ETD Lifecycle Management Workshop

Jimma University (Ethiopia)
July 7, 2015
By
Daniel Gelaw Alemneh
University of North Texas

This workshop series is made available from the above organizations under a Creative Commons CC-BY 4.0 International License
Workshop Overview

• Module 1: ETD Lifecycle Management Interactive Overview

• Module 2: Guidance Documents for Lifecycle Management

• Module 3: Lifecycle Management Tools

• Discussions of Local Context
Workshop Background

• IMLS-funded *Lifecycle Management of ETDs* project (2011-2014)

• **Project Objectives:** Documenting lifecycle curation practices for ETDs and improving implementations for curation tools

• **Project Deliverables:**
  • Guidance Documents
  • Workshop & Training Materials
  • Lifecycle Management Tools

http://educopia.org/research/etd
Module 1
ETD Lifecycle Management

Interactive Overview
Learning Outcomes

1. Attendees will gain a clear understanding of information lifecycle management models and how they apply to ETD programs.

2. Attendees will understand how their current ETD program responsibilities fit within a larger lifecycle for managing ETDs.

3. Attendees will understand how they can better coordinate with other ETD program stakeholders at key stages of curation activity.
Lifecycle Management for ETDs

• ETD programs provide policies, workflows, and services around crucial lifecycle curation functions:
  – creation guidelines
  – deposit/submission
  – documenting approvals
  – metadata capture
  – rights management
  – ingest into commercial and/or library-based repositories
  – cataloging, access and usage monitoring
Lifecycle Management for ETDs

U.S. Federal Law 44 U.S.C. 2901
ISO 15489

DCC Curation Lifecycle Model
Lifecycle Management for ETDs

http://www.dcc.ac.uk/resources/curation-lifecycle-model
ETD Lifecycle Workflows

Create & Receive

Appraise & Select

Manage

Store, Access & Re-Use

U.S. Federal Law 44 U.S.C. 2901
ISO 15489
ETD Stakeholders on the Lifecycle

Student authors create & submit ETDs

Graduate Schools process, approve & embargo ETDs

Graduate Schools & Libraries/Vendors also update ETDs

Libraries/IT/Vendors disseminate ETDs

Libraries/IT/Vendors catalog & archive ETDs
Interactive Exercise

• What Phase is your ETD Program/Service in right now?
  – Planning
  – Development
  – Testing
  – Implementation
  – Assessment/Improvement

• **What questions/issues are you grappling with? Write one or two down on paper**

• Where do these questions/issues fall on the lifecycle?
Lifecycle Management for ETDs
Questions?
Module 2
ETD Guidance Documents

Overview & Usage
Learning Outcomes

1. Attendees will gain an understanding of the full range of lifecycle management activities for ETDs

2. Attendees will understand how each Guidance document is most relevant to various ETD program stakeholders and lifecycle management stages

3. Attendees will understand how to make use of the documents in practical ways depending on their stage of ETD program planning or implementation
Guidance Documents for ETD Lifecycle Management

1. Guidelines for Implementing ETD Programs - Roles & Responsibilities
2. Guide to Access Levels and Embargoes of ETDs
3. Briefing on Copyright Issues and Fair Use in ETDs
4. Guidelines for Collecting Usage Metrics & Demonstrations of Value for ETD Programs
5. Managing the Lifecycle of ETDs: Curatorial Decisions and Practices
6. Metadata for ETD Lifecycle Management
8. Guide to Options for ETD Programs
Guidelines for Implementing ETD Programs – Roles & Responsibilities

• 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
Guidelines for 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
Guidelines for 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
Guide to Access Levels & Embargoes of ETDs

• 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.
Guide to Access Levels & Embargoes of ETDs

- 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?
    - Some 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 from:
  - No embargo
  - Blanket fixed-length embargo
  - Limited list of fixed-length embargo
  - Embargo renewal
Briefing on Copyright Issues & Fair Use in ETDs

• 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?
Briefing on Copyright Issues & Fair Use in ETDs

• 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) agreements 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?
Guidelines for Collecting Usage Metrics & Demonstrations of Value

• 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.
Guidelines for Collecting Usage Metrics & Demonstrations of Value

- **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.
Guidelines for Collecting Usage Metrics & Demonstrations of Value

- Benefits of Usage Metrics
  - Authors, faculty and graduate students
    - Impacts and importance of a title
    - Impacts of their ETDs
- Institution
  - Impacts of the ETD collection
  - Return on Investment (ROI)
  - Satisfaction of graduate students
- Scholarly Society
  - Impacts of ETDs vs. other content
Managing the Lifecycle of ETDs:
Curatorial Decisions & Practices

• 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.
Managing the Lifecycle of ETDs: Curatorial Decisions & Practices

• 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
Metadata for ETD Lifecycle Management

- 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, professional supervision provides quality control.
- The ETD repository software heavily influences the creation of metadata, but metadata should be thorough and comprehensive.
Metadata for ETD Lifecycle Management

- NDLTD has developed and maintains an ETD specific descriptive metadata schema.

- PREMIS stands for "PREservation Metadata: Implementation Strategies."
  - 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.
Guide to ETD Program Planning & Cost Estimation

• 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.
Guide to ETD Program Planning & Cost Estimation

• Case Study Interviews
  – Large and Small
  – Public and Private
  – Rural and Urban
  – Portland State University
  – Rice University
  – University of Arizona
  – University of North Texas
  – Virginia Tech
  – ETD program personnel?
  – Key technologies?
  – Resource allocation?
  – Etc.
Guide to Options for ETD Programs

• Benefits & Impacts of ETD Programs
  – Moving past debates and to implementations

• Information Resources
  – Lifecycle Management Project Bibliographies
  – NDLTD

• Access Policies & Intellectual Property Issues
  – Copyright violations & disputes
  – Access violations
  – Quality control
  – Embargo issues

• Deposit Procedures
  – Mandatory vs. optional
  – File formats
Guide to Options for ETD Programs

• 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
1. Guidelines for Implementing ETD Programs

2. Guide to Access Levels and Embargoes of ETDs

3. Briefing on ETD Copyright Issues and Fair Use in ETDs

4. Guidelines for Collecting Usage Metrics and Demonstrations of Value for ETD Programs

5. Managing the Lifecycle of ETDs: Curatorial Decisions and Practices

6. Metadata for ETD Lifecycle Management

7. Guide to ETD Program Planning and Cost Estimation

8. Guide to Options for ETD Programs

Lifecycle Guidance
Lifecycle Guidance

1. Guidelines for Implementing ETD Programs
   Roles & Responsibilities

2. Guide to Access Levels and Embargoes of ETDs

3. Briefing on ETD Copyright Issues and Fair Use in ETDs

4. Guidelines for Collecting Usage Metrics and Demonstrations of Value for ETD Programs

5. Managing the Lifecycle of ETDs: Curatorial Decisions and Practices

6. Metadata for ETD Lifecycle Management

7. Guide to ETD Program Planning and Cost Estimation

8. Guide to Options for ETD Programs

Graduate School
Library/IT

Vendors
Institutional Repository

Submission System
Discussion

• Break into groups of 2-3
• Talk to each other using the following questions
  1. Which topic seems most useful for your area of work and focus? Why?
  2. Using the Module 2 Handout, talk through a few questions based on your preferred document
  3. Who do you coordinate most closely with at your institution on ETDs? What document(s) might be most helpful to them?
How to Use These Documents

1. Share them with your ETD Program Manager(s)
2. Coordinate a targeted document review with individual ETD Program stakeholders
3. Organize a brown-bag lunch discussion w/ETD Program stakeholders
4. Document recommendations & next steps for program implementation or improvement
Questions?
Module 3
Lifecycle Management Tools

Overview
Learning Outcomes

1. Attendees will gain an understanding of the importance of modular curation for ETDs

2. Attendees will gain an understanding of how to begin making use of the project’s documented modular curation tools

3. Attendees will leave with a set of demonstrated examples of the tools in order to encourage their uptake and implementation
Exploring Modular Curation

- Modular curation refers to the use of single-function standalone services that can be used alone or incorporated into larger repository systems.
- These services are often referred to as micro-services.
- Micro-services are an approach to curation pioneered by the California Digital Library and now heavily used by groups like UNT, MetaArchive, Archivematica, and the Florida Digital Archive.
Exploring Modular Curation for ETDs

- **Methodology**
  - Hold interviews with project partners and external ETD programs
  - Research open source modular curation tools and micro-services
  - Research existing open source submission and repository software systems for interoperability with curation tools
Investigating Open IR Systems

- Dspace
  - Built-in support for format validation, virus checking & some preservation metadata (plugin architecture)
- EPrints
  - Built-in support for format validation, virus checking & some preservation metadata (plugin architecture)
- Open-ETD
  - No built-in support for any curation tools (no plugin architecture)
- ETD-db
  - No built-in support for any curation tools (no plugin architecture)
Overall Findings

• Only a few submission & IR software systems have plugin architectures.
• By the time ETDs arrive at the IR it can often be too late to apply curation services.
• Existing curation services have clear standalone usage instructions and available APIs.
• Many ETD Programs have technical staff at their disposal to help implement modular curation.
Exploring Modular Curation for ETDs

• Proposed Modular Curation Tools
  – ETD Format Recognition
  – Virus Checking
  – PREMIS Metadata Event Record-keeping
  – Digital Drop Box with Metadata Submission Functionality
Modular Curation Tools for ETDs

ETD Submission
ETD Drop • ETD Drop

Format Recognition
UNIX • Unix command-line programs
JHOVE2 • JHOVE/2
DROID • DROID
FITS • FITS

Virus Checking
Clam AV

Preservation Metadata
PREMIS Event Service (UNT)
Lifecycle Management Tools

- Conceptualise
- ETD Drop
- Submission
- Virus Checking
- DISPOSE
- UNIX
- JHOVE2
- FITS
- Format Recognition
- PREMIS Event Record-Keeping
- Archiving
- Reference Link
Get the Tools

ETD Lifecycle Management Tools Manual

Now available at:
http://www.educopia.org/research/etd/etdlmtm
Lifecycle Management Tools

Discussions of Issues in Ethiopian Higher Education Context
Discussion and Local Contexts

• Break into groups of 2-3
• Talk to each other using the following questions
  1. Which tools are most needed at your institution?
  2. Where do you turn for technical resources?
  3. What are the greatest barriers to your adoption of lifecycle management tools such as these?
Discussion and Local Contexts

• **Discuss the need and/or roles of:**
  – National and regional consortiums
    • *Consortium of Ethiopia Academic and Research Libraries (CEARL)*
  – National research and academic networks
    • *EthERNet*
  – National ETD Portal
    • *Ethiopian Theses and Dissertations (ETHD)*
Questions?
Thank You
አመሰግናለሁ

Daniel Gelaw Alemneh
University of North Texas
Daniel.Alemneh@unt.edu

This workshop series is made available from the above organizations under a Creative Commons CC-BY 4.0 International License