An Introduction to Digital Forensics for Archivists

Porter Olsen
BitCurator Community Lead
email: polsen@umd.edu
twitter: @pwolsen @bitcurator

MITH MARYLAND INSTITUTE FOR TECHNOLOGY IN THE HUMANITIES

UNC SCHOOL OF INFORMATION AND LIBRARY SCIENCE

Overview and Acknowledgment

- The bulk of this lecture is drawn from the SAA DAS course on digital forensics designed & taught by Cal Lee (BitCurator PI) & Kam Woods (BitCurator Technical Lead)

- Overview
  1. Defining digital forensics and its role in digital preservation
  2. Brief introduction to the BitCurator project
  3. Layers of abstraction: multiple ways to interact with digital objects
  4. Disk imaging vs. Logical copy
  5. How data is stored on digital media
  6. An introduction to file systems

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: Simson Garfinkel

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

Same Goals as When Acquiring Analog Materials

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

Same Fundamental Archival Principles Apply

Provenance

- Reflect “life history” of records
- Records from a common origin or source should be managed together as an aggregate unit

Original Order

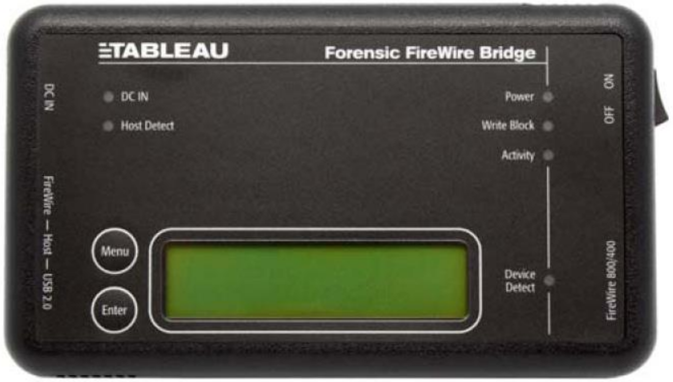
Organize and manage records in ways that reflect their arrangement within the creation/use environment

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:



AFFLIB

Open Source Computer Forensics Software

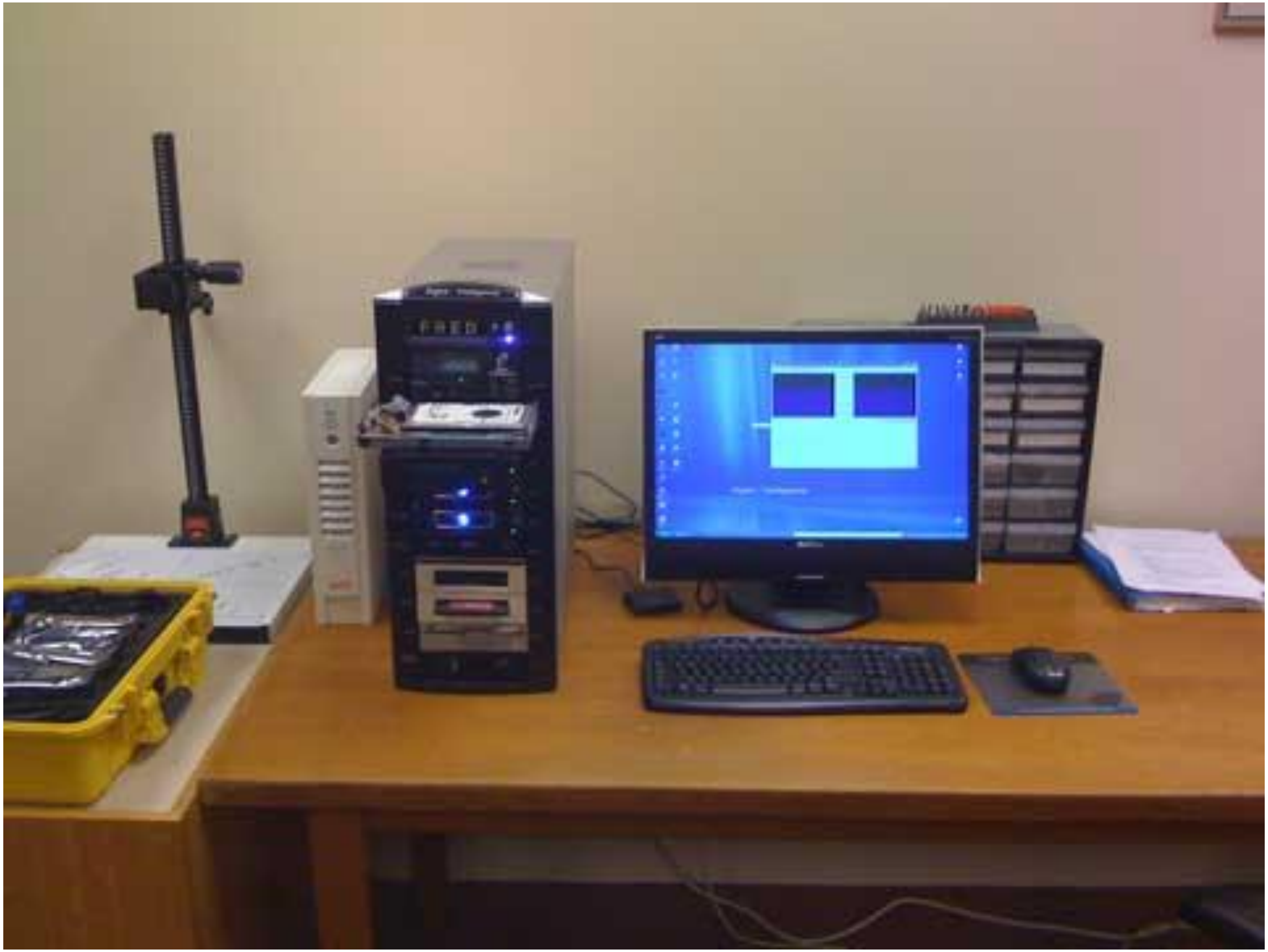
Luckily, there are a lot of people with expertise in using such tools in places like this:

El Paso County Sheriff’s Office (Colorado)

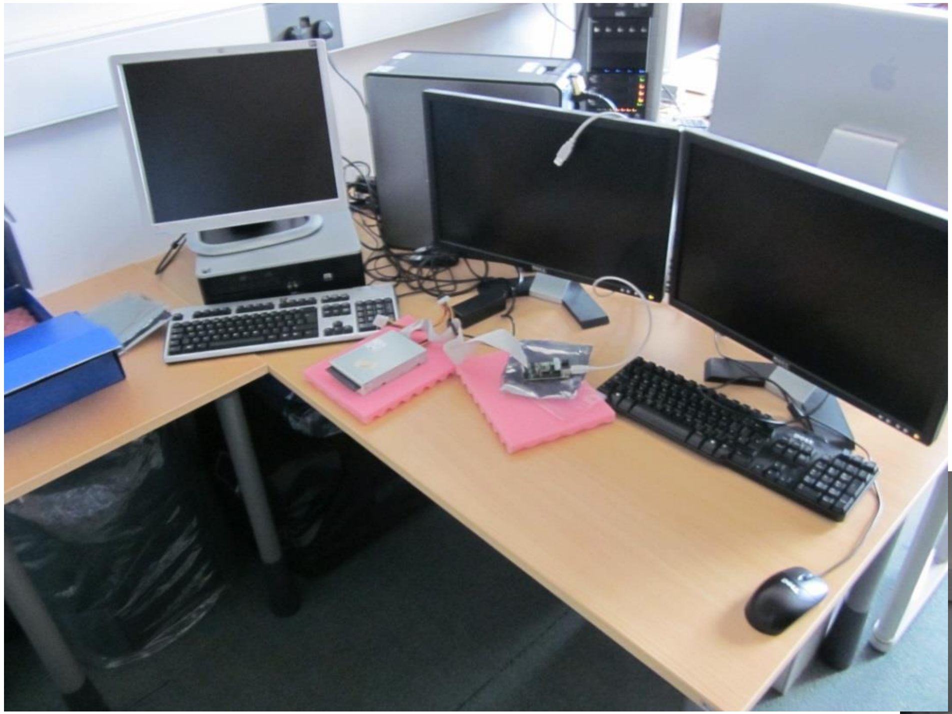
http://shr.elpasoco.com/Law+Enforcement+Bureau/Investigations+Division/Computer+Crime+Lab.htm

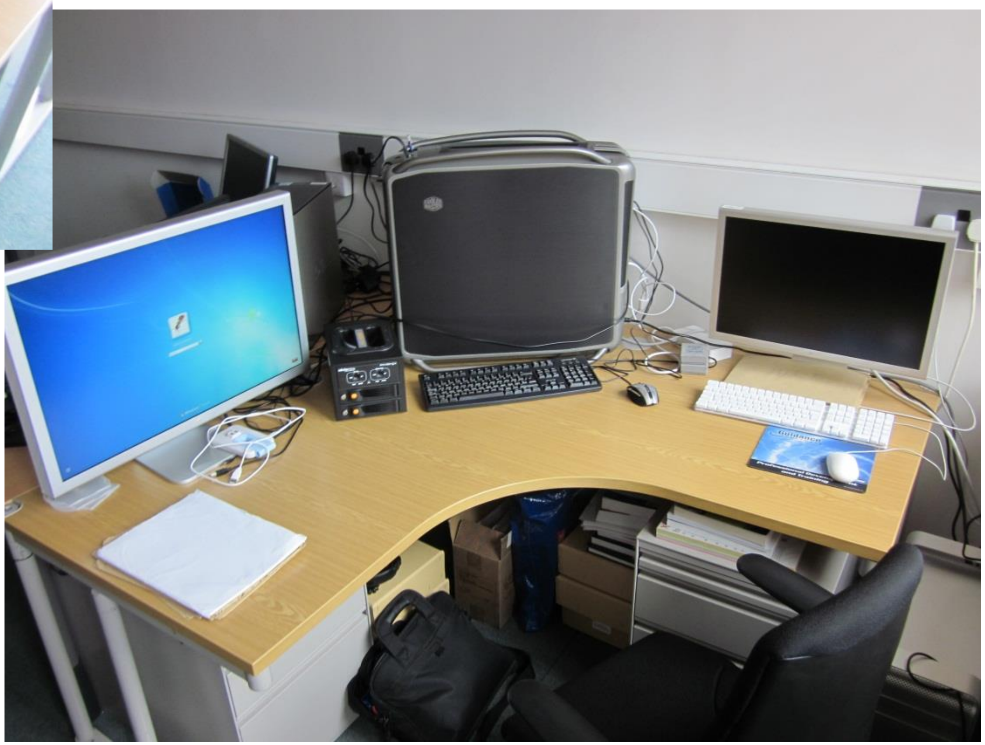
Here’s what it looks like in libraries and archives:

Stanford University Libraries and Academic Information Resources (SULAIR)



British Library, London





UNC School of Information and Library Science

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

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

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)

- Cal Lee, PI
- Matt Kirschenbaum, Co-PI
- Kam Woods, Technical Lead
- Porter Olsen, Community Lead
- Alex Chassanoff, Project Manager
- Sunitha Misra, GA (UNC)
- Amanda Visconti, GA (MITH)

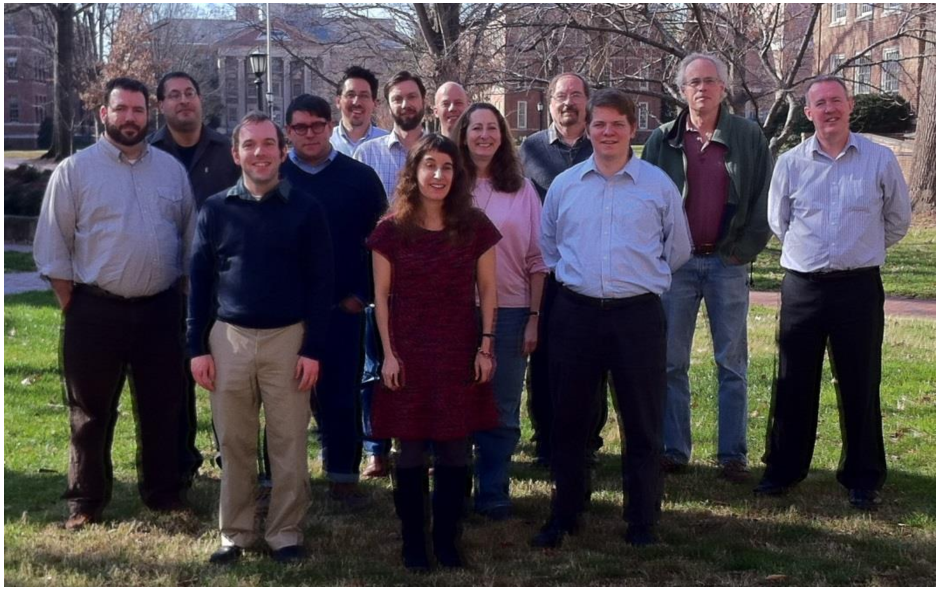
Two Groups of Advisors

Professional Experts Panel

- Bradley Daigle, University of Virginia Library
- Erika Farr, Emory University
- Jennie Levine Knies, University of Maryland
- Jeremy Leighton John, British Library
- Leslie Johnston, Library of Congress
- Naomi Nelson, Duke University
- Erin O’Meara, Gates Archive
- Michael Olson, Stanford University Libraries
- Gabriela Redwine, Harry Ransom Center, University of Texas
- Susan Thomas, Bodleian Library, University of Oxford

Development Advisory Group

- Barbara Guttman, National Institute of Standards and Technology
- Jerome McDonough, University of Illinois
- Mark Matienzo, Yale University
- Courtney Mumma, Artefactual Systems
- David Pearson, National Library of Australia
- Doug Reside, New York Public Library
- Seth Shaw, University Archives, Duke University
- William Underwood, Georgia Tech

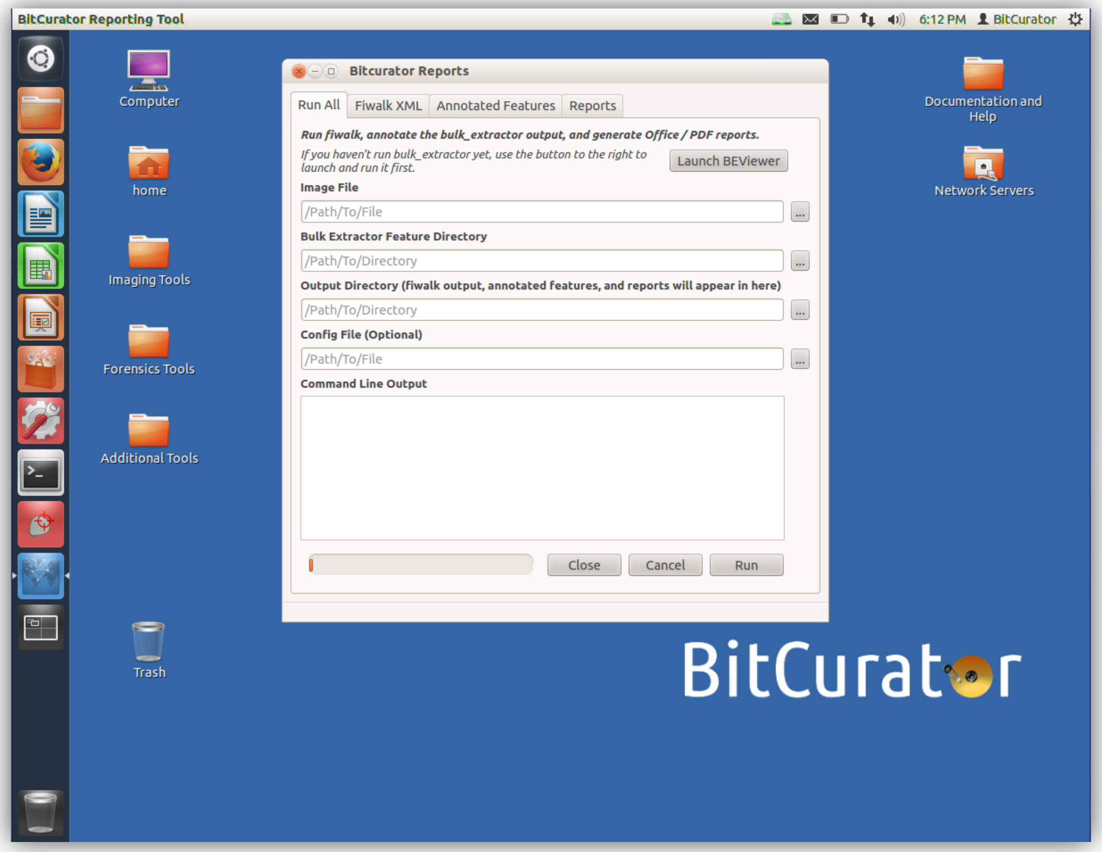


BitCurator Goals

- Develop a system for collecting professionals that incorporates the functionality of open-source digital forensics tools
- Address two fundamental needs not usually addressed by the digital forensics industry:
  - Incorporation into the workflow of archives/library ingest and collection management environments
  - Provision of public access to the data

The BitCurator Environment

- Ubuntu Linux 12.04
- Open source digital forensics tools (Guymager, The Sleuth Kit, bulk_extractor, etc.)
- BitCurator interface and reporting tool
- Digital forensics plug-ins for Nautilus (Ubuntu file browser)



Run fiwalk, annotate the bulk_extractor output, and generate Office / PDF reports.

If you haven't run bulk_extractor yet, use the button to the right to launch and run it first.

Image File

/Path/To/File

Bulk Extractor Feature Directory

/Path/To/Directory

Output Directory (fiwalk output, annotated features, and reports will appear in here)

/Path/To/Directory

Config File (Optional)

/Path/To/File

Command Line Output

Nature of Digital Materials

Layers and Abstraction

"Computer science is largely a matter of **abstraction**: identifying a wide range of applications that include some overlapping functionality, and then working to **abstract out** that shared functionality into a distinct service layer (or module, or language, or whatever). That new service layer then becomes a platform on top of which many other functionalities can be built that had previously been impractical or even unimagined. How does this activity of abstraction work as a practical matter? It's technical work, of course, but it's also **social work**. It is unlikely that any one computer scientist will be an expert in every one of the important applications areas that may benefit from the abstract service. So **collaboration** will be required." (emphasis added)

- Phil Agre, Red Rock Eater, March 25, 2000

Translations Across Layers

- Users view, read, write and click on things
- Programmers usually write & reuse source code

```
#include <iostream>
int main()
{
        std::cout << "Hello, world!\n";
}
```

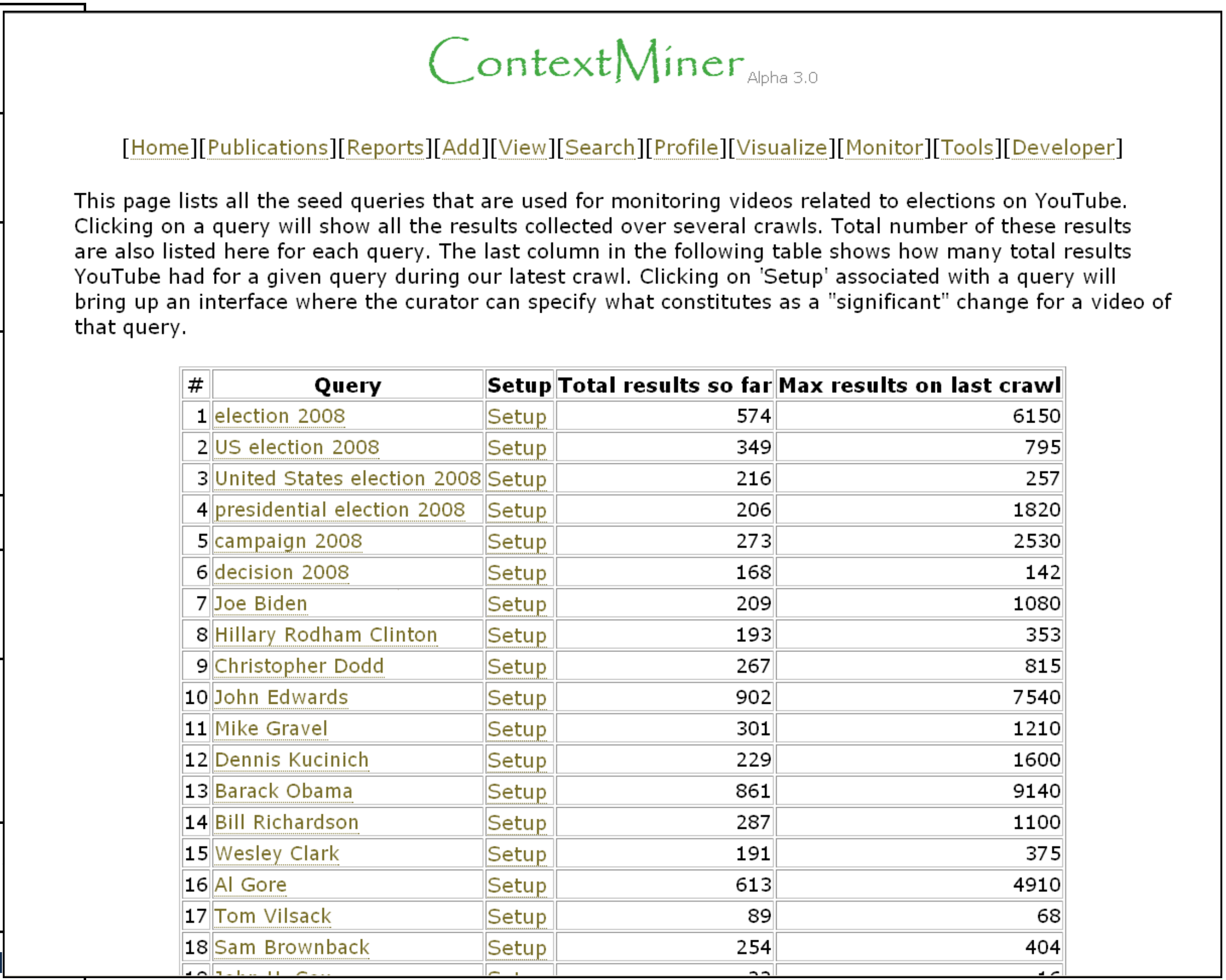
- Software & firmware manipulates data and instructions as bits (10100001110101)
- Physical equipment deals with magnetic charges, holes in optical disks, holes in punch cards

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

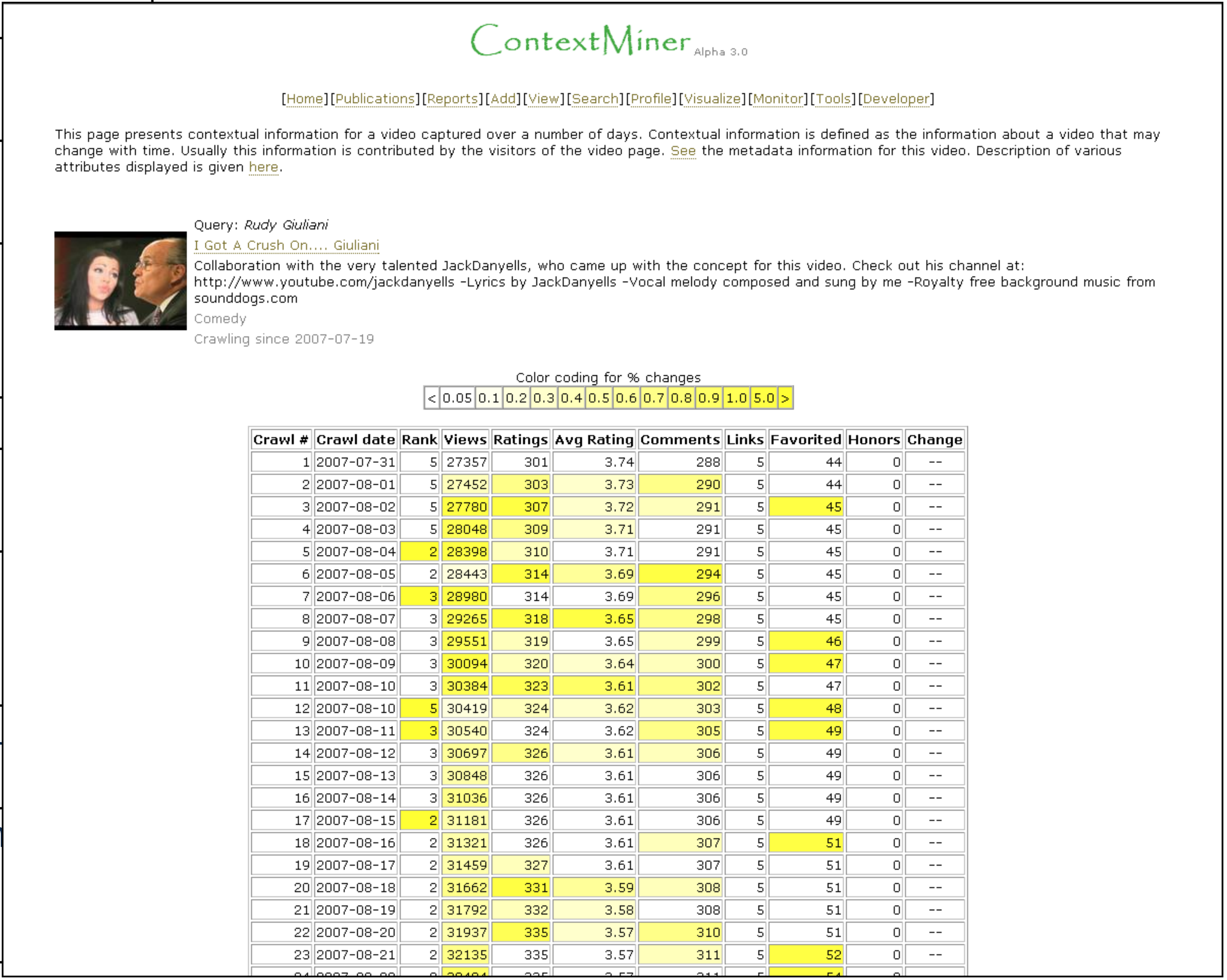
[Home][Publications][Reports][Add][View][Search][Profile][Visualize][Monitor][Tools][Developer]

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

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

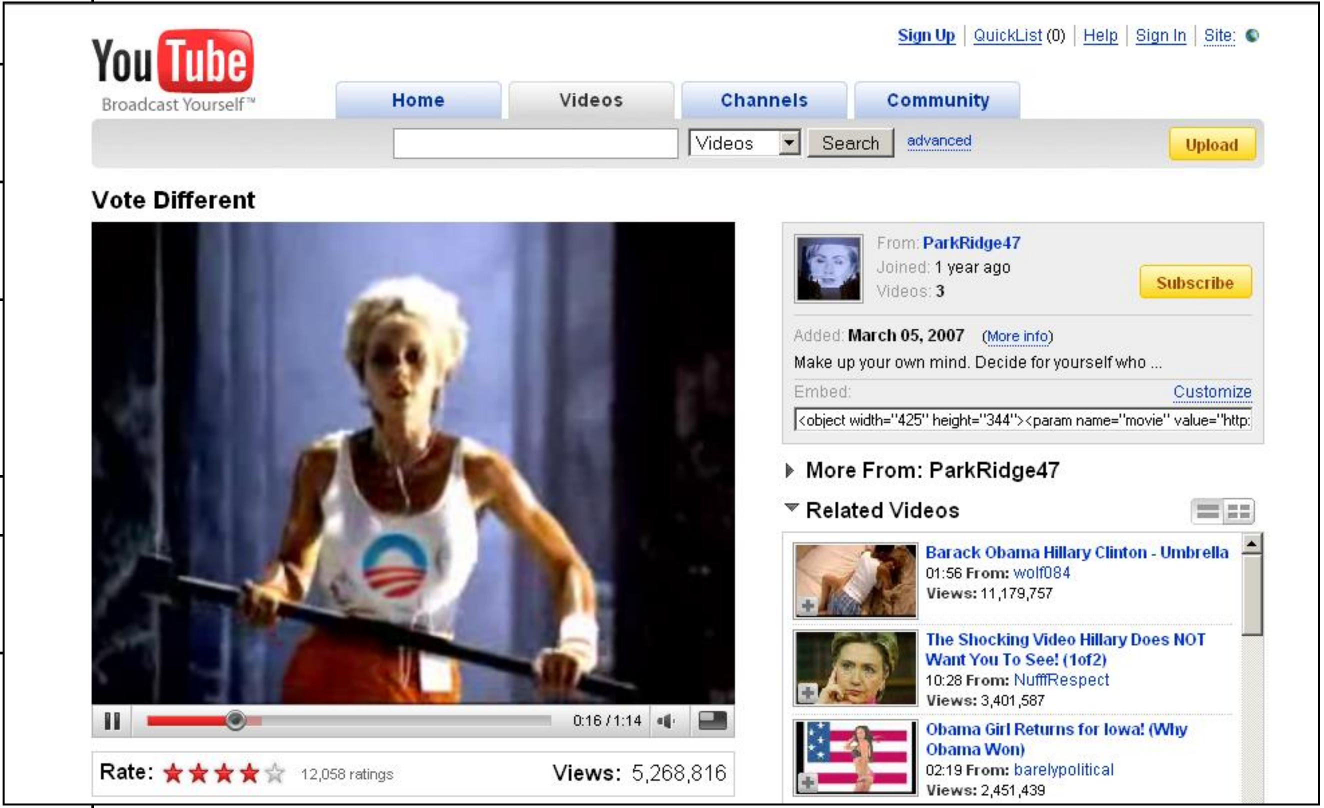
Color coding for % changes

<	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	--

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



Vote Different

From: ParkRidge47
Joined: 1 year ago
Videos: 3

Added: March 05, 2007 (More info)

Make up your own mind. Decide for yourself who ...

Embed: <object width="425" height="344"><param name="movie" value="http:

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

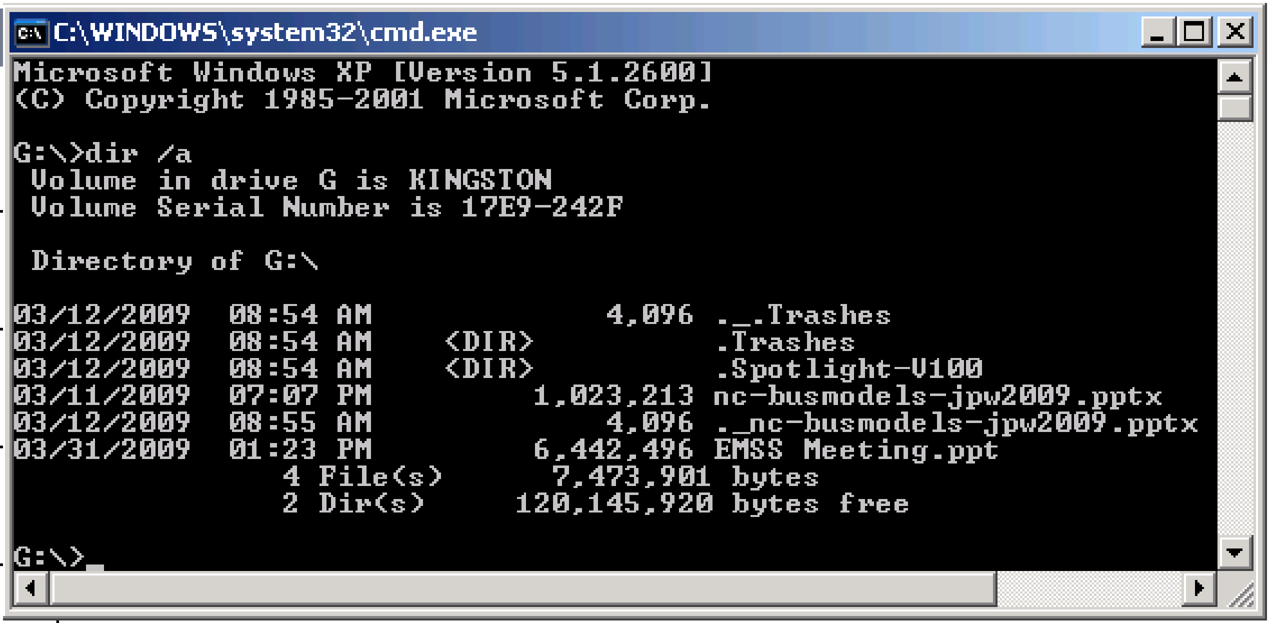
The Shocking Video Hillary Does NOT Want You To See! (1of2)
10:28 From: NuffRespect
Views: 3,401,587

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

Rate: ★★★★☆ 12,058 ratings

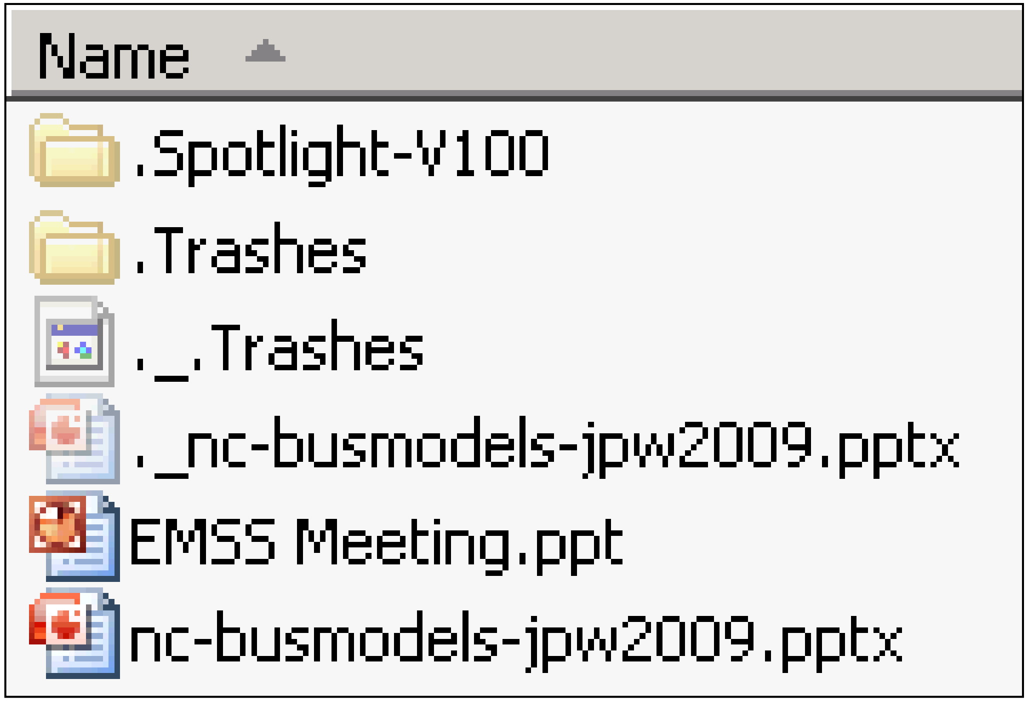
Views: 5,268,816

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



C:\WINDOWS\system32\cmd.exe

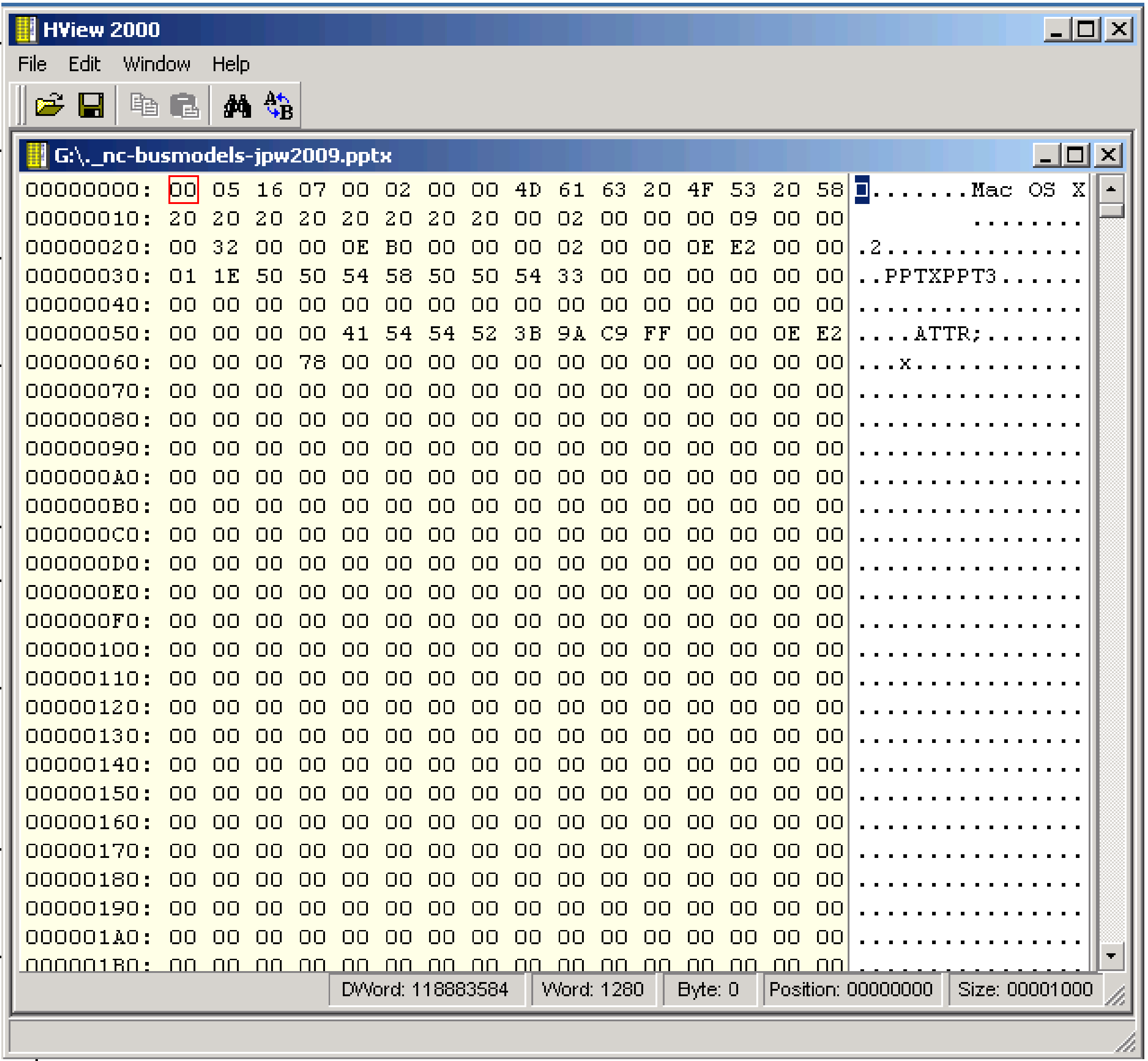
```
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

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

 Directory of G:\

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

G:\>_
```



```
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

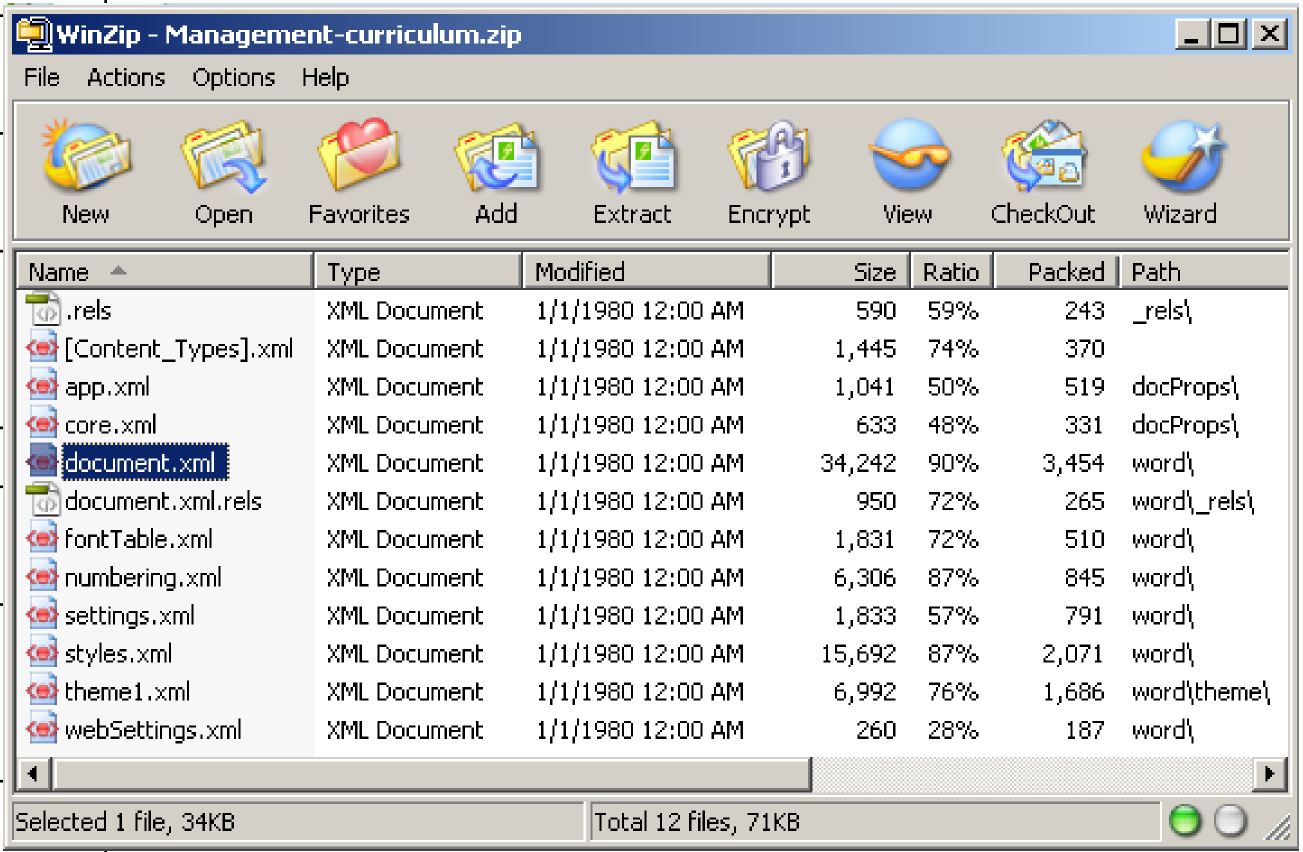


```
00000000: 00 05 16 07 00 02 00 00 4D 61 63 20 4F 53 20 58 ........Mac OS X
00000010: 20 20 20 20 20 20 20 20 00 02 00 00 00 09 00 00         ........
00000020: 00 32 00 00 0E B0 00 00 00 02 00 00 0E E2 00 00 .2..............
00000030: 01 1E 50 50 54 58 50 50 54 33 00 00 00 00 00 00 ..PPTXPPT3......
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



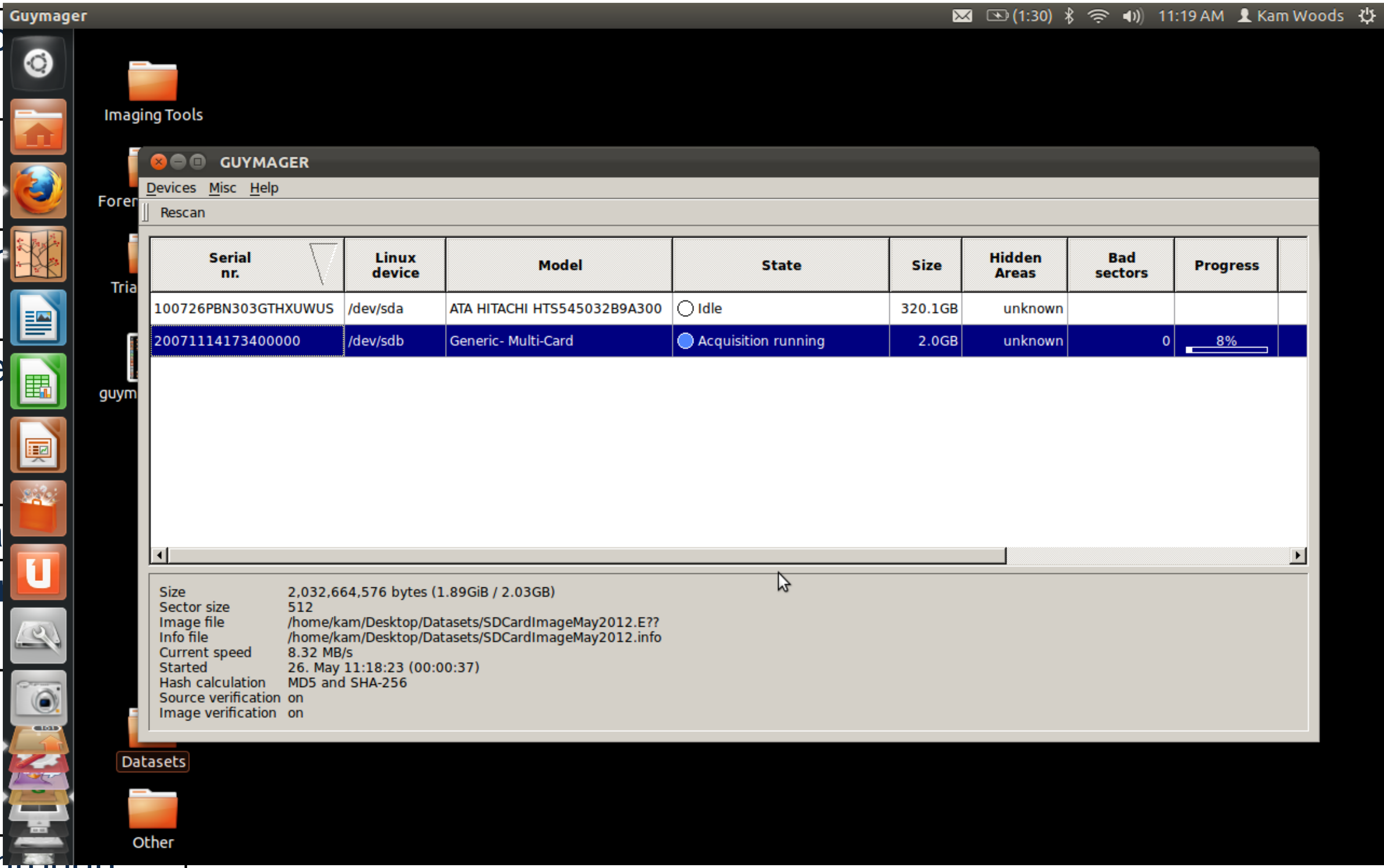
Name	Type	Modified	Size	Ratio	Packed	Path
.rels	XML Document	1/1/1980 12:00 AM	590	59%	243	_rels\
[Content_Types].xml	XML Document	1/1/1980 12:00 AM	1,445	74%	370	
app.xml	XML Document	1/1/1980 12:00 AM	1,041	50%	519	docProps\
core.xml	XML Document	1/1/1980 12:00 AM	633	48%	331	docProps\
document.xml	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

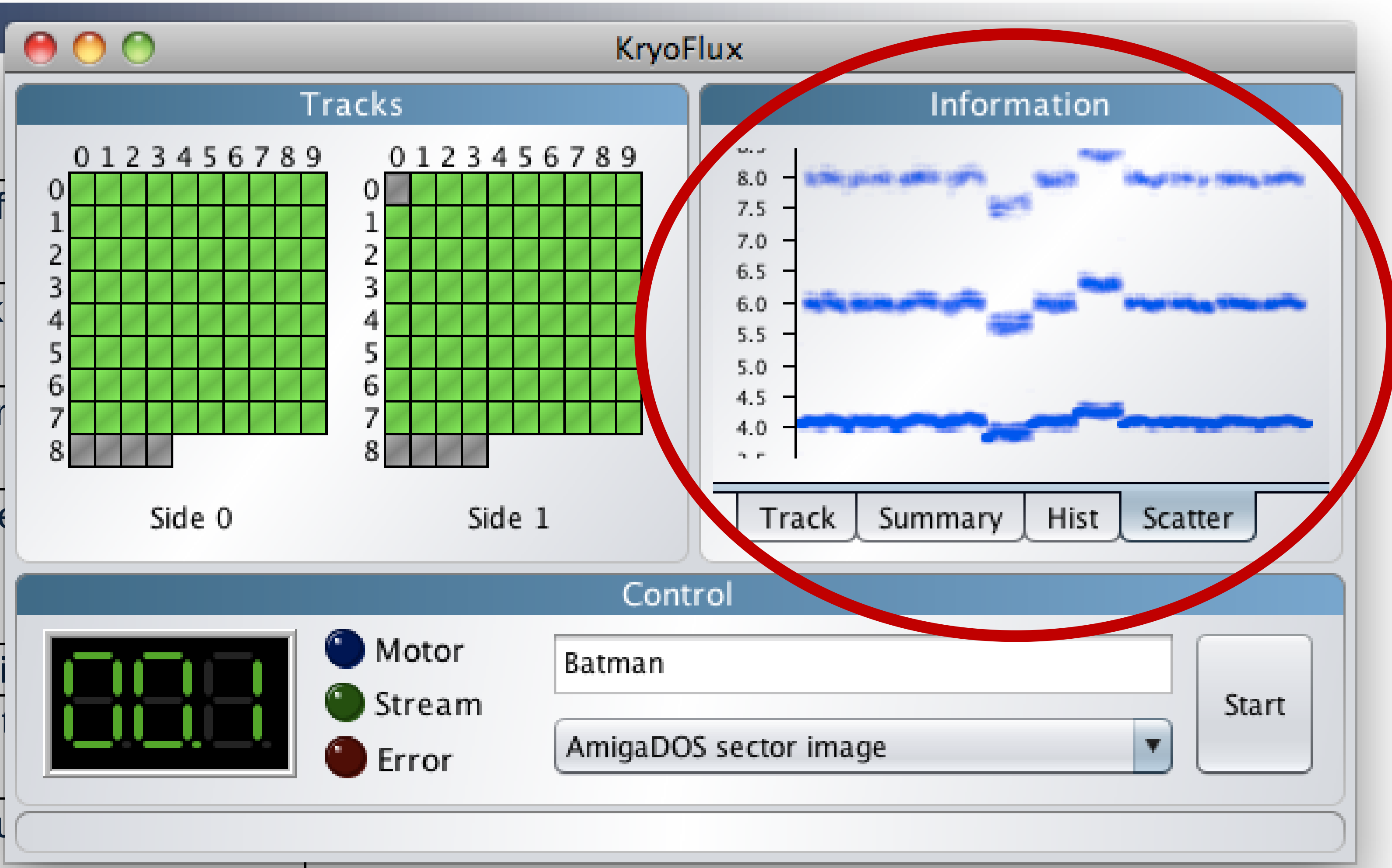
Level
Aggregation of objec
Object or package
In-application render
File through filesyste
File as “raw” bitstrea
Sub-file data structu
Bitstream through I/O equipment
Raw signal stream through I/O equipment
Bitstream on physical medium



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

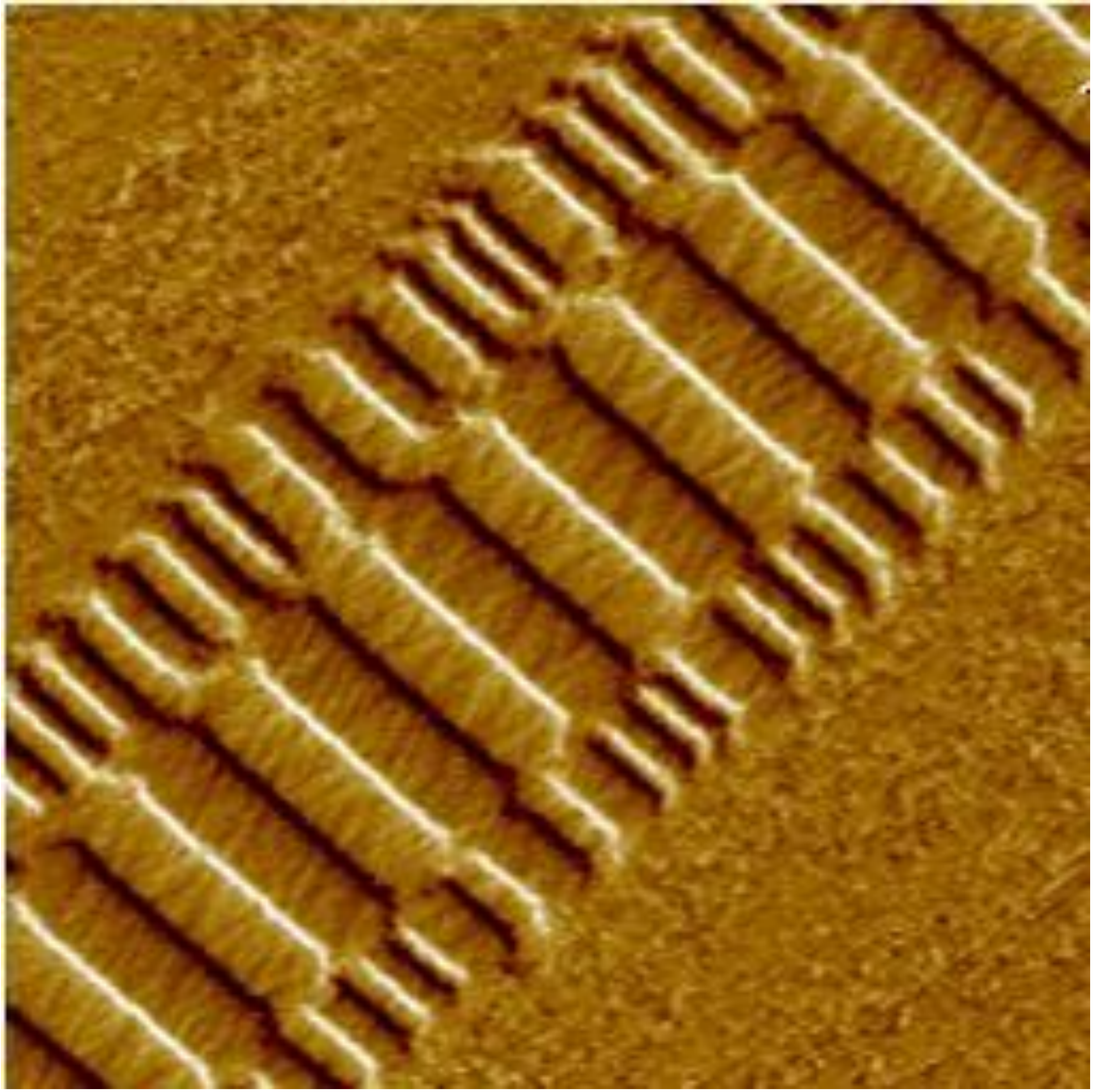


Batman

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

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

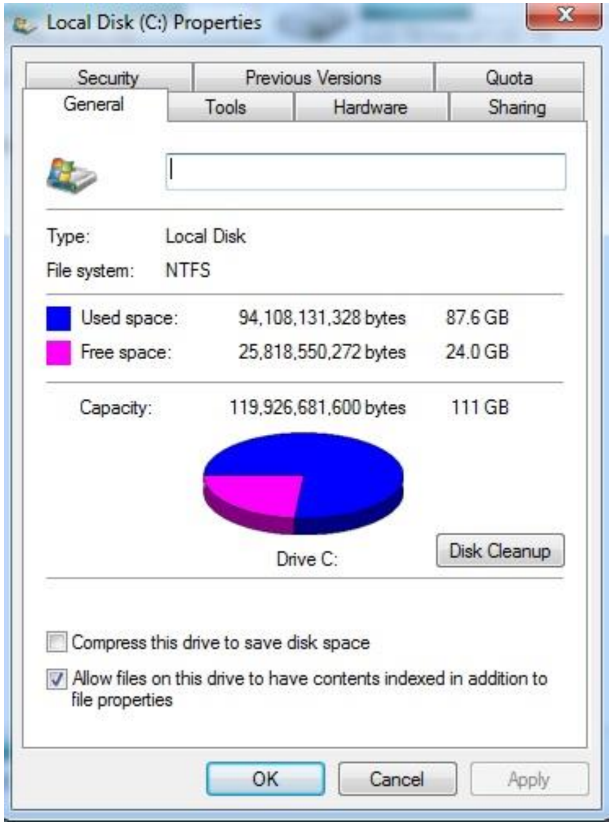
The technology associated with interpreting the representation at each of the layers can change or become less available

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

~~The~~ A First Rule for Digital Forensics

Empty disk space is rarely empty



Local Disk (C:) Properties

Type: Local Disk

File system: NTFS

🟦 Used space: 94,108,131,328 bytes 87.6 GB

🟪 Free space: 25,818,550,272 bytes 24.0 GB

Capacity: 119,926,681,600 bytes 111 GB

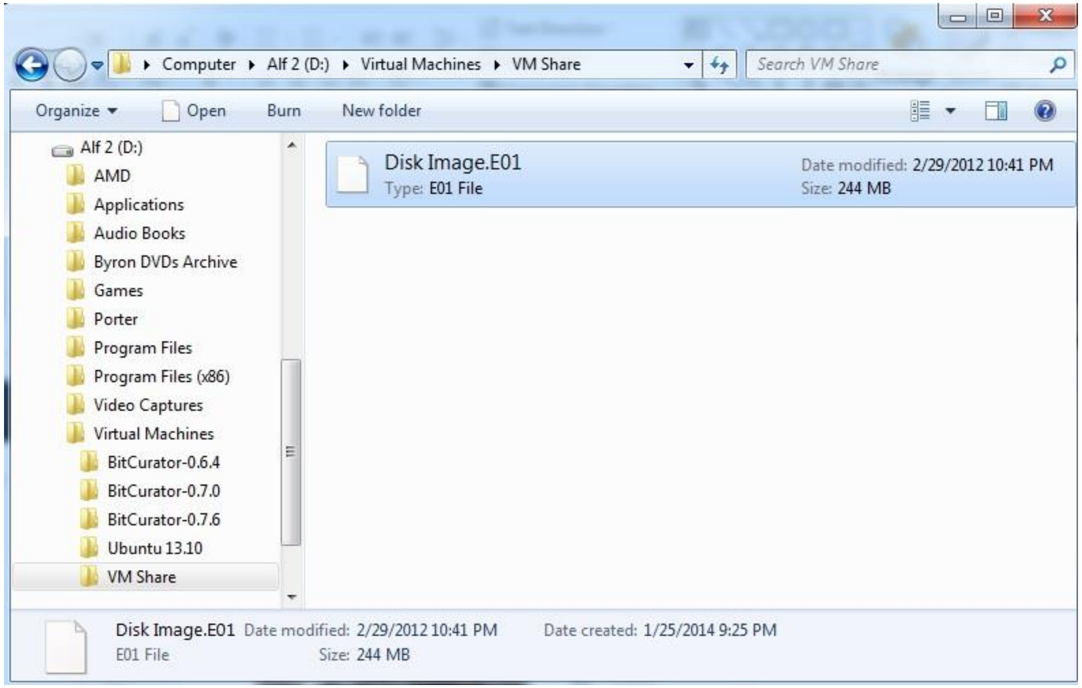
Drive C:

☐ Compress this drive to save disk space

☑ Allow files on this drive to have contents indexed in addition to file properties

- Deleted files
- File level metadata
- Swap space
- Temporary files
- Auto recovery files
- System files

Disk Imaging vs. Logical Copy



Alf 2 (D:)
AMD
Applications
Audio Books
Byron DVDs Archive
Games
Porter
Program Files
Program Files (x86)
Video Captures
Virtual Machines
BitCurator-0.6.4
BitCurator-0.7.0
BitCurator-0.7.6
Ubuntu 13.10
VM Share

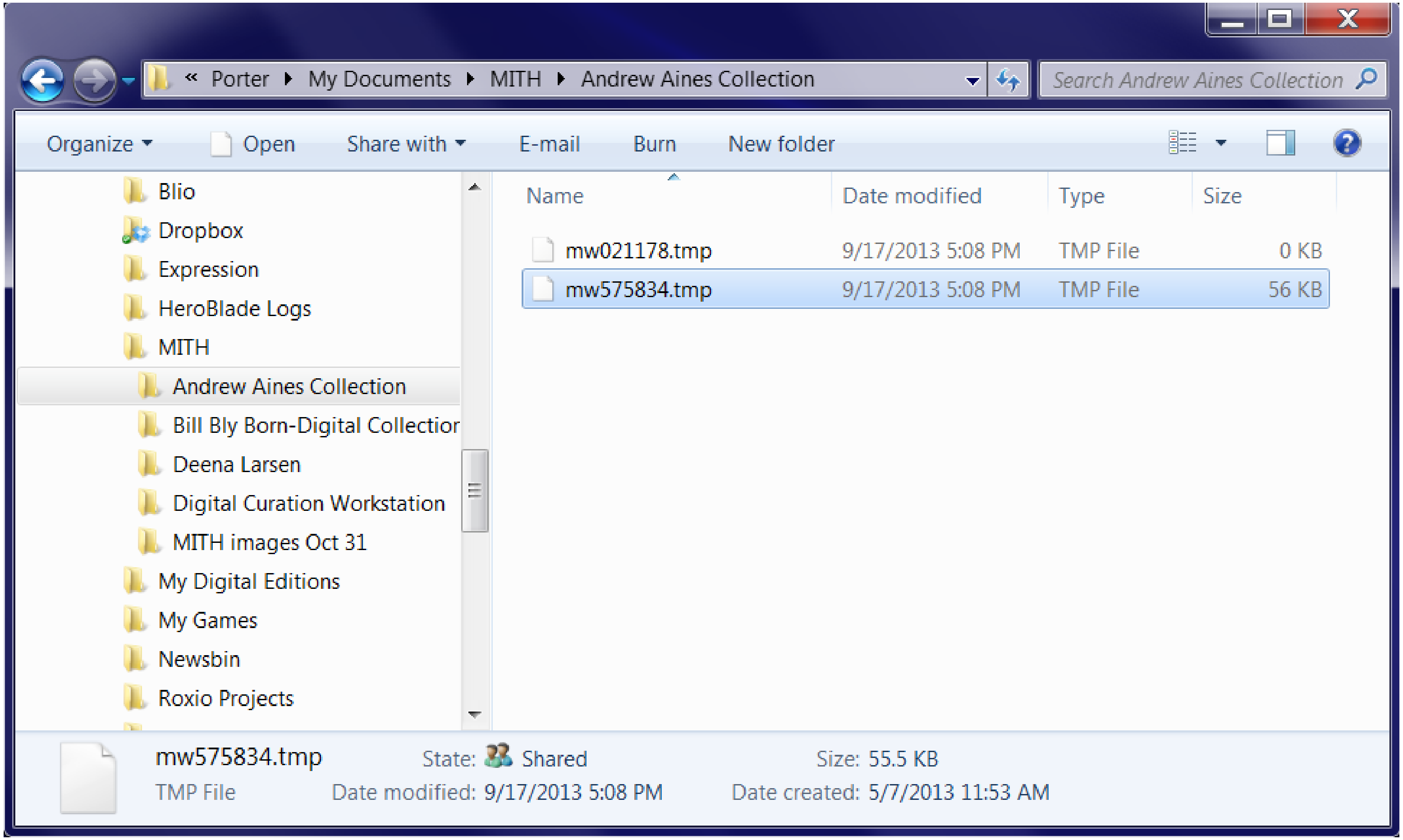
Disk Image.E01
Type: E01 File
Date modified: 2/29/2012 10:41 PM
Size: 244 MB

Disk Image.E01 Date modified: 2/29/2012 10:41 PM Date created: 1/25/2014 9:25 PM
E01 File Size: 244 MB

"[A] single file that contains an exact, sector-by-sector bitstream copy of the disk's content and ensures that various forms of essential metadata and technical dependencies will be retained."

- Ricky Erway, "You've Got to Walk Before You Can Run: First Steps
for Managing Born-Digital Content Received on Physical Media"

Disk Imaging vs. Logical Copy


« Porter ▸ My Documents ▸ MITH ▸ Andrew Aines Collection
Search Andrew Aines Collection
Organize ▾
Open
Share with ▾
E-mail
Burn
New folder

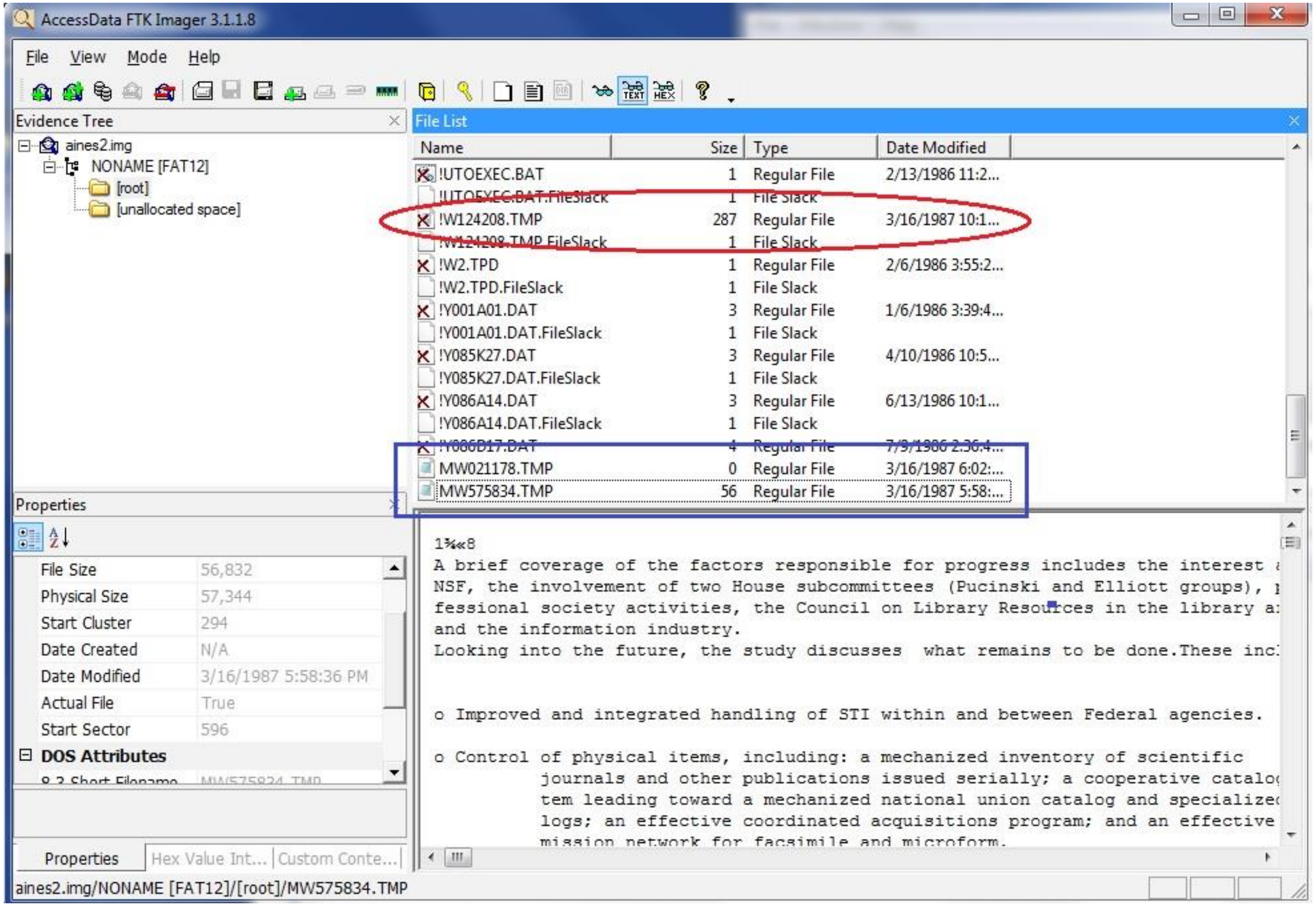
Blio
Dropbox
Expression
HeroBlade Logs
MITH
Andrew Aines Collection
Bill Bly Born-Digital Collectior
Deena Larsen
Digital Curation Workstation
MITH images Oct 31
My Digital Editions
My Games
Newsbin
Roxio Projects

Name	Date modified	Type	Size
mw021178.tmp	9/17/2013 5:08 PM	TMP File	0 KB
mw575834.tmp	9/17/2013 5:08 PM	TMP File	56 KB

mw575834.tmp
TMP File
State: Shared
Date modified: 9/17/2013 5:08 PM
Size: 55.5 KB
Date created: 5/7/2013 11:53 AM

Contents of a “logical” disk image—just visible files

Disk Imaging vs. Logical Copy



Name	Size	Type	Date Modified
!UTOEXEC.BAT	1	Regular File	2/13/1986 11:2...
!UTOEXEC.BAT.FileSlack	1	File Slack	
!W124208.TMP	287	Regular File	3/16/1987 10:1...
!W124208.TMP.FileSlack	1	File Slack	
!W2.TPD	1	Regular File	2/6/1986 3:55:2...
!W2.TPD.FileSlack	1	File Slack	
!Y001A01.DAT	3	Regular File	1/6/1986 3:39:4...
!Y001A01.DAT.FileSlack	1	File Slack	
!Y085K27.DAT	3	Regular File	4/10/1986 10:5...
!Y085K27.DAT.FileSlack	1	File Slack	
!Y086A14.DAT	3	Regular File	6/13/1986 10:1...
!Y086A14.DAT.FileSlack	1	File Slack	
!Y086D17.DAT	4	Regular File	7/9/1986 2:36:4...
MW021178.TMP	0	Regular File	3/16/1987 6:02:...
MW575834.TMP	56	Regular File	3/16/1987 5:58:...

File Size	56,832
Physical Size	57,344
Start Cluster	294
Date Created	N/A
Date Modified	3/16/1987 5:58:36 PM
Actual File	True
Start Sector	596
DOS Attributes	
8.3 Short Filename	MW575834.TMP

```
1¾«8
A brief coverage of the factors responsible for progress includes the interest a
NSF, the involvement of two House subcommittees (Pucinski and Elliott groups), p
fessional society activities, the Council on Library Resources in the library ai
and the information industry.
Looking into the future, the study discusses  what remains to be done.These incl


o Improved and integrated handling of STI within and between Federal agencies.

o Control of physical items, including: a mechanized inventory of scientific
          journals and other publications issued serially; a cooperative catalog
          tem leading toward a mechanized national union catalog and specialized
          logs; an effective coordinated acquisitions program; and an effective
          mission network for facsimile and microform.
```

aines2.img/NONAME [FAT12]/[root]/MW575834.TMP

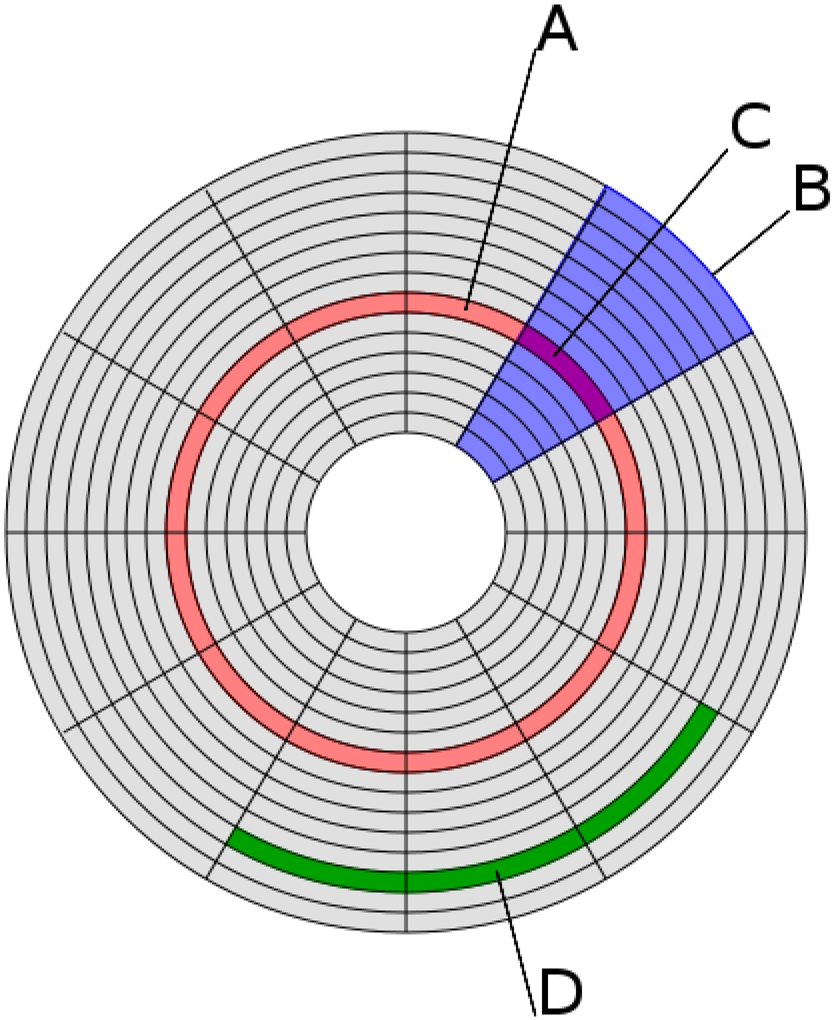
Contents of a “forensic” disk image—deleted files, slack space, system files, and more

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



Hard Drive Structure:

A = track

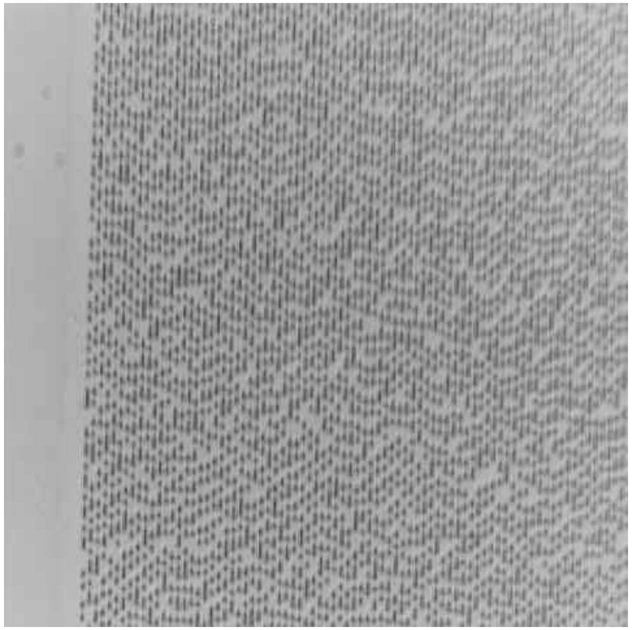
B = sector

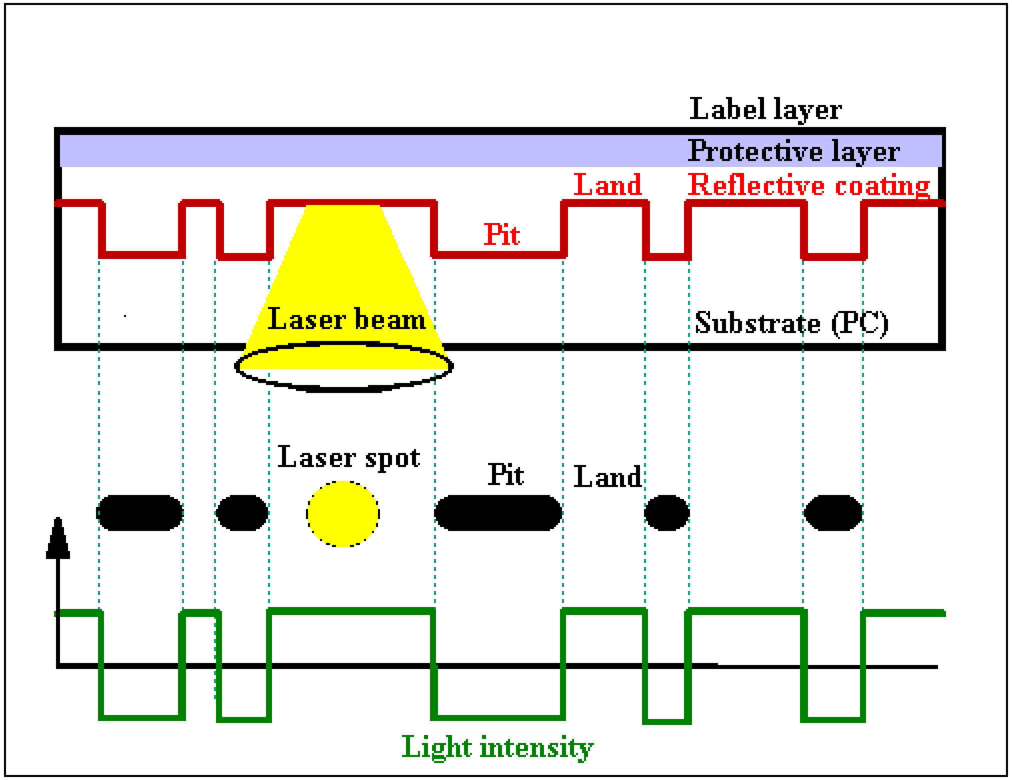
C = sector of a track

D = cluster

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

Optical Media – CD-ROM as Example



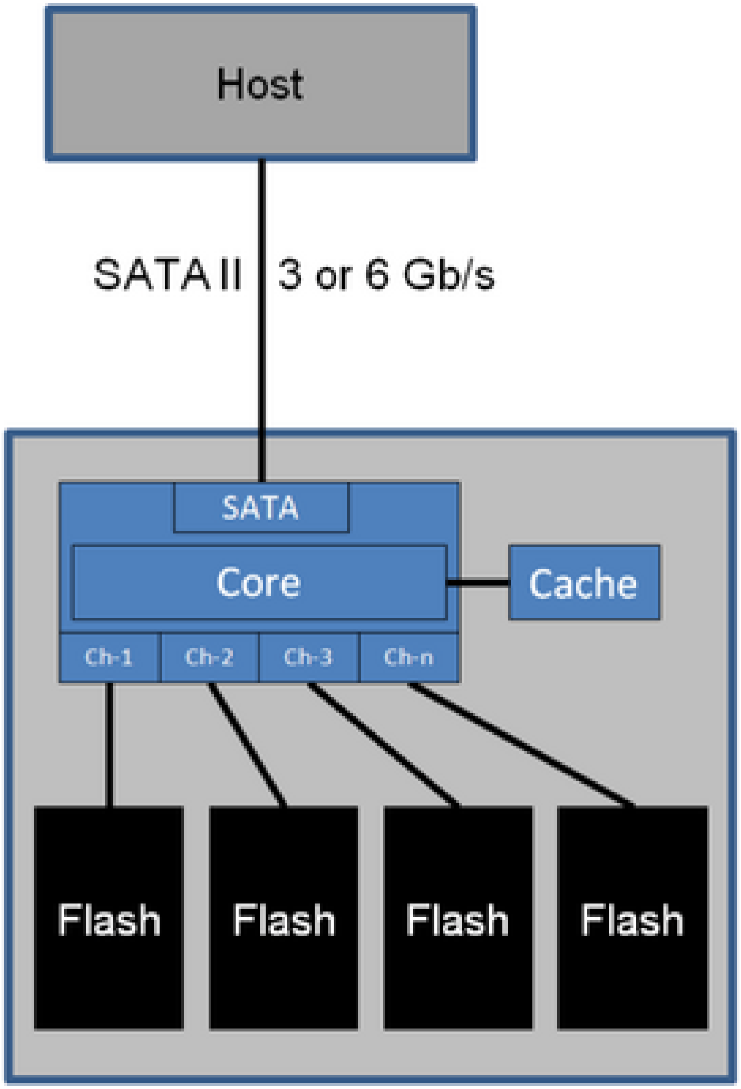


Source of Images: Compact Disk (CD). USByte.
http://www.usbyte.com/common/compact_disk_3.htm

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



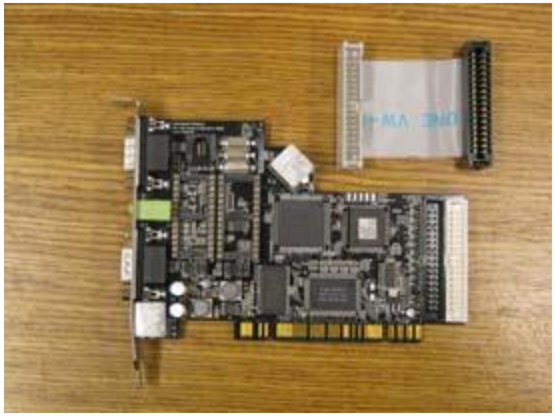
Source: http://www.tomshardware.com/gallery/ssd-controller-external-cache,0101-260900-0-0-0-0-jpg-.html

Floppy Disks

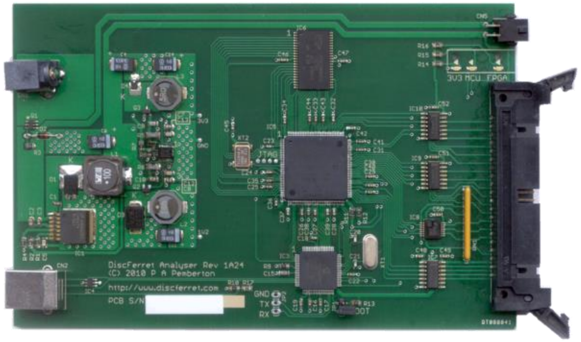
- Physical storage is similar to hard drives described above (magnetic charges in a spinning disk)
- Various types and sizes, e.g. high density, double density, 3.5 inch, 5.25 inch, 8 inch
- 3.5 inch floppies are relatively easy to read using a USB drive, but older ones are more complicated…

Floppy Controller Hardware

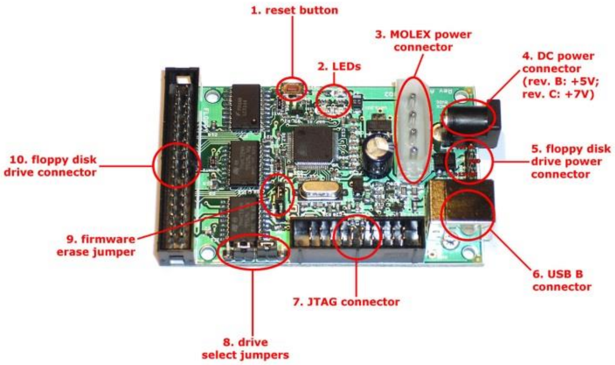
CatWeasel[1] (no longer available)



Disc Ferret[2]

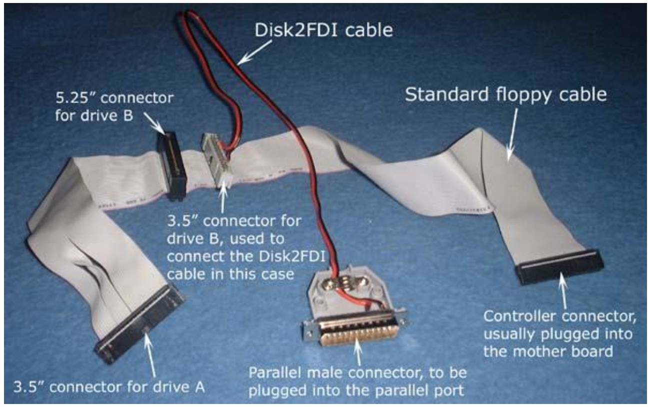


Kryoflux[3]

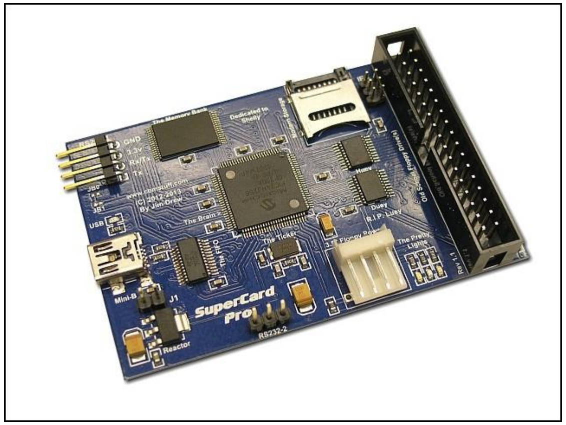


FC 5025[4]

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
Parallel male connector, to be plugged into the parallel port
3.5″ connector for drive A

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

Checksums – Compact Representations of Bitstreams

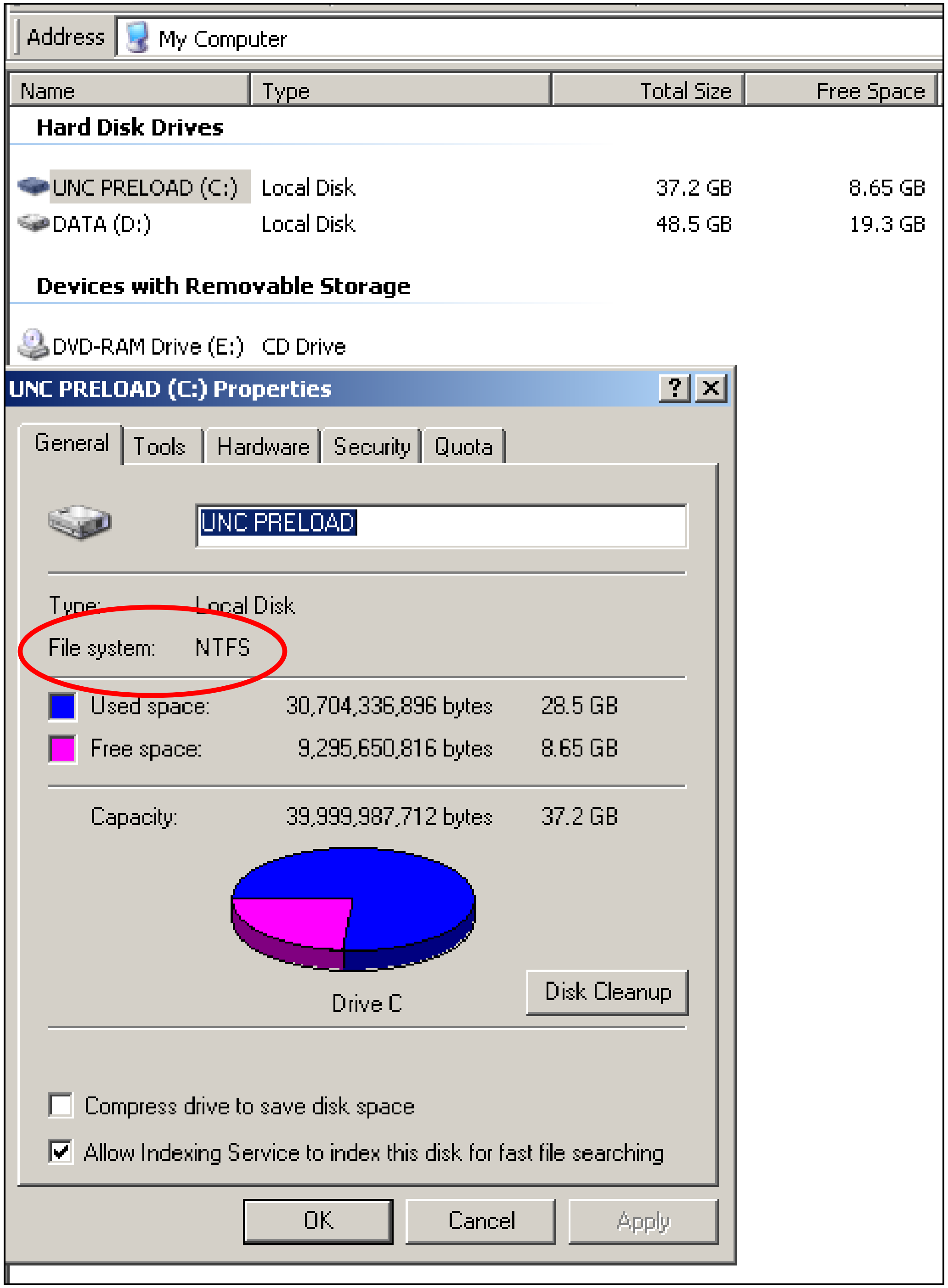
- A given bitstream, fed into an algorithm, will generate a short string of characters that is **extremely** unlikely to be generated by a different bistream fed into that same algorithm
- Most common = MD5, SHA-1
- Can determine:
  - □ If bits have changed after a transfer
  - □ If bits have flipped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

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

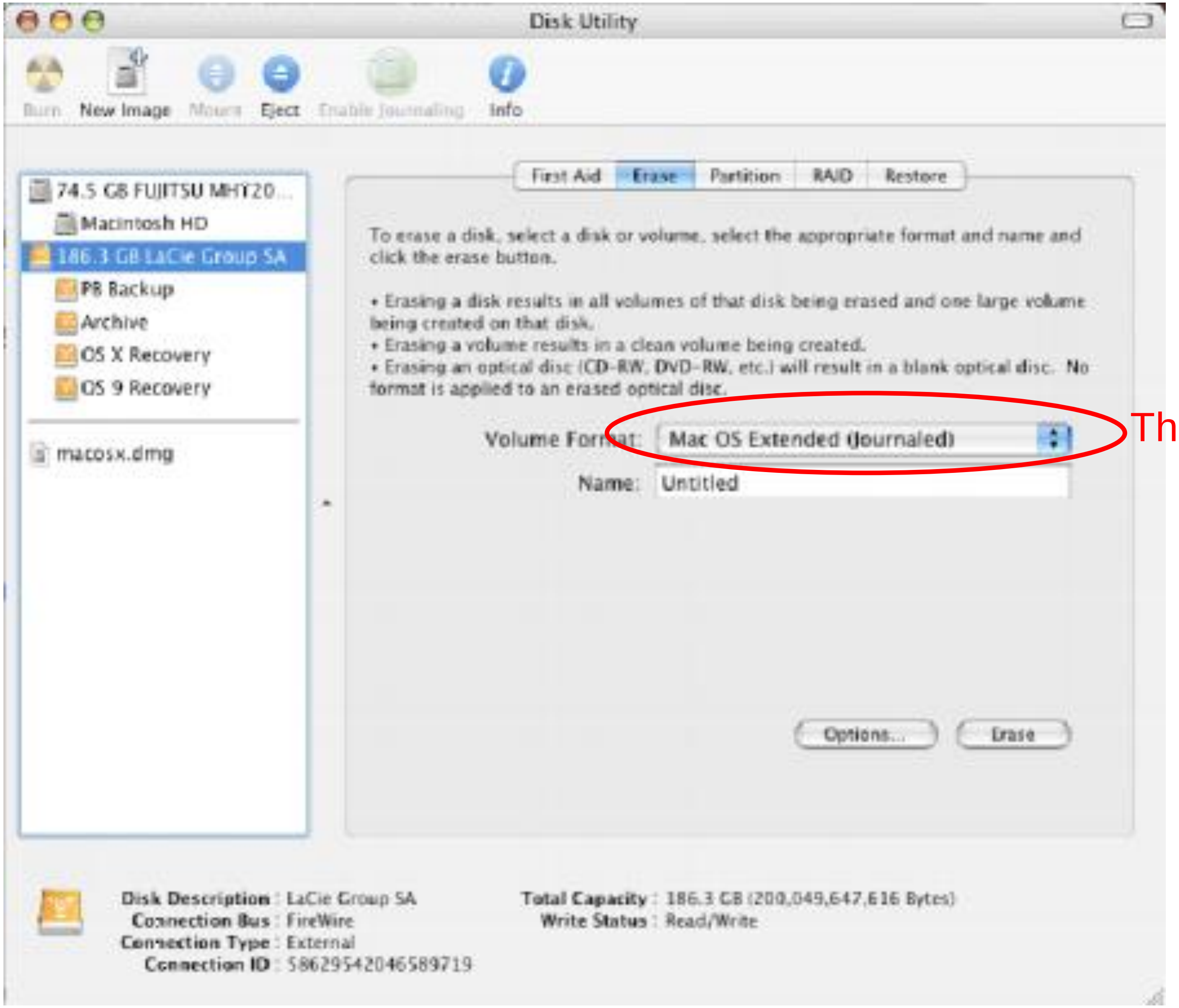
Type: Local Disk

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 : 5862954204658719

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]
ext, ext2, ext3 (Extended File System)	Linux
FAT16	MS-DOS
FAT32 (VFAT)	Windows 95, 98
HFS (Hierarchical File System)	Macintosh System 4-8
HFS+	Macintosh System 8.1-X
HPFS (High Performance File System)	OS/2
ISOFS (ISO 9660)	Any OS that reads data from a CD
JFS1 (Journaled File System)	AIX (IBM)
MFS (Macintosh File System)	Macintosh System 1-3
NTFS	Windows NT, 2000, XP, Server 2003, Server 2008, Vista
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]
ext, ext2, ext3 (Extended File System)	
FAT16	
FAT32 (VFAT)	
HFS (Hierarchical File System)	
HFS+	
HPFS (High Performance File System)	
ISOFS (ISO 9660)	
JFS1 (Journaled File System)	
MFS (Macintosh File System)	
NTFS	ver 2008, Vista
ReiserFS	
UFS (Unix File System) aka FFS (Fast File System)	Various flavors of Unix

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

Three Key Takeaways

1. Digital Forensics tools and methods allow us to work with files at a variety of layers (Application, OS, File System, Bitstream, etc.)
2. Disk imaging allows us to capture metadata and files not made visible by the operating system
3. File Systems are the means by which the operating system organizes data on a disk… and there are a lot of them.

Disable Automount

- Windows: At the command line type “mountvol /N”
- OS X: ?