

# Forensication of Everything\*

Preservation and Archiving Special Interest Group (PASIG)

April 4-5, 2011

London, UK

Cal Lee

University of North Carolina at Chapel Hill



The Andrew W. Mellon Foundation

\* Apologies to Harold Wilensky ["The Professionalization of Everyone?" *American Journal of Sociology* 70, no. 2 (1964): 137-58.]

Let's define our terms

# Digital Forensics (aka Forensic Computing):

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

\*Source: McKemmish, R. “What is Forensic Computing?” *Trends and Issues in Crime and Criminal Justice* 118 (1999).

Subtext: This is to catch and prosecute evil doers

# Forensication:

The incorporation of digital forensics methods, tools and concepts in contexts other than criminal investigations

# Forensics of Archives

- Recover data when layers of technology fail or are no longer available
- Capturing evidence from places that are not always immediately visible
- Ensuring that actions don't make irreversible changes to essential characteristics (e.g. MAC values)
- Attending to order of volatility
- Documenting what we do, so others will know what we might have changed
- Taking advantage of the information associated with files to ensure that users of the files understand their context of creation

Consider two streams of activity:

(1) Collecting institutions are getting removable media and need workflows for dealing with them

(2) The digital forensics field/industry provides training and tools, primarily focused on law enforcement, government intelligence and criminal investigation

Let's cross the streams



# Big Goals:

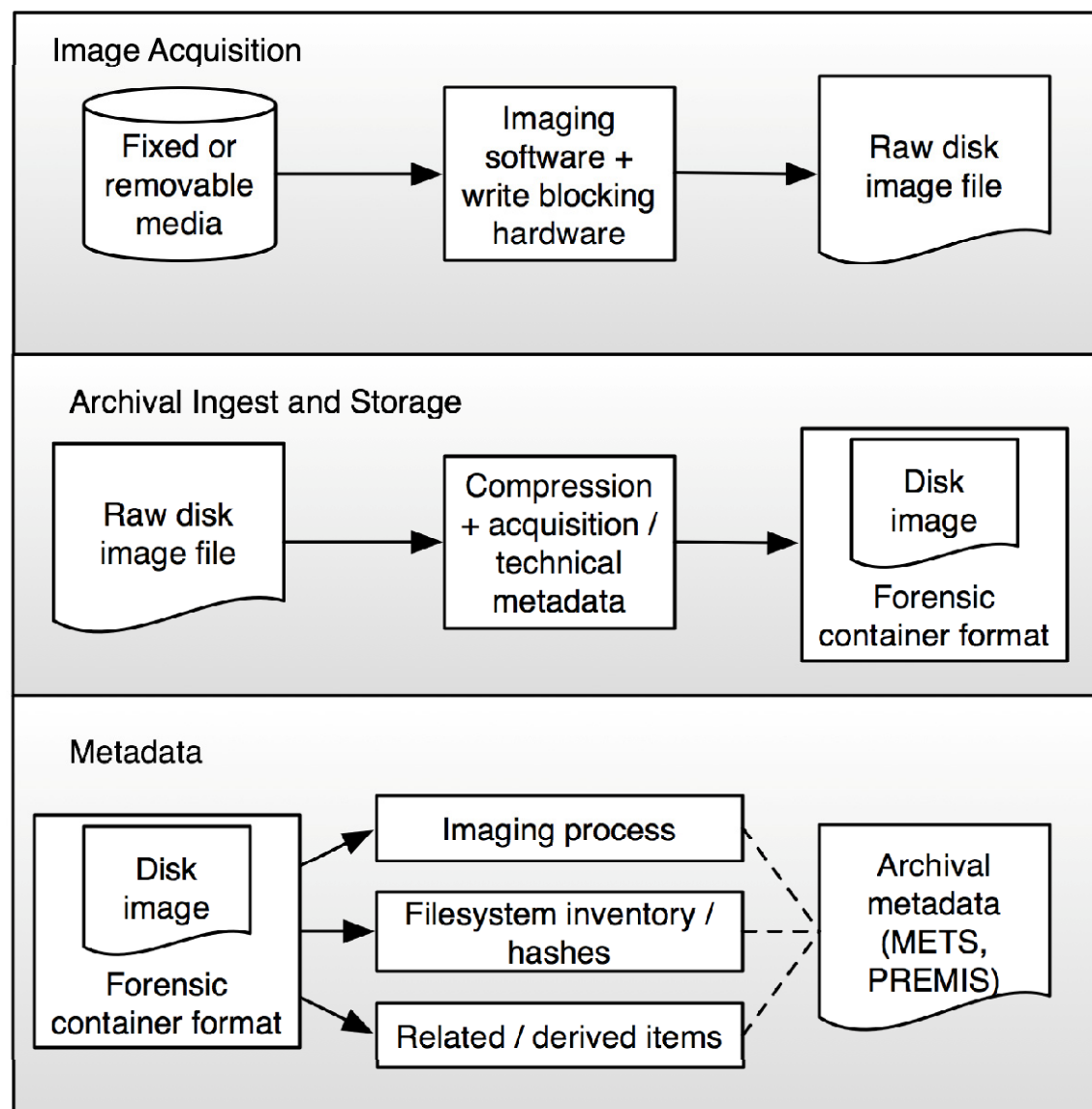
- Translation – established industry models, methods, techniques and resources into forms that are appropriate to information professionals
- Capacity building – training of the trainers, sustainable computer lab arrangements, incorporation into ongoing curriculum, collaboration with collecting-institution partners

# Big Questions and Issues:

- Which features of the available software are most useful for digital acquisition training?
- What do they need to know to make the right decisions about applying the tools appropriately in collecting contexts?
- How much and what types of adaptation are required to make the translation from the industry's target audience to the cultural heritage audience?
- Ethics of access: Can't be avoided, because users can exploit forensic methods, even if we don't

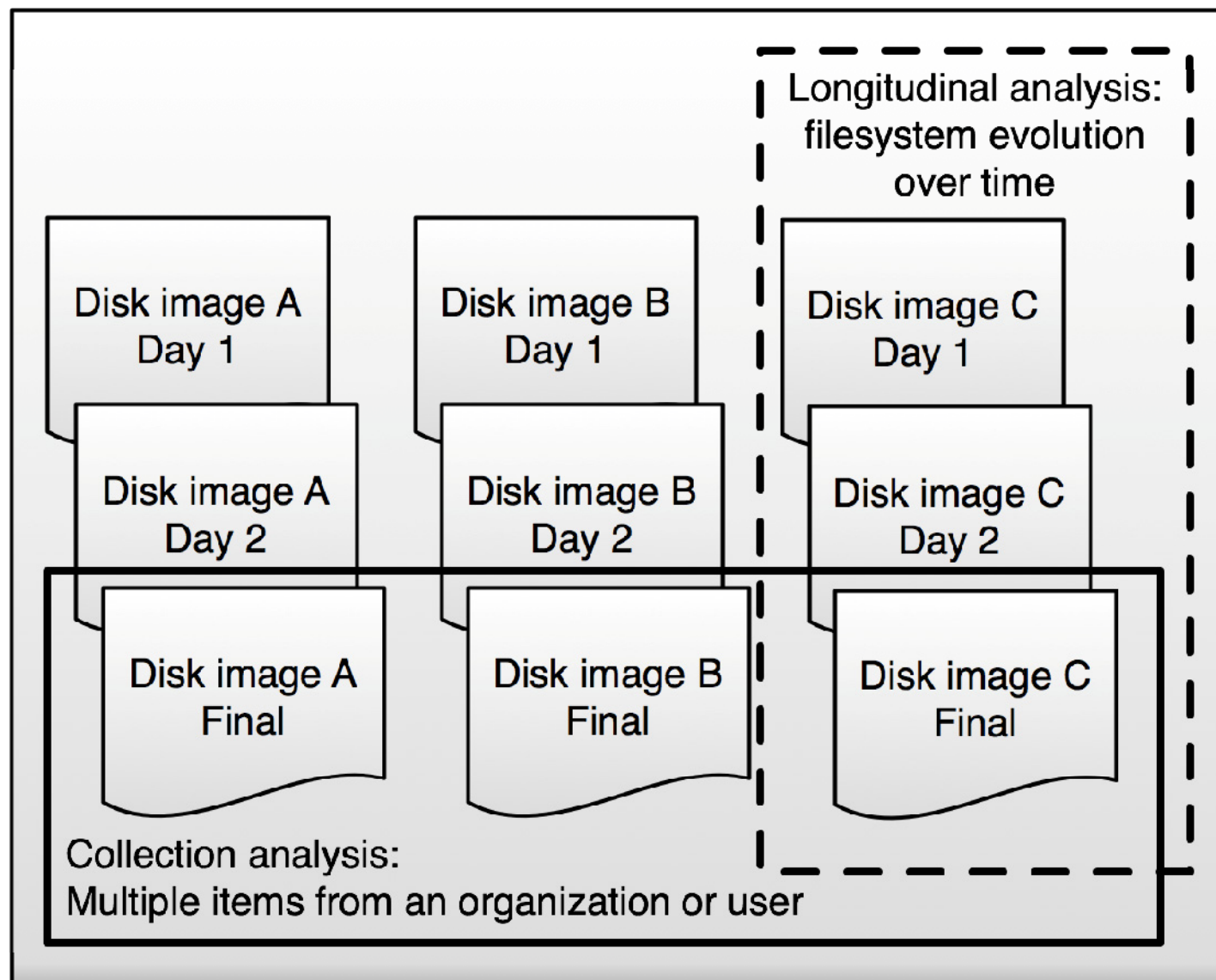
# Curation of Disk Images

# Storage Media Acquisition and Handling Profile for Digital Repositories\*



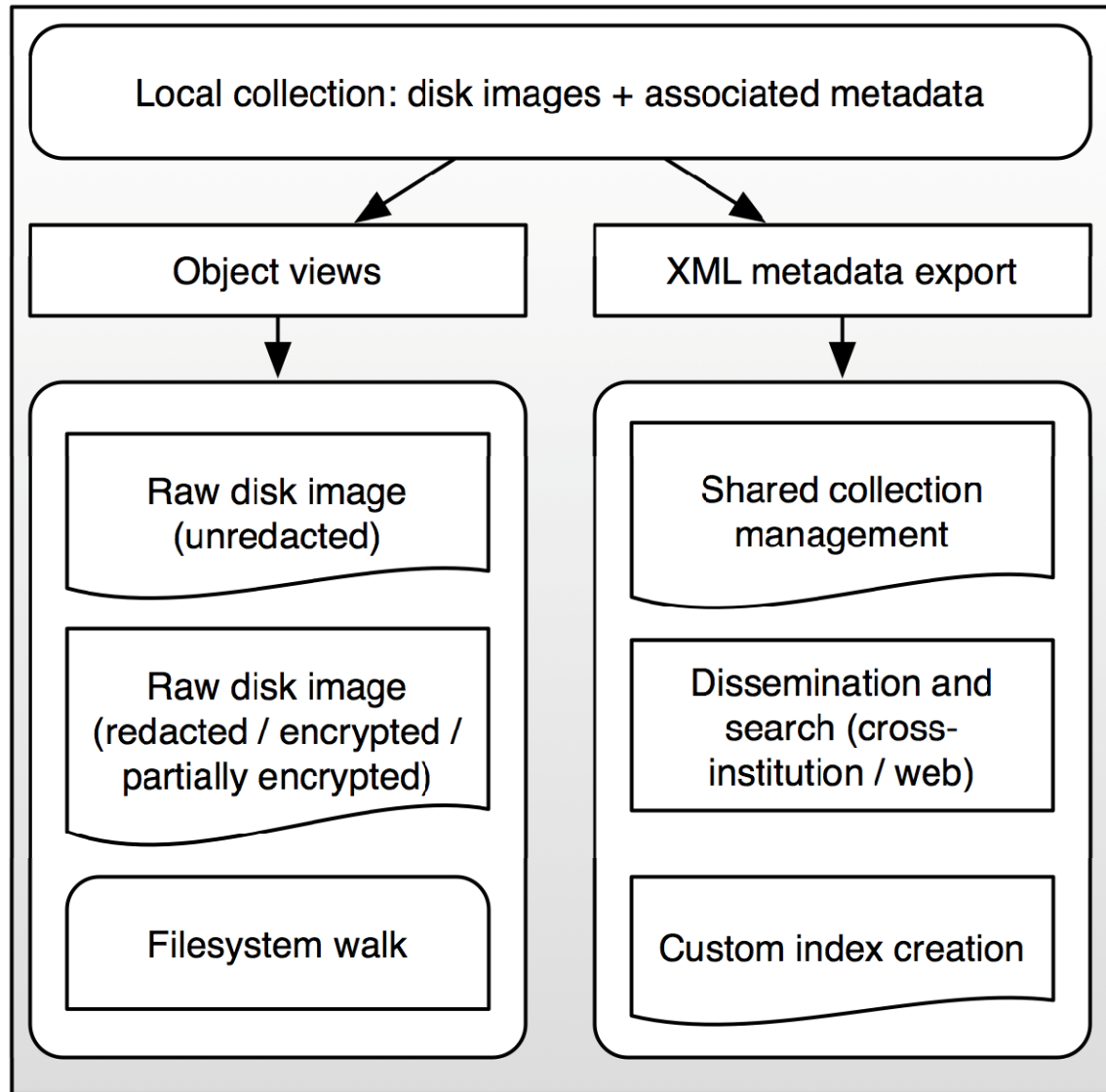
\*Woods, Kam, Christopher A. Lee, and Simson Garfinkel. "Extending Digital Repository Architectures to Support Disk Image Preservation and Access." In *Proceedings of the 2011 Joint Conference on Digital Libraries* (forthcoming).

## Modes of Disk Image Collection Analysis\*



\*Woods, Kam, Christopher A. Lee, and Simson Garfinkel. "Extending Digital Repository Architectures to Support Disk Image Preservation and Access." In *Proceedings of the 2011 Joint Conference on Digital Libraries* (forthcoming).

## Distribution Paths\*



\*Woods, Kam, Christopher A. Lee, and Simson Garfinkel. "Extending Digital Repository Architectures to Support Disk Image Preservation and Access." In *Proceedings of the 2011 Joint Conference on Digital Libraries* (forthcoming).

## Raw Image Metadata Captured by aimage

| Segment<br>=====        | arg<br>===== | length<br>===== | data<br>=====                              |
|-------------------------|--------------|-----------------|--------------------------------------------|
| afflib_version          | 0            | 7               | "3.3.3"                                    |
| aff_file_type           | 0            | 3               | AFF                                        |
| acquisition_commandline | 0            | 36              | aimage /dev/sda<br>/mnt/charlie-002        |
| acquisition_device      | 0            | 8               | /dev/sda                                   |
| sectorsize              | 1024         | 0               |                                            |
| pagesize                | 16777216     | 0               |                                            |
| devicesectors           | 2            | 8               | = 9999864 (64-bit value)                   |
| acquisition_macaddr     | 0            | 18              | 00:0b:db:4f:6b:10.                         |
| acquisition_dmesg       | 0            | 27298           | [ 0.000000]<br>Initializing cgro           |
| image_gid               | 0            | 16              | 7256 F895 DE4F E304<br>233E 21C0 2347 CCC5 |
| acquisition_date        | 0            | 20              | 2009-11-12 19:12:18.                       |
| md5                     | 0            | 16              | 0609 2DFE AA4F B183<br>946F 95D8 AD84 519E |
| acquisition_seconds     | 1570         | 0               | = 00:26:10 (hh:mm:ss)                      |
| imagesize               | 2            | 8               | = 10239860736 (64-bit value)               |

## Raw Image Metadata Captured by aimage

| Segment<br>=====        | arg<br>===== | length<br>===== | data<br>=====                              |
|-------------------------|--------------|-----------------|--------------------------------------------|
| afflib_version          | 0            | 7               | "3.3.3"                                    |
| aff_file_type           | 0            | 3               | AFF                                        |
| acquisition_commandline | 0            | 36              | aimage /dev/sda<br>/mnt/charlie-002        |
| acquisition_device      | 0            | 8               | /dev/sda                                   |
| sectorsize              | 1024         | 0               |                                            |
| pagesize                | 16777216     | 0               |                                            |
| devicesectors           | 2            | 8               | = 9999864 (64-bit value)                   |
| acquisition_macaddr     | 0            | 18              | 00:0b:db:4f:6b:10.                         |
| acquisition_dmesg       | 0            | 27298           | [ 0.000000]                                |
|                         |              |                 | Initializing cgr                           |
| image_gid               | 0            | 16              | 7256 F895 DE4F E304<br>233E 21C0 2347 CCC5 |
| acquisition_date        | 0            | 20              | 2009-11-12 19:12:18.                       |
| md5                     | 0            | 16              | 0609 2DFE AA4F B183<br>946F 95D8 AD84 519E |
| acquisition_seconds     | 1570         | 0               | = 00:26:10 (hh:mm:ss)                      |
| imagesize               | 2            | 8               | = 10239860736 (64-bit value)               |

## Raw Image Metadata Captured by aimage

| Segment<br>=====        | arg<br>===== | length<br>===== | data<br>=====                              |
|-------------------------|--------------|-----------------|--------------------------------------------|
| afflib_version          | 0            | 7               | "3.3.3"                                    |
| aff_file_type           | 0            | 3               | AFF                                        |
| acquisition_commandline | 0            | 36              | aimage /dev/sda<br>/mnt/charlie-002        |
| acquisition_device      | 0            | 8               | /dev/sda                                   |
| sectorsize              | 1024         | 0               |                                            |
| pagesize                | 16777216     | 0               |                                            |
| devicesectors           | 2            | 8               | = 9999864 (64-bit value)                   |
| acquisition_macaddr     | 0            | 18              | 00:0b:db:4f:6b:10.                         |
| acquisition_dmesg       | 0            | 27298           | [ 0.000000]<br>Initializing cgro           |
| image_gid               | 0            | 16              | 7256 F895 DE4F E304<br>233E 21C0 2347 CCC5 |
| acquisition_date        | 0            | 20              | 2009-11-12 19:12:18.                       |
| md5                     | 0            | 16              | 0609 2D5E AA4F B183<br>946F 95D8 AD84 519E |
| acquisition_seconds     | 1570         | 0               | = 00:26:10 (hh:mm:ss)                      |
| imagesize               | 2            | 8               | = 10239860736 (64-bit value)               |

## Raw Image Metadata Captured by aimage

| Segment<br>=====        | arg<br>===== | length<br>===== | data<br>=====                              |
|-------------------------|--------------|-----------------|--------------------------------------------|
| afflib_version          | 0            | 7               | "3.3.3"                                    |
| aff_file_type           | 0            | 3               | AFF                                        |
| acquisition_commandline | 0            | 36              | aimage /dev/sda<br>/mnt/charlie-002        |
| acquisition_device      | 0            | 8               | /dev/sda                                   |
| sectorsize              | 1024         | 0               |                                            |
| pagesize                | 16777216     | 0               |                                            |
| devicesectors           | 2            | 8               | = 9999864 (64-bit value)                   |
| acquisition_macaddr     | 0            | 18              | 00:0b:db:4f:6b:10.                         |
| acquisition_dmesg       | 0            | 27298           | [ 0.000000]<br>Initializing cgro           |
| image_gid               | 0            | 16              | 7256 F895 DE4F E304<br>233E 21C0 2347 CCC5 |
| acquisition_date        | 0            | 20              | 2009-11-12 19:12:10.                       |
| md5                     | 0            | 16              | 0609 2DFE AA4F B183<br>946F 95D8 AD84 519E |
| acquisition_seconds     | 1570         | 0               | = 00:26:10 (hh:mm:ss)                      |
| imagesize               | 2            | 8               | = 10239860736 (64-bit value)               |

# Output from fiwalk (XML)

```
<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>
  <ctime>1257729636</ctime>
  <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'>
    4eee44b18576e84de7b163142b537d2fe6231845
  </hashdigest>
</fileobject>
```

# Vision for the Future

- Widespread incorporation of the right bits and pieces of forensics methods into routine processing of acquisitions by collecting institutions
- BitCurator – a modular software environment that implements various batch processes on bitstreams to support two contexts:
  - Collecting institutions that are already generating forensic images of removable media
  - Individuals and institutions that are just getting started (BitCurator in a Box)

# A Word from Our Sponsors:

- National Science Foundation: Creating Realistic Forensic Corpora for Undergraduate Education and Research (DUE-0919593)
- Andrew W. Mellon Foundation: Digital Acquisition Learning Laboratory
- Institute for Museum and Library Services: DigCCurr II: Extending an International Digital Curation Curriculum to Doctoral Students and Practitioners (RE-05-08-0060-08) and Educating Stewards of Public Information in the 21st Century (ESOPI-21) (RE-05-09-0085-09)

Thank you!