Archival Application of Digital Forensics Methods for Authenticity, Description and Access Provision

Cal Lee
School of Information and Library Science
University of North Carolina, Chapel Hill

International Council on Archives Congress
August 20-24, 2012
Brisbane, Australia
What is an archivist to do with things like this?

Source: Simson Garfinkel

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”
- Ideal recordkeeping system would provide “an unblemished line of responsible custody”

But digital is different...
“No computation without representation”

# Digital Resources - Levels of Representation

<table>
<thead>
<tr>
<th>Level</th>
<th>Label</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Aggregation of objects</td>
<td>Set of objects that form an aggregation that is meaningful encountered as an entity</td>
</tr>
<tr>
<td>7</td>
<td>Object or package</td>
<td>Object composed of multiple files, each of which could also be encountered as individual files</td>
</tr>
<tr>
<td>6</td>
<td>In-application rendering</td>
<td>As rendered and encountered within a specific application</td>
</tr>
<tr>
<td>5</td>
<td>File through filesystem</td>
<td>Files encountered as discrete set of items with associate paths and file names</td>
</tr>
<tr>
<td>4</td>
<td>File as “raw” bitstream</td>
<td>Bitstream encountered as a continuous series of binary values</td>
</tr>
<tr>
<td>3</td>
<td>Sub-file data structure</td>
<td>Discrete “chunk” of data that is part of a larger file</td>
</tr>
<tr>
<td>2</td>
<td>Bitstream through I/O equipment</td>
<td>Series of 1s and 0s as accessed from the storage media using input/output hardware and software (e.g. controllers, drivers, ports, connectors)</td>
</tr>
<tr>
<td>1</td>
<td>Raw signal stream through I/O equipment</td>
<td>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)</td>
</tr>
<tr>
<td>0</td>
<td>Bitstream on physical medium</td>
<td>Physical properties of the storage medium that are interpreted as bitstreams at Level 1</td>
</tr>
</tbody>
</table>
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

ContextMiner Alpha 3.0

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.

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

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

ContextMiner

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 Oulkami
Lent & Created by Giulani
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

<table>
<thead>
<tr>
<th>Crawl #</th>
<th>Crawl date</th>
<th>Color</th>
<th>Views</th>
<th>Ratings</th>
<th>Avg Rating</th>
<th>Comments</th>
<th>Links</th>
<th>Favorited</th>
<th>Honors</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2007-07-31</td>
<td>&lt;0.05</td>
<td>27942</td>
<td>301</td>
<td>3.74</td>
<td>298</td>
<td>$5</td>
<td>44</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2007-08-01</td>
<td>0.1</td>
<td>27742</td>
<td>303</td>
<td>3.73</td>
<td>299</td>
<td>$5</td>
<td>44</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2007-08-02</td>
<td>0.2</td>
<td>27760</td>
<td>307</td>
<td>3.72</td>
<td>291</td>
<td>$5</td>
<td>45</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>2007-08-03</td>
<td>0.3</td>
<td>27860</td>
<td>309</td>
<td>3.71</td>
<td>291</td>
<td>$5</td>
<td>45</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>2007-08-04</td>
<td>0.4</td>
<td>27960</td>
<td>310</td>
<td>3.71</td>
<td>291</td>
<td>$5</td>
<td>45</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2007-08-05</td>
<td>0.5</td>
<td>28060</td>
<td>314</td>
<td>3.59</td>
<td>294</td>
<td>$5</td>
<td>45</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>2007-08-06</td>
<td>0.6</td>
<td>28960</td>
<td>314</td>
<td>3.59</td>
<td>296</td>
<td>$5</td>
<td>45</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>2007-08-07</td>
<td>0.7</td>
<td>29265</td>
<td>318</td>
<td>3.58</td>
<td>298</td>
<td>$5</td>
<td>45</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>2007-08-08</td>
<td>0.8</td>
<td>29961</td>
<td>319</td>
<td>3.55</td>
<td>299</td>
<td>$5</td>
<td>46</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>2007-08-09</td>
<td>0.9</td>
<td>30064</td>
<td>320</td>
<td>3.54</td>
<td>300</td>
<td>$5</td>
<td>42</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>2007-08-10</td>
<td>1.0</td>
<td>30164</td>
<td>323</td>
<td>3.61</td>
<td>302</td>
<td>$5</td>
<td>45</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>2007-08-11</td>
<td>0.1</td>
<td>30410</td>
<td>324</td>
<td>3.62</td>
<td>302</td>
<td>$5</td>
<td>49</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>2007-08-12</td>
<td>0.2</td>
<td>30540</td>
<td>324</td>
<td>3.52</td>
<td>305</td>
<td>$5</td>
<td>49</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>2007-08-13</td>
<td>0.3</td>
<td>30697</td>
<td>326</td>
<td>3.61</td>
<td>306</td>
<td>$5</td>
<td>49</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>2007-08-14</td>
<td>0.4</td>
<td>30740</td>
<td>325</td>
<td>3.60</td>
<td>306</td>
<td>$5</td>
<td>49</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>2007-08-15</td>
<td>0.5</td>
<td>30123</td>
<td>326</td>
<td>3.61</td>
<td>306</td>
<td>$5</td>
<td>49</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>2007-08-16</td>
<td>0.6</td>
<td>31181</td>
<td>326</td>
<td>3.61</td>
<td>306</td>
<td>$5</td>
<td>49</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>2007-08-17</td>
<td>0.7</td>
<td>31512</td>
<td>327</td>
<td>3.61</td>
<td>307</td>
<td>$5</td>
<td>49</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>2007-08-18</td>
<td>0.8</td>
<td>31666</td>
<td>321</td>
<td>3.59</td>
<td>309</td>
<td>$5</td>
<td>51</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>2007-08-19</td>
<td>0.9</td>
<td>31744</td>
<td>322</td>
<td>3.59</td>
<td>309</td>
<td>$5</td>
<td>51</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>2007-08-20</td>
<td>1.0</td>
<td>31792</td>
<td>322</td>
<td>3.59</td>
<td>309</td>
<td>$5</td>
<td>51</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>2007-08-21</td>
<td>0.1</td>
<td>31937</td>
<td>333</td>
<td>3.57</td>
<td>310</td>
<td>$5</td>
<td>51</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>2007-08-22</td>
<td>0.2</td>
<td>32133</td>
<td>335</td>
<td>3.57</td>
<td>311</td>
<td>$5</td>
<td>52</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Level</td>
<td>Interaction Examples</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggregation of objects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Object or package</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-application rendering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>File through filesystem</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>File as “raw” bitstream</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sub-file data structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bitstream through I/O equipment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw signal stream through I/O equipment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bitstream on physical medium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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
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
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
## Interaction Examples

### Level

<table>
<thead>
<tr>
<th>Aggregation of objects</th>
<th>Object or package</th>
<th>In-application rendering</th>
</tr>
</thead>
<tbody>
<tr>
<td>File through filesystem</td>
<td>File as “raw” bitstream</td>
<td>Sub-file data structure</td>
</tr>
<tr>
<td>Bitstream through I/O equipment</td>
<td></td>
<td>Raw signal stream through I/O equipment</td>
</tr>
<tr>
<td>Bitstream on physical medium</td>
<td></td>
<td>Bitstream on physical medium</td>
</tr>
</tbody>
</table>

### Segment

<table>
<thead>
<tr>
<th>Segment</th>
<th>arg</th>
<th>length</th>
<th>data</th>
</tr>
</thead>
<tbody>
<tr>
<td>afflib_version</td>
<td>0</td>
<td>7</td>
<td>&quot;3.3.3&quot;</td>
</tr>
<tr>
<td>aff_file_type</td>
<td>0</td>
<td>3</td>
<td>AFF</td>
</tr>
<tr>
<td>acquisition_commandline</td>
<td>0</td>
<td>36</td>
<td>aimage /dev/sda/mnt/charlie-002</td>
</tr>
<tr>
<td>acquisition_device</td>
<td>0</td>
<td>8</td>
<td>/dev/sda</td>
</tr>
<tr>
<td>sectorsize</td>
<td>1024</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>pagesize</td>
<td>16777216</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>devicesectors</td>
<td>2</td>
<td>8</td>
<td>= 9999864 (64-bit value)</td>
</tr>
<tr>
<td>acquisition_macaddr</td>
<td>0</td>
<td>18</td>
<td>00:0b:db:4f:6b:10.</td>
</tr>
<tr>
<td>acquisition_dmesg</td>
<td>0</td>
<td>27298</td>
<td>[ 0.000000]</td>
</tr>
<tr>
<td>image_gid</td>
<td>0</td>
<td>16</td>
<td>7256 F895 DE4F E304</td>
</tr>
<tr>
<td>acquisition_date</td>
<td>0</td>
<td>20</td>
<td>2009-11-12 19:12:18.</td>
</tr>
<tr>
<td>md5</td>
<td>0</td>
<td>16</td>
<td>0609 2DFE AA4F B183</td>
</tr>
<tr>
<td>acquisition_seconds</td>
<td>1570</td>
<td>0</td>
<td>= 00:26:10 (hh:mm:ss)</td>
</tr>
<tr>
<td>imagesize</td>
<td>2</td>
<td>8</td>
<td>= 10239860736 (64-bit value)</td>
</tr>
</tbody>
</table>
Level
Aggregation of objects
Object or package
In-application
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

Interaction Examples
## Interaction Examples

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aggregation of objects</strong></td>
<td>In-application rendering</td>
</tr>
<tr>
<td><strong>Object or package</strong></td>
<td>File through filesystem</td>
</tr>
<tr>
<td><strong>In-application rendering</strong></td>
<td>File as “raw” bitstream</td>
</tr>
<tr>
<td><strong>File through filesystem</strong></td>
<td>Sub-file data structure</td>
</tr>
<tr>
<td><strong>File as “raw” bitstream</strong></td>
<td>Bitstream through I/O equipment</td>
</tr>
<tr>
<td><strong>Sub-file data structure</strong></td>
<td>Raw signal stream through I/O equipment</td>
</tr>
<tr>
<td><strong>Bitstream on physical medium</strong></td>
<td></td>
</tr>
</tbody>
</table>

---

Digital Forensics Principles 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
Previous Work*

• Some library/archives literature on recovering data from media
• Report by Ross and Gow (1999) on relevance of advances in data recovery and digital forensics to collecting institutions
• More recently, stream of literature related to use of forensic tools and methods for acquiring and managing digital collections
• Related projects:
  – Computer Forensics and Born-Digital Content in Cultural Heritage Collections
  – Born Digital Collections: An Inter-Institutional Model for Stewardship (AIMS)
  – Digital Records Forensics project

Renewed Energy and Attention around Personal Archives

• Arguably under-represented in first few decades of literature about electronic records
• Technical challenges of unruly acquisitions from individuals are channeling new energy into personal archives issues
1, Digital

Personal Collections in the Digital Era

Edited by Christopher A. Lee
What is Digital Forensics (aka Forensic Computing)?

• “The process of identifying, preserving, analysing and presenting digital evidence in a manner that is legally acceptable.”¹

• “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”²

2. Brad Glisson, Introduction to Computer Forensics & E-discovery, University of Glasgow, Week 1 Lecture, September 2008. (emphasis mine)
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.”
• “Minimise 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
Write Blocking – One-Way Streets for Data

File System – An Essential Layer for Metadata

• 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)
Getting below the File System – Low-Level Copying

- Getting an “image” of a storage medium involves working at a level below the file system
- Image is a copy of all of the storage sectors from the drive, rather than just copying the files
- Can get at file attributes and deleted files not visible through higher-level copy operations
Digital Forensics Tools – Hardware and Software
Digital forensics tools are designed primarily to be used in places like this:

El Paso County Sheriff’s Office (Colorado)
But they’re also be used in places like this:
Stanford University Libraries and Academic Information Resources (SULAIR)
British Library, London
BitCurator Project

- Funded by Andrew W. Mellon Foundation - October 1, 2011 – September 30, 2013
- Partners: SILS and Maryland Institute for Technology in the Humanities (MITH)
- Core Team:
  - Cal Lee, PI
  - Matt Kirschenbaum, Co-PI
  - Kam Woods, Technical Led
  - Alex Chassonoff, Project Manager (UNC), Sunitha Misra, GA (UNC), Porter Olsen, GA (MITH)

Professional Experts Panel
- Bradley Daigle, University of Virginia Library
- Erika Farr, Emory University
- Jeremy Leighton John, British Library
- Leslie Johnston, Library of Congress
- Courtney Mumma, Artefactual Systems
- 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
- Geoffrey Brown, Indiana University
- Barbara Guttman, National Institute of Standards and Technology
- Jerome McDonough, University of Illinois
- Mark Matienzo, Yale University
- David Pearson, National Library of Australia
- Doug Reside, New York Public Library
- Seth Shaw, University Archives, Duke University
- William Underwood, Georgia Tech
- Peter Van Garderen, Artefactual Systems
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
BitCurator Environment*

• Bundles, integrates and extends functionality of open source software: fiwalk, bulk extractor, Guymager, The Sleuth Kit, sdhash and others

• Can be run as:
  – Self-contained environment (based on Ubuntu Linux) running directly on a computer (download installation ISO)
  – 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/
Main Acquisition Interface for Guymager
Guymager Showing Technical Metadata about an SD Card (Right Click)
<fileobject>
  <filename>Documents and Settings/All Users/Documents/My Pictures/Sample Pictures/Blue hills.jpg</filename>
  ...
  <filesize>28521</filesize>
  <alloc>1</alloc>
  <used>1</used>
  <inode>6245</inode>
  ...
  <uid>0</uid>
  <gid>0</gid>
  <mtime>1208174400</mtime>
  <ctime>1257729636</ctime>
  <atime>1257729636</atime>
  <crt ime>1257729636</crt ime>
  <seq>2</seq>
  <libmagic>JPEG image data, JFIF standard 1.02</libmagic>
  <byte_runs>
    <run file_offset='0' fs_offset='0' img_offset='363200512' len='0'/>
  </byte_runs>
  <hashdigest type='MD5'>6fb2a38dc107eacb41cf1656e899cf70</hashdigest>
  <hashdigest type='SHA1'>4ee44b18576e84de7b163142b537d2fe6231845</hashdigest>
</fileobject>

*fiwalk created by Simson Garfinkel
Identifying “Features” of Interest in Disk Images

Bulk Extractor
(Created by Simson Garfinkel)
<table>
<thead>
<tr>
<th>Scanner</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>scan-accts</td>
<td>Looks for phone numbers, credit card numbers, etc</td>
</tr>
<tr>
<td>scan_base64</td>
<td>Decodes BASE64 text</td>
</tr>
<tr>
<td>scan_kml</td>
<td>Detects KML (Keyhole Markup Language) files – used to identify geographic locations</td>
</tr>
<tr>
<td>scan_gps</td>
<td>Detects XML from Garmin GPS devices</td>
</tr>
<tr>
<td>scan_aes</td>
<td>Detects in-memory AES (Advanced Encryption Standard) keys from the key schedules</td>
</tr>
<tr>
<td>scan_json</td>
<td>Detects JavaScript Object Notation files</td>
</tr>
<tr>
<td>scan_exif</td>
<td>Detects EXIF structures from JPEG files</td>
</tr>
<tr>
<td>scan_zip</td>
<td>Detects and decompresses ZIP files and zlib streams</td>
</tr>
<tr>
<td>scan_gzip</td>
<td>Detects and decompresses GZIP files and gzip streams</td>
</tr>
<tr>
<td>scan_pdf</td>
<td>Extracts text from some kinds of PDF files</td>
</tr>
<tr>
<td>scan_hiber</td>
<td>Detects and decompresses Windows hibernation file fragments</td>
</tr>
<tr>
<td>scan_winprefetch</td>
<td>Detects and extracts fields from windows prefetch files and file fragments</td>
</tr>
</tbody>
</table>
Email Addresses Identified - With Repeats Highlighted
Nautilus Scripts

• In addition to the specialized forensics tools in the BitCurator environment, there are a variety of scripts that can be run using the GNOME file manager called Nautilus (Linux analog to Windows Explorer or Mac OS X Finder)

• Can be used in the BitCurator environment or your own Linux environment
File Details for Word Documents in a Directory (Nautilus Script)
MD5 Hashes of Files (Nautilus Script)
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
Thank you!

http://bitcurator.net

http://wiki.bitcurator.net

Twitter: @bitcurator