Educating for Digital Archiving at UT Austin

Patricia Galloway
School of Information
University of Texas at Austin

ALA-Midwinter 2012
Digital Preservation Interest Group

Digital Archiving at UT: Philosophy

- Digital archivists must know their materials and how they are and have been created, valued, and used
- Where materials are digital, archivists must understand and use both the technology used to create the materials and that used to preserve and access them

Digital Archiving at UT: Teaching Goals

- Create a suite of courses to provide students with the skills to work independently with digital objects of any vintage
- Include the history of the relevant technologies
- Include the history of the development of the digital problematic in archival research and practice
- Set students up to tackle novel problems within a framework designed to protect original materials against damage
- Develop multiskilled teams that crosstrain
- Teach reflective research-in-practice
- Promote multigenerational mentoring by archiving case study reports

Digital Archiving at UT: Infrastructure Goals

- Build a repository and ancillary infrastructure (laboratory space, equipment) where both experimentation and real work can be done
- Integrate the virtual and physical laboratory into the life of the institution at the local and broader levels
    - This aids sustainability: the laboratory serves the institutions recordkeeping and dissemination needs
    - "Incubator" projects can grow digital archiving within other institutions on campus
    - Thus provides service learning opportunities on one's doorstep

Models

- Frederick Brooks, *The Mythical Man-Month*--and his iconic software engineering course at UNC: seeing the work of the laboratory as the emergence--and documentation--of the "theory of the program" for specific archiving tasks

- Gerhard Fischer and the Center for Lifelong Learning at University of Colorado and their research in participatory design for CSCW: framing the development of a virtual/physical laboratory as an evolutionary process
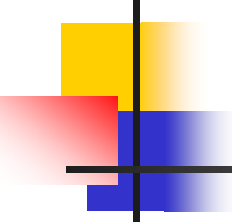
Experience: MDAH

- Automating the Mississippi Department of Archives and History
  - Minimal staff
  - Little funding for equipment
  - Spreading skills everywhere through gradualism
- Building a digital archives in that context: grant money truly "seeded" since the bed was already cultivated

2000-2002

- Course creation
  - 2000: Introduction to digital records (student in first class: Hannah Frost)
  - 2001: Digital Archiving
  - 2001: Lifecycle Metadata for Digital Objects
- Beginning infrastructure
  - First used LAMP (Linux, Apache, MySql, php)
  - Emergence of DSpace winter 2002-3
- Early experiments: email, text processing

2003-2004

- DSpace implemented, spring 2003
- Digital Archiving course, 2003 (4 students)
  - Website archive planning
  - Publication of infrastructure plan
  - Planning documents ingested
- Digital Archiving course, 2004 (6 students)
  - AAA Anthrosource (ASSC) experiments
  - Development of plan for archiving of AAA publications (eventual adoption by AAA of Portico)
  - Planning documents ingested

2005: Departmental IR, incubator for HRC

- Digital Archiving course (19 students)
  - Four faculty members: Davis, Dillon, Hallmark, Wyllys
    - Publications, Syllabi, Webpages, Other learning objects
  - School of Information website
    - Experiment with crawling (Heretrix)
  - Michael Joyce collection, HRC
    - Protocol for processing legacy formats
- Age of media: current
- Copying procedure
  - Write block, Virus scan, Copy, Message digest
- Students in class
  - Catherine Stollar [now Peters] (Joyce project; first digital archivist for HRC)
  - Zach Vowell (Davis project; first digital archivist for BCAH)

2006

- Digital Archiving course (14 students)
  - ASSC: email attachments
  - Cochineal: multimedia documents
  - School of Information tutorials: complex objects
  - School of Information website: apply records schedule
- Age of media: current
- Copying procedure
  - Write block, Virus scan, Copy, Message digest
- Metadata course, 2006
  - Developed METS SIP profiles for DSpace object types
- Student in Introduction class
  - Gabriela Redwine (beginning of Mailer digital issues; second digital archivist for HRC)

2005-2006 Self-study

- ALA RRT-Ingenta Grant: studied work thus far as an effort at institutionalization
- Examined
  - Adoption of participatory model
  - Involvement of School stakeholders
- Measured
  - Citation impact
  - Cost/benefit
- Interviewed faculty
- Reported
  - New Skills for the Digital Era
  - iPRES 2006

2007

- Digital Archiving course (18 students)
  - Wesker: naming conventions, privacy (HRC)
  - Uris and Crowley: corrupt and unreadable files (HRC)
  - Dillon redux: copyright
  - School of Information commencement videos: originals and derivatives
  - School of Information tutorials: complex objects
  - ANAGPIC publications: digitization
- Age of media: current
- Copying procedure
  - Write block, Virus scan, Copy, Message digest

2008

- Digital Archiving course (13 students)
  - Gracy: range of learning objects
  - Mailer: extracting from Iomega OneStep (HRC)
  - George Sanger Videogame sound: IP issues, music formats (BCAH)
- Age of media: (relatively) current
- Copying procedure
  - Write block, Virus scan, Copy, Message digest

2009

- Digital Archiving course (23 students)
  - Paul Banks digital materials (3.5″, Zip; BCAH)
  - Warren Spector email (Apple DVD-R: BCAH)
  - Warren Spector design documents (Apple and Kaypro 5.25″: BCAH)
  - George Sanger games: ATF, Putt-Putt (3.5″, Zip, DAT, SVHS: BCAH)
  - Heather Kelley, Redbeard's Pirate Quest (Jaz: BCAH)
  - Terrence McNally (3.5″, CD-ROM: HRC)
- Age of media: Current, legacy
- Copying procedure
  - Write block, Virus scan
  - In 5.25″ cases: Disk image, clone extraction (using original hardware at Goodwill Computer Museum)
  - Message digest

2010

- Digital Archiving course (24 students)
  - George Sanger email (DVD: BCAH)
  - George Sanger ADAT (SVHS: BCAH)
  - 1988 campaign interviews (3.5″: BCAH)
  - Faculty members Lukenbill, Davis (Zip, USB key, 5.25″)
  - Tutorials (available on server)
  - Maya database materials (5.25″, 3.5″: AAA)
- GCM collaboration (Frankenstein I)
- Copying procedure (offline)
  - Write block
  - Media image (dd), message digest
  - Replicate image, clone extraction, message digests
  - Tar directory tree content where relevant

2011

- Digital Archiving (20 students)
  - George Sanger ADAT (BCAH)
  - *Ultima* II significant property testing (BCAH)
  - Architectural catalog recovery (AHC)
  - Zenith fonds definition (GCM)
  - Frankenstein II/DAL project
- Copying procedure (offline): same as 2010; used Frankenstein I, Dracula, and FRED-L

2012

- Digital Archiving and Preservation (25 students)
    - George Sanger 3.5″ diskettes (BCAH)
    - Hard drive backups (BCAH: several projects)
    - Videogame significant properties testing: *Advanced Dungeons and Dragons* (BCAH)
    - Setup of image database in digital repository (LBJ Library)
    - Reorganization of museum catalog into DSpace (Goodwill Computer Museum)
    - Loriene Roy faculty "papers"
- Copying procedure (offline): further developed since 2011; refined SIP agreement to deal with erased files

Technologies: Hardware

- Standard School laboratory PC and Mac platforms as well as Linux server environments
- FRED-L laptop forensic workstation with suite of connectors, write-blockers, and stand-alone hard drive
- Frankenstein I modified Dell server
- Dracula modified Dell PC

Frankenstein I (2010: antelope on veldt)

Frankenstein I (2011: antelope in zoo)

Frankenstein (F1): Basic Topology

Profile View of F1

Face-on View of F1

Rear View of F1

Top View of F1

1. Additional cooling fan
2. IDE cable, behind which the IDE ports are located
3. Adaptec AIC-7899W onboard SCSI controller
4. Random access memory
5. CPU heat sinks
6. RAID card
7. Power/reset panel
8. 3.5 inch drive
9. F1's 6 swappable hard drives
10. CD-ROM drive
11. F1's power supplies
12. printer port
13. USB ports
14. DE9 video serial port
15. mouse & keyboard ports
16. monitor cable
17. power supply for the CD-ROM drive
18. IDE connector for the CD-ROM drive

Technologies: Software

- DSpace repository software: 1.2, 1.3, 1.5, 1.6, 1.7.2 from 2003-present (15,056 items)

- Linux "clean-room" environment

- Target OSs: CP/M, ProDOS, Mac OS, DOS, Windows, Linux

- Utilities: focus on open-source, standard -ix environment tools

Current Technology Interests

- Creating digital archaeology lab in new School of Information space (established in 2010, elaborated in 2011: detailed protocol manual)

- Establish collecting strategy for legacy drives, software, media, collaborating with IT staff and GCM

- Acquire, configure, and begin using forensic workstation for current media

- Collaborate with Goodwill Computer Museum to build Frankenstein II, develop Ditto machine

- Note that new requirements arise in the process of classwork-as-discovery; every collection is unique in the technological details, but students can build on what previous students did, hence reports are archived as case studies.

Growing a Digital Archiving Laboratory Program

- Stakeholders
- Community of practice
- Evolutionary process
- Service learning
- Participatory research
- Assimilating to infrastructure