IMLS-funded Lifecycle Management of ETDs Project

Overview

The University of North Texas Libraries together with the Networked Digital Library of Theses and Dissertations, The Educopia Institute/MetaArchive Cooperative, and the libraries of Virginia Tech, Rice University, Boston College, Indiana State University, Pennsylvania State University, and the University of Arizona are engaged in an IMLS-funded, two-year project (Oct 2011-Sept 2013) to develop and disseminate guidelines, educational materials and an associated workshop, and a set of software tools for life-cycle data management and preservation of Electronic Theses and Dissertations (ETDs).

The intended audience for this project includes academic libraries currently managing or prospectively considering programs for ETD preservation.

Project Activities and Products

The project team seeks to better understand, document and address the preservation challenges presented by ETDs to ensure that colleges and universities have the requisite knowledge to properly curate these new collections. To that end, the project team will accomplish the following:

1. Dissemination of Guidance Documents for Lifecycle Management of ETDs: These products will provide assistance to information professionals seeking to effectively manage the lifecycle of their ETDs. Based on collaborative research between members of the worldwide NDLTD consortium and the MetaArchive Cooperative, we will develop a series of concise and informative documents concerning the issues of specific concern to the curators of ETD collections, and address key considerations for curators who seek to implement data management and preservation solutions for ETD collections.

2. Production of ETD Lifecycle Management Tools: The project will develop and disseminate a set of software tools to address targeted needs in managing ETDs throughout their lifecycle. These tools will be created as completely modular micro-services that can be used alone or incorporated into larger repository systems. These micro-service tools will be targeted at individual functions such as ETD format recognition and PREMIS metadata event record-keeping.

3. Creation of Educational Materials and Associated Workshop: A set of educational materials on the topic of lifecycle management of ETDs will be prepared and made freely available. These materials will be utilized in a workshop that will be offered in the second year of the project. Materials will include curriculum syllabi, training handouts, PowerPoint presentations, exercises, and other relevant items.

Project Outcomes

The success criteria for this project include the following measurable changes in audience members:

1. ETD lifecycle management knowledge and skills gained by workshop attendees and others nationwide that use the project products will be assessed.

2. Utilization of ETD lifecycle management software tools created will be assessed for a nationwide group of institutional users who download and install the tools.

3. Attitudes and adoption behavior regarding project products of NDLTD members concerning the guidelines will be tracked after dissemination of the guidelines and use cases.
Project Highlights (Oct 2011-Sept 2012)

Highlights of this research and development project’s accomplishments during the first year of work include the following:

**Project Planning and Outreach**
- Kick-off meeting at TXETD 2011
- Monthly meetings as conference calls/videoconferences using WebEx
- Presentations and a publication (TXETD, USETDA, Digital Directions, NDLTD, iPres)

**Guidance Documents**
- Drafted, approved, and assigned the Guidelines sections to Steering Committee members
- Developed and reviewed and approved bibliographies and outlines for each Guidelines section
- Conducted focus group on ETD practices/concerns
- Produced initial drafts of all Guidelines sections

**Lifecycle Management Tools**
- Engaged a programmer consultant to make progress on microservices
- Performed a thorough interoperability tools inventory & gap analysis
- Developed a draft Project Development Plan for review and approval by the Steering Committee
- Hired and began orientation for the project programmer
- Began preparations for microservices integration and development work

**Summary of Findings**

**Project Planning and Outreach**
The first year of our project work has focused on start-up activities (a kick-off meeting with the Steering Committee in Texas, establishing a regular meeting schedule, hiring staff members), research (literature reviews, focus groups), technical planning (developing a tools inventory, gap analysis, and Tools Development Plan), and outreach (presentations in a wide variety of venues). We have also worked to foster a close, collaborative environment for all project partners and staff members.

**Guidance Documents**
Research in preparation for documenting the Guidance Documents has been a heavy focus in the first year of project activities. The Lifecycle Management of ETDs project research seeks to answer the question: How will institutions address the entire lifecycle of ETDs, ensuring that the electronic theses and dissertations they acquire from students today will be available to future researchers?

The project team selected experts to serve on a Steering Committee that is researching and drafting the Guidance documents. This Steering Committee has performed a thorough literature scan, authored bibliographies, and developed outlines for each of their respective Guidance documents. These draft materials were reviewed and discussed at the project’s first Steering Committee meeting, which was held in conjunction with the 1st annual TxETDA conference in February 2012. During this meeting, the Steering Committee and project staff also led an important focus group with the broader assembled ETD community, including representatives from a variety of stakeholder communities, including faculty, college/graduate school administration, and library representatives. The meeting and focus group were highly successful, and provided each of the expert authors with the basis they needed to begin drafting their Guidance documents. The Guidance documents are approaching near-final draft completion in advance of our second project Steering Committee meeting, scheduled for the end of September 2012 (27th-28th).

**Lifecycle Management Tools**
In advance of the second project Steering Committee meeting work has also been underway to clearly model the use case implementations and necessary development work for the proposed lifecycle management tools. Specifically, the project manager has worked closely with a programmer consultant to research existing microservice tools and perform a gap analysis. This work has culminated in a draft Project Development Plan that will be reviewed at the upcoming Steering Committee meeting in September 2012. Currently three main tools are proposed: 1) a format recognition service, 2) virus checking service, and a 3) PREMIS event tracking service. In addition, work will be completed to document and model the proper development of ETD submission tools and plagiarism tools. A project programmer has now been hired to produce the microservices for later installation and testing at project partner sites. At the upcoming Steering Committee meeting, the project programmer will provide a brief functional overview of the proposed tools and their application in an ETD program environment. The project staff and Steering Committee will discuss any needed improvements and work will then begin to functionalize the tools this fall.
Appendix A: Guidance Documents

Introduction to ETDs
Prepared by Dr. Katherine Skinner and Matt Schultz (Educoopia, MetaArchive), this document introduces the “Guidelines” and chronicles the history of ETDs. Using survey data and research findings, it describes the evolving and maturing set of practices in this area. It discusses the philosophical and political issues for this genre, including what to do with digitized vs. born-digital objects, how to make outsourcing decisions, and how to deal with lifecycle concerns about future publications and embargoed materials. It includes a conceptual overview of a lifecycle model for ETDs.

Access Levels and Embargoes
Prepared by Geneva Henry (Rice University), this document provides information about the ramifications of campus policy decisions for or against different kinds of access restrictions. It defines access restriction and embargo, and discusses reasons for each, including publishing concerns, sensitivity of data, research sponsor restrictions, and patent concerns. It discusses how institutions may provide consistent policies in this area and how policies might impact an institution’s lifecycle management practices. It also reviews and compares existing university policies and makes policy recommendations.

Copyright Issues and Fair Use
Patricia Hswe (Penn State) chronicles ETD copyright and fair use issues that arise both in the retrospective digitization and the born-digital acquisition of theses and dissertations. It discusses institutional stances and guidelines for sponsored research and student work, and also reviews copyright and fair use issues with respect to commercial publishers (including e-book publishers) and vendors such as ProQuest. It seeks to provide clarifying information concerning publisher concerns and issues, providing a concise summary of the relevant information for stakeholders.

Implementation: Roles & Responsibilities
Xiaocan (Lucy) Wang (Indiana State University) documents the variety of stakeholders who impact and are impacted by the transition to electronic theses and dissertations, including such internal stakeholders as institutional administration (e.g., president, provost, CIO, general counsel), graduate schools (administrators, students, faculty), libraries (administrators, systems divisions, technical services, reference), and offices of information technology, and such external stakeholders as commercial vendors/publishers, NDLTD, access harvesters (e.g., OCLC), and digital preservation service providers (e.g., MetaArchive, FCLA, DuraCloud). It emphasizes the range of functions played by stakeholders in different management phases and institutions.

Demonstrations of Value
Dr. Yan Han (University of Arizona) provides guidance for institutions concerning assessment of ETD usage, and how communicating such assessment metrics can demonstrate a program’s benefits to stakeholders. Han also documents practical examples of documenting and conveying usage metrics for stakeholder audiences, including the university, the students, and the research community more generally. He provides practical guidance for collecting, evaluating, and interpreting usage metrics in support of ETD programs, and discusses how it may be used to refine and promote this collections area.

Formats and Migration Scenarios
What factors should be considered by colleges and universities to determine what formats they should accept? How can they manage on an ongoing basis the increasingly complex ETDs that are now being produced by students? Bill Donovan (Boston College) discusses these format issues, including “data wrangling” practices for legacy content and migration scenarios for simple and complex digital objects in ETD collections.

PREMIS Metadata and Lifecycle Events
Another issue revealed in the needs assessment process was that most institutions do not have workflows and systems in place to capture the appropriate levels of metadata needed to manage ETDs over their entire lifecycle. Daniel Alemneh (University of North Texas) describes the critical issues to be aware of in gathering and maintaining preservation metadata for ETDs, not just at the point of ingestion, but subsequently, during transitional events (embargo releases, redactions, etc.).

Cost Estimation and Planning
Gail McMillan (Virginia Tech) provides institutions with information on costs and planning, laying out the critical paths that many ETD programs have charted to date. This document provides cost-benefit analyses of multiple scenarios to give institutions a range of options to consider for their local needs.

Options for ETD Programs
Our surveys and focus group have demonstrated that many institutions are delayed in ETD program planning simply because they do not have a clear understanding of the range of options to consider in implementing an ETD program. Restricted or open access? Implement an ETD repository or lease a commercial service? Who has responsibility for what functions? Dr. Martin Halbert (University of North Texas) explains the relevant decisions institutions must make as they set up an ETD program and clarifies the pros and cons of different options.