The Return of Lost Content: Born-digital processing of 5.25 inch floppy disks

Karen Ballinger
University of Texas at Austin
SAA Research Forum 2011
The class: Problems in the Permanent Retention of Electronic records a.k.a. Problems

Taught by Dr. Patricia Galloway

Groups process legacy media from archival repositories around Austin and try to put as much information as possible on the UT Austin DSpace installation.
Our team was charged with getting the processing environment ready, which previously fell on individual groups with the most difficult media.
The Digital Archeology Lab

Processing workstations available and administered with archival policies, such as no food or drink, since it stores collection material.
Frankenstein

Donated from the Goodwill Computer Museum and originally a Dell server. Has a RAID card and multiple hard drives with different operating systems in the front. Connected 5.25 inch drive directly to motherboard after changing BIOS. Went down often; tension between extending machine and having it available as processing workstation.
Used to image 5.25 inch floppy disks of specific formats with modern computers. Communicated and collaborated with creator.
Toolkit

Disk dump

Dcfldd

dcfldd if=[/dev/devicename] of=nameofdiskette.img hash=md5 hashlog=nameofdiskette.md5

Checksum

md5sum nameofdiskette.img > nameofdiskette.img.md5

Disktype

sudo disktype [/dev/devicename] | tee -a nameofdiskette_sessionmmddyy.log

The Sleuth Kit

open source digital forensics analysis tools
Workflow

1. Write protect
2. Bitstream copy
3. Checksum
4. Make copies of captured image files
5. Virus scan
6. Mount and access, if possible
7. Metadata extraction
8. DSpace ingest
The George Sanger Games Project, 2011
Schmandt-Besserat (Denise) Papers
3.5 inch floppy disks and CDs relatively straightforward to image.
Imaged Apple disks with floppy controller. Encountered copy protection issues.

Ultima I, II, III:

Ultima Trilogy I – II – III front and back cover, original box

Ultima II:

Ultima II front and back cover, original box

Three 5.25" floppy disks. Disk one, side one Ultima I and side two, Ultima II Galactic disk. Disk two Ultima II, side one Program Master, side two Player Master. Disk three Ultima III, side one Program Master, side two Player Master.
Head stopped moving after imaging 30 disks. Eventually discovered heads were getting dirty by 31st disk, couldn’t read first track, stopped moving.
Discovered legacy virus
The future of disk imaging: Ditto

Being developed by Goodwill Computer Museum and grant from NHPRC
Minimize hardware, maximize software

Robust vintage hardware

Custom modern circuitry

Exact disk copy

XML Container

Analysis at all levels of disk information

Potential error correction

Collaboration with archival goals
Potential research areas

How do you clean magnetic media?
Interoperability between hardware and software
Access and mounting
Old viruses and environments
How does this legacy media equipment really work?
Lessons learned

Document!! Especially failures, for yourself and for the future.

Deliberate testing Plan your trial and error testing.

Communication and collaboration Lots of opportunity for collaboration with the vintage computing community.

Learning through practice We learned through doing! We encourage all repositories to set up their own born-digital legacy media processing environments and discover Problems of their own!
Questions?
Legacy Media Born-digital Processing Resources

University of Texas School of Information DSpace Digital Archiving Repository
https://pacer.ischool.utexas.edu/

All project documentation from the 392K Problems in the Permanent Retention of Electronic Records class is available on pacer. Specific reports include:

Digital Archeology Lab manual: https://pacer.ischool.utexas.edu/handle/2081/23283
UT-iSchool Digital Archeology Lab collection: https://pacer.ischool.utexas.edu/handle/2081/21808
George Sanger Game Projects 2011: https://pacer.ischool.utexas.edu/handle/2081/21810
Schmandt-Besserat (Denise) Papers: https://pacer.ischool.utexas.edu/handle/2081/21807
Ultima II and Published Video Games https://pacer.ischool.utexas.edu/handle/2081/21815
Curtis Riggs Zenith Data System Collection: https://pacer.ischool.utexas.edu/handle/2081/21958
Austin History Center Librarian's Helper Project: https://pacer.ischool.utexas.edu/handle/2081/21850

Goodwill Computer Museum: http://www.goodwillcomputermuseum.org/
If you are interested in collaborating on the Ditto project, please contact the Goodwill Computer Museum curator Russell Corley at Russell.Corley@austingoodwill.org.

Device Side Data FC5025 Floppy Controller: http://www.deviceside.com/fc5025.html

The Sleuth Kit Open Source Digital Forensics tools: http://www.sleuthkit.org/sleuthkit/