

Morphing Metadata: Automated Support for Cataloging Electronic Theses and Dissertations

Mike Kreyche
 Systems Librarian, Associate Professor

Sevim McCutcheon
 Monographs Cataloger, Assistant Professor

 Kent State University
 Libraries and Media Services

Eastern Great Lakes IUG Annual Conference
October 20, 2006

1

Comparison of Print and Electronic TDs

CHARACTERISTIC	PRINT	ELECTRONIC
Processing time	4-6 weeks	Instant access
Cataloging time	Same	Same
Cost (student's)	\$106.50	None to minimal
Shelf space	Yes	No
Media of content	Limited: text, illustrations	Multimedia possible
Availability	Library hours and building	24/7; worldwide
Searchability	Visual only	Keyword

2

- ### Cataloging ETDs: Differences and Similarities
- Differences:
 - No physical object to track or examine
 - Author-supplied metadata as record basis
 - Similarities:
 - It's still a monograph with a title page
 - It's still a dissertation, requiring complex subject analysis
- 3

- ### ETD Project Timeline and Today's Topics
- OhioLINK ETD Center established, 2000
First KSU ETD, November 2004
1. Automated process created, 2005-2006 (Cataloging bot)
 2. Statewide cataloging standards developed, 2005-2006
 3. Established local standards based on the statewide baseline, 2006
- 4

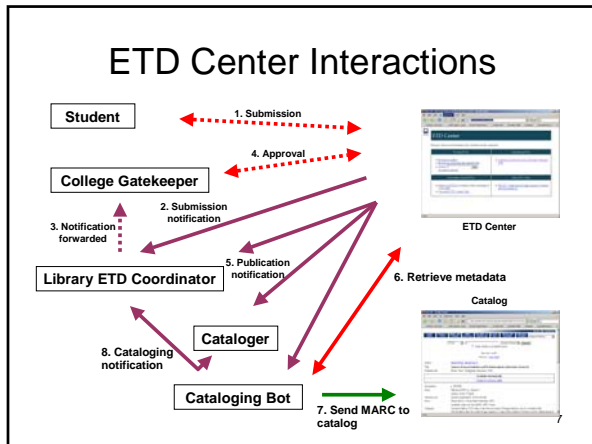
OhioLINK ETD Center

[ETD Center web site](#) [NDLTD web site](#)

5

OAI-PMH

6



Publication Notification

8

ETD Center OAI-PMH URLs

TWO DIFFERENT FORMATS

ETD-MS: ETD Metadata Standard
http://www.ohiolink.edu/etd/oai.php?verb=GetRecord&metadataPrefix=oai_etdms&identifier=kent1122136806

DC: Dublin Core
http://www.ohiolink.edu/etd/oai.php?verb=GetRecord&metadataPrefix=oai_dc&identifier=kent1122136806

9

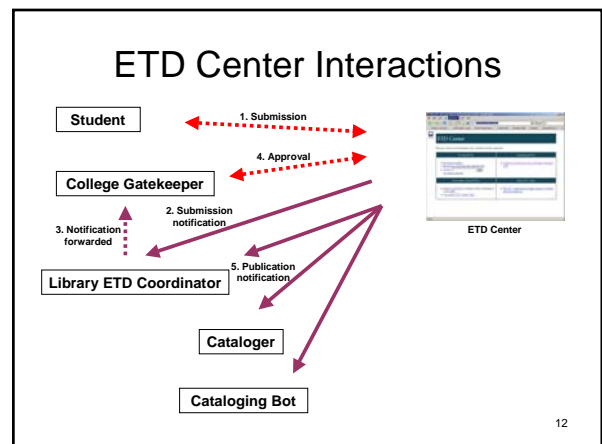
ETD Center Data Sources

Dublin Core	ETD-Metadata Standard	view.cgi	marcview.cgi
Title	Title	245	
Creator	Creator	Author	100
Description	Description	Abstract	520
Subject	Subject	Subject headings	650
Publisher	Publisher		
Date	Date	Degree	008, 260c, 502
	Contributor (advisor)	Advisor	500
Type	Type		
	Format		538
Language	Language	008	
	Degree name	Degree	502
	Degree discipline	Degree	
	Degree grantor	Degree	502
Identifier	Identifier		035, 856
Rights (G)	Rights (z)		
	Pages		300
	Keywords		256 (file size)

- Dublin Core and ETD-MS: XML amenable to automated harvesting
- view.cgi: HTML
- marcview.cgi: Text formatted with white space

10

- ### How The Bot Works
1. Waits for email
 2. Parses email for document ID
 3. Retrieves metadata
 4. Parses record for useful data
 5. Builds MARC record
 6. Sends record to local catalog
 7. Notifies staff
- 11



Email Notification

Subject: OhioLINK ETD: New paper published - kent1155924832
 From: etd@kent.edu
 Date: 2006-10-19PM
 To: etd@kent.edu, etd@kent.edu, etd@kent.edu

This paper has been published on the OhioLINK Electronic Thesis and Dissertation Center:
 kent1155924832

Direct URL:
http://www.ohiolink.edu/etd/view_of?ident=1155924832

Prepared by: etd@kent.edu
 MARC record for this ETD:
<http://www.ohiolink.edu/etd/marcview.cgi?kent1155924832>

[You are receiving this message because your address is on file with the OhioLINK ETD Center either as the grant recipient, administrative contact or as the library tech services contact.]

13

OAI-PMH Metadata

http://www.ohiolink.edu/etd/oai.php?verb=GetRecord&metadataPrefix=oai_etdms&identifier=kent1155924832

14

Builds MARC Record

- Data extracted from XML file
 - I used regular expressions
 - There are better ways to do it!
- Local standards document
 - “Constant” data and formatting
- MARC/Perl module
 - Routine for easy processing of MARC records
 - <http://marcpm.sourceforge.net/>

15

Send MARC Record to Catalog

- Direct communication between Bot and local system
 - “OCLC” interactive interface on Ill system
 - “Bidirectional forking client” from Perl Cookbook, slightly modified
- Record download a possible alternative
 - Import to OCLC or local system
 - Manual process

16

Cataloging Notification

Subject: ETD brief cataloging notice kent1155924832
 From: etd@kent.edu
 Date: 2006-10-19PM
 To: etd@kent.edu, etd@kent.edu, etd@kent.edu

ETD kent1155924832 has been processed.
 Log on at http://www.ohiolink.edu/etd/view_of?ident=1155924832
 Record added to KentLINK as 330820601:
<http://kentlink.kent.edu/record=330820601>

LDR \$m \$d \$s
 004 \$m \$d
 007 \$c \$m
 008 060824e2006 \$bu \$m 000 0 \$mg \$d
 010 \$a(OCLC)kent1155924832
 100 1 \$aMarunowski, Kenneth Ray.
 245 14 \$1The Euro :
 \$1electronic resource
 \$2Multinational Study in Presence.
 260 \$fKent, Ohio :
 \$fKent State University,
 \$c2006.
 \$p.
 300 \$pPDF file.

17

Brief Record in KentLINK

LANG	eng	English	CAT DATE	...	DISPLAY	DIS LOC	DCINT	
SNP	4							
LOCATION	kent: Kent Campus							
MARC Leader	##### 229999ha \$500							
MARC	006	\$m	\$d					
MARC	007	\$c	\$m					
MARC	010	\$a(OCLC)kent1155924832						
BNPNTD	005	OCLC#kent1155924832						
AUTHOR	100	1	Marunowski, Kenneth Ray					
TITLE	245	1	4	The Euro [electronic resource] in a Multinational Study in Presence				
SUBJECT	700			Kent, Ohio : Kent State University, : 2006				
DESCRIPTION	300			p. PDF file				
NOTE	500			Title from PDF file, keyword :				
NOTE	500			Author: Christina Haas				
NOTE	530			System requirements: Adobe Reader				
NOTE	502			Thesis (Ph.D.)--Kent State University, 2006				
NOTE	500			Available online via OhioLINK's ETD Center				
NOTE	520	3		Through the exemplary case study of the euro, the official currency of the European Union, this dissertation explores Perelman and Orlitzky's (1985) historical construct, presence as a multinational phenomenon. Presence, according to Perelman and Orlitzky, is fundamental to				

18

Brief Record in KentLINK

00820601 Last Update: 08-24-2006 Created: 08-24-2006 Revisions: 1

is a material object constructed of co-occurring sign systems, each of which contributes to the aggregate meaning of a text in its own way. To examine presence as a multimodal phenomenon is to consider the ways that different modes of representation and communication are, at times, enhanced and, at other times, suppressed so as to focus the attention of an audience with the intent of securing that audience's adherence. Presence, in this respect, is a cumulative effect resulting from the interaction of the different semantic modes drawn together in a particular line of argumentation. I examine presence as a multimodal phenomenon to extend current studies in presence that focus primarily on linguistic techniques for its enhancement or suppression. Moreover, through an examination of presence as a multimodal phenomenon, I extend contemporary investigations of multimodal discourse analysis to include how different modes of representation and communication not only mean but also persuade. Through my analysis of the presence conferred to the television spots and print ads of the Euro 2002 Information Campaign and for each headline and comic, I discover the ways in which the Campaign Union attempts to secure the adherence of its culturally and linguistically diverse citizens as it prepares them for the largest monetary changes in history. Together, the systematic enhancement and suppression of national and supranational references drawn together in the multimodal texts that serve as my objects of study results in a compelling superordinate presence that speaks to the compatibility of nationalism and European integration and the ideals and values that both perspectives embody.

NOTE 504 Includes bibliographical references (p. 3)
 AL1 NAME 710 2 OhioLINK Electronic Theses and Dissertations Center
 AL1 TITLE 793 0 KSU dissertations (Dept. of English)
 AL1 TITLE 793 0 ETD Word Record
 MARC 856 4 0 http://on.athens.edu/etd/viewDoc.cfm?doc=00820601:Connect to resource online

OhioLINK Committee Discussions

- Born digital vs. reproduction
 - Available first, before book or microfiche = born digital
- Published vs. unpublished
 - Born digital thus published. Jay Weitz's email

When ETDs considered published...

- FF: Record Type (Leader/06) "a," not "t"
 - KSU uses z in Mat Type (BCODE2) for both electronic and print Theses and Dissertations
- FF and 260 \$a, \$b: Place of publication and publisher
- FF: in KSU's case, GPub (008/28) = s

Other Standards Issues

- Physical description (300): We include
- Size of file: 300? 256? 516?
- 538: wording, content
- Author-supplied keywords: Include or not? Where?
- UMI subject descriptors: 650 4, 653, elsewhere?

Options KSU Considered

- Not Used
 - UMI subject descriptors
 - 653 for keywords (500 instead)
- Used
 - 500 with advisor's name
 - 793 local access point, serves as a series "KSU Dissertation (Dept. of Physics)"

Once the Bot's work ends, the Cataloger...

- Exports brief record from KentLINK to OCLC save file
- Upgrades to full record
- Contributes record to WorldCat
- Overlays KentLINK brief record with OCLC Full Record

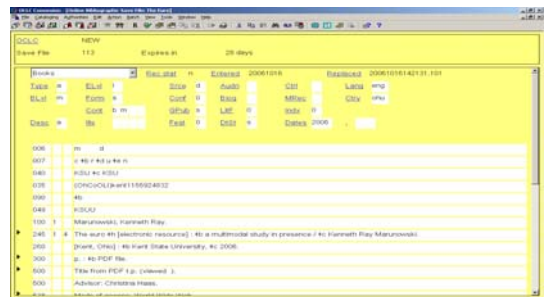
Putting it into practice



Josh Nickerson, MLS, available for employment

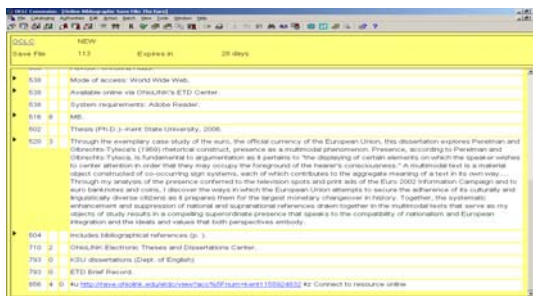
25

Brief Record in OCLC save file



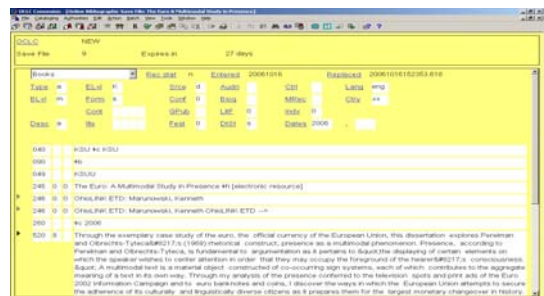
26

Brief Record in OCLC save file



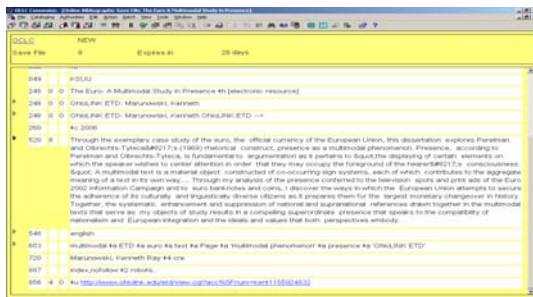
27

Record made via OCLC's Metadata Extraction process



28

Record made via OCLC's Metadata Extraction process



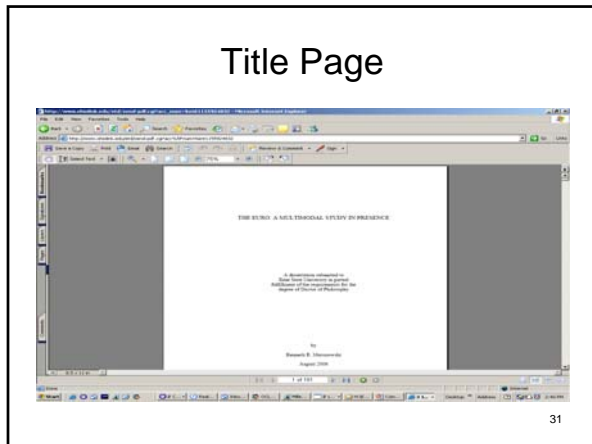
29

Abstract



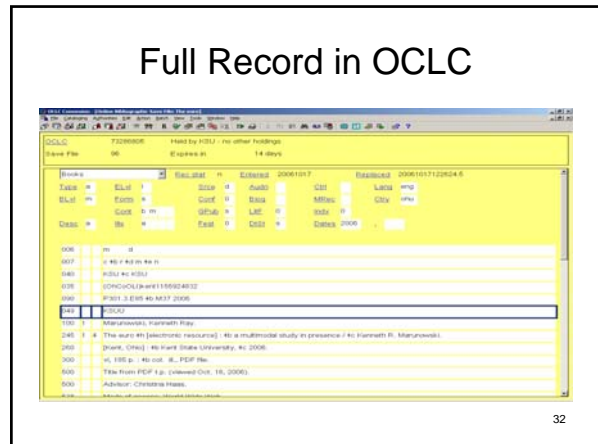
30

Title Page



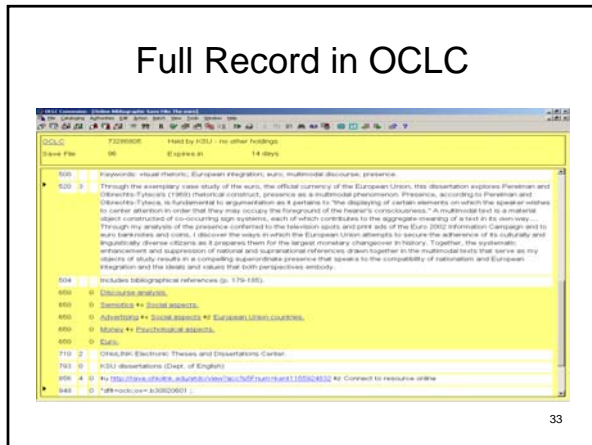
31

Full Record in OCLC



32

Full Record in OCLC



33

Full Record in KentLINK



34

Final words of advice

- Be flexible
- Be visionary
- Be practical
 - Meet local needs
 - Is brief/preliminary record adequate?

35

Any Questions?

Mike Kreyche – mkreyche@kent.edu
 330-672-1918
 Sevim McCutcheon Lmccutch@kent.edu
 330-672-1703

36