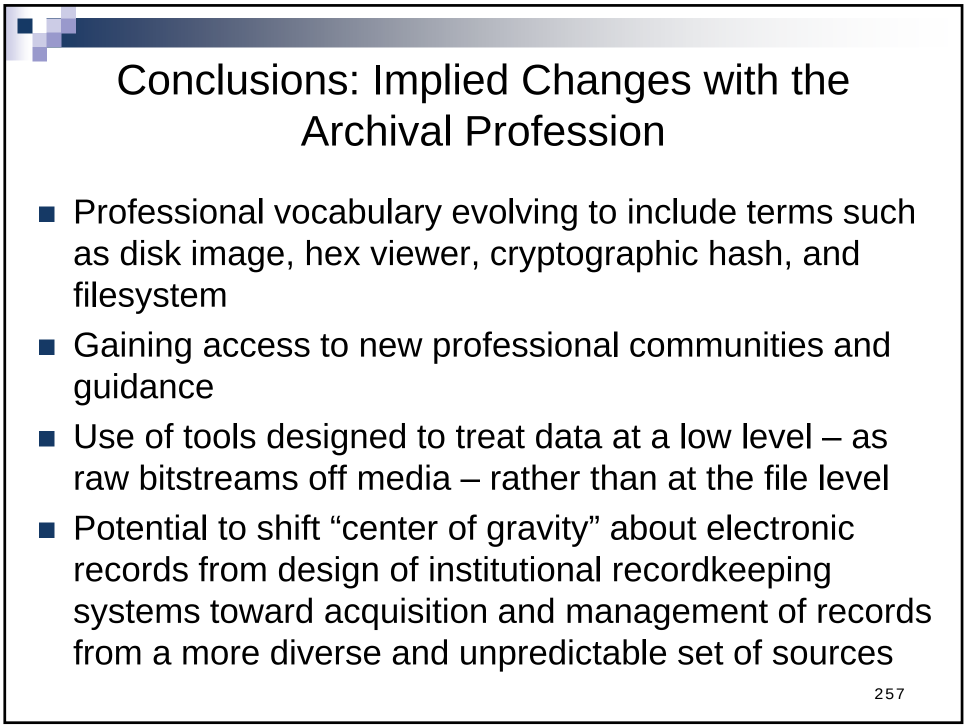
Feature Filter □ Match case

Feature File pii.txt

```
SSN: [redacted]
SSN: [redacted]
SSN: [redacted]
SSN: [redacted]
SSN: [redacted]
SSN: [redacted]
SSN: [redacted]
SSN: [redacted]
SSN: [redacted]
SSN: [redacted]
SSN: [redacted]
SSN: [redacted]
```

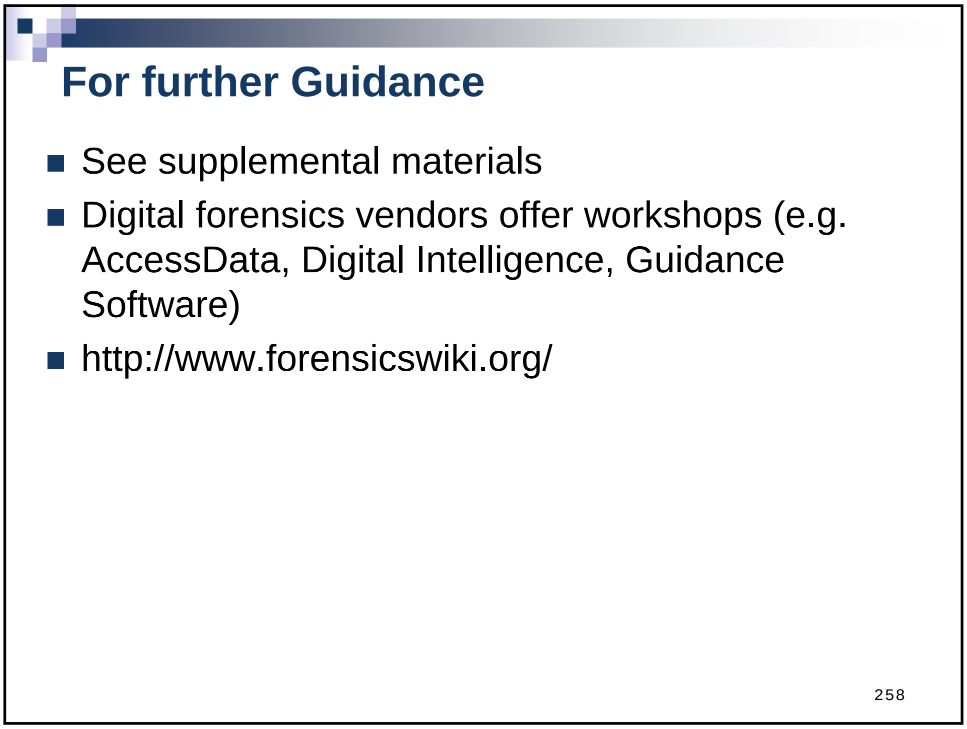
Image File
Feature File
Forensic Path
Feature

Image



Conclusions: Implied Changes with the Archival Profession

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



For further Guidance

- See supplemental materials
- Digital forensics vendors offer workshops (e.g. AccessData, Digital Intelligence, Guidance Software)
- http://www.forensicswiki.org/

BitCurator Resources

Welcome to the BitCurator Wiki

Downloads

Documentation

News

BitCurator This Week (February 24th)

BitCurator 0.7.4 Now Available

Call for Proposals: BitCurator On-Site Visit and Training

Video Spotlight

Screencast Tutorial: Using Bulk Extractor to find potentially sensitive information

Events

Alex Chassanoff speaking at IDCC 2014

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

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

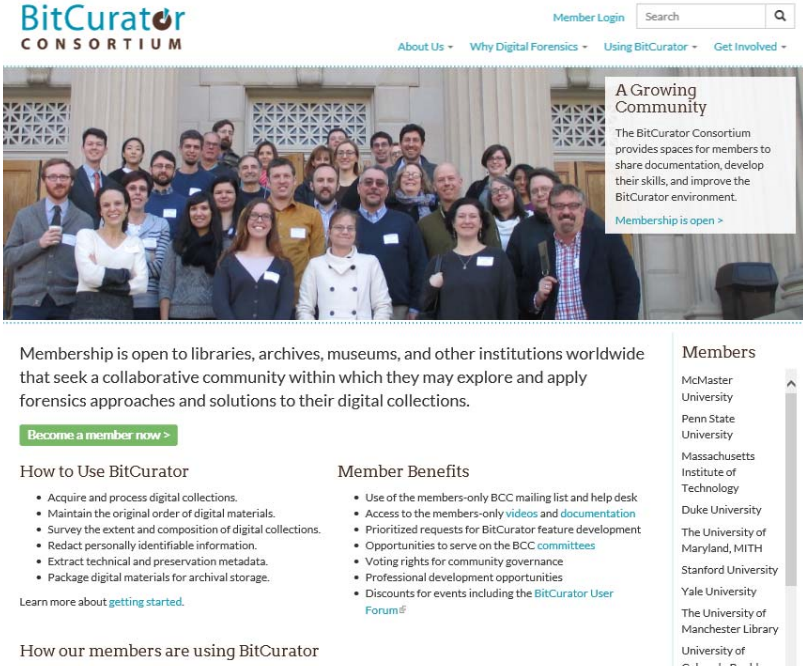
Twitter: @bitcurator

259

BitCurator Access

- Web-based access to raw and forensically packaged disk images
- Transforming and using digital forensics metadata in collecting environments
- Redaction of file items, metadata and hidden data from disk images
- OS and executable virtualization for legacy disk images

260



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

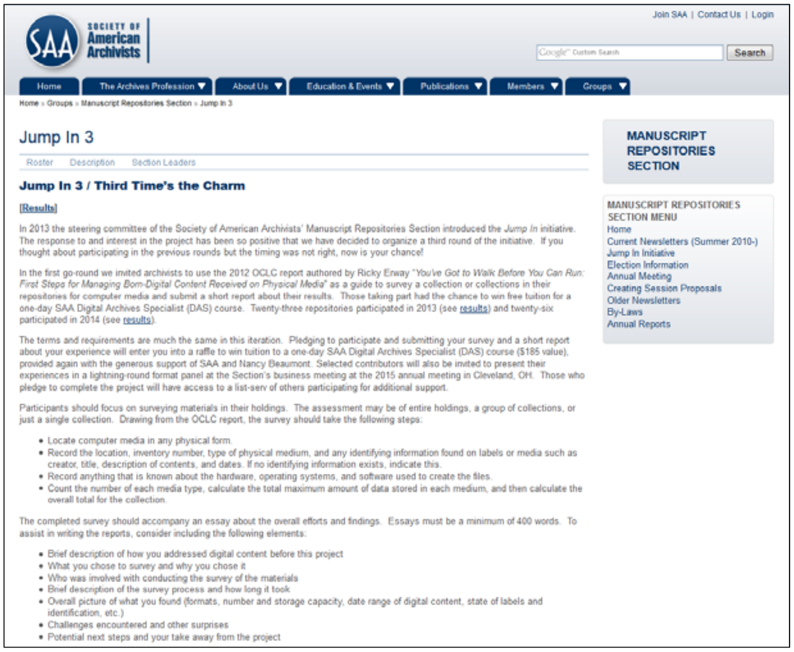
Stanford University

Yale University

The University of Manchester Library

University of

http://bitcuratorconsortium.org



SAA SOCIETY OF American Archivists
Join SAA | Contact Us | Login

The Archives Profession ▼

Home › Groups › Manuscript Repositories Section › Jump In 3

Jump In 3

Jump In 3 / Third Time's the Charm

[Results]

In 2013 the steering committee of the Society of American Archivists' Manuscript Repositories Section introduced the *Jump In* initiative. The response to and interest in the project has been so positive that we have decided to organize a third round of the initiative. If you thought about participating in the previous rounds but the timing was not right, now is your chance!

In the first go-round we invited archivists to use the 2012 OCLC report authored by Ricky Erway "*You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media*" as a guide to survey a collection or collections in their repositories for computer media and submit a short report about their results. Those taking part had the chance to win free tuition for a one-day SAA Digital Archives Specialist (DAS) course. Twenty-three repositories participated in 2013 (see results) and twenty-six participated in 2014 (see results).

The terms and requirements are much the same in this iteration. Pledging to participate and submitting your survey and a short report about your experience will enter you into a raffle to win tuition to a one-day SAA Digital Archives Specialist (DAS) course ($185 value), provided again with the generous support of SAA and Nancy Beaumont. Selected contributors will also be invited to present their experiences in a lightning-round format panel at the Section's business meeting at the 2015 annual meeting in Cleveland, OH. Those who pledge to complete the project will have access to a list-serv of others participating for additional support.

Participants should focus on surveying materials in their holdings. The assessment may be of entire holdings, a group of collections, or just a single collection. Drawing from the OCLC report, the survey should take the following steps:

- Locate computer media in any physical form.
- Record the location, inventory number, type of physical medium, and any identifying information found on labels or media such as creator, title, description of contents, and dates. If no identifying information exists, indicate this.
- Record anything that is known about the hardware, operating systems, and software used to create the files.
- Count the number of each media type, calculate the total maximum amount of data stored in each medium, and then calculate the overall total for the collection.

The completed survey should accompany an essay about the overall efforts and findings. Essays must be a minimum of 400 words. To assist in writing the reports, consider including the following elements:

- Brief description of how you addressed digital content before this project
- What you chose to survey and why you chose it
- Who was involved with conducting the survey of the materials
- Brief description of the survey process and how long it took
- Overall picture of what you found (formats, number and storage capacity, date range of digital content, state of labels and identification, etc.)
- Challenges encountered and other surprises
- Potential next steps and your take away from the project

MANUSCRIPT REPOSITORIES SECTION

MANUSCRIPT REPOSITORIES SECTION MENU

Home
Current Newsletters (Summer 2010-)
Jump In Initiative
Election Information
Annual Meeting
Creating Session Proposals
Older Newsletters
By-Laws
Annual Reports

http://www2.archivists.org/groups/manuscript-repositories-section/jump-in-3

To Learn More about Available Software

Forensics Wiki. http://forensicswiki.org

BitCurator environment. http://wiki.bitcurator.net

BitCurator Access software. http://access.bitcurator.net

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

Information Guides on Tools for Electronic Records. Minnesota State Archives.
http://www.mnhs.org/preserve/records/tools.php

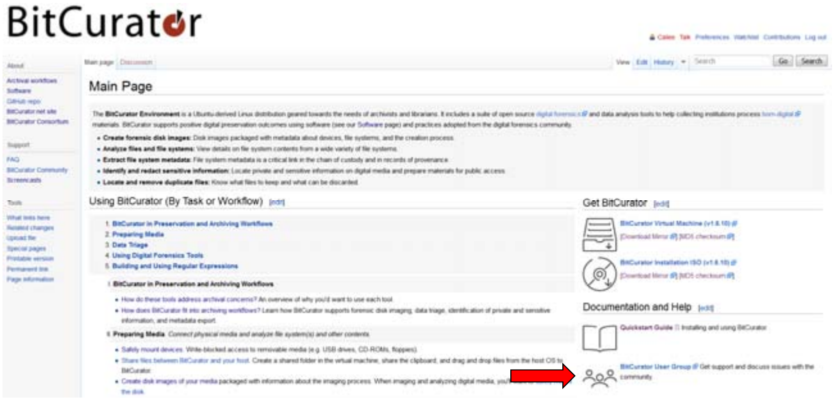
Lifecycle Tools for Archival Email Stewardship.
https://docs.google.com/spreadsheets/d/1V1N22xnr5e0EbDlZWx58bjYO6rkrMrYH9wGX9-CK8c4/

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



Main Page

The **BitCurator Environment** is a Ubuntu-derived Linux distribution geared towards the needs of archivists and librarians. It includes a suite of open source digital forensics and data analysis tools to help collecting institutions process born-digital materials. BitCurator supports positive digital preservation outcomes using software (see our Software page) and practices adopted from the digital forensics community.

- **Create forensic disk images:** Disk images packaged with metadata about devices, file systems, and the creation process.
- **Analyze files and file systems:** View details on file system contents from a wide variety of file systems.
- **Extract file system metadata:** File system metadata is a critical link in the chain of custody and in records of provenance.
- **Identify and redact sensitive information:** Locate private and sensitive information on digital media and prepare materials for public access.
- **Locate and remove duplicate files:** Know what files to keep and what can be discarded.

Using BitCurator (By Task or Workflow) [edit]

1. BitCurator in Preservation and Archiving Workflows
2. Preparing Media
3. Data Triage
4. Using Digital Forensics Tools
5. Building and Using Regular Expressions

I. **BitCurator in Preservation and Archiving Workflows**

- How do these tools address archival concerns? An overview of why you'd want to use each tool.
- How does BitCurator fit into archiving workflows? Learn how BitCurator supports forensic disk imaging, data triage, identification of private and sensitive information, and metadata export.

II. **Preparing Media** Connect physical media and analyze file system(s) and other contents.

- Safely mount devices. Write-blocked access to removable media (e.g. USB drives, CD-ROMs, floppies).
- Share files between BitCurator and your host. Create a shared folder in the virtual machine, share the clipboard, and drag and drop files from the host OS to BitCurator.
- Create disk images of your media packaged with information about the imaging process. When imaging and analyzing digital media, you'll [illegible] the disk.

Get BitCurator [edit]

BitCurator Virtual Machine (v1.8.10)
[Download Mirror] [MD5 checksum]

BitCurator Installation ISO (v1.8.10)
[Download Mirror] [MD5 checksum]

Documentation and Help [edit]

Quickstart Guide – Installing and using BitCurator

BitCurator User Group Get support and discuss issues with the community.

Digital Curation List

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

Further Education

Digital Forensics for Archivists: Advanced (SAA)

http://www2.archivists.org/prof-education/course-catalog/digital-forensics-for-archivists-advanced#.V1SSg-S0OzA

Training Overview

AccessData Training is conducted by Syntricate, [illegible]

COMPUTER FORENSICS

Syntricate will educate you in technology and prepare you with innovative ideas and workflows to improve and strengthen your skills to identify, respond, investigate, prosecute, and adjudicate cases. The Syntricate computer forensic team focuses on how to properly collect, process, review and report case data toward successful case resolution. View Syntricate's Computer Forensic courses and syllabi, course calendar, on-demand training videos, annual training passes custom training and certifications. Or, simply contact Syntricate and we'll set up a call to discuss our offerings and which type of training and delivery method suits your specific needs.

Complete Schedule	Digital Archive Boot Camp	Cellebrite	JTAG
DFF	EnCase 7 / DFFER	FTK Bootcamp / DFFAD	NUIX

DIGITAL INTELLIGENCE COMPUTER FORENSICS COURSE LISTING

Digital Archiving Boot Camp

The Digital Archivist Boot Camp is as 3 day course is designed to provide the attendee with an overview of digital forensic workstations and equipment, as well as how to utilize them for forensic triage and duplication/archiving. The foundation of how to build a case within forensic tools will also be discussed and covered with practical exercises.

The course should be attended by new digital forensic archivists, new forensic or eDiscovery practitioners or first responders tasked with preserving various types of digital media and tasked with basic data recovery, organization of data or password cracking.

Digital Archivist Boot Camp $1795.00 REGISTER / ORDER

DIGITAL INTELLIGENCE
17165 W. Glendale Drive
New Berlin, WI 53151
866-DIGINTEL (866-344-4683)
Outside the US: 262-782-3332

Site Contents Copyright © 2016
www.DigitalIntelligence.com

REQUEST A QUOTE

Courses

Level	Title	Delivery Method
	EnCase v7 OnDemand Forensics and Enterprise Fundamentals Training	[illegible]
	EnCase® v7 OnDemand Forensics and Enterprise Fundamentals Training	[illegible]
	EnCase® OnDemand eDiscovery for the Legal Team	[illegible]
	Digital Media – Acquisition and Triage	[illegible]
	EnCase® v7 Computer Forensics I	[illegible]
Introductory	EnCase® v7 vClass Computer Forensics I	[illegible]

Thank you!

Digital Archives Specialist (DAS) Questions?

http://www2.archivists.org/prof-education/das/FAQs

education@archivists.org