Lifecycle Management of ETDs:

Providing the ETDS of today to the researchers of tomorrow

Presenters: Katherine Skinner,
Bill Donovan, & Matt Schultz

Agenda

- Introduction and Project Overview
- Guidance Documents Overview
  - Options for ETD Programs
  - Implementing ETD Programs - Roles & Responsibilities
  - Formats, Complex Content Objects, and Format Migration Scenarios for ETDs
  - Metadata & Lifecycle Event Record-Keeping for ETDs
  - Collecting Usage Metrics & Demonstrations of Value for ETD Programs
  - ETD Program Cost Estimation and Planning
  - Access Levels and Embargoes of ETDs
  - ETD Copyright Issues and Fair Use
- Lifecycle Management Tools
- Summary and Future Plans

Why ETDs?

The Calf-Path by Sam Walter Foss

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 three 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 bellwether sheep
Pursued the trail o'er vale and steep,
And drew the flock behind him, too,
As good bellwethers 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 migrations of that calf…

…A hundred thousand men were led
By one calf near three centuries dead.
They follow still his crooked way,
And lose one hundred years a day,
For thus such reverence is lent
To well-established precedent.

Project Overview

- Project Name: Lifecycle Management of ETDs
- IMLS National Leadership Grant
  - Duration: October 2011 – September 2013
- Project Objective
  - To promote best curatorial practices and to increase the capacity of academic libraries to reliably preserve ETDs

Project Partners

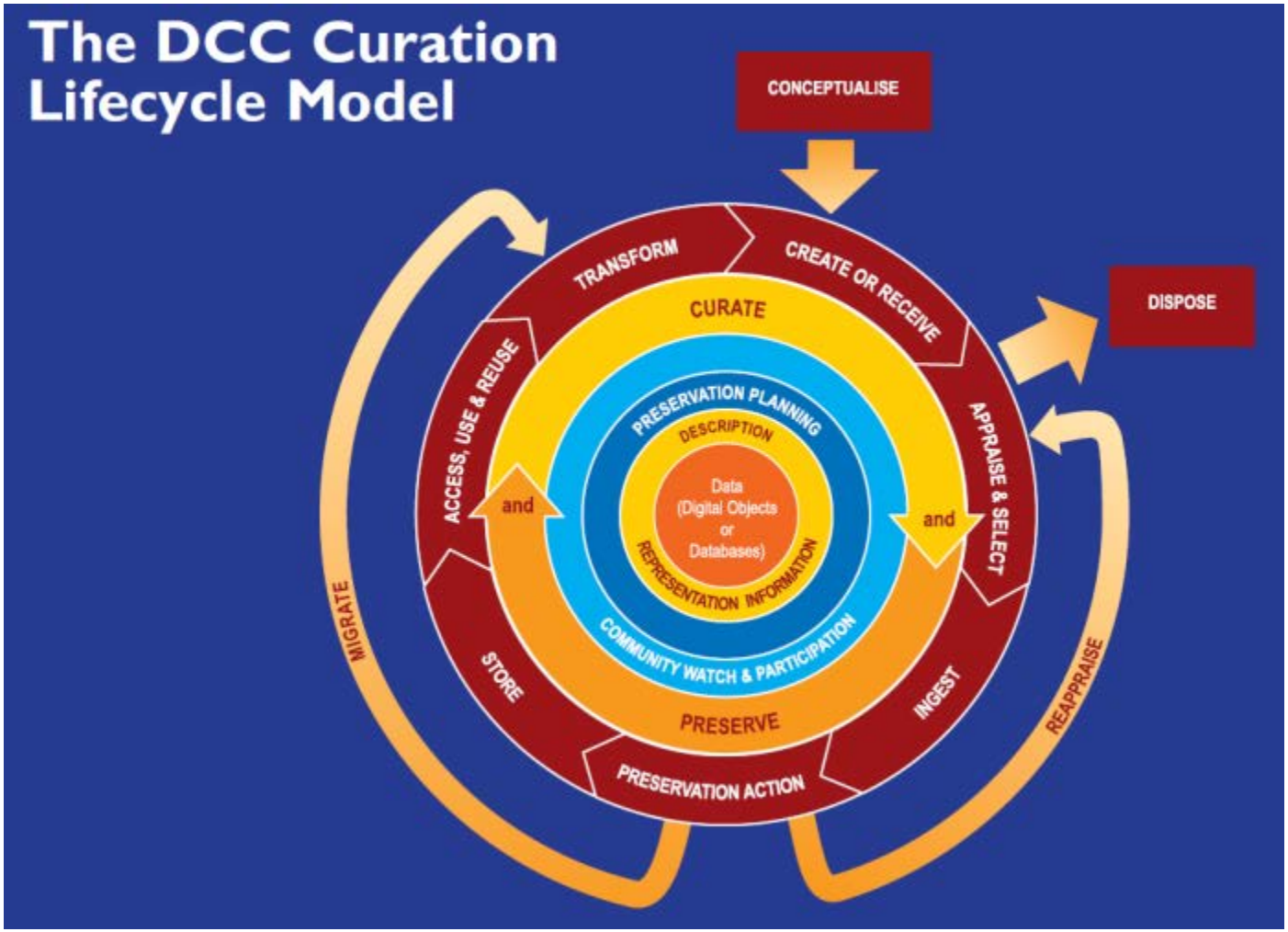
- University of North Texas
- Educopia
- MetaArchive
- Virginia Tech
- NDLTD
- Boston College
- Rice University
- Indiana State University
- Penn State University
- University of Arizona

Project Activities and Products

- Guidance Documents for Lifecycle Management of ETDs
- Lifecycle Management Tools
- Educational Materials
- Workshop

Guidance Documents

“Lifecycle management”



Introduction to ETDs

Principal Investigator: Katherine Skinner

katherine.skinner@metaarchive.org

Project Manager: Matt Schultz

matt.schultz@metaarchive.org

Introduction to ETDs & Guidance Docs

- Defining ETDs
- History of ETDs
- Stages of Development
- Interest Issues
  - Digitized vs. Born-Digital
  - Outsourcing vs. In-House
  - Lifecycle curation & challenges
- Guide to the Guidance Docs
  - Background, purpose & overview
  - How to read and use

Options for ETD Programs

Steering Committee Chair: Martin Halbert

martin.halbert@unt.edu

Options for ETD Programs

- Serves as a basic guide for decision-makers to key options in creating or overhauling an ETD service program
- Describes Option Pros and Cons
  - Restricted or Open Access?
  - Implement a repository or lease a commercial service?
  - Where does responsibility fall?
- This document will reference and integrate all the other documents in the series

Implementing ETD Programs

Steering Committee Chair: Xiaocan (Lucy) Wang

xiaocan.wang@indstate.edu

Implementing ETD Programs

- Types of Stakeholders:
  - Internal
  - External
- Roles and Responsibilities:
  - Program Planning
  - Program Implementation
    - Creation, Submission, and Ingest
    - Access
    - Archiving and Preservation
  - Program Assessment

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

Steering Committee Chair: Bill Donovan

bill.donovan@bc.edu

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

- ETD formats issues and considerations
  - Complex multimedia content objects
  - Metadata
  - Hyperlinks
  - Research data
- Guidance for what format to be used
  - Text
  - Images
  - Audio
  - Video
  - Others
- Format migration scenarios
  - Versioning
  - Fixity checking
  - Migration from one repository or system to another

Media-specific file formats

media	authoring software	original format(s)	open format(s)	comments
text	Microsoft Word LaTeX	.doc .docx .tex	.pdf (version 1.7)	PDF version 1.7 is an ISO standard, thereby making it effectively non-proprietary. See Appendix A below. For archival purposes, a PDF file should have no security features enabled, and should have all of its original fonts embedded (as a subset).
images	Adobe Photoshop	.psd	.tif .jp2	A small amount of compression applied by JPEG2000 can be "visually lossless", and may be acceptable for archival purposes.
audio		.mp3, .aif, .m4a	.wav	
video			.mp4 .avi	Further details at….
spreadsheet	Microsoft Excel	.xls .xlsx	.ods	See ODF (open document format)

Metadata & Lifecycle Event Record-Keeping for ETDs

Steering Committee Chair: Daniel Gelaw Alemneh

Daniel.Alemneh@unt.edu

Metadata & Lifecycle Event Record-Keeping for ETDs

- Overview of ETDs practices
  - Repository platforms , tools, standards, and guidelines
- Metadata for ETDs lifecycle management
  - Supporting Graduate students (restriction, embargoes)
  - Supporting Graduate Schools (submission requirements,)
  - Facilitating access, use, reuse, copyright management, and preservations activities
- Issues and considerations
  - What information needs to be captured? When?
  - Who creates metadata?
  - Quality assurance mechanisms

Collecting Usage Metrics & Demonstrations of Value for ETD Programs

Steering Committee Chair: Yan Han

hany@u.library.arizona.edu

Collecting Usage Metrics & Demonstrations of Value for ETD Programs

- Collecting Usage Metrics
  - Quantitative approaches
  - Qualitative approaches
- Demonstrating ETD value in terms of:
  - Benefits to scholarly communications
  - Benefits to university
  - Benefits to students
- Evaluation and use of data
  - Analysis and interpretation of statistics
  - Comparisons with other collections (licensed, open access, etc.)

ETD Program Cost Estimation and Planning

Steering Committee Chair: Gail McMillan
gailmac@vt.edu

ETD Program Cost Estimation and Planning

- Overview of the costs of the entire life cycle of ETDs
  - cost components in the initial (i.e., set up) stages of operation,
  - during the early stages of development, over a yearly time frame
- Context and costs associated with the full range of ETD initiatives
  - born digital vs. digitization of analog
  - new (i.e. initial set up) vs. an existing initiative,
  - university-wide vs. a single university unit strategy.
- Anticipated Expenses
  - Putting the process (submission, approval, access: SOP) in place
  - Maintenance for ensuring the long-term access

Access Levels and Embargoes

Steering Committee Chair: Geneva Henry

ghenry@rice.edu

Access Levels and Embargoes

- What are access restrictions for ETDs?
  - Embargoes
  - Restricted access based on permissions
- Why restrict access?
  - A multitude of reasons given
  - Need to understand policies and consistency in applying them
- How are access restriction decisions made and enforced?
  - Look at whether or not decisions are on a case-by-case basis or based on general policy
  - Understand how enforcement occurs and what notification, if any, there is upon release

Copyright Issues and Fair Use

Steering Committee Chair: Patricia Hswe

phswe@psu.edu

Copyright Issues and Fair Use

- Overview of ETD copyright and fair use
  - What can be copyrighted and what is not copyright protected?
  - International copyright considerations
- Who provides copyright and fair use guidance?
  - Graduate students need to know who to consult with on these issues
  - Is the guidance for students, their advisors?
- What are the institution's policies regarding intellectual property rights?
  - Do student's own their copyright or does the university?
  - Are there funder mandates impacting the copyright? Publisher concerns?

Lifecycle Management Tools

Lifecycle Management Tools

- Deliverable: Develop and disseminate a set of software tools to address specific needs in managing ETDs throughout their lifecycle
- Created as completely modular micro-services
  - Single-function standalone services
  - Can be used alone; or
  - Incorporated into larger repository systems
- Pioneered by CDL & Library of Congress
  - Now being used by UNT, UKY, Penn State, Chronopolis & MetaArchive

Lifecycle Management Tools

- ETD Format Recognition Service
  - Accurate identification of ETD component format types
- PREMIS Metadata Record-Keeping Service
  - PREMIS Event semantic units to track a set of transitions in the lifecycle of particular ETDs
- Virus Checking Service
  - Check ETD component files using ClamAV, etc.
- Digital Drop Box
  - Deposit ETDs via a webform & gather requisite submission information
- Plagiarism Tool (under consideration)

Summary and Directions

Summary

- Project completion scheduled for end of 2013
  - Nine guidance documents will be produced
  - Educational Materials and Workshops
  - Software tools for Lifecycle Management
- Workshop at International ETD conference in Lima, Peru in September: http://www.etd2012.edu.pe
- Training materials will be available in 2013, including workshop syllabi, training handouts and exercises and PowerPoint presentations

Project Contacts

- PI: Dr. Martin Halbert
  Martin.Halbert@unt.edu
- Co-PI: Katherine Skinner
  katherine.skinner@metaarchive.org
- Project Manager: Matt Schultz
  matt.schultz@metaarchive.org
- Assistant Coordinator: Shannon Stark
  Shannon.Stark@unt.edu
- Project Wiki:
  http://metaarchive.org/imls/index.php/Main_Page