Providing the ETDs of Today for the Researchers of Tomorrow

Martin Halbert, Shannon Stark, Daniel Alemneh, Geneva Henry

2013 TxETDA Conference
Austin, Texas
Thursday, February 28, 2013
Welcome!

- 9:00am – 10:00am: Project Background & Guidance Document Overview
- 10:00am-10:15am: Questions
- 10:15am-10:30am: Break
- 10:30am-11:45am: Focus Group Questions
- 11:45am-12:00pm: Questions & Wrap-Up
The move to digital theses and dissertations greatly enhances the accessibility and sharing of graduate student research, but it also raises grave concerns about the potential ephemerality of these digital resources.

How can institutions best ensure that the electronic theses and dissertations that they acquire from students today will be available to both current and future researchers?

This IMLS-funded project is investigating ETD lifecycle management, focusing on both guidance documents and technology.
Presenters

- Dr. Martin Halbert (Dean of Libraries, University of North Texas)
- Shannon Stark (University of North Texas)
- Daniel Alemneh (University of North Texas)
- Geneva Henry (Rice University)
Key Challenges

- 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 identify and institutionalize the best long-term curatorial practices for this important genre of digital content?
Digital Preservation Coalition (DPC) Definition of “Life-Cycle Management”

“...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
There is a 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 permanently.

A project to identify and document these best practices has been funded by the U.S. Institute of Museum and Library Services.

The project is now underway, and will be complete in December 2013. Early project findings are being shared in this pre-conference workshop.
1. University of North Texas
2. Networked Digital Library of Theses and Dissertations (NDLTD)
3. Educopia Institute/MetaArchive 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

A. Dissemination of Guidance Documents for Lifecycle Management of ETDs
B. Production of ETD Lifecycle Management Tools
C. Creation of Educational Materials and Associated Workshop
Guidance Documents for Lifecycle Management of ETDs

1. Access Levels and Embargoes of ETDs
2. ETD Copyright Issues and Fair Use
3. ETD Metadata & Lifecycle Event Record-Keeping for ETDs
4. Collecting Usage Metrics & Demonstrations of Value for ETD Programs
5. Formats, Complex Content Objects, and Format Migration Scenarios for ETDs
6. Implementing ETD Programs - Roles & Responsibilities
7. ETD Program Cost Estimation and Planning
8. Options for ETD Programs
Access Levels and Embargoes

- One of the most contested topics in ETD program planning is the question of ETD embargoes and levels of access restriction
  - As evidenced by both the NDLTD/MetaArchive surveys and the 2010 CNI survey.

- An “embargo” of an ETD means delaying public access to the ETD, either temporarily or permanently.
• Different stakeholders are particularly concerned about embargoes.
  • For instance, does depositing an ETD in a public repository constitute publication and hinder future development/publication of the work?
    • Publishers do not think so. (NDLTD survey)
    • The US Patent Office might consider an ETD to be prior art.

• Policy options for the embargo of ETDs range:
  • No embargo
  • Blanket fixed-length embargo
  • Limited list of fixed-length embargo
  • Embargo renewal
Copyright Issues and Fair Use

- ETD program may introduce students to issues of copyright
  - Author rights
  - Types of licenses
  - Fair use
  - Commercial publishers

- An ETD program does disservice to both students and institution if it does not provide information to make informed decisions on copyright and fair use.

- Who on the campus can provide guidance about ETD copyright and fair use?
Copyright Issues and Fair Use

- ETD programs must know
  - What understanding or agreement is in place at your institution regarding intellectual property rights of students?
    - What rights does the university/college exert on student work?
    - When does research belong to the university, and when does it not?
  - Does sponsored research require students to sign University Intellectual Property (IP) agreement or embargo their results?
  - What qualifies as plagiarism?
  - When does publication violate copyright agreements of work reproduced in the ETD?
  - What should students know regarding ETDs and future scholarly publishing efforts?
ETD Metadata & Lifecycle Management for ETDs

- Metadata structures much of the ongoing management of ETDs.
- Metadata may be assigned by a mix of: Librarians, Student authors, system, even users.
  - No matter the mix, professionals supervision provides quality control.
- The ETD repository software heavily influences the creation of metadata, but metadata should be thorough and comprehensive.
NDLTD has developed and maintains an ETD specific descriptive metadata schema.

**PREMIS** stands for "**PRE**servation **M**etadata: **I**mplementation **S**trategies."

- The PREMIS data model consists of five interrelated entities: Intellectual, Object, Event, Agent, and Rights.
- It allows repositories to implement varying workflow and submission models. In light of tracking transitions in the lifecycle of digital objects, this project is experimenting with updating ETD records.

An effective ETD metadata management approach can help institutions improve consistency, clarity of data lineage, and relationships so that they can better integrate related resources and ensure long-term access to ETDs.
Collecting Usage Metrics & Demonstrations of Value for ETD Programs

- Libraries have a long history of evaluating and studying use of library resources and collections, and ETDs should be no exception.

- Usage data can make a strong case for ETD program support to university administrations.

- Usage reports of all kinds should be prominently featured on the ETD program website, and easily reviewed by all users of the service.
Collecting Usage Metrics & Demonstrations of Value for ETD Programs

- Quantitative statistics
  - Easily gathered
  - Frequently used
  - Provide compelling indication of utility, e.g. download statistics
  - Analysis can be performed on the data. e.g. demographics, behavior (new, returning users), technology, and devices.

- Qualitative evaluation
  - Case studies, surveys, interviews, and visual observations
  - Less commonly performed
  - Provide more nuanced information.

- Recommendation
  - Minimal level
  - Advanced Level: Gathering more data, but requiring more resources.
Collecting Usage Metrics & Demonstrations of Value for ETD Programs

Benefits of Usage Metrics
• Authors, faculty and graduate students
  o Impacts and importance of a title
  o Impacts of their ETDs
• Institution
  o Impacts of the ETD collection
  o Return on Investment (ROI)
  o Satisfaction of graduate students
• Scholarly Society
  o Impacts of ETDs v.s. other content
• Many ETD programs mandate that the primary item deposited be some form of PDF, sometimes with format checking of the specific characteristics of the PDF.

• ETDs can contain non-textual supplementary files
  • Music
  • 3D Renderings
  • Datasets
  • For these files, some flexibility is necessary, but the institution should also provide guidelines.

• The data and preferred file formats won’t remain stable over time.
Upon deposit, both primary and supplementary files should be checked for:

- Format validity
- Viruses
- Fixity

Format migrations are anticipated by many ETD repositories. Depending on the migration policy, they may be:

- Manually batched
- Automated
Implementing an ETD program requires identifying the various stakeholders and specifying their roles and responsibilities throughout the entire course of ETD management.

- **Internal Stakeholders:** Institutional Administrators, Graduate Schools, Academic Libraries, Offices of Information Technology

- **External Stakeholders:** Commercial Companies, ETD Organizations, Library Consortia, Access Harvesters/Facilitators, Digital Repository Services, and Digital Preservation Services
Implementing ETD Programs - Roles & Responsibilities

• ETD Program Planning
  ▪ Provide a rationale for establishing an ETD program
  ▪ Advocate the program
  ▪ Propose an implementation plan

• ETD Creation, Submission and ingestion
  ▪ Construct, format and submit ETDs, may make embargo requests
  ▪ Develop submission policies, procedures and standards
  ▪ Administer the process, review and approve ETDs
  ▪ Offer assistance including legal services
  ▪ Prepare submission systems
  ▪ Catalog and ingest ETDs, digitize retrospective theses and dissertations

• ETD access
  ▪ Develop ETD access policies and ETD end user license
  ▪ ETD access management
Implementing ETD Programs - Roles & Responsibilities

- ETD access (cont.)
  - Distribute ETDs in multiple access venues
  - Assist with ETD visibility, accessibility, and searchability

- ETD Archiving and Preservation
  - Develop a formal preservation plan
  - Organize and preserve in reliable media or systems
  - Preserve ETD contents, format, metadata and URLs

- ETD program evaluation and assessment
  - Evaluate ETD submission, support services, and program impact
  - Evaluate ETD cataloging, archiving and preservation practices
  - Evaluate ETD accessibility and usability
  - Evaluate ETD systems: operational, sustainable, and viable
ETD Program Options and Cost Planning

- Planning an ETD program depends on a number of decisions with varying costs.
  - Staffing
  - Repository platform
  - Equipment main
  - Preservation Strategy
  - Outsourcing

- Prepare a careful plan involving a full range of stakeholders. There will be unexpected changes, but it gives you a place to start.
ETD Program Options and Cost Planning

- Incorporating ETDs into larger institutional repositories
  - Engages economies of scale
  - Reduces focus on ETD specific issues

- Partnering with other institutions or alliances
  - Improves access to information
  - Shares access to digital preservation resources

- Completely outsourcing to external vendors,
  - Reduces work required locally
  - Relinquishes some control to the vendor
Focus Group on ETD Guidance Documents
1. What process is used to determine embargoes at your university?

2. Do you do any redactions of theses; if so, how are these managed? Do you maintain two separate copies of the thesis?

3. Do you do partial embargoes of theses? How is this managed?
Copyright & Fair Use

1. What guidance or resources (including workshops) on copyright and fair use does your institution provide for students writing ETDs?

2. Who is involved in developing such copyright and fair use guidance & resources (and conducting workshops, if applicable)? What aspects of copyright and fair use issues are addressed?

3. How do you assess the impact of what are, ultimately, outreach activities for copyright and fair use?
ETD Metadata & Record-Keeping

1. Do any of the current metadata practices support your institution's ETDs specific requirements? If not, where would you like to see improvements?

2. What software or tools are used (or required) to facilitate your institution's ETDs metadata management?
Collecting Usage Metrics

1. Does your institution track ETD usage? If so, what are the typical measures?

2. How does your institution measure the success of your ETD program? What are the criteria?
ETD Format Issues

1. What restrictions (if any) do you encourage/enforce regarding what file formats can be submitted for the ETD itself or for any supplementary files?

2. How do you expect to use emulation or migration or some combination as part of your digital preservation program?

3. What do you consider to be the greatest threat to long term access to ETDs and how do you plan to deal with it?
ETD Program Roles & Responsibilities

1. Who are the primary stakeholders involved in your ETD program? Inside and outside your institution?

2. Does your ETD program apply any curation activities to preserve ETDs for the long-term?
ETD Program Costing & Planning

1. Has your institution evaluated the costs associated with its ETD program? If so, what form did this evaluation take and who initiated it? What are/might the key cost factors be?

2. Are there any areas of your institution’s ETD program that could benefit from additional resources?
Contact information:

- Martin Halbert, Principal Investigator (martin.halbert@unt.edu)
- Katherine Skinner, Principal Investigator & Project Manager (katherine.skinner@metaarchive.org)
- Matt Schultz, Co-Project Manager (matt.schultz@metaarchive.org)