The Life Cycle of Electronic Theses and Dissertations

Texas ETD Association Meeting 2012 Keynote

Dr. Martin Halbert, Dean of Libraries
Texas ETD Association Conference
Friday, February 24, 2012
University of North Texas
Organization of Presentation

- ETD programs today
- Definitions of lifecycle management
- ETD lifecycle management challenges
  - Problems of format and ingestion
  - Problems of curatorial process
  - Problems of intellectual property
- IMLS ETD Lifecycle Management Project
A 2010 survey by the Coalition for Networked Information (CNI) of its members by Joan Lippincott resulted in 88 responses from 142 institutions contacted.

73% of respondents reported that they had instituted an ETD program.

In 89% of the institutions, ETDs were reported to be part of their institutional or consortial repository holdings.

43% of institutions reported that the ETD program was mandatory for both doctoral and masters students.
The respondents to the 2010 CNI survey reported that:
- 69% reported preserving ETDs in an institutional or consortial repository
- 47% relied on Proquest to preserve a digital copy
- 5% reported using LOCKSS
- Lippincott observed from these numbers that some institutions were obviously using multiple strategies for preservation.
ETD Associations

101 institutional members

21 institutional members
Can we ensure that ETDs acquired from students today will be available to future researchers? In 10 years? In a century?

How will institutions address the entire life cycle of ETDs?

How will libraries and universities identify and institutionalize the best long-term curatorial practices for this important genre of digital content?
Library of Congress Definition of “Life-Cycle Management”

“...the progressive technology and workflow requirements needed to ensure long-term sustainability of and accessibility to digital objects and/or metadata.”

- http://www.loc.gov/preservation/about/prd/presdig/preslifecycle.html
“...the need actively to manage the resource at each stage of its life-cycle and to recognise the inter-dependencies between each stage and commence preservation activities as early as practicable.”

http://www.dpconline.org/advice/preservationhandbook/introduction/definitions-and-concepts
National/International Digital Preservation Alliances

netpreserve.org
International Internet Preservation Consortium

NATIONAL DIGITAL INFORMATION INFRASTRUCTURE AND PRESERVATION PROGRAM

National Digital Stewardship Alliance

Chronopolis

Digital Preservation Coalition

University of North Texas
Discover the power of ideas
Problems of Format and Ingestion
“Standard” ETDs deposited

- Most ETD repositories are primarily collections of PDFs
- PDFs as format of choice for long term preservation
But not all dissertations are comprised solely of traditional print documents

- An example:
  - [http://digital.library.unt.edu/ark:/67531/metadc30516/](http://digital.library.unt.edu/ark:/67531/metadc30516/)
But then...the Pop Baroque

- POP BAROQUE, by Jennifer Gassiraro, B.F.A
- “Problem in Lieu of Thesis” (?) Prepared for the Degree of MASTER OF FINE ARTS
- “…vivid colors, textures, patterns and designs collected from my environment...”
...and here’s all we get...
And then, the *Pleroma*...

- **Critical Discussion of Pleroma: A Digital Drama and Its Relevance to Tragic Form in Music**, by Stephen Lucas, B.M. Thesis Prepared for the Degree of MASTER of MUSIC
- “Pleroma is a digital drama: a work composed of digital animation combined with electroacoustic music, presenting an original dramatic narrative. Pleroma's dramatic elements evoke both the classical form of tragedy and the concept of perceptual paradox...”
...Pleroma is depicted as a floating eyeball with half of a blue, translucent eyelid encircling its form... Abraxas and Gnosis are two sides of a double-headed creature and serve as advisers to Pleroma. Abraxas has a head that resembles some type of bird, with hair made of fire, and Gnosis's head resembles a metallic robot, with hair made of wires...

http://digital.library.unt.edu/ark:/67531/metadc33228/
The actual, relevant content of theses and dissertations may not be captured, or *capturable*, in a PDF. Performative or functional works are often not preserved in ETD repositories.
Problems of Curatorial Process
One day, through the primeval wood,
A calf walked home, as good calves should;

But made a trail all bent askew,
A crooked trail as all calves do.

Since then two hundred years have fled,
And, I infer, the calf is dead.

But still he left behind his trail,
And thereby hangs my moral tale.
The trail was taken up next day
   By a lone dog that passed that way;

And then a wise bell-wether sheep
   Pursued the trail o'er vale and steep,

   And drew the flock behind him, too,
   As good bell-wethers always do.

And from that day, o'er hill and glade,
   Through those old woods a path was made.
And many men wound in and out,
    And dodged, and turned, and bent about;

And uttered words of righteous wrath,
    Because 'twas such a crooked path.

But still they followed - do not laugh -
    The first migration of that calf.

And through this winding wood-way stalked,
    Because he wobbled when he walked.
This forest path became a lane,
    That bent, and turned, and turned again.

This crooked lane became a road,
    Where many a poor horse with his load,

Toiled on beneath the burning sun,
    And traveled some three miles in one.

And thus a century and a half,
    They trod the footsteps of that calf.
The years passed on in swiftness fleet,
    The road became a village street;

And this, before men were aware,
    A city's crowded thoroughfare;

And soon the central street was this,
    Of a renowned metropolis;

And men two centuries and a half,
    Trod the footsteps of that calf.
Each day a hundred thousand rout,
Followed the zigzag calf about;

And o'er his crooked journey went,
The traffic of a continent.

A hundred thousand men were led,
By one calf near three centuries dead.
They followed still his crooked way,
And lost one hundred years a day;

For thus such reverence is lent,
To well-established precedent.

-Sam Walter Foss,
“The Calf-Path” 1896
Unfortunate legacies of prior decisions, or lack of decisions/re-examination of processes, are omnipresent and cause enormous problems.

Many (most?) efforts of organizational process remediation are aimed at addressing calf-path issues.

Avoiding following calf-paths should be a goal whenever establishing new precedents.
Some Calf-Path Case Studies

- The following are some representative actual examples distilled from many years of consulting with ETD and other cultural memory organization clients that were seeking to address digital preservation issues
- Selected because they are emblematic of particular calf-path problems
- These are examples of bad things happening to good bellwethers who followed practices set down by (someone) before them
- The names have (mostly) been withheld to protect the embarrassed
Case Study:
We know where things are (kindasorta)

- A pilot electronic thesis and dissertation (ETD) program of a major research university reposited incoming ETD’s for several years, through a series of ad hoc workflow processes emulating the original print submission process.
- These “echo” processes were created ad hoc by a series of library paraprofessionals, storing ETD’s in one directory, then another, and keeping metadata in non-standardized text files resembling the paper forms...
- ...after a decade we overhauled the whole program and its workflow because we wanted to actually preserve the data.
- Observation: consistently structured workflow makes for digital preservation readiness.
Case Study:
We have a backup program (kindasorta)

- A mid-size cultural memory organization engaged in active digitization of its archival collections for a full decade
- They scanned at high quality and kept reasonable metadata
- They backed up the accumulated scans to tape and stored the tapes...
- ...in the room next to the server...
- ...and one afternoon had a fire and lost it all.
- Observation: local backups are not a preservation program
Case Study:
We know what we’ve digitized (kindasorta)

- The archives of a major research library engaged in selective digitization of its collections for years...
- ...by a series of junior archivists with no official mandate or standardized workflow process in place...
- ...the minimal “metadata” captured was recorded idiosyncratically in the image filenames...
- ...after almost a decade they gave up and started over because no one could now make any sense out of the mass of unorganized files they had accumulated.
- **Observation:** a digital collection without consistent metadata and workflow may become completely unusable
Libraries have “deep infrastructure” (professionally internalized practices and organizational systems) with long term management of analog collections.

We lack such deep infrastructure for ETD collections because:

- digital technologies are unfamiliar; they are relatively new and have quite different properties and dynamics
- digital technologies are unpredictably changing
- our institutional mandates and funding are often still oriented toward analog, not digital resources

In an unfamiliar and shifting wilderness, one tends to follow any kind of path encountered, and lack the critical perspective to analytically question such choices.
2010 CNI survey shows that most research universities have by now implemented some form of ETD program

Many (most?) of these programs inevitably incorporated calf-path processes during their startup

When on them, it is hard to see calf-paths (because “that’s the way everyone has always done it”) and even harder to remediate them (metaphor of fixing the airplane’s wings while in flight)

Many ETD programs are now preparing to take on calf-path straightening exercises, *which may be harder than the original implementation*
Problems of Intellectual Property
87% of the institutions responding to the CNI survey in 2010 had a policy allowing students to request embargo of access to their ETD.

Widely repeated anecdotes by students and faculty to the effect that “if they allow open access to their work, it will preclude future publication of the content in certain journals or as a monograph.”

These kind of fears have led to “knee jerk” policies of comprehensive embargo of all ETDs deposited.
IMLS ETD Lifecycle Management Project 2011-2013
Prior Collaboration: NDLTD/MetaArchive ETD Program

- Started in 2008 with the establishment of a partnership between MetaArchive and NDLTD
- Project study of the genre-specific preservation issues that arise with ETD collections
- Initial partners: Virginia Tech, Boston College, Georgia Tech, Rice U, Emory U, and Auburn U.
- Highly successful--preserving ETDs for NDLTD/MetaArchive members, including overseas members (such as a consortia of 20 institutions in Spain)
MetaArchive Member Work on ETD Preservation

- Studied the "calf path" issues that arise in ETD programs
- Analyzed a range of ETD repository structures and developed exchange mechanisms between those and LOCKSS (CONTENTdm, ETD-db, DSpace)
- Provided simple addition mechanisms so that as new and embargoed ETDs are added, members are able to easily add them to the archive
- Developed mechanisms to version content, so that if ETDs are changed/replaced, reflected in preservation copies
- Determined the need for documented best practices for ETD preservation readiness (IMLS project)
How will institutions ensure that the electronic theses and dissertations they acquire from students today will be available to future researchers?

- We need to better understand, document, and address the lifecycle management challenges presented by ETDs to ensure that colleges and universities have the requisite knowledge to properly curate these new collections.
- Project will take place over a two year period from November 2011 to October 2013.
IMLS ETD Lifecycle Management Project Partners

1. University of North Texas,
2. Networked Digital Library of Theses and Dissertations (NDLTD),
3. Educopia Institute/Meta Archive Cooperative,
4. Virginia Tech,
5. Rice University,
6. Boston College,
7. Indiana State University,
8. Pennsylvania State University,
9. University of Arizona
IMLS ETD Lifecycle Management Project Goals

1. Dissemination of Guidance Documents for Lifecycle Management of ETDs
2. Production of ETD Lifecycle Management Tools
3. Creation of Educational Materials and Associated Workshop
Guidance Documents for Lifecycle Management of ETDs

1. Briefing on Access Levels and Embargoes of ETDs
2. Briefing on ETD Copyright Issues and Fair Use
3. Guidelines for Implementing ETD Programs - Roles & Responsibilities
4. Guidelines for Collecting Usage Metrics & Demonstrations of Value for ETD Programs
5. Overview of Formats, Complex Content Objects, and Format Migration Scenarios for ETDs
6. Overview of ETD Metadata & Lifecycle Event Record-Keeping for ETDs
8. Guide to Options for ETD Programs
The project will develop and disseminate a set of software tools to address specific needs in managing ETDs throughout their lifecycle.

a) ETD format recognition
b) PREMIS metadata event record-keeping
c) Virus checking
d) Digital drop box with metadata submission functionality
Educational Materials and Associated Workshop

- Educational Materials
  - Workshop Syllabi
  - Training Handouts and Exercises
  - PowerPoint presentations

- Workshop
  - Will use these educational materials
  - Will be held in 2013, perhaps in conjunction with another major ETD conference
As an emerging field, ETD programs and practitioners need to internalize habits of reflection, critical thinking, and documentation of best practices.

As institutions we need to identify needed improvements to problematic practices and plan to systematically undertake them.

We need to always keep in mind the big picture of why we are seeking to preserve and make accessible ETDs, and be prepared to assess and clearly articulate the value of these programs.