Accessioning Born-Digital Materials

Northern California Fall Workshop
Society of California Archivists

Peter Chan, Digital Archivist
Nov. 8, 2012
Agenda

- Literature review
- Put accessioning in context
- Stanford work flow
- Hands-on
- Demonstration
- Questions
- Forensic Lab Tour
Oxford and Manchester


http://www.paradigm.ac.uk/workbook/index.html

• One section on “Accessioning digital and hybrid personal archives”

• Very comprehensive (except delivery)
Hull, Stanford, Yale, and UVa

AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship (2011)

http://www2.lib.virginia.edu/aims/whitepaper/

• High level
• Covering pre-accessioning, accessioning, processing and delivery
OCLC

You’ve Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media (2012)


• Very basic
• More to follow
Stanford


https://sites.google.com/site/workflowdocumentation/home

• Used in processing the born digital component of the STOP Aids Project Records

• Detail

• Still work-in-progress
Best Practices

- Harder to agree (compare to processing of paper archives) because the vast difference among institutions in software and hardware platforms
- Not enough people with necessary knowledge to contribute
- Still early stage of development (especially in processing and delivery)
Don’t Act Now

• Stanford spent $10K to recovery files from 4 hard drives received ~10 years ago.
• Only 47 percent of the recordable DVDs tested indicated an estimated life expectancy beyond 15 years. Some had a predicted life expectancy as short as 1.9 years. (note 1)
• Manufacturing of 5.25 inch floppy drive stopped for some years already.

Note 1: http://www.thexlab.com/faqs/opticalmedialongevity.html
Born-Digital Workflow – Textual Files

Pre-Accessioning
- Survey collection
- Analyze feasibility
- Enhanced curation
- Legal agreement
- Prepare for accessioning

Accessioning
- Physical control
- Create accession record – physical count
- Photograph media & labels
- Quarantine & run virus check
- Create disk image
- Generate summaries
- Create/update accession record
- Transfer data to secured storage

Processing
- Restricted files
- Extract technical metadata
- Assign descriptive metadata
- Assign rights metadata
- Create display derivative
- Create finding aids
- Create category record
- Transfer files with metadata and display derivative to delivery platform

Delivery & Access
- Users view collections (remote or in-person)
- Users register & agree to terms of use
- Users browse and search materials
- Users tag & annotate items
- User submit requests

Preservation
- Multiple copies
- Multiple locations – earthquake, flood, Tsunami
- Periodic integrity check
- Migration if necessary
- Emulation if appropriate
- Technology watch
**Accessioning**

### Activities
- “Physical” control
- Create accession record
- Photograph media & labels
- Quarantine & run virus check
- Create disk image
- Generate summaries
- Create/update accession record
- Transfer data to secured storage

### Tools / Persons
- Manual / SecureFX™ / Others
- Archivist’s Toolkit (AT)
- Canon EOS T1i with copy stand
- Sophos™
- FTK™ Imager
- AccessData FTK
- Archivists’ Toolkit (AT)
- SecureFX™

### Deliverables
- Control of media, files held locally
- Accession record – physical count
- Photographs of labels and media
- Virus free files
- Disk image, audit log, image log
- Summary report
- AT accession record – no. of files and size
- Files on secure network
Accession: MSS.2011-119

Date Expression

Date

Begin [ ] End [ ]

Bulk Dates

Begin [ ] End [ ]

Deaccession Date

Extent

Extent Measurement

Location

Note

Repository: SPECCOLL

Accession Notes

no listing available - therefore closed till processed.
no restrictions.

Created: Aug 24, 2011 by ohanluin | Modified: Mar 30, 2012 by pchan3 | Record Number: 32571
AT Accession Records

• Physical
  – Media count
  – Computer count
Multiple Extent Plugin

Submitted by michael_vandermillen on Thu, 08/11/2011 - 08:29

Independent of collection size, archivists count and report on a variety of container and media types as they manage, store, and seek to understand the nature of their collections and meet preservation and access challenges.

In answer to this need, Harvard developed a two-fold approach in the Archivists’ Toolkit. First, we re-labeled the AT default extent as "Primary extent." Second, we developed a plug-in* to allow multiple alternate statements of extent and labeled this "Alternate extent." The alternate extent statements have the same data elements as the default AT extent statements, but the alternate extents have a separate drop-down list of extent types. Alternate extents are not required, but multiple alternate extents may be applied to an accession or resource record.

This plugin is an extension of BYU's Multiple Date and Physical Description custom panels plugin.

Attachment	Size
Harvard_Multiple_extent_plugin.pdf 239.88 KB

Author: Harvard University Library
Zip File:
multipleExtent.zip
Source File:
multipleExtentSrc.zip
Description:
Plugin to allow primary and alternate extent entry for accession and resource records (based on the BYU Multiple Date and Phys. Desc. plugin)
<table>
<thead>
<tr>
<th>List Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition type</td>
</tr>
<tr>
<td>Actuate</td>
</tr>
<tr>
<td>Alternate Extent type</td>
</tr>
<tr>
<td>Calendar</td>
</tr>
<tr>
<td>Container types</td>
</tr>
<tr>
<td>Date type</td>
</tr>
<tr>
<td>Description rules</td>
</tr>
<tr>
<td>Digital object types</td>
</tr>
<tr>
<td>Era</td>
</tr>
<tr>
<td>Extent type</td>
</tr>
<tr>
<td>File use attributes</td>
</tr>
<tr>
<td>Finding aid status</td>
</tr>
<tr>
<td>Index Value Types</td>
</tr>
<tr>
<td>Instance types</td>
</tr>
<tr>
<td>Language codes</td>
</tr>
<tr>
<td>List Numeration Types</td>
</tr>
<tr>
<td>Name Description Type</td>
</tr>
<tr>
<td>Name link creator / subject role</td>
</tr>
<tr>
<td>Name link form</td>
</tr>
<tr>
<td>Name link function</td>
</tr>
<tr>
<td>Name link source role</td>
</tr>
<tr>
<td>Name rules</td>
</tr>
<tr>
<td>Name source</td>
</tr>
<tr>
<td>Name type</td>
</tr>
<tr>
<td>Processing Priorities</td>
</tr>
</tbody>
</table>
## Administration Lookup Lists

**List Name:** Alternate Extent type

- **Paired Values**
- **Restrict to NMOKEM**

<table>
<thead>
<tr>
<th>List Item</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>album(s)</td>
<td></td>
</tr>
<tr>
<td>architectural drawing(s)</td>
<td></td>
</tr>
<tr>
<td>audicassette(s)</td>
<td></td>
</tr>
<tr>
<td>audicassette(s) (DAT)</td>
<td></td>
</tr>
<tr>
<td>audicassette(s) (microcassette)</td>
<td></td>
</tr>
<tr>
<td>audiotape(s) (reel-to-reel)</td>
<td></td>
</tr>
<tr>
<td>boxes</td>
<td></td>
</tr>
<tr>
<td>computer cartridge(s)</td>
<td></td>
</tr>
<tr>
<td>computer file(s)</td>
<td></td>
</tr>
<tr>
<td>computer media</td>
<td></td>
</tr>
<tr>
<td>computer tape(s)</td>
<td></td>
</tr>
<tr>
<td>computer, desktop</td>
<td></td>
</tr>
<tr>
<td>computer, portable</td>
<td></td>
</tr>
<tr>
<td>cubic feet</td>
<td></td>
</tr>
<tr>
<td>film reel(s) (16mm)</td>
<td></td>
</tr>
<tr>
<td>film reel(s) (8mm)</td>
<td></td>
</tr>
<tr>
<td>floppy disk(s) (3.5 inch)</td>
<td></td>
</tr>
<tr>
<td>floppy disk(s) (5.25 inch)</td>
<td></td>
</tr>
<tr>
<td>floppy disk(s) (8 inch)</td>
<td></td>
</tr>
<tr>
<td>folder(s)</td>
<td></td>
</tr>
<tr>
<td>folio(s)</td>
<td></td>
</tr>
<tr>
<td>gigabyte(s)</td>
<td></td>
</tr>
<tr>
<td>hard drive(s) (external)</td>
<td></td>
</tr>
</tbody>
</table>

- **Items in Red can't be modified**
- **Items in Blue are AT initial values**

**Created:** Sep 12, 2011 by dhartwig | **Modified:** Oct 19, 2012 by dhartwig | **Record Number:** 67
Media Count

• Media count by
  – 3, 3.5, 5.25, 8 inch. floppy diskettes
  – Zip disk
  – Open reel, cartridge tape
  – CD, DVD, Optical disk
  – External hard drive

• Computer
  – Desktop / Portable
  – Mac / PC / Others
8-inch, 5.25-inch, and 3.5-inch floppy
Smith Corona DataDisk 3-inch floppy
100MB Zip Disc for Iomega Zip
Open Reel Tape
Cartridge Tape
Punch Cards
Apple II
Commodore C64 SX-64
Kaypro 10 portable computer
Media Label /Rehouse

• Label
  – Use “Call No._CMxxx” as label name, 0.5 x 1.875 inch. label

• Re-House
  – Follow the same box no. naming convention as other materials (paper, av, etc.).
Media Sleeve
Rehouse Media
# Floppy Disc Box

<table>
<thead>
<tr>
<th>Cat #</th>
<th>Description</th>
<th>Size</th>
<th>3 or more</th>
<th>10 or more</th>
<th>20 or more</th>
<th>Qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDBB</td>
<td>CD/DVD Box</td>
<td>5 x 4 7/8 x 5 5/8&quot;</td>
<td>$5.40</td>
<td>$5.05</td>
<td>$4.85</td>
<td></td>
</tr>
<tr>
<td>CDBT</td>
<td>CD/DVD Box</td>
<td>5 x 4 7/8 x 5 5/8&quot;</td>
<td>$4.65</td>
<td>$4.40</td>
<td>$4.20</td>
<td></td>
</tr>
<tr>
<td>FDB442</td>
<td>Floppy Disc Box</td>
<td>3 1/2&quot;</td>
<td>$3.60</td>
<td>$3.35</td>
<td>$3.20</td>
<td></td>
</tr>
<tr>
<td>FDB552</td>
<td>Floppy Disc Box</td>
<td>5 1/2&quot;</td>
<td>$4.05</td>
<td>$3.85</td>
<td>$3.70</td>
<td></td>
</tr>
</tbody>
</table>

[Buy](hollingermetaledge.com)
hollingermetaledge.com
Live View Shooting
Filename

• Photographing media
  – Use ““Call No._CMxxx “ as filename
  – If more than 1 photo is taken, add _1 for first and _2 for second photo, etc. (e.g. front, back, box, etc.)
  – Use computer to control the camera if you have more than 20 media to photo; otherwise, just use stand alone camera.
  – Store all photos in “Media Photo” folder
Preference

**File Name**

Prefix + Number

File Prefix
M0662_CM

Assign Sequence No.
- Number of Digits: 3
- Start: 1

Example:
M0662_CM_0001.xxx

*Note:* file extension will be the same as the original file name
Media Photo
Media Photo

Greeley Files
(email)
"Greeley" John
2/25/03
AT Accession Records

- Provide finer information than physical storage media count
  - Size in MB/TB, etc.
  - No. of files
  - Link to
    - Image log spreadsheet
    - Collection summary from FTK
Virus Check

• Quarantine for 30 days (count from the day the media arrive at Stanford)
• Run Sophos
• Remove virus, if any, before creating disk/logical image (unless you are capturing the disk for a researcher in computer virus!)
Forensic vs. Logical Image vs. File Copy

• Do you want to copy deleted files?
• Are you sure you didn’t change file dates (creation, modified, last accessed) when copying the files?
• Are you sure you copied all files (files in different partitions)?
• Are you sure you copied all required files (fonts for design files, etc.)
• Are you sure the source and copied files are same?
Disk Partition

• Disk partitioning is the act of dividing a physical hard disk drive into multiple logical storage units.

• The logical units can have different file systems so that we can set the same physical machine as a Window based PC and/or a Linux based PC and/or a Mac PC.
Computer Management
GParted

GParted window displaying partition details for `/dev/sda`. The partitions include:

- `/dev/sda1` (ntfs) mounted as `/media/disk`, size 101.56 GB, used 96.17 GiB, unused 5.39 GiB, boot
- `/dev/sda2` (linux-swap)
- `/dev/sda3` (ext3) mounted as `/`, size 8.82 GB, used 4.76 GiB, unused 4.06 GiB

0 operations pending
File System

• Microsoft: FAT12 (File Allocation Table), FAT16, FAT32, exFAT (FAT64), NTFS (New Technology File System)

• Apple: Apple ProDOS, HFS (Hierarchical File System), HFS+

• CPM file system

• Linux: ext (extended filesystem), ext2, ext3, ext4

• Optical Discs: ISO 9660, UDF (Universal Disk Format)
<table>
<thead>
<tr>
<th>file name</th>
<th>file ext</th>
<th>file attr</th>
<th>reserv</th>
<th>create time</th>
<th>create date</th>
<th>1st cluster</th>
<th>file length</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEST</td>
<td>DAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

- **FAT**: 10, 9, 8, 7, 6, 6
- **Disk**: CHS (Connects to BIOS)
- **LBN**: LBN (Connects to BIOS)
- **BIOS**: Connects to Disk

The diagram illustrates the relationship between the file system components (file name, file extension, file attributes, etc.) and the storage organization (FAT, LBN, Disk).
Disk / Forensic image

• A complete bit-by-bit copy of a storage medium or device, such as a hard drive, SSD (solid state drive), tape drive, floppy disk, CD/DVD/BD, or flash memory device.

• The image can be stored in one or more files.

• Deleted files, if any, will be copied in this process.

• All partitions will be copied.
Logical image

• A copy of the files in the directory(folder) / directories(folders) specified in the copy process.

• The full path of each file is recorded and the files are embedded in one or more files in AD1 format.

• Since deleted files and unpartitioned space are not represented in a directory, they are not copied in the process.

• The host operating system has to recognize the target directory (folder) for the operation. (You cannot perform a logical copy of an Apple disk using FTK Imager under Windows perform).
Capture 5.25 Floppy Diskette
Compare 4 5.25 Floppy Drive Solutions

• https://docs.google.com/document/d/1TLY3mn3duadGBLqqb2_XnYoF3jakjmo1hXfzvmWCtZE/edit?hl=en_US
Hard Disk Interface

- SCSI
- IDE
- Firewire 1394
- SATA
- USB
Internal Write Blocker
Image Log Spreadsheet

• Use the following columns
  – Box#
  – CM#
  – Media Details (3.5, 5.25, single/double sided, single/double/high density, etc.)
  – Image Result (successful / unreadable)
  – Note (bad sector, etc.)
  – Manufacturer
  – Earliest modification/creation year of files
  – Folder title (if removed from collection folder)

• Use “Call No. Imaging Log” as file name
FTK Imager
Create Disk Image

• Physical Drive – hard / flash memory drive
• Logical Drive – floppy / CD / DVD drive
• Contents of a Folder – logical image
• Fernico Device – a device which backs up forensic data from network locations or from locally attached hard drives, automatically spanning the content over a series of discs.
Fernico Device
Tick “Verify images after they are created”
Tick “Create directory listings of all files in the image after they are created” if you need one.
Image Type

Select Image Type

Please Select the Destination Image Type

- Raw (dd)
- SMART
- E01
- AFF

< Back  Next >  Cancel  Help
Image Type

- Raw (dd) – commonly used disk image format created by the UNIX command dd
- SMART – proprietary disk image format created by ASR Data.
- E01 – proprietary disk image format created by EnCase.
- AFF – an open source disk image format, allow encryption of disk image
- AD1 - proprietary logical disk image format created by AccessData.
Evidence Item Information
Evidence Item Information

• Use “Raw (dd)” as image type
• Use call no. as “Case Number”
• Use CMxxx as “Evidence Number”
• Put your name as “Examiner”
• Use the following in the “Notes” field:
  – 3.5 inch floppy disk; 5.25 inch floppy disk; Zip disk;
  – External hard disk; Internal hard disk;
  – Optical Disk
Select Image Destination

![Select Image Destination dialog box](image)

- **Image Destination Folder**: C:\Images from Imager 2.9\AD1 Image
- **Image Filename (Excluding Extension)**: Mantooth1
- **Image Fragment Size (MB)**: 1500
- **Compression (0=None, 1=Fastest, ..., 9=Smallest)**: 6
- **Use AD Encryption**: On
Select Image Destination

- Store all files under “Disk Image” folder in Desktop
- Use “Case No._CMxxx” as Image Filename for disk image
- Default Image Fragment Size = 1500 MB
- To save images segments that can be burned to a CD, specify 650 MB.
- To save image segments that can be burned to a DVD, specify 4000 MB.
- Selecting 0 (zero) produces the largest file, with no compression.
Storage Locations for Files
Title: Rebecca Solnit papers - captured data from portable computers

Accession Number: MSS 2012-056
Accession Date: 2012-4-2

Date Expression
- Date
- Begin: [ ]
- End: [ ]

Bulk Dates
- Begin: [ ]
- End: [ ]

Deaccessions
- Deaccession Date
- Extent
- Extent Measurement

Locations
- Location
- Note

Physical Description
- Primary Extent
- Alternate Extent
- Extent Number: 68.54 gigabyte(s)
- Container Summary:
  - megabytes: 72,570
  - number of files: 732,812
  - captured from: 3 portable computers

Repository SPECCOLL

Accession Notes
- related ACCN - MSS.2011-119.
  Closed until processed.

Created: Apr 2, 2012 by pchan3 | Modified: Jul 5, 2012 by pchan3 | Record Number: 33677
AT Accession Record
External Documents

<table>
<thead>
<tr>
<th>Accessions</th>
<th>Inventory</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title:</strong> Stephen Jay Gould papers</td>
<td><strong>Description:</strong> Over-size artwork and decorative awards/objects not created by Gould: 3 typewriters; stenograph adding machine; chair with Harvard seal; various toys; trunk containing childhood toys and games. Most materials oversize and present preservation/storage issues, items reviewed by KGT and Henry Lowood, and approved for deaccessioning. KGT to communicate with Rhonda Roland Sherer to determine final disposition of these items. 2011 Aug 10. 3 boxes of miscellaneous music LPs, cassettes, and CDs were given to the Archive of Recorded Sound 2011 Aug 10.</td>
</tr>
<tr>
<td><strong>Related collections:</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Related documents:</strong></td>
<td></td>
</tr>
</tbody>
</table>

**External Documents**

<table>
<thead>
<tr>
<th>HREF</th>
<th>Type</th>
<th>Title</th>
<th>Content</th>
</tr>
</thead>
</table>

Created: Dec 16, 2008 by amorgan | Modified: Aug 10, 2011 by jejohnst | Record Number: 220995
External Documents Location

- `\\sul-wallaby\\Special Collections\\Manuscripts\\Collections\\M1437 Gould`

- M1437 FTK Accession Report

- M1437 Computer Media Imaging Log
Problems in Capturing Floppy Disk

- Sophos under Win 7 will claim the completion of scanning a floppy disk even though it don’t recognize the file format.
- FTK Imager under Win 7 will claim the imaging of a floppy disk successful even though it don’t recognize the file format.
Floppy Disk Capture

• Virus Check
  – Make sure the system you use can understand the filesystem of the floppy disk by double click the floppy disk
Ensure Accruate Virus Check

• List the directory of the floppy disk
  – Double click on the floppy disk drive
Floppy Disk Capture
FTK
Virus Scan

- Open Sophos Antivirus
- “+” Custom Scan
- Start Scan
FTK Imager

- click on "Terminal" icon
- cd ~/Desktop
diskutil list (find disk identifier that corresponds to NAME of disk)
- diskutil unmountDisk /dev/disk1 (match device found from diskutil list)
- ./ftkimager /dev/disk1 floppy1.dd --verify
Outsourcing

• I have never worked with the following vendors. Don’t know the quality of the service.

• Punch Cards (http://punchcardreader.com/)
  – 6.5¢ per card + $3 setup fee + return postage

• 9-track Tape (http://www.emaglink.com)
  – $150 per tape plus the output media and shipping. $25 for a DVD.
Get Hard drives out from Mac

- PowerBook G4

- Macintosh PowerBook 165c
Take Home

• How to identify a variety of storage media, both obsolete and current: floppy diskettes (3.5 and 5.25 inch), computer tapes, optical disks, and hard drives.

• An introduction to computer storage system interfaces: IDE, SATA, SCSI (HDI, 40 pin, 68 pin, and xx pin connectors).

• An introduction to file systems, including: File Allocation Table (FAT), New Technology File System (NTFS), and Hierarchical File System (HFS).

• An introduction to the difference between logical and forensic images.

• How to create an image log using spreadsheet software.

• How to create basic accession records in Archivists' Toolkit.
Take Home

• How to install FTK Imager for Windows (free/low cost software), and how to create logical and disk images using Access Data FTK Imager software.

• How to image disks using write blocker (acquisition of information on a drive without creating the possibility of accidentally damaging the drive contents). -How to view files using FTK Imager.

• How to use the command line FTK Imager for Mac files.

• Demonstrations will include: how to remove a hard drive from a desktop computer, as well as how to remove a hard drive from and external hard drive enclosure. Peter will lead a discussion about outsourcing (tapes, punch cards, data recovery).