OURTSOURCE VS. DIY:

STRADDLING THE BORDER BETWEEN USING VENDORS AND IN-HOUSE DATA EXTRACTION

HARVARD | BUSINESS | SCHOOL

Tessa Beers | Baker Library Special Collections | Harvard Business School

ABSTRACT

In reviewing Baker Library Special Collection's various storage media within current collections, it has become apparent that the library has materials that are at-risk of being lost because they exist on obsolete media formats (i.e., floppy disks). In addition, the recent acquisition of a private firm's collection, containing hundreds of floppy disks and other born-digital materials, would require significant amount of temporal and financial resources to extract and preserve the data, prompting the library to explore available options to accomplish the task.

This poster outlines the library's decision-making process in selecting between contracting an appropriate vendor to extract the data, or investing in the hardware and software required to establish a forensic workstation, or a combination of both. It will also explain the reasons behind the Library's chosen option of creating an in-house **digital forensic workstation**.

SUMMARY

DECISION: The decision to create a designated digital forensic workstation gives Baker Library complete control of security and access to data contained within born-digital collections. The DIY approach allows the archivist to authenticate the data through checksums and ensures that the library is compliant with the OAIS functional model so the original SIP is not lost through the creation of the AIP. Finally, the cost and time of the workstation's creation is less than the total cost and risk of outsourcing, as the total time to complete the disk imaging process is 6:14 minutes, and the actual disk imaging creation time is approximately 20 seconds.

OPTION 1: OUTSOURCE

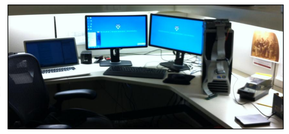
PROS: Dedicated machines and personnel with experience and training
CONS: Cost: assuming no disk errors and without file format conversion, cost of a simple transfer is anywhere from approx. $10/disk (Data Recovery Masters) to $40/disk (Tech Fusion). Forensic transfer, which does not include any checksum verification, is an extra $400-2,000. The vendor gaps in knowledge of archival practices leaves some concern about quality of the final product. Also, despite signing a standard nondisclosure agreement with the vendor, concerns about confidentiality and security breaches are high.

OPTION 2: DIY

PROS: Library can ensure secure method of extracting the data from the disks and other media while limiting the number of staff members who have contact with the collection in accordance with the donor's confidentiality agreement. Time: Takes 29 seconds to 10 minutes to execute a functional disk.
CONS: Potential complications with converting older file formats; difficulties involved in installing hardware and software for the workstation.

EQUIPMENT AND COST BREAKDOWN:

Floppy Disk Drives (IBM 3.5" and internal 5.25")=$0 (acquired in-house)
FC5025 USB 5.25" Floppy Controller = $55.25
AccessData FTK Imager 3.0 = $0
MD5summer Software = $0
} Open Source Software
Computer (Tower + Screen) = $3,000
TOTAL COST: $3,055.25



COMPUTER SPECS:
Dell Precision T3500
Mini-tower form factor
Six Core Intel Xeon Processor
64-bit OS
12GB memory (6 DIMMS)
Entry 3D graphics
1.5 TB disk
16X DVD+/-RW drive

OPTION 3: COMBINATION OF 1 AND 2

PROS: Ability to capture disk image and then outsource the file format conversion. Use vendors to extract data from media where the hardware is difficult to obtain, i.e. 8" floppy disks.
CONS: Confidentiality would still be a risk.

DATA EXTRACTION WORKFLOW FOR 3.5" FLOPPIES

GOALS

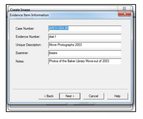
Transform submission information package into archival information package

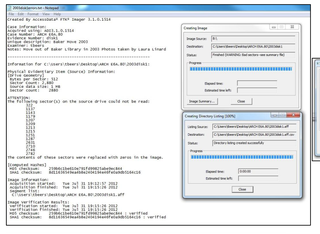
Successfully create disk image and check for accuracy

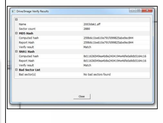
Run a checksum to save with disk image and verify data authenticity

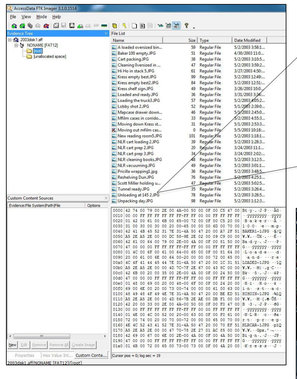
Deposit Disk Image and associated files into the Digital Repository

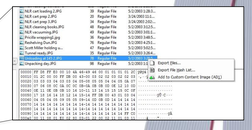
- Take a photo of the media to capture donor-written metadata
- Flip the write-block lock tab on the back of the disk to prevent data destruction
- Insert disk into the floppy drive and open to view contents
- Open AccessData FTK Imager and select : CREATE DISK IMAGE
- Choose Physical or Logical Drive to complete a bit-by-bit transfer
- Select Computer Drive and Add Image Destination Path
- Select Image Type AFF and enter Evidence Metadata
- Click Finish – view evidence tree in hex view by selecting ADD EVIDENCE and uploading the AFF image file
- Run MD5summer Checksum and save with disk image. Right click and export file to see document or image

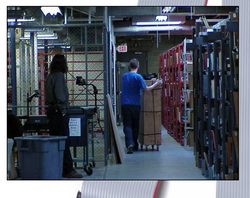


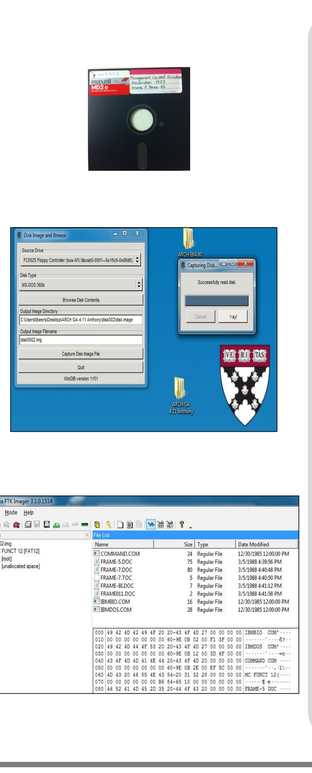












DATA EXTRACTION WORKFLOW FOR 5.25" FLOPPIES

- Take a photo of the media to capture donor-written metadata
- Insert disk into the floppy drive and close – a "hardware ready" popup should show that the disk is recognized.
- Open and Run FC5025 Software "Disk Image and Browse"
- Select the Source Drive and appropriate Disk Type
- Click "Capture Disk Image File."
- Progress bar will say "Yay!" if successful, or "BUMMER!" if the image capture failed
- Follow the 3.5" steps to view evidence and generate MD5summer Checksum

Note: Disk Type MS-DOS 360k seems to work well for regular 3M DS, DD, RH disks. If you choose the wrong disk type, the drive will keep cranking without results

Acknowledgements | Thanks to Laura Linard, Director of Special Collections, Knowledge and Library Services, Harvard Business School