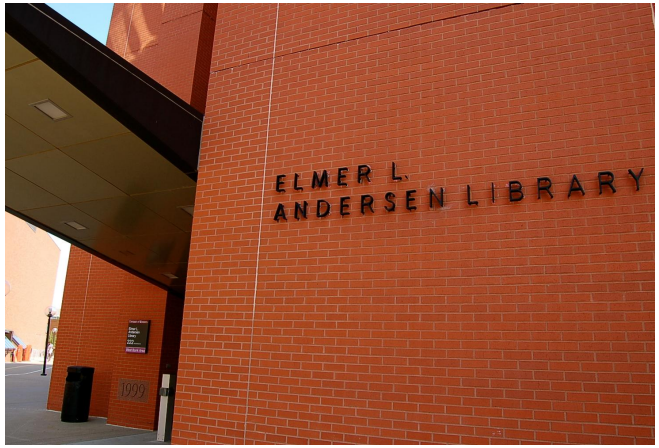
Fugitive Bits

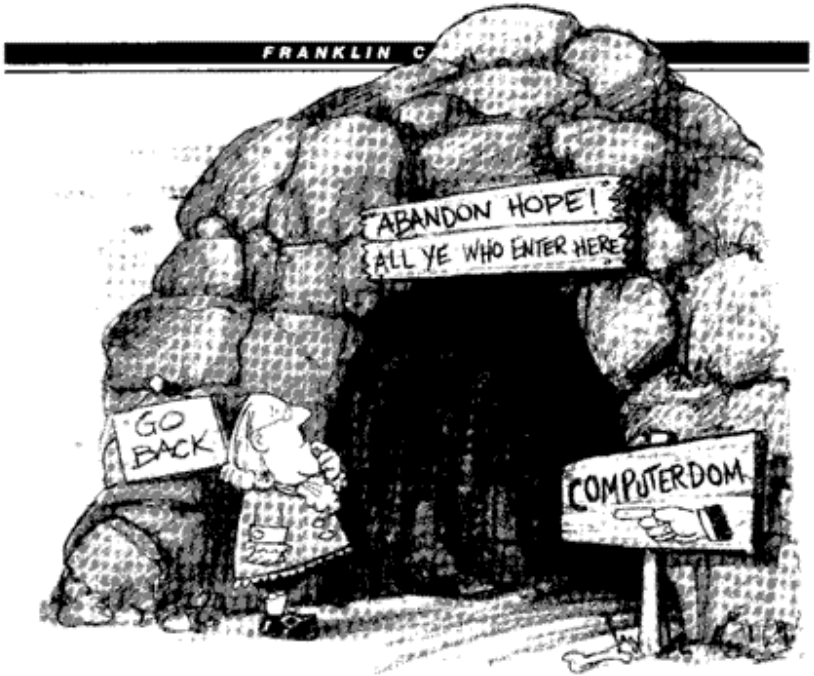
Taking Born Digital Records from Up in the Clouds Down to Earth

Upper Midwest Digital Collections (UMDC) Conference
19 August 2014
Lara Friedman-Shedlov and Arvid Nelsen

Archives & Special Collections at the University of Minnesota



Starting at Zero



http://www.ironicsans.com/images/ace09.gif

Jumping In: Leap of Faith!

- Assessment
- Investigation
- Training
- Discussions
- Strategizing
- Seeking Support

Demonstrate Need

A. Identify and Speak to People You Think Can Help and Provide resources
A. Identify and Speak to People You Think Can Help and Provide resources

Identify Resources

Digital Preservation Framework
University of Minnesota Libraries

UNIVERSITY OF MINNESOTA
LIBRARIES

1. Purpose

The Digital Preservation Framework formalizes the University of Minnesota Libraries' commitment to the long-term preservation of its diverse and extensive range of digital resources, thereby assuring enduring access to these resources. This document outlines the Libraries' approach to the preservation of digital resources and the associated information used to effectively manage these resources, especially for digital content that resides uniquely under the Libraries' stewardship.

The Libraries' digital resources are subject to the same overarching criteria for curation -- selection, management, and preservation -- as other resources in the Libraries' collections. Collection specialists, experts on the enduring value of the content, in consultation with preservation and information technology experts, make these decisions. Digital preservation decisions are made on the basis of this framework document, the Libraries' strategy and direction, the regard for the enduring value of specific digital resources, and the feasibility of providing the necessary preservation services. Ideally, decisions about the need for and level of preservation are made at the time of creation, acquisition, or licensing of digital resources. In general, the Libraries retains primary preservation responsibility for all digital resources directly under its stewardship.

Submitted July 2013
Reviewed by Cabinet and LLC, July 2013

Digital Repository Task Force Findings

Summary of Task Force Activities

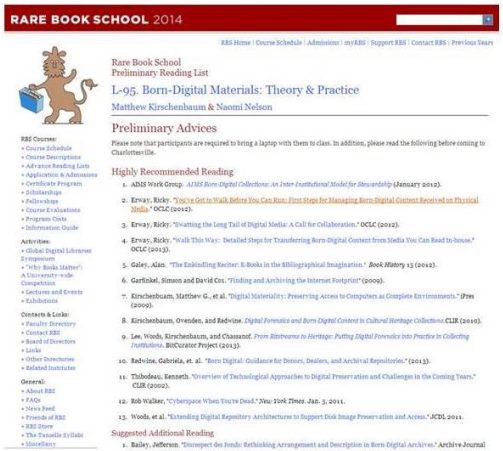
The Task Force was initially chartered and launched in September 2012 by John Butler and Karen Williams. Members included Erik Moore (chair), Chad Fennell, Cody Hanson, Stephen Hearn, Lisa Johnston, Julia Kelly, Arvid Nelsen, Jon Nichols, Jason Roy, and Bill Tantzen.

The Task Force met a total of thirteen times as a group and most members attended three separate webinars on repository topics. Meetings included presentations on digital repositories by the repository managers (AgEcon, UDC, UMedia), discussions of related repository issues (born-digital records, metadata, discoverability, preservation), review of architectural systems and their component parts (Fedora, Drupal, Solr, APIs), as well as a look at locally developed systems that have repository-like functionality (EthicShare, Ojibwe People's Dictionary). All members of the Task Force provided their expert opinions and added to robust, thought provoking discussions.

Start Where You Can

Training & Self Education

DAS
Digital Archives Specialist
Society of American Archivists


RARE BOOK SCHOOL 2014

RBS Home | Course Schedule | Admissions | myRBS | Support RBS | Contact RBS | Previous Years

Rare Book School
Preliminary Reading List

L-95. Born-Digital Materials: Theory & Practice

Matthew Kirschenbaum & Naomi Nelson

Preliminary Advices

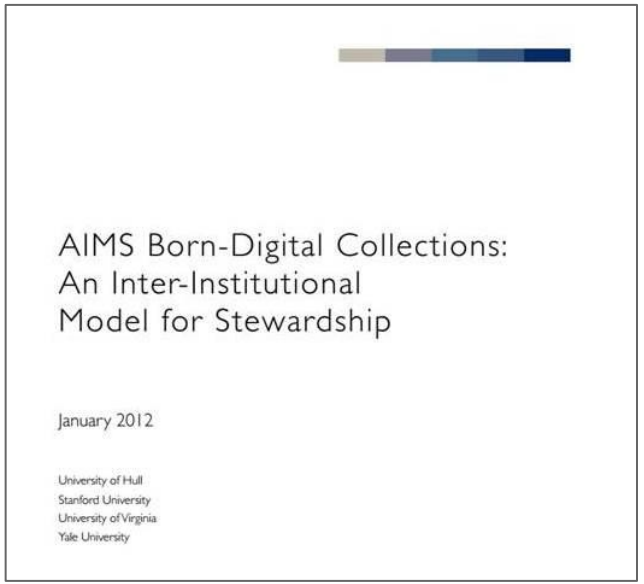
Please note that participants are required to bring a laptop with them to class. In addition, please read the following before coming to Charlottesville.

Highly Recommended Reading

1. ADHS Work Group. *AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship* (January 2012).
2. Erway, Ricky. "You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media." OCLC (2012).
3. Erway, Ricky. "Swatting the Long Tail of Digital Media: A Call for Collaboration." OCLC (2012).
4. Erway, Ricky. "Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read In-house." OCLC (2013).
5. Galey, Alan. "The Enkindling Reciter: E-Books in the Bibliographical Imagination." *Book History* 13 (2012).
6. Garfinkel, Simson and David Cox. "Finding and Archiving the Internet Footprint" (2009).
7. Kirschenbaum, Matthew G., et al. "Digital Materiality: Preserving Access to Computers as Complete Environments." iPres (2009).
8. Kirschenbaum, Ovenden, and Redwine. *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*. CLIR (2010).
9. Lee, Woods, Kirschenbaum, and Chassanof. *From Bitstreams to Heritage: Putting Digital Forensics into Practice in Collecting Institutions*. BitCurator Project (2013).
10. Redwine, Gabriela, et. al. "Born Digital: Guidance for Donors, Dealers, and Archival Repositories." (2013).
11. Thibodeau, Kenneth. "Overview of Technological Approaches to Digital Preservation and Challenges in the Coming Years." CLIR (2002).
12. Rob Walker, "Cyberspace When You're Dead." *New York Times*. Jan. 5, 2011.
13. Woods, et al. "Extending Digital Repository Architectures to Support Disk Image Preservation and Access." JCDL 2011.

Suggested Additional Reading

1. Bailey, Jefferson. "Disrespect des Fonds: Rethinking Arrangement and Description in Born-Digital Archives." Archive Journal


AIMS Born-Digital Collections:
An Inter-Institutional
Model for Stewardship

January 2012

University of Hull
Stanford University
University of Virginia
Yale University

Priority Setting

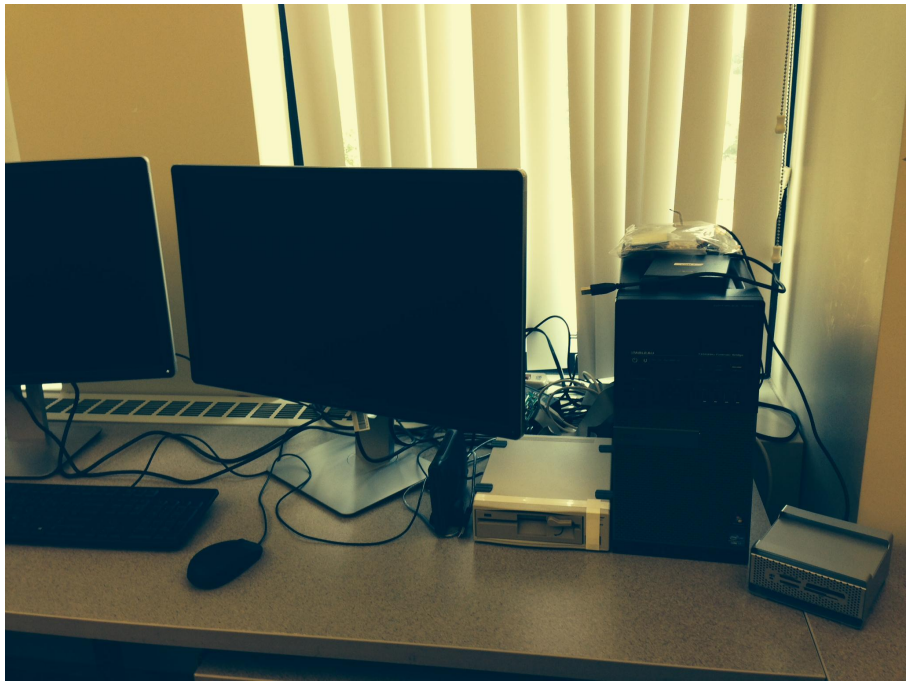
- ~~Legacy media~~
- New acquisitions
- ~~End User Access~~
- Ingest

- Where do we store it?
- How do we transfer it (software/equipment)
- What procedures and workflows do we use?

	MITH program at the University of Maryland OPTIMAL	MITH program at the University of Maryland MINIMAL	Penn State	Johns Hopkins	ASC Task Force Recommendations, If blank we are looking for discussion and input
System					
Processor					
RAM					
Storage					
Network Storage					
OTHER					
Memory Card Reader					
Optical drive					
Zip Disks					
3.5" floppy disks					
5.25" floppy disks					
Write Blockers					
Software					
Administrative Rights					
Network Connectivity					
COMMENTS					

Blank comparative evaluation chart used to formulate recommendation for a born-digital ingest workstation

U of M ASC Workstation



U of M ASC Workstation

Dell Optiplex 7010, minitower

- Processor: Intel Quad-core i7-3770 CPU @ 3.4-GHz
- Memory: 16 GB
- Storage:
  C: = 500 GB hybrid drive with 8 GB of SSD (operating system)
  D: = 4 TB hybrid drive with 8 GB of SSD (ingest, analysis, interim storage)
- 4 front facing USB ports (2 are USB 3.0)

Video: AMD Radeon HD 7570 / dual monitors / Dell 27-inch

Operating System:
Dual-boot = Windows 7 64-bit (primary) **&** Ubuntu / BitCurator (secondary)

Write-blockers:
Tableau Forensic USB Bridge in second bay of minitower

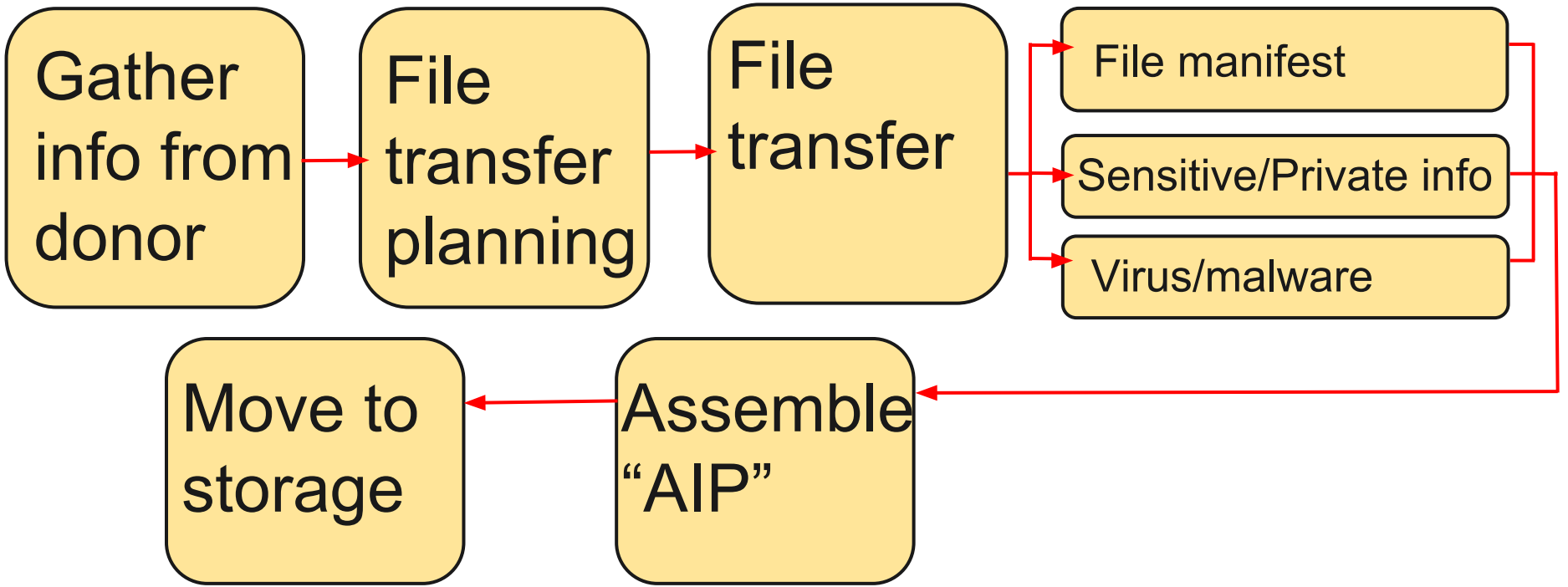
Additional Hardware:

- 3.5 inch USB floppy drive
- USB zip disk drive
- Internal CD/DVD drive

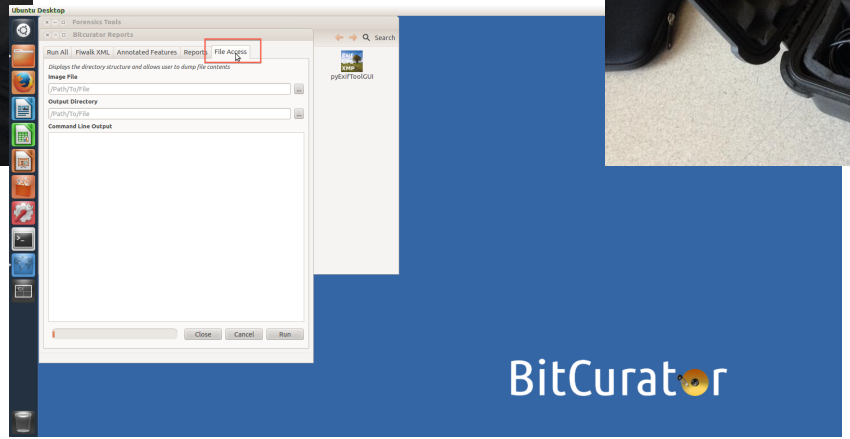
Donor Relations

- Copyright/intellectual property concerns
- Sensitive or private information?
- Deleted information
- Donor-imposed restrictions
- File formats
- Provenance & contextual info

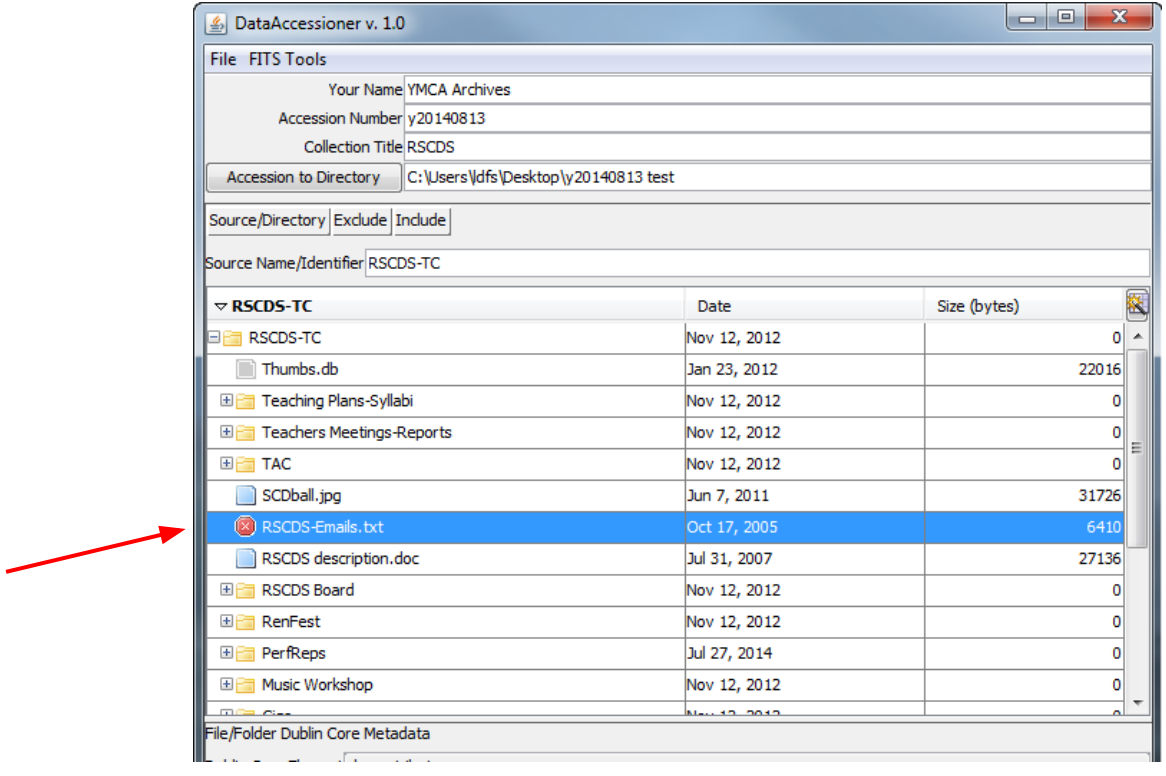
Accessioning Procedures



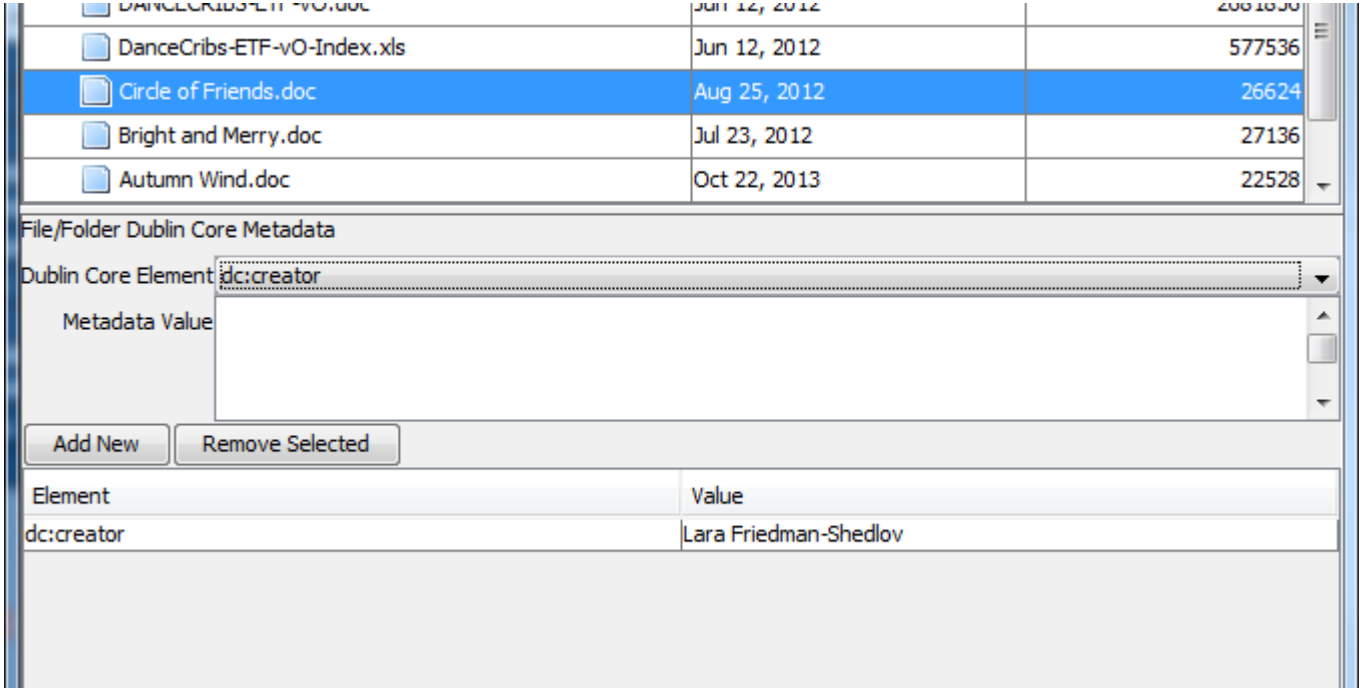
Tools: BitCurator



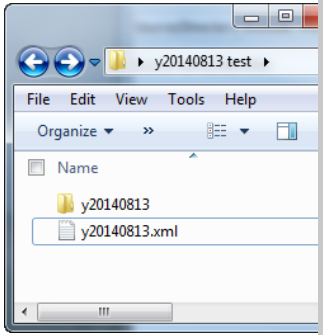
Tools: DataAccesioner



Tools: DataAccesioner

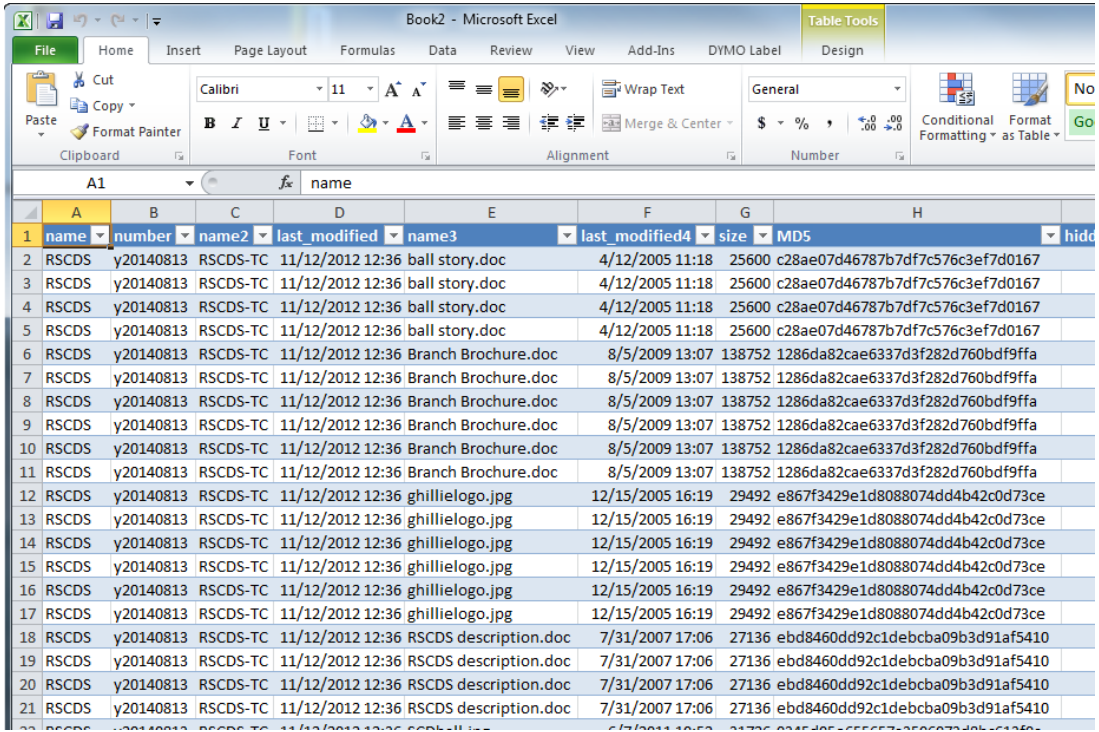


Tools: DataAccesioner



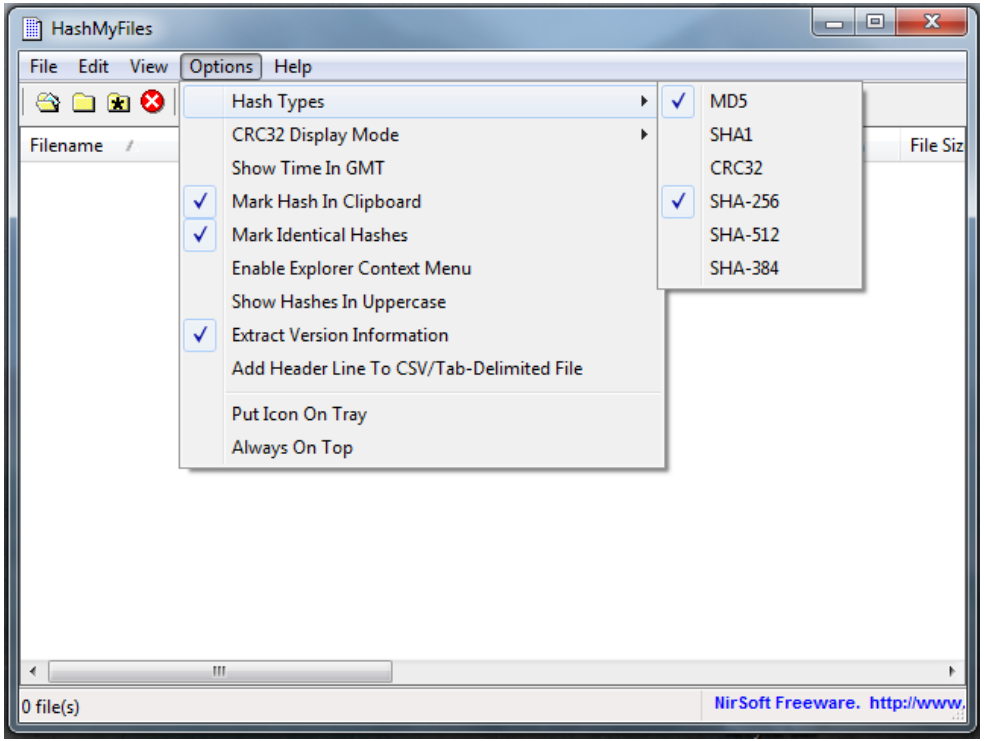

```
<?xml version="1.0" encoding="UTF-8"?>
<collection xmlns="http://dataaccessioner.org/schema/dda-0-3-1" name="RSCDS">
  <accession number="y20140813">
    <folder name="RSCDS-TC" last_modified="2012-11-12T12:36:10.000">
      <file name="ball story.doc" last_modified="2005-04-12T11:18:40.000" size="25600" MD5="c28ae07d46787b7df7c576c3ef7d0167">
        <premis:object xmlns:premis="info:lc/xmlns/premis-v2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:fits="ht
          <premis:objetIdentifier>
            <premis:objectIdentifierType>uuid</premis:objectIdentifierType>
            <premis:objectIdentifierValue>86b3b974-6f23-4ced-9129-2ce9d8ffe304</premis:objectIdentifierValue>
          </premis:objetIdentifier>
          <premis:objectCharacteristics>
            <premis:compositionLevel>0</premis:compositionLevel>
            <premis:fixity>
              <premis:messageDigestAlgorithm>MD5</premis:messageDigestAlgorithm>
              <premis:messageDigest>c28ae07d46787b7df7c576c3ef7d0167</premis:messageDigest>
              <premis:messageDigestOriginator>OIS File Information</premis:messageDigestOriginator>
            </premis:fixity>
            <premis:size>25600</premis:size>
            <premis:format>
              <premis:formatDesignation>
                <premis:formatName>Microsoft Word Document</premis:formatName>
              </premis:formatDesignation>
              <premis:formatNote>Identified by: Exiftool v9.13</premis:formatNote>
              <premis:formatNote>Identified by: ffident v0.2</premis:formatNote>
              <premis:formatNote>Identified by: Tika v1.3</premis:formatNote>
            </premis:format>
          </premis:objectCharacteristics>
          <premis:originalName>ball story.doc</premis:originalName>
        </premis:object>
      </file>
      <file name="Branch Brochure.doc" last_modified="2009-08-05T13:07:00.000" size="138752" MD5="1286da82cae6337d3f282d760bdf9ff
        <premis:object xmlns:premis="info:lc/xmlns/premis-v2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:fits="ht
          <premis:objetIdentifier>
```

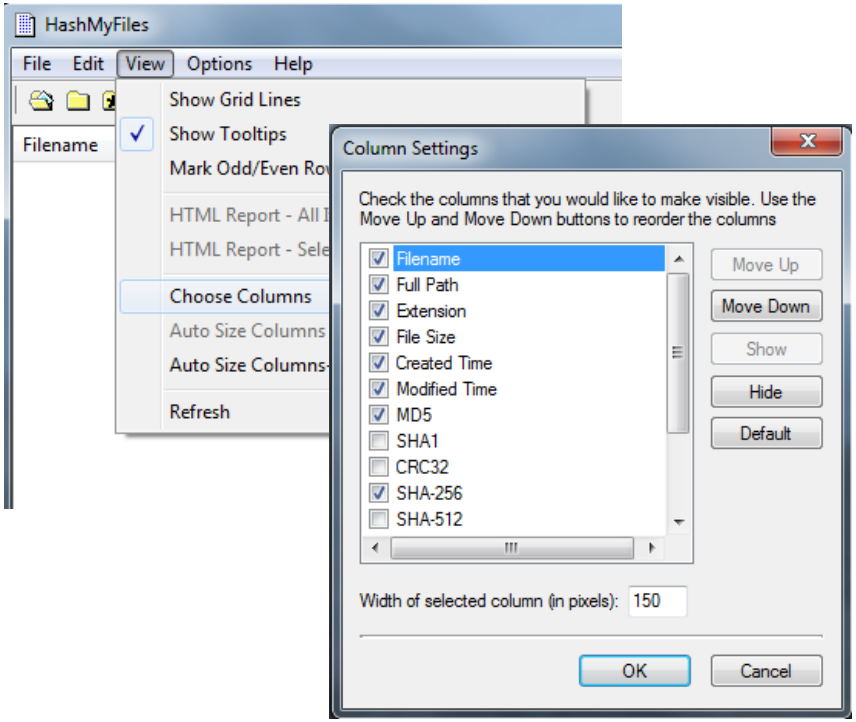
Tools: DataAccesioner



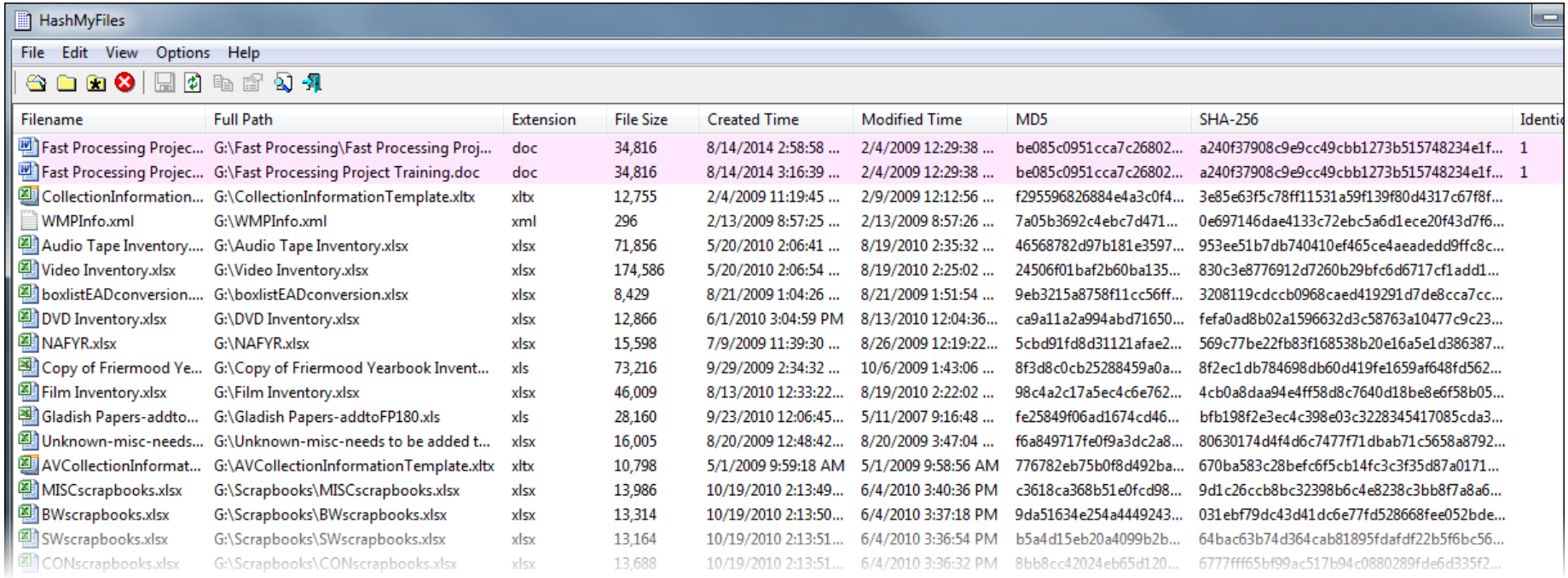
	A	B	C	D	E	F	G	H
1	name	number	name2	last_modified	name3	last_modified4	size	MD5
2	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	ball story.doc	4/12/2005 11:18	25600	c28ae07d46787b7df7c576c3ef7d0167
3	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	ball story.doc	4/12/2005 11:18	25600	c28ae07d46787b7df7c576c3ef7d0167
4	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	ball story.doc	4/12/2005 11:18	25600	c28ae07d46787b7df7c576c3ef7d0167
5	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	ball story.doc	4/12/2005 11:18	25600	c28ae07d46787b7df7c576c3ef7d0167
6	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	Branch Brochure.doc	8/5/2009 13:07	138752	1286da82cae6337d3f282d760bdf9ffa
7	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	Branch Brochure.doc	8/5/2009 13:07	138752	1286da82cae6337d3f282d760bdf9ffa
8	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	Branch Brochure.doc	8/5/2009 13:07	138752	1286da82cae6337d3f282d760bdf9ffa
9	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	Branch Brochure.doc	8/5/2009 13:07	138752	1286da82cae6337d3f282d760bdf9ffa
10	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	Branch Brochure.doc	8/5/2009 13:07	138752	1286da82cae6337d3f282d760bdf9ffa
11	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	Branch Brochure.doc	8/5/2009 13:07	138752	1286da82cae6337d3f282d760bdf9ffa
12	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	ghillielogo.jpg	12/15/2005 16:19	29492	e867f3429e1d8088074dd4b42c0d73ce
13	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	ghillielogo.jpg	12/15/2005 16:19	29492	e867f3429e1d8088074dd4b42c0d73ce
14	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	ghillielogo.jpg	12/15/2005 16:19	29492	e867f3429e1d8088074dd4b42c0d73ce
15	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	ghillielogo.jpg	12/15/2005 16:19	29492	e867f3429e1d8088074dd4b42c0d73ce
16	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	ghillielogo.jpg	12/15/2005 16:19	29492	e867f3429e1d8088074dd4b42c0d73ce
17	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	ghillielogo.jpg	12/15/2005 16:19	29492	e867f3429e1d8088074dd4b42c0d73ce
18	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	RSCDS description.doc	7/31/2007 17:06	27136	ebd8460dd92c1debcba09b3d91af5410
19	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	RSCDS description.doc	7/31/2007 17:06	27136	ebd8460dd92c1debcba09b3d91af5410
20	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	RSCDS description.doc	7/31/2007 17:06	27136	ebd8460dd92c1debcba09b3d91af5410
21	RSCDS	y20140813	RSCDS-TC	11/12/2012 12:36	RSCDS description.doc	7/31/2007 17:06	27136	ebd8460dd92c1debcba09b3d91af5410

Tools: HashMyFiles





Tools: HashMyFiles



Filename	Full Path	Extension	File Size	Created Time	Modified Time	MD5	SHA-256	Identic
Fast Processing Projec...	G:\Fast Processing\Fast Processing Proj...	doc	34,816	8/14/2014 2:58:58 ...	2/4/2009 12:29:38 ...	be085c0951cca7c26802...	a240f37908c9e9cc49cbb1273b515748234e1f...	1
Fast Processing Projec...	G:\Fast Processing Project Training.doc	doc	34,816	8/14/2014 3:16:39 ...	2/4/2009 12:29:38 ...	be085c0951cca7c26802...	a240f37908c9e9cc49cbb1273b515748234e1f...	1
CollectionInformation...	G:\CollectionInformationTemplate.xltx	xltx	12,755	2/4/2009 11:19:45 ...	2/9/2009 12:12:56 ...	f295596826884e4a3c0f4...	3e85e63f5c78ff11531a59f139f80d4317c67f8f...	
WMPInfo.xml	G:\WMPInfo.xml	xml	296	2/13/2009 8:57:25 ...	2/13/2009 8:57:26 ...	7a05b3692c4ebc7d471...	0e697146dae4133c72ebc5a6d1ece20f43d7f6...	
Audio Tape Inventory....	G:\Audio Tape Inventory.xlsx	xlsx	71,856	5/20/2010 2:06:41 ...	8/19/2010 2:35:32 ...	46568782d97b181e3597...	953ee51b7db740410ef465ce4aeadedd9ffc8c...	
Video Inventory.xlsx	G:\Video Inventory.xlsx	xlsx	174,586	5/20/2010 2:06:54 ...	8/19/2010 2:25:02 ...	24506f01baf2b60ba135...	830c3e8776912d7260b29bfc6d6717cf1add1...	
boxlistEADconversion....	G:\boxlistEADconversion.xlsx	xlsx	8,429	8/21/2009 1:04:26 ...	8/21/2009 1:51:54 ...	9eb3215a8758f11cc56ff...	3208119cdccb0968caed419291d7de8cca7cc...	
DVD Inventory.xlsx	G:\DVD Inventory.xlsx	xlsx	12,866	6/1/2010 3:04:59 PM	8/13/2010 12:04:36...	ca9a11a2a994abd71650...	fefa0ad8b02a1596632d3c58763a10477c9c23...	
NAFYR.xlsx	G:\NAFYR.xlsx	xlsx	15,598	7/9/2009 11:39:30 ...	8/26/2009 12:19:22...	5cbd91fd8d31121afae2...	569c77be22fb83f168538b20e16a5e1d386387...	
Copy of Friermood Ye...	G:\Copy of Friermood Yearbook Invent...	xls	73,216	9/29/2009 2:34:32 ...	10/6/2009 1:43:06 ...	8f3d8c0cb25288459a0a...	8f2ec1db784698db60d419fe1659af648fd562...	
Film Inventory.xlsx	G:\Film Inventory.xlsx	xlsx	46,009	8/13/2010 12:33:22...	8/19/2010 2:22:02 ...	98c4a2c17a5ec4c6e762...	4cb0a8daa94e4ff58d8c7640d18be8e6f58b05...	
Gladish Papers-addto...	G:\Gladish Papers-addtoFP180.xls	xls	28,160	9/23/2010 12:06:45...	5/11/2007 9:16:48 ...	fe25849f06ad1674cd46...	bfb198f2e3ec4c398e03c3228345417085cda3...	
Unknown-misc-needs...	G:\Unknown-misc-needs to be added t...	xlsx	16,005	8/20/2009 12:48:42...	8/20/2009 3:47:04 ...	f6a849717fe0f9a3dc2a8...	80630174d4f4d6c7477f71dbab71c5658a8792...	
AVCollectionInformat...	G:\AVCollectionInformationTemplate.xltx	xltx	10,798	5/1/2009 9:59:18 AM	5/1/2009 9:58:56 AM	776782eb75b0f8d492ba...	670ba583c28befc6f5cb14fc3c3f35d87a0171...	
MISCscrapbooks.xlsx	G:\Scrapbooks\MISCscrapbooks.xlsx	xlsx	13,986	10/19/2010 2:13:49...	6/4/2010 3:40:36 PM	c3618ca368b51e0fcd98...	9d1c26ccb8bc32398b6c4e8238c3bb8f7a8a6...	
BWscrapbooks.xlsx	G:\Scrapbooks\BWscrapbooks.xlsx	xlsx	13,314	10/19/2010 2:13:50...	6/4/2010 3:37:18 PM	9da51634e254a4449243...	031ebf79dc43d41dc6e77fd528668fee052bde...	
SWscrapbooks.xlsx	G:\Scrapbooks\SWscrapbooks.xlsx	xlsx	13,164	10/19/2010 2:13:51...	6/4/2010 3:36:54 PM	b5a4d15eb20a4099b2b...	64bac63b74d364cab81895fdafdf22b5f6bc56...	
CONscrapbooks.xlsx	G:\Scrapbooks\CONscrapbooks.xlsx	xlsx	13,688	10/19/2010 2:13:51...	6/4/2010 3:36:32 PM	8bb8cc42024eb65d120...	6777fff65bf99ac517b94c0880289fde6d335f2...	

Questions & Considerations

- Who is involved in each step/process?
- How critical are forensic processes?
- If disk image is captured, how long to keep initial bitstream?
- When and what file formats should be “normalized”

Summary Points

- Don’t be afraid
- Determine what you are contending with
- Find friends & allies
- Plan short term -- build capacity over time
- Get some training; build your vocabulary
- Expand your circle of friends & allies -- Reach out!
- Identify your resources -- administrators and IT

Thank You

U of M Libraries Electronic Records Task Force Web Page

z.umn.edu/libertf

Lara Friedman-Shedlov | ldfs@umn.edu | @ldfs

Arvid Nelsen | nels0307@umn.edu

UNIVERSITY OF MINNESOTA

LIBRARIES