Integrating digital forensics techniques into curatorial tasks: a case study

Sam Meister, University of Montana-Missoula
Alex Chassanoff, BitCurator Project Manager
Or,
A tale of tiny data.
Sort of.
Archives meet...digital forensics?

- Increased acquisition of born-digital materials
- Technical, administrative + intellectual challenges
  - Processing
  - Preservation
  - Access
Introduction to BitCurator

- Open source digital forensics environment for archivists

- Arose from recognized need for tools that
  - Integrate easily into digital curation workflows
  - Provide public access to data

- Year 3 of iterative software development
Primary Needs of a Practitioner
Get data off of media
Maintain authenticity and integrity
Generate preservation metadata
Analyze / investigate / understand data
To support future processing / description preservation / access / use
Case Study
Stuff 5/9/99
PG 10
pub invasion
Document media
Create disk image
Configure Hardware

Attach Write-Blocker

Create Disk Image

Verify Disk Image

Export files

Generate file hashes

Virus Check
Born Digital Workstation 2.0
Initial analysis
Extract filesystem metadata → Identify personal / private info → Generate reports
Extract filesystem metadata
<?xml version='1.0' encoding='UTF-8'?>
<dfxml version='1.0'>
  <metadata>
    <xmlns:='http://www.forensicswiki.org/wiki/Category:Digital_Forensics_XML'
    <xmlns:xsl='http://www.w3.org/2001/XMLSchema-instance'
    <xmlns:dc='http://purl.org/dc/elements/1.1/'>
      <dc:type>Disk Image</dc:type>
    </metadata>
  <creator version='1.0'>
    <program>fiwalk</program>
    <version>4.1.0</version>
    <build_environment>
      <compiler>GCC 4.6</compiler>
      <library name='afflib' version='3.7.1'/>
      <library name='libewf' version='20130416'/>
    </build_environment>
    <execution_environment>
    <start_time>2013-08-09T19:35:17Z</start_time>
    </execution_environment>
  </creator>
  <source>
    <image_filename>/media/sf_borndigital/2007_038/2007_038_001/disk_image/2007_038_001.aff</image_filename>
  </source>

<!-- fs start: 0 -->
  <volume_offset='0'>
    <partition_offset>0</partition_offset>
    <sector_size>512</sector_size>
    <block_size>512</block_size>
    <ftype>2</ftype>
    <ftype_str>fat12</ftype_str>
    <block_count>2880</block_count>
    <first_block>0</first_block>
    <last_block>2879</last_block>
    <fileobject>
      <parent_object>
        <inode>2</ inode>
      </parent_object>
      <filename>junk holder.wpd</filename>
      <partition>1</partition>
      <id>1</ id>
      <name_type>l</name_type>
      <filesize>2328</filesize>
      <alloc>1</alloc>
      <used>1</used>
    </fileobject>
  </volume_offset>
Identify private / personal / sensitive info
### 32 Files Found - Search for Files

**Contains the text:** harcourt

<table>
<thead>
<tr>
<th>Name</th>
<th>Folder</th>
<th>Size</th>
<th>Type</th>
<th>Date Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>MILKPR.rtf</td>
<td>files/2007_038_015/[root]</td>
<td>5.3 kB</td>
<td>RTF document</td>
<td>Wednesday, September 17 1997 at 4:26:04 AM</td>
</tr>
<tr>
<td>PGBI0.rtf</td>
<td>files/2007_038_015/[root]</td>
<td>11.9 kB</td>
<td>RTF document</td>
<td>Wednesday, September 17 1997 at 4:42:42 AM</td>
</tr>
<tr>
<td>MILKPR</td>
<td>files/2007_038_025/[root]</td>
<td>2.7 kB</td>
<td>unknown</td>
<td>Tuesday, December 17 1991 at 7:42:58 AM</td>
</tr>
<tr>
<td>MWPRIH95</td>
<td>files/2007_038_028/[root]</td>
<td>65.3 kB</td>
<td>unknown</td>
<td>Wednesday, October 11 1995 at 12:17:58 PM</td>
</tr>
<tr>
<td>PGBI0</td>
<td>files/2007_038_028/[root]</td>
<td>7.7 kB</td>
<td>unknown</td>
<td>Friday, July 28 1995 at 12:34:06 PM</td>
</tr>
</tbody>
</table>
Generate reports
Disk Image: 2007_038_001.aff File counts (by format)

Counts

- Co_Wp: 4
- Sys_e: 2
- emp_plx: 1
- DOS_tor: 1
# Report: File System Statistics and Files

**Technical Metadata**

*Found 1 Partition in this disk*

Disk Image: 2007_038_001'aff

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partition</td>
<td>1</td>
</tr>
<tr>
<td>PARTITION OFFSET</td>
<td>0</td>
</tr>
<tr>
<td>BLOCK COUNT</td>
<td>2880</td>
</tr>
<tr>
<td>FIRST BLOCK</td>
<td>0</td>
</tr>
<tr>
<td>LAST BLOCK</td>
<td>2879</td>
</tr>
<tr>
<td>BLOCK SIZE</td>
<td>2880</td>
</tr>
<tr>
<td>FTYPE</td>
<td>2</td>
</tr>
<tr>
<td>FTYPE STR</td>
<td>fat12</td>
</tr>
<tr>
<td>Number of Files</td>
<td>4</td>
</tr>
<tr>
<td>Total Directories</td>
<td>1</td>
</tr>
<tr>
<td>Total Deleted Files</td>
<td>0</td>
</tr>
<tr>
<td>Total Unused Files</td>
<td>0</td>
</tr>
<tr>
<td>Files with Nlinks &gt; 1</td>
<td>0</td>
</tr>
<tr>
<td>Empty Files</td>
<td>1</td>
</tr>
<tr>
<td>Big Files (&gt; 1 MB)</td>
<td>0</td>
</tr>
</tbody>
</table>
# Report: File System Statistics and Files

## File Format Table

Disk Image: 2007_038_001.aff

<table>
<thead>
<tr>
<th>Format</th>
<th>Short Form</th>
<th>Files</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOS floppy 1440k, x86 hard disk boot sector</td>
<td>DOS_tor</td>
<td>1</td>
</tr>
<tr>
<td>empty</td>
<td>emp_p ty</td>
<td>1</td>
</tr>
<tr>
<td>SysEx File - (Corel/WP)</td>
<td>Sys_e-</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>-Co_WP-</td>
<td>4</td>
</tr>
</tbody>
</table>
Conclusion / future
Future Work for BitCurator

- Metadata reporting on specific preservation events
- Phased community uptake through training and outreach activities
- Documentation and wiki expansion
Thanks!

sam.meister@mso.umt.edu
@samalanmeister

achass@email.unc.edu
@amchass