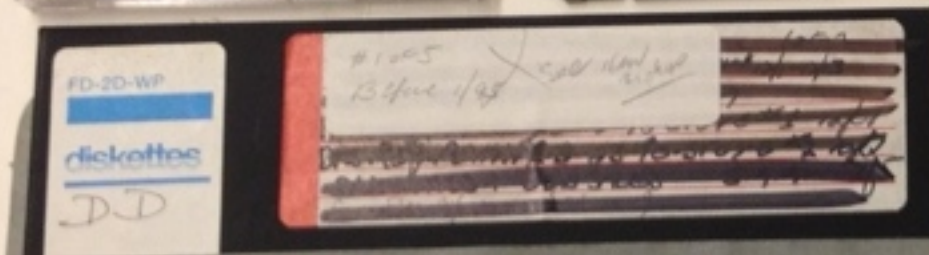
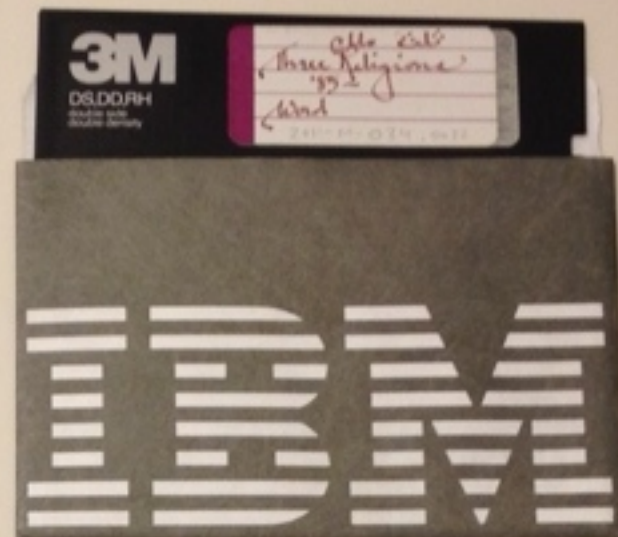
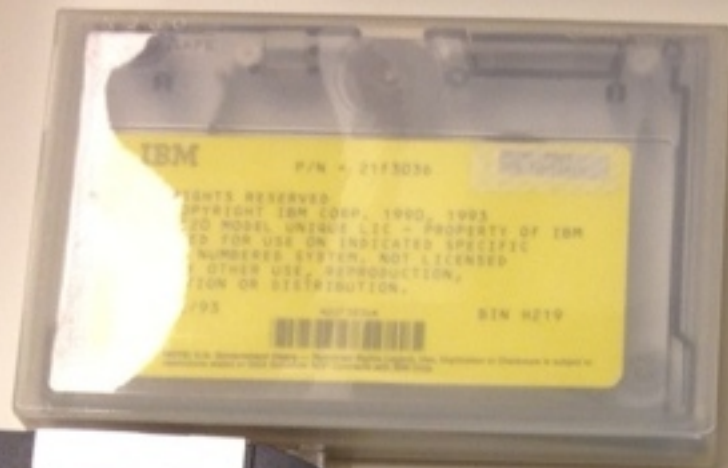


Digital Forensics & Born-Digital Archives at Yale

Mark A. Matienzo
YUL Manuscripts and Archives
SCOPA Conference Recap
September 19, 2012



Digital forensics in the archival domain

- Increasing use of digital forensics tools/methodologies within the context of digital archives programs (Kirschenbaum et al. 2010)
- Technology-focused work (John 2008; Woods & Brown 2009; AIMS Work Group 2012, Lee et al. 2012)
- Methodology-focused work (Duranti 2009; Xie 2011)

Significant barriers to use of digital forensics in archives

- Cost (Kirschenbaum et al. 2010; Daigle 2012)
- Complexity (Kirschenbaum et al. 2010; Daigle 2012)
- Digital archives as an emerging market for forensics

Potential of open source digital forensics software

- Requires additional tool development work to be useful for archivists (Kirschenbaum et al. 2010)
- Requires additional integration work (Lee et al. 2012)

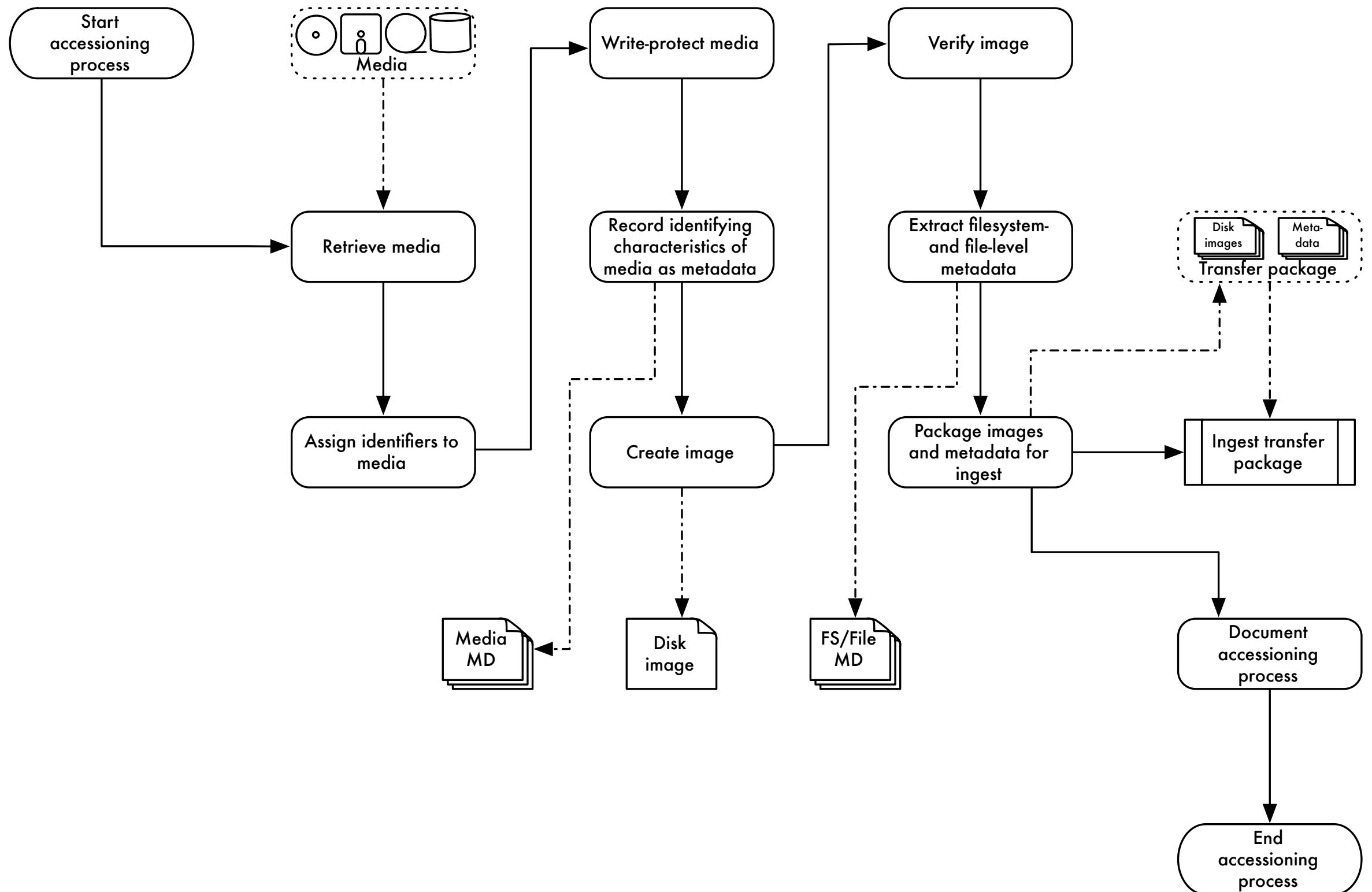
Institutional Context

- Focus on implementation of and development with digital forensics software at YUL
- Work must support accessioning, processing, and management of born-digital archival material
- Primary focus are records received on legacy media

Design Principles

- Whenever possible, use and develop with open source digital forensics software to support accessioning, arrangement, and description of born-digital archival records
- Focus on first two phases (preservation and searching) of Carrier's (2005) model of digital investigation process
- Curation micro-services (Abrams, et al. 2010) as philosophical basis to guide development and implementation
- Recognition of both disk images as digital object (Woods, Lee, and Garfinkel 2011) and objects within disk images as needing management
- Intention of forensic soundness, but assume much of state is lost

Workflow













Disk Image Acquisition

- Requires a combination of hardware (drives/media readers, controller cards, write blockers) and software
- In some cases, software depends on particular hardware
- Software tested: FTK Imager (proprietary/gratis), hardware-specific solutions (FC5025 WinDIB; KryoFlux DTC/GUI; Catweasel Imagetool3)
- Goal: sector image interpretable by multiple tools



File List

Name	Size	Type	Date Modified
 !w0000	59 KB	Regular File	7/21/1997 7:58...
 Hangman.bak	56 KB	Regular File	6/10/1997 5:23...
 Hangman.bak	58 KB	Regular File	6/10/1997 5:36...
 Hangman.bak	59 KB	Regular File	7/21/1997 7:57...
 Hangman.tex	0 KB	Regular File	6/10/1997 5:23...
 Hangman.tex	59 KB	Regular File	7/21/1997 7:58...
 KENNEDY.WPD	41 KB	Regular File	10/17/1997 11:...
 Q3.DIR	1 KB	Regular File	7/5/1999 11:18...
 Q3.DIR.FileSlack	1 KB	File Slack	
 ODATA.ABD	10 KB	Regular File	7/5/1999 11:12...

×

0000	43	39	31	30	39	32	37	41-30	30	31	28	00	00	00	00	C910927A001 (....
0010	00	00	00	00	00	00	4C	77-1A	21	00	00	00	00	00	00Lw~!.....
0020	51	44	41	54	41	20	20	20-51	44	46	20	00	43	48	5A	QDATA QDF .CHZ
0030	E5	26	31	3D	00	00	48	5A-E5	26	02	00	C0	7D	04	00	â&1=...HZâ&...À}..
0040	51	44	41	54	41	20	20	20-51	53	44	20	00	B4	4C	5A	QDATA QSD .`LZ
0050	E5	26	31	3D	00	00	48	5A-E5	26	92	02	60	49	00	00	â&1=...HZâ&...`I..
0060	51	44	41	54	41	20	20	20-51	45	4C	20	00	66	4D	5A	QDATA QEL .fmZ
0070	E5	26	31	3D	00	00	65	B7-B6	26	B7	02	00	3C	00	00	â&1=...e·Ŧ&...<..
0080	51	44	41	54	41	20	20	20-41	42	44	20	00	16	4E	5A	QDATA ABD .·NZ
0090	E5	26	31	3D	00	00	8B	59-E5	26	D5	02	97	26	00	00	â&1=...Yâ&Õ...&..
00a0	51	33	20	20	20	20	20	20-44	49	52	20	00	88	4E	5A	Q3 DIR .·NZ
00b0	E5	26	31	3D	00	00	4E	5A-E5	26	E9	02	17	00	00	00	â&1=...NZâ&é.....
00c0	E5	41	4E	47	4D	41	4E	20-42	4B	21	20	00	13	F9	85	âANGMAN BK! .·ù·
00d0	CA	22	CA	22	00	00	FA	85-CA	22	23	00	77	F7	00	00	Ê"Ê"··ú·Ê"#·w÷..
00e0	E5	41	4E	47	4D	41	4E	20-42	4B	21	20	00	2F	41	89	âANGMAN BK! ·/A·
00f0	CA	22	CA	22	00	00	44	89-CA	22	02	00	91	F7	00	00	Ê"Ê"··D·Ê"···÷..
0100	E5	41	4E	47	4D	41	4E	20-54	58	54	20	00	6B	65	89	âANGMAN TXT ·ke·
0110	CA	22	CA	22	00	00	66	89-CA	22	03	00	41	F9	00	00	Ê"Ê"··f·Ê"··Aù..
0120	E5	48	00	61	00	6E	00	67-00	6D	00	0F	00	B6	61	00	âH·a·n·g·m···Ŧa·
0130	6E	00	2E	00	74	00	65	00-78	00	00	00	00	00	FF	FF	n··t·e·x·····ÿÿ
0140	E5	41	4E	47	4D	41	4E	20-54	45	58	20	00	A0	F6	8A	âANGMAN TEX · ö·
0150	CA	22	CA	22	00	00	F6	8A-CA	22	00	00	00	00	00	00	Ê"Ê"··ö·Ê"·····
0160	E5	57	30	30	30	30	20	20-20	20	20	20	08	1C	F7	8A	âW0000 ···÷·
0170	CA	22	CA	22	00	00	F9	8A-CA	22	04	00	8C	DD	00	00	Ê"Ê"··ù·Ê"···Ý·
0180	E5	48	00	61	00	6E	00	67-00	6D	00	0F	00	22	61	00	âH·a·n·g·m···"a·

Properties | Hex Value Inter... Custom Content...

Cursor pos = 0; log sec = 19

Tracks

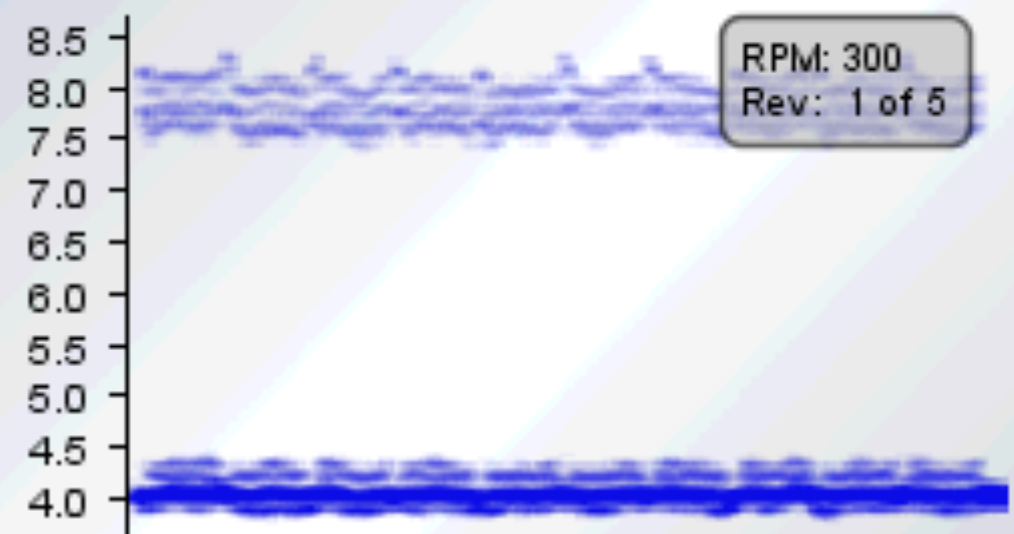
	0	1	2	3	4	5	6	7	8	9
0	H		H		H		H		H	
1	H		H		H		H		H	
2	H		H		H		H		H	
3	H		H		H		H		H	
4	H		H		H		H		H	
5	H		H		H		H		H	
6	H		H		H		H		H	
7	H		H		H		H		H	
8										

Side 0

	0	1	2	3	4	5	6	7	8	9
0										
1										
2										
3										
4										
5										
6										
7										
8										

Side 1

Information

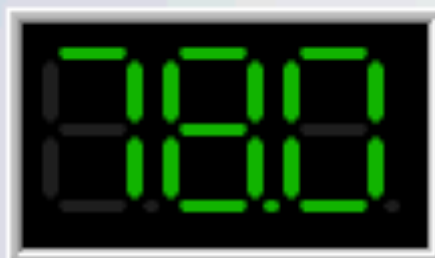


Track

Histogram

Scatter

Control



Motor

Stream

Error















testimage

FM sector image; 40 track. single sided, 2...

Start





Error: Extra data was found hidden in unused parts of the block header.

Electronic Records on Media Accessioning Log

New ▾ Actions ▾ Settings ▾						
	Type	Media number	Media Format	Imaging Date	Imaging Successful?	Bag Create
		2011-M-075.0001	CD-R		No	No
		2011-M-075.0002	DVD-R		Yes	No
		2011-M-075.0003	DVD-R		Yes	No
		2011-M-075.0004	DVD-R		Yes	No
		2011-M-075.0005	DVD-R		Yes	No
		2011-M-075.0006	DVD-R		Yes	No
		2011-M-075.0007	CD-R		Yes	No
		2011-M-075.0008	CD-R		Yes	No
		2011-M-075.0009	CD-R		Yes	No
		2011-M-075.0010	DVD-R		Yes	No
		2011-M-075.0011	CD-R		Yes	No
		2011-M-075.0012	CD-R		Yes	No
		2011-M-075.0013	Zip disk		Yes	No

Electronic Records on Media Accessioning Log: 2011-M-075.0008

Close

 New Item
  Edit Item
  Delete Item
  Manage Permissions
 Alert Me

Media number	2011-M-075.0008
Media Format	CD-R
Media Density (floppies only)	N/A
Interface	N/A
Label text	Osaka Monograph Final Images Aug 29 2003 Monograph Latest Files
Manufacturer	
Serial Number (hard drives only)	
Examiner	Glick, Kevin
Imaging Successful?	Yes
Imaging Date	
Image filename	2011-M-075.0008.ISO
Source File System	ISO9660, Joliet
Image format	ISO
Imaging Software	ImgBurn
Image Fixity Function	MD5
Image Fixity Value	dbca43c94690edff07329b6687550f60
Notes	mam54 04/28/2011: Could not extract metadata using fiwalk; log file from imaging process says that the block structure is Mode 2/Form 1
Metadata Extracted?	No
Bag Created?	No
Transfer to Storage Date	
Fiscal Year	2010-11

Created at 4/27/2011 9:35 AM by Glick, Kevin
 Last modified at 4/28/2011 4:26 PM by Matienzo, Mark

Close

Analysis Process

- Multiple levels of analysis within digital forensics based on layers of abstraction (Carrier 2003)
- Conceptual linkages with metadata extraction/analysis processes with digital curation/archival domain

Physical Media			Media Management		File System			Application	
Head	Cyl	Etc.							
Sectors			Partition Table						
			Partition		Boot Sector	FAT	Data Area		
					...				
					File			ASCII	
								HTML	

Carrier, 2003

Metadata Extraction

- Use open source digital forensics software (Sleuth Kit, fiwalk) and other open source tools to characterize media, volume, file system, and file information
- Attempt to repurpose this information as descriptive, structural, and/or technical metadata to support accessioning, appraisal, and processing

The Sleuth Kit

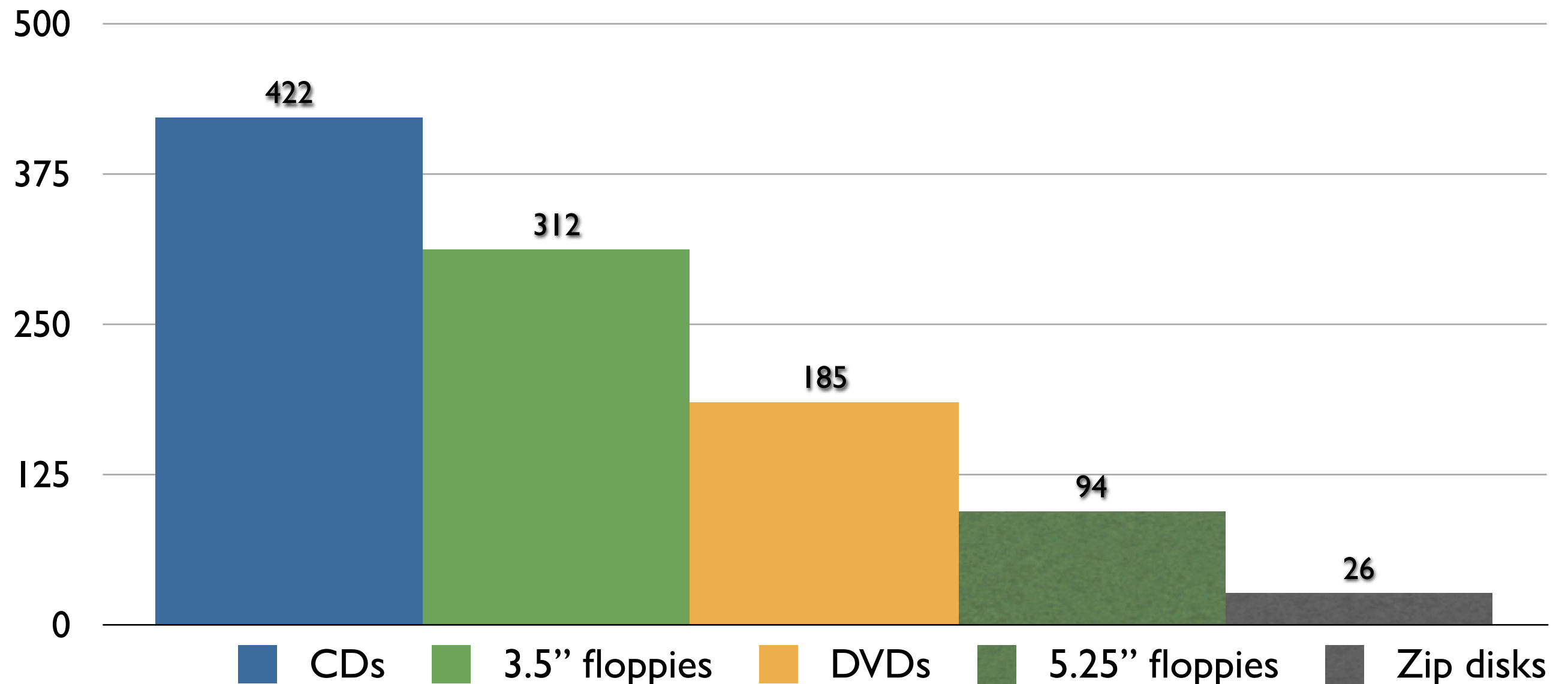
- Open source C library, command line tools, and GUI application (Autopsy) for forensic analysis
- Supports analysis of FAT, NTFS, ISO9660, HFS+, Ext2/3, UFS1/2
- Splits tools into layers: volume system, file system, file name, metadata, data unit ("block")
- Additional utilities to sort and post-process extracted metadata

Digital Forensics XML

- Representation in XML of structured forensic information developed by Simson Garfinkel
- Produced by tools including fiwalk (Garfinkel 2012), which uses Sleuth Kit for volume, file system, file, and application-level analysis
- Easily extensible (local plugin development as focus)
- Straight forward to process

Disk Images

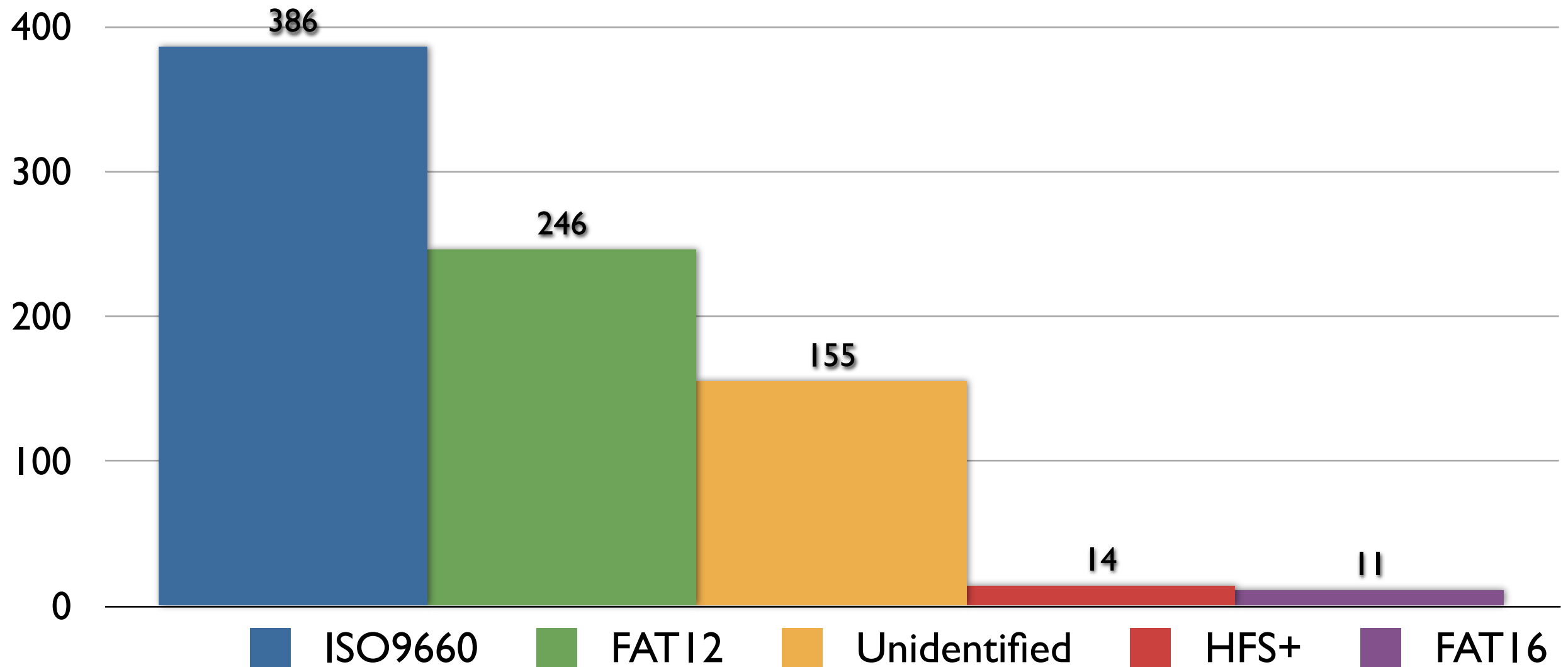
- Acquired 1,039 disk images from across 69 accessions at Manuscripts and Archives



Metadata Extraction

- Ran metadata extraction on 812 images

File Systems within Images



Metadata Extraction

- Ran enhanced metadata extraction on 619 images (users plugins for fiwalk developed during research)
- Performed analysis on 49,724 files within images
- Successfully identified 43,729 files (147 unique file types) against PRONOM format registry
- Identified 9 files as containing virus signatures (2 unique virus signatures)

Software Development

- Created Fiwalk plugins to perform additional analysis and evaluation of files/bitstreams within disk images (virus identification and file format identification)
- Gumshoe: prototype interface (using Blacklight and Solr) to provide search/browse access to disk image metadata

Advantages

- Faster (and more forensically sound) to extract metadata once rather than having to keep processing an image
- Possibility of developing better assessments during accessioning process (significance of directory structure, accuracy of timestamps)
- Integrating additional extraction processes and building supplemental tools is simple
- Performance of tools correlates to complexity of analysis

Limitations

- Use of tools limited to specific types of file systems
- Additional software (particularly to document imaging process) requires additional integration and data normalization
- DFXML is not (currently) a metadata format common within domains of archives/libraries and requires an domain-specific application profile
- Extracted metadata maybe harder to repurpose for descriptive purposes based on level of granularity

Work in Progress

- BitCurator project under development; early release available for testing: <http://wiki.bitcurator.net>
- The Sleuth Kit and related tools under continuing development (Autopsy, fiwalk, etc.): <http://sleuthkit.org>
- Additional testing, development integration under work at Yale and NYPL

Thanks!

Mark A. Matienzo

mark.matienzo@yale.edu

References

- Abrams, S., et al. (2011). "Curation Micro-Services: A Pipeline Metaphor for Repositories." *Journal of Digital Information* 12(2). <http://journals.tdl.org/jodi/article/view/1605>
- AIMS Work Group (2012). *AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship*. <http://www2.lib.virginia.edu/aims/whitepaper/>
- Carrier, B. (2003). "Defining Digital Forensic Examination and Analysis Tools Using Abstraction Layers." *International Journal of Digital Evidence* 1(4).
- Carrier, B. (2005). *File System Forensic Analysis*. Boston and London: Addison Wesley.
- Daigle, B.J. (2012). "The Digital Transformation of Special Collections." *Journal of Library Administration* 52(3-4), 244-264.
- Duranti, L. (2009). "From Digital Diplomats to Digital Records Forensics." *Archivaria* 68, 39-66.
- Garfinkel, S. (2012). "Digital Forensics XML and the DFXML Toolset." *Digital Investigation* 8, 161-174.
- John, J.L. (2008). "Adapting Existing Technologies for Digitally Archiving Personal Lives: Digital Forensics, Ancestral Computing, and Evolutionary Perspectives and Tools." Presented at iPRES 2008. http://www.bl.uk/ipres2008/presentations_day1/09_John.pdf
- Kirschenbaum, M.G., et al. (2010). *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*. Washington: Council on Library and Information Resources.
- Lee, C.A., et al. (2012). "BitCurator: Tools and Techniques for Digital Forensics in Collecting Institutions." *D-Lib Magazine* 18(5/6).
- UC Curation Center/California Digital Library (2019). "UC3 Curation Foundations." Revision 0.13. <https://confluence.ucop.edu/download/attachments/13860983/UC3-Foundations-latest.pdf>
- Woods, K. and Brown, G. (2009). "From Imaging to Access: Effective Preservation of Legacy Removable Media." In *Archiving 2009*. Springfield, VA: Society for Imaging Science and Technology.
- Woods, K., Lee, C.A., and Garfinkel, S. (2011). "Extending Digital Repository Architectures to Support Disk Image Preservation and Access." In *JCDL '11*.
- Xie, S.L. (2011). "Building Foundations for Digital Records Forensics: A Comparative Study of the Concept of Reproduction in Digital Records Management and Digital Forensics." *American Archivist* 74(2), 576-599.