Providing the ETDs of Today for the Researchers of Tomorrow

Martin Halbert, Katherine Skinner, Matt Schultz
2012 CNI Fall Membership Meeting
Washington, D.C.
Tuesday, December 11, 2012

Overview

- 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.

UNT
UNIVERSITY OF
NORTH·TEXAS

ETD Lifecycle Management   NDLTD   12/11/12   Slide 2

Meta·Archive
C O O P E R A T I V E

Presenters

- Dr. Martin Halbert (Dean of Libraries, University of North Texas)
- Matt Schultz (Program Manager, Educopia Institute and MetaArchive Cooperative)
- Dr. Katherine Skinner (Executive Director, Educopia Institute and MetaArchive Cooperative)

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

NDLTD Digital Preservation Initiative

- In 2007 the NDLTD board of directors became concerned about the need for digital preservation of ETDs
- Surveys were conducted to assess the status of ETD programs and preservation
- Informed by these surveys, the NDLTD leadership approached the MetaArchive Cooperative for digital preservation services

NDLTD/MetaArchive Alliance

- Based on 2008 survey responses, NDLTD leadership negotiated a strategic alliance with the MetaArchive Cooperative for digital preservation purposes
- A follow-up survey jointly conducted in 2009 by MetaArchive and NDLTD identified more specific needs for collaborative distributed digital preservation (DDP) efforts
- Based on the findings of these two surveys, a DDP network for ETDs from NDLTD members was created by MetaArchive starting in 2008-2009

IMLS ETD Lifecycle Management Project

- 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 October 2013. Early project findings will be shared in this section of the workshop. We would like your advice on how to best share this information with the international community.

UNT
UNIVERSITY OF
NORTH·TEXAS

ETD Lifecycle Management   NDLTD   12/11/12   Slide 8

Meta·Archive
C O O P E R A T I V E

IMLS ETD Lifecycle Management Project Partners

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. 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
7. Guide to ETD Program Cost Estimation and Planning
8. Guide to Options for ETD Programs

Implementing ETD Programs - Roles & Responsibilities

- Carefully consider how to engage all stakeholders from the beginning.

- Different types of stakeholders have different interests and concerns:
  - Graduate schools
  - Academic libraries
  - University IT Office
  - Students
  - Faculty members

Access Levels and Embargoes

- Different stakeholders are particularly concerned about embargoes. There is concern by some academic fields (notably the humanities) that depositing a thesis or dissertation in a public repository somehow constitutes publication and prevents students from subsequently developing their work into a book. (A recent NDLTD survey of publishers indicates that publishers do not consider this to be the case.)
- Options for embargo of ETDs can range from extremes of none to all ETDs stored in the repository. This is one of the most important policy decisions for ETD program developers.

Copyright Issues and Fair Use

- Who on the campus can provide guidance about ETD copyright and fair use?
- ETD program may introduce students to notion of themselves as authors, and their rights as authors, experience with licenses, fair use, commercial publishers, etc.
- ETD program does disservice to both students and institution if we don't provide information to make informed decisions on copyright

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 reports of all kinds should be prominently featured on the ETD program website, and easily reviewed by all users of the service.

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

Formats, Complex Content Objects, and Format Migration Scenarios for ETDs

- Many institutions may wish to mandate which formats are to be used for ETD deposits. For example, 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.
- For any files that are included as supplementary files to the ETD itself, while some flexibility is necessary, the institution should consider providing guidelines to the students.

ETD Metadata & Lifecycle Event Record-Keeping for ETDs

- An ETD specific metadata scheme has been developed by NDLTD and is being updated on an ongoing basis for use by ETD repositories. Metadata for deposited ETDs should be as thorough as possible.
- PREMIS is a metadata standard for tracking transitions in the lifecycle of digital objects. It can be used to update ETD records (project is experimenting with this).

ETD Program Options and Cost Planning

- There are a large number of options to consider in planning an ETD program, and costs of the program will depend on these many decisions.

- Some institutions have completely outsourced the ETD program to external vendors. This may or may not be an attractive option; while it reduces the work required locally it has significant costs and requires relinquishing some control to the vendor.

UNT
UNIVERSITY OF
NORTH·TEXAS

Meta·Archive
COOPERATIVE

ETD Program Options and Cost Planning

- Many institutions manage ETDs as part of their larger institutional repositories, thereby combining infrastructures. Digital preservation programs for ETDs can be accomplished by partnering with other institutions, either international alliances (like MetaArchive) or consortia.
- Under any circumstance, a careful plan should be prepared in advance, with staffing, system, and other costs identified. There will undoubtedly be unexpected changes, but it gives you a place to start.

Software Micro-Services

The project will develop and disseminate a set of software tools to address specific needs in managing ETDs throughout their lifecycle.

- ETD format recognition
- PREMIS metadata event record-keeping
- Virus checking
- Digital drop box with metadata submission functionality

Educational Materials and Associated Workshop

Educational Materials

- Workshop Syllabi

- Training Handouts and Exercises

- PowerPoint presentations

Full Workshop on ETD Lifecycle Best Practices

- Will use these educational materials

- Will be held in 2013, perhaps in conjunction with USETDA 2013

Discussion, Questions, and Answers

Contact Information

Contact information:

- Martin Halbert   (martin.halbert@unt.edu)

- Matt Schultz   (matt.schultz@metaarchive.org)

- Katherine Skinner (katherine.skinner@metaarchive.org)