



Customized Mapping and Metadata Transfer from DSpace/SOAR to OCLC to Improve ETD Work Flow

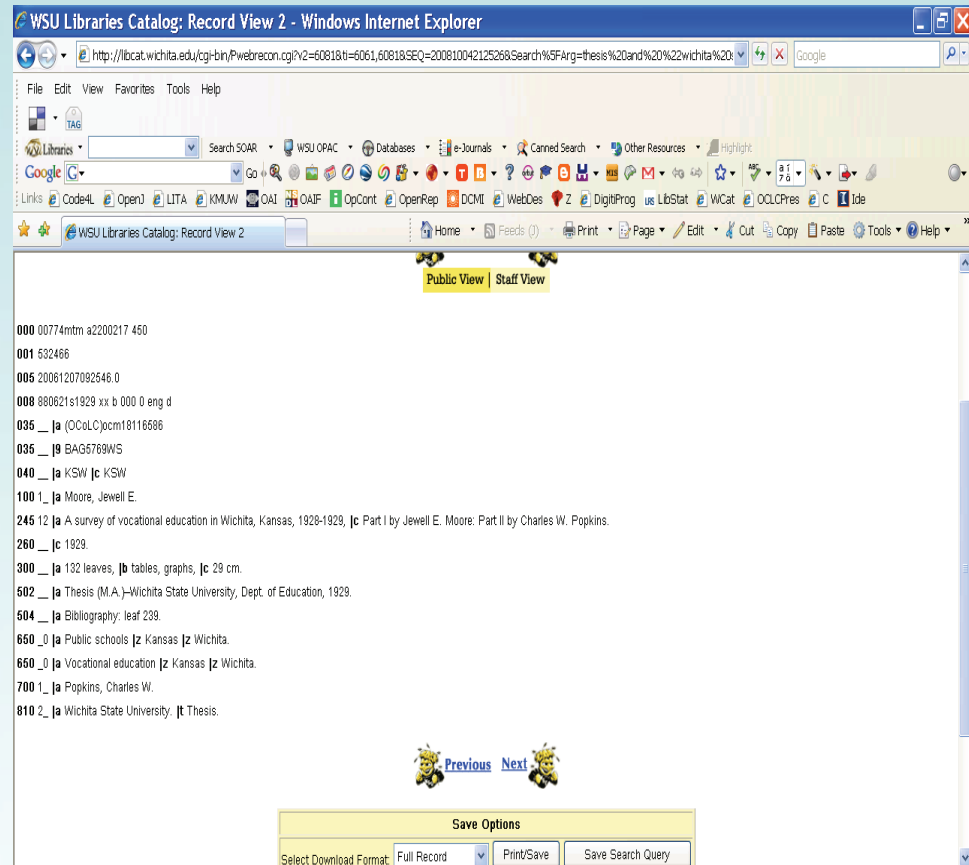
- Sai Deng, Susan Matveyeva, Tse-Min Wang, Wichita State University Libraries
- Consultant: Terry Reese, Oregon State University Libraries

Outlines

- ❑ Thesis Cataloging Workflow Dynamics: overview of changes
- ❑ Cataloging ETDs in SOAR and OCLC/Voyager: records & workflow
- ❑ Improving ETD Workflow through metadata harvesting, customized mapping and metadata transfer

Workflow for Paper Theses

- ❑ 1929-2002 – over 80% records (~ 5000)
- ❑ 70 year range: stable record's structure
- ❑ Workflow: (1) original cataloging (2) item's marking/ labeling
- ❑ Cataloging efficiency: constant data
- ❑ Labor intensive: SH



The screenshot shows a Windows Internet Explorer browser window displaying a record from the WSU Libraries Catalog. The address bar shows the URL: <http://libcat.wichita.edu/cgi-bin/Pwebrecon.cgi?v2=60818&t=6061,60818&SEQ=200810042125268&Search%5FArg=thesis%20and%20%22wichta%20>. The page title is "WSU Libraries Catalog: Record View 2 - Windows Internet Explorer". The browser's menu bar includes File, Edit, View, Favorites, Tools, and Help. The toolbar contains various icons for navigation and search. The main content area displays a MARC record for a thesis, with fields such as 000, 001, 005, 008, 035, 040, 100, 245, 260, 300, 502, 504, 650, 660, 700, and 810. The record details include the title "A survey of vocational education in Wichita, Kansas, 1928-1929", the author "Jewell E. Moore", and the publisher "Wichita State University". At the bottom of the page, there is a "Save Options" section with a "Select Download Format" dropdown menu set to "Full Record", and buttons for "Print/Save" and "Save Search Query".

Thesis MARC Record (till 2002)

_000 01093nam a2200277 i 450 001 331612 005 19991028065706.0 008 780705s1977 ksu 000 0 eng d
035 __ |a (OCoLC)ocm04023056
035 __ |9 ABK7544WS
040 __ |a KSW |c KSW
099 __ |a LD|a 2667 |a .T4 |a V871d
100 1_ |a Vliet, Martha Tasheff.
245 12 |a A descriptive study of obstetric patients' knowledge of and self reported attitudes toward the prenatal experience / |c by Martha Tasheff Vliet.
246 3_ |a Patients' perceptions of prenatal experience
260 __ |a Wichita, Kan. : |b WSU, |c 1977.
300 __ |a viii, 75 leaves ; |c 29 cm.
490 1_ |a Wichita State University. Theses
500 __ |a Also in University Archives: THESIS.
500 __ |a Title on spine: Patients' perceptions of prenatal experience.
502 __ |a Thesis (M. Ed.) - Wichita State University, December 1977. Department of Instructional Services.
504 __ |a Bibliography: leaves 48-52.
650 _0 |a Pregnancy.
650 _0 |a Pregnancy |x Psychological aspects.
650 _0 |a Prenatal care.
810 2_ |a Wichita State University. |t Thesis.

Theses Digitization, Workflow & Records

- ❑ 2003-2004 digitization of WSU Theses began
- ❑ UMI/ProQuest effects workflow
- ❑ Linking Voyager records to UMI/ProQuest

WSU Libraries Catalog: Record View 1 - Windows Internet Explorer

http://libcat.wichita.edu/cgi-bin/Pwebrecon.cgi?BBID=1172115

File Edit View Favorites Tools Help

Libraries TAG Search SOAR WSU OPAC Databases e-Journals Canned Search Other Resources Highlight

Go Google

Links Code4L Open1 LITA KMIJW OAL OAJF OpCont OpenRep DCMI WebDes Z DigitProg us LibStat WICat OCLCPres C Ide

WSU Libraries Catalog: Record View 1 - Windows Internet Explorer

Home Feeds () Print Page Edit Cut Copy Paste Tools Help

Main Author: [Perera, Bupani Asiri.](#)

Title: A comparison of multiple-stage tandem MS of protonated and metal cationized peptides in the context of direct sequencing and sequence tag generation / by Bupani Asiri Perera.

Electronic Resource: [Click here for available full-text of this dissertation via Current Research@Gateway.](#)

Published: 2004.

Description: xiv, 136 leaves : ill. ; 29 cm.

Series: [Wichita State University. Thesis.](#)

Thesis/Dissertation: Thesis (Ph.D.)--Wichita State University, College of Liberal Arts and Sciences, Dept. of Chemistry.

Summary: [Author abstract] We have examined the multiple stage collision induced dissociation (CID) of protonated and metal cationized peptides and the N acetylated and N versions of the peptides using ion trap mass spectrometry. An assessment of the accuracy of the MSⁿ experiment for sequence determination was made. Thus far, our survey shows several advantages to the MSⁿ of metal cationized peptides. In most cases, MSⁿ of metal cationized peptides is more accurate than experiments involving protonated versions. However, MSⁿ of LH⁺ and Na⁺ cationized peptides produces accurate sequence information only when the C terminus is not modified: CID of alkali cationized peptides with C terminal ester and amide moieties causes the loss of H₂O as the dominant fragmentation process, and MSⁿ leads to dissociation schemes that are difficult to interpret. Ag⁺ cationized peptides are less sensitive to a modified C terminus because of a lower dependence on the (b_n-17+Cat)⁺ product (the main product ion generated from alkali metal cationized peptides) for sequence information. Modification of the N terminus, by acetylation or benzylation, improves the efficiency, and reduces the ambiguity, of the MSⁿ experiment primarily by increasing the intensity of (b₁+17+Cat)⁺ and (b₁+Cat)⁺ species at high CID stages. For the former, the positive influence is likely attributable to the addition of a carbonyl group to coordinate the metal ion. In the latter case, the added carbonyl group can function as a nucleophile to produce the cyclic oxazolone structure necessary for the formation of the (b_n-1) type ions (b₅). Also we have compared the MS/MS spectra of a range of native and acetylated Ag⁺ cationized peptides to determine the influence of the derivatization step on the intensity of the (b_n-17+Ag)⁺ product ions. Using tripeptides, the smallest for which the mechanisms to generate (b₂-1+Ag)⁺ and (b₂-17+Ag)⁺ products are both operative, we found that in most cases acetylation causes an increase in the intensity of the C terminal rearrangement ion, (b₂-17+Ag)⁺, relative to the rival N terminal rearrangement ion, (b₂-1+Ag)⁺. The presence of a free amino group to bind to the metal ion significantly influences the relative intensities of the product ions. We propose a mechanism for the formation of the (b_n-17+Ag)⁺ that is based on the formation of a 5 membered oxazolidin 5 one and tetrahedral carbon intermediate that may collapse to a peptide upon release of CO and an imine, aided by the fact that the ring formed during C terminal rearrangement is both a hemiacylal and hemiaminal.

Notes: "July 2004."
Thesis advisor: Michael J. Vanstipdonk.
Includes bibliographical references (leaves 128-136).

Other Author(s), etc.: [Vanstipdonk, Michael J.](#)

Location: Abiah (A-L on Floor3)(M-QR on Floor2)(R-ZA on LowerLevel)

OCCLC/Voyager- UMI/ProQuest

- ❑ Record enhancements (fields /contents)
- ❑ 856 -links from a catalog to full text in UMI
- ❑ 520 – author abstracts
- ❑ 500 & 700 -- advisor's name
- ❑ Workflow changes: Special projects: a repetitive data entry goes to students
- ❑ Cataloger creates procedure; MACRO for speedy processing; trains students, and review their work

Thesis Bib Record 2004 (MARC)

000 03794ctm a22002891a 45 001 1172115 005 20070208132604.0 008 050201s2004 xx a bm 000 0 eng d

035 __ |a (OCoLC)ocm57545066

035 __ |a 1172115 040 __ |a KSW |c KSW

049 __ |a KSWA

050 _4 |a LD2667.T42 |b P437733

099 _9 |a Microfilm 1391

100 1_ |a Perera, Bupani Asiri.

245 12 |a A comparison of multiple-stage tandem MS of protonated and metal cationized peptides in the context of direct sequencing and sequence tag generation / |c by Bupani Asiri Perera.

260 __ |c 2004.

300 __ |a xiv, 136 leaves : |b ill. ; |c 29 cm.

502 __ |a Thesis (Ph.D.)--Wichita State University, College of Liberal Arts and Sciences, Dept. of Chemistry.

500 __ |a "July 2004."

500 __ |a Thesis advisor: Michael J. Vanstipdonk.

504 __ |a Includes bibliographical references (leaves 128-136).

520 8_ |a [Author abstract] We have examined the multiple stage collision we bind to the metal ion significantly

700 12 |a Vanstipdonk, Michael J.|e advisor

810 2_ |a Wichita State University. |t Thesis.

856 40 |u <http://proxy.wichita.edu:2048/login?url=http://wwwlib.umi.com/cr/wichita/fullcit?p3137654> |z Click here for available full-text of this dissertation via Current Research@Gateway.

994 __ |a C0 |b KSW

Transitional Period: 2004-2006

- ❑ e-Theses in four places: OCLC/Voyager; ProQuest; a temporary web site and SOAR
- ❑ Paper theses are still submitted
- ❑ Development of a new workflow for ETDs
- ❑ e-docs, paper docs, inventory table
- ❑ Naming convention, ETD file preparation
- ❑ MARC and DC manual input; further changes in records (identifiers)

00003279ctm a2200433la 450

001 1245843

00520080422003723.0

006m d

007cr m|||||

008070423s2005 xx a sbm 000 0

020 __ |a **9780542757921**

020 __ |a **0542757923**

0247 _ |a **AAT 1436580 |2 UMI**

0248 _ |a **778 SOAR**

035 __ |a (OCoLC)ocn**123426976**

035 __ |a 1245843

040 __ |a KSW |c KSW

049 __ |a KSWA

099 _9 |a Microfilm 1502

099 __ |a **t05040**

1001 _ |a Radhakrishnan, Preetha.

24510 |a Enhanced routing protocol for graceful degradation in wireless sensor networks during attacks |h [electronic resource] / |c by Preetha Radhakrishnan.

260 __ |c 2005.

300 __ |a xii, 50 leaves : |b ill., digital, PDF file.

500 __ |a "December 2005."

504 __ |a Includes bibliographic references (leaves 48-50).

500 __ |a Title from PDF title page (viewed on April 23, 2007).

533 __ |a Electronic reproduction. |b Ann Arbor, MI : |c ProQuest Information and Learning Company, |d c2006.

538 __ |a System requirements: Adobe Acrobat Reader.

538 __ |a Mode of access: World Wide Web.

502 __ |a Thesis (M.S.)--Wichita State University, College of Engineering, Dept. of Electrical and Computer Engineering.

500 __ |a Thesis adviser: Ravi Pendse.

500 __ |a UMI Number: AAT 1436580

5203 _ |a [Author's abstract] With the deployment of Sensor networks gaining some ...

655 _0 |a Electronic dissertations.

70012 |a Pendse, Ravindra.|e advisor

85640 |u <http://proxy.wichita.edu:2048/login?url=http://www.lib.umi.com/cr/wichita/fullcit?p1436580> |z Click here for available full-text of this thesis via Current Research@Gateway.

856 40 |u <http://soar.wichita.edu/dspace/handle/10057/778> |z A link to full text of this thesis in SOAR

New additions to ETD record: identifiers of several databases that have this thesis

Record consists of 30 fields



ETD Program 2006-2008

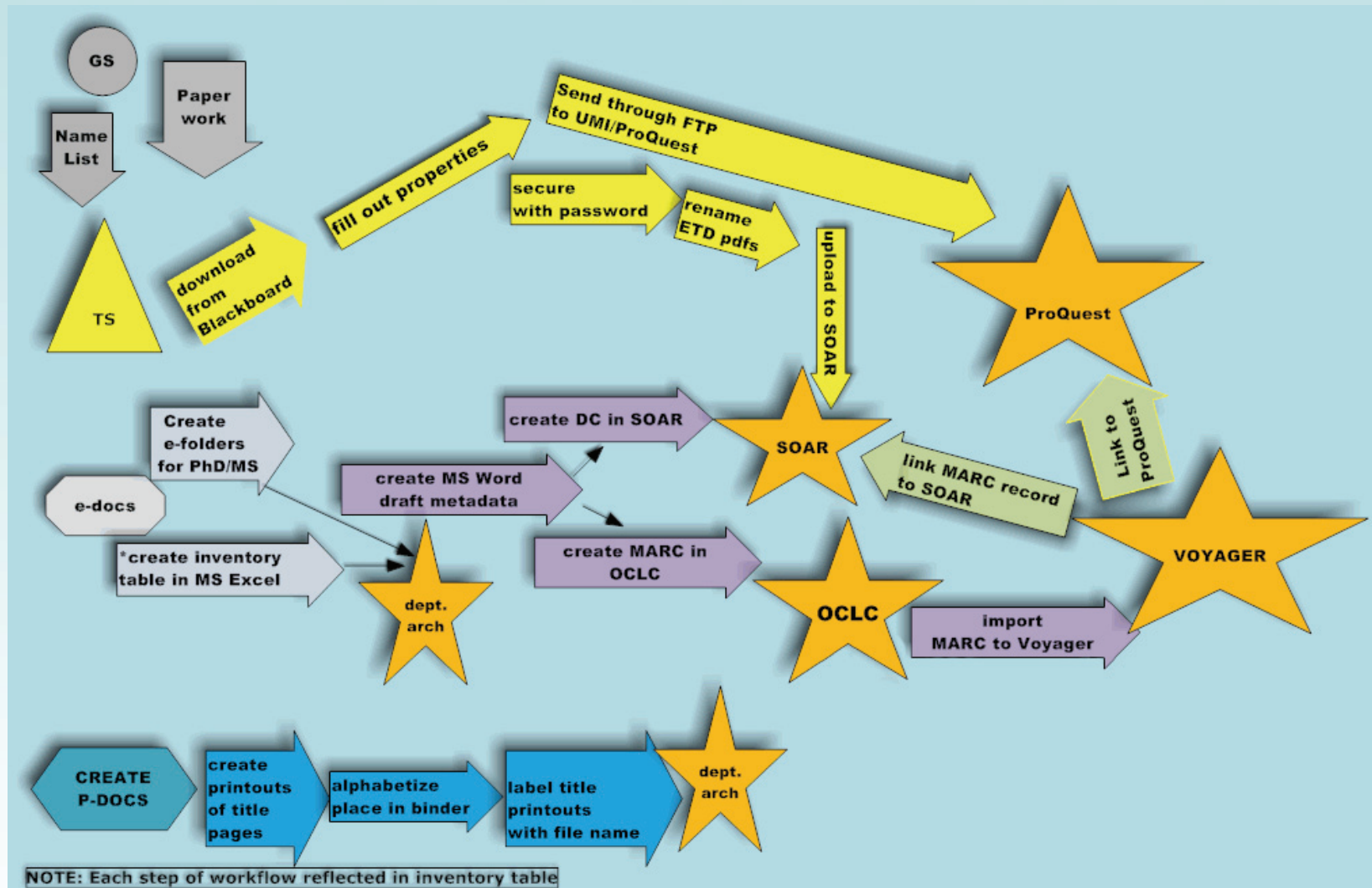
- ❑ From 2006, WSU have a full scale ETD program (400 records, 2005-2007)
- ❑ eTheses (no paper); no ProQuest or a temporary access to ETD via a web site
- ❑ eTheses are in three databases: SOAR and OCLC/Voyager
- ❑ Work Flow includes the number of operations with a digital file (thesis) and metadata records (MARC and DC)

Inventory Table

Pdf ID	No	Last First Name	Year	Mon.	GS send list	PDF Harvested	PDF Property filled	PDF Subm To UMI	PDF secured
d07001	1	Smith John	2007	May	date	date	date	date	date

PDF re-named	GS Paper work received	Soar ID	Voyager Bib	UMI ID	UMI Link	Soar Link	Micr film No	Link Checked	Note
date	date	1074	1262388	32408 65	Yes/no	Yes/no	2740	date	

ETD Workflow: Manual Input DC & MARC



A Wider Context of ETD Workflow

- ❑ ETD workflow in different institutions
 - ❑ University of Virginia (1999), Texas A & M (2004)
 - ❑ Home-grown scripts, site-specific harvesters
 - ❑ Kent State University (2007)
 - ❑ Harvest from OhioLINK ETD Center, ETD-MS to Marc...

- ❑ XSLT Transformation
 - ❑ LC MARC 21 XML schema with MarcXML toolkit
 - ❑ [Dublin Core to MARCXML Stylesheet](#)
 - ❑ OAI community developed tools, mostly for IT staff
 - ❑ [MarcEdit](#) (Terry Reese)
 - ❑ Metadata Harvester, MARC Editor
 - ❑ Low-barrier harvester, can be used by catalogers

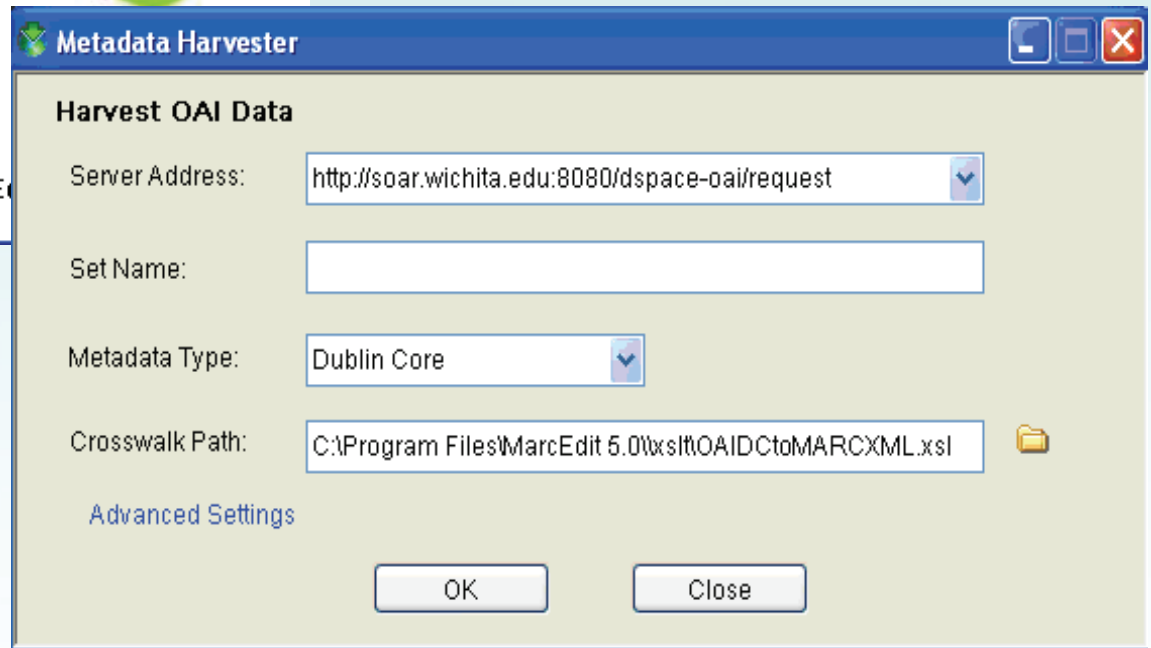
Sample Record in SOAR (Dublin Core)

DC Field	Value
<input type="checkbox"/> dc.contributor.author	Niles, Rae-
<input type="checkbox"/> dc.date.accessioned	2006-12-24T14:56:10Z
<input type="checkbox"/> dc.date.available	2006-12-24T14:56:10Z-
<input type="checkbox"/> dc.date.copyright	2006
<input type="checkbox"/> dc.date.issued	2006-05
<input type="checkbox"/> dc.identifier.other	d06005
<input type="checkbox"/> dc.identifier.uri	http://hdl.handle.net/10057/373-
<input type="checkbox"/> dc.description	Thesis (Ed.D.)--Wichita State University, College of Education.en
<input type="checkbox"/> dc.description	"May 2006."
<input type="checkbox"/> dc.description	Includes bibliographic references (leaves 129-145).en
<input type="checkbox"/> dc.description.abstract	The purpose of this study was to describe and identify Sedgwick High School's teacher and student perceptions of the impact of one-to-one laptop computer access using an appreciative inquiry theoretical research perspective and the theoretical frameworks of change and paradigm shift...
<input type="checkbox"/> dc.format.extent	xiv, 167 leaves : digital, PDF file.
<input type="checkbox"/> dc.format.extent	1174852 bytes-
<input type="checkbox"/> dc.format.mimetype	application/pdf-
<input type="checkbox"/> dc.language.iso	en_US
<input type="checkbox"/> dc.rights	Copyright Rae Niles, 2006. All rights reserved.
<input type="checkbox"/> dc.subject.lcsh	Educational technology
<input type="checkbox"/> dc.subject.lcsh	Education--Data processing
<input type="checkbox"/> dc.subject.lcsh	Electronic dissertations
<input type="checkbox"/> dc.title	A study of the application of emerging technology: teacher and student perceptions of the impact of one-to-one laptop computer access
<input type="checkbox"/> dc.type	Dissertation
<input type="checkbox"/> dc.thesis.adviser	Calabrese, Raymond L.
<input type="checkbox"/> dc.identifier.oclc	71805797-
<input type="checkbox"/> Appears in Collections:	EL Theses and Dissertations COE Theses and Dissertations Dissertations

Dublin Core to MARC Mapping

Fields in DSpace	Transformed MARC fields in OCLC (What we want)
<input type="checkbox"/> dc.contributor.author →	100 1 _ Author.
<input type="checkbox"/> dc.date .accessioned →	
<input type="checkbox"/> dc.date.available →	
<input type="checkbox"/> dc.date.copyright →	
<input type="checkbox"/> dc.date.issued →	260 ≠c year.
<input type="checkbox"/> dc.identifier .other →	099
<input type="checkbox"/> dc.identifier.uri →	856 4 0 ...
<input type="checkbox"/> dc.description →	502 Thesis (Ed.D.)--Wichita State University, College of ...
<input type="checkbox"/> dc.description →	500 "Month year."
<input type="checkbox"/> dc.description →	504 Includes bibliographic references...
<input type="checkbox"/> dc.description.abstract →	520 3 _ ...
<input type="checkbox"/> dc.format .extent →	300
<input type="checkbox"/> dc.format.extent →	
<input type="checkbox"/> dc.format.mimetype →	
<input type="checkbox"/> dc.language.iso →	546 en_US
<input type="checkbox"/> dc.rights →	540 Access restricted to WSU students, faculty and staff (delete)
<input type="checkbox"/> dc.subject →	690 (keywords, non CV, delete)
<input type="checkbox"/> dc.subject.lcsh →	650 _ 0
<input type="checkbox"/> dc.title →	245 1 _ ...
<input type="checkbox"/> dc.type →	655 _ 7 Dissertation ≠2 local
<input type="checkbox"/> dc.thesis.adviser →	700 1 2 ... ≠e advisor
<input type="checkbox"/> dc.identifier.oclc →	856 4 1 ...
<input type="checkbox"/> Appears in Collections: →	

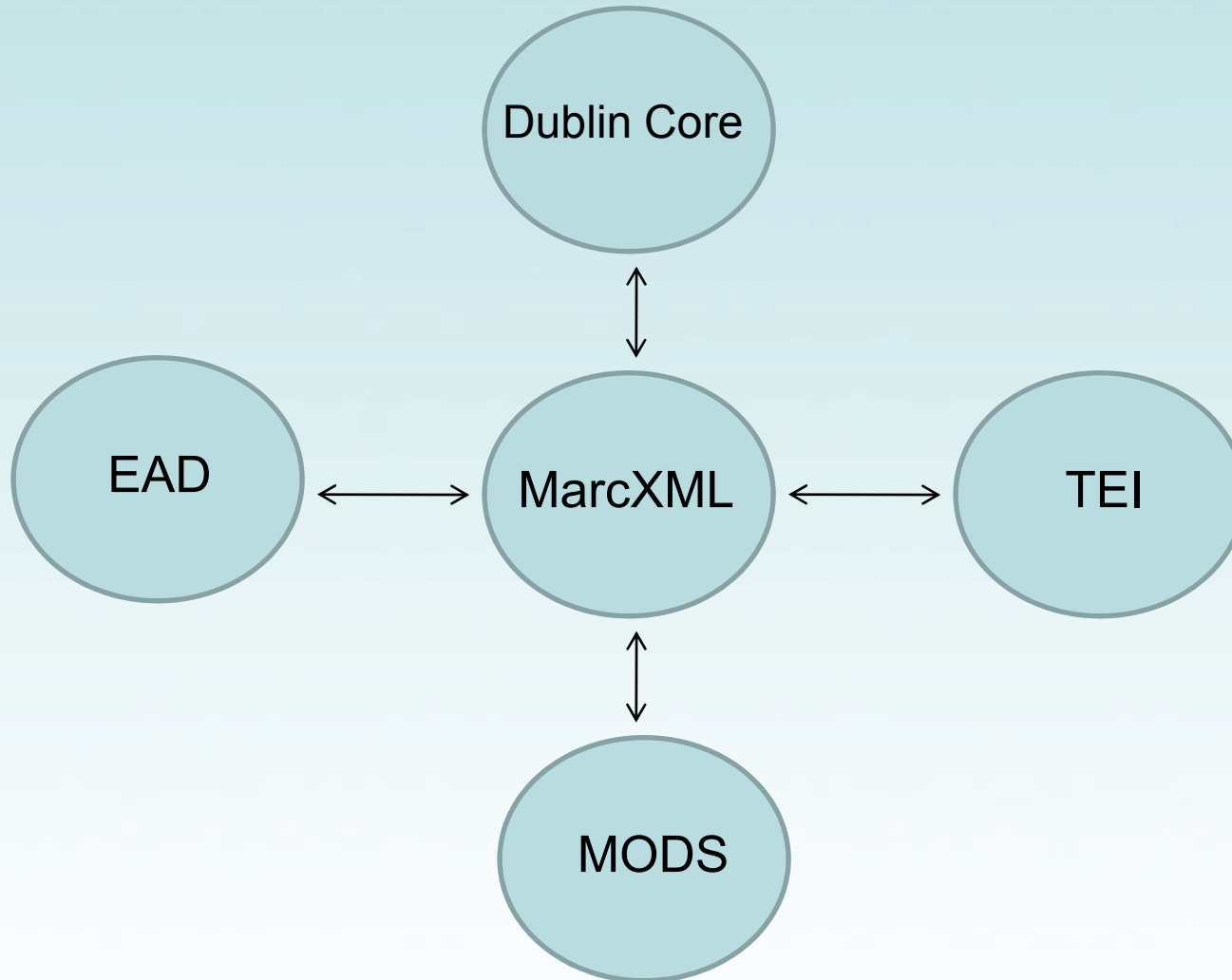
Using MarcEdit



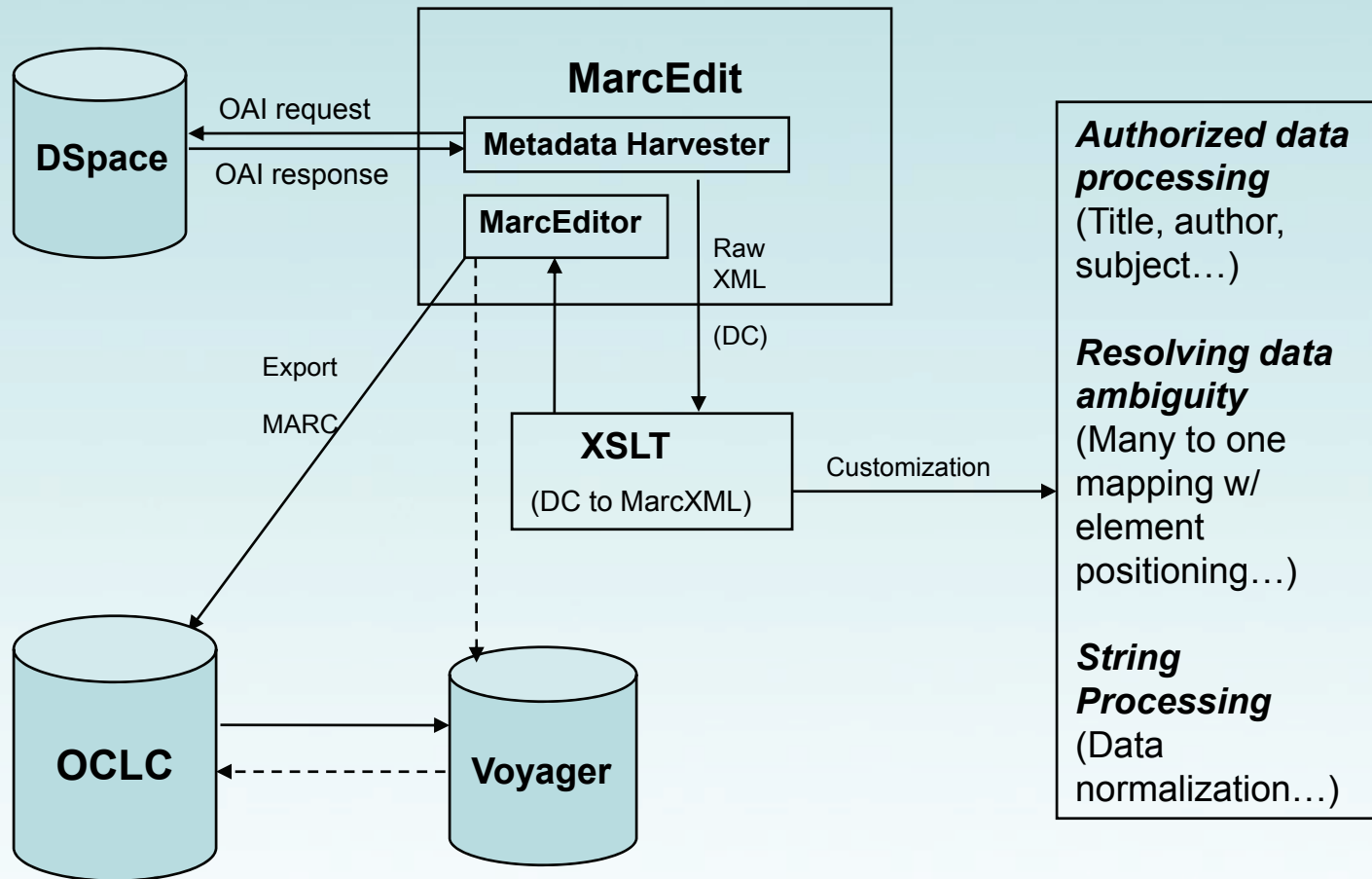
MarcEdit Interface

Metadata transformation in MarcEdit

- ❑ The wheel and spoke design for metadata transformation (by Reese)



Data Flow Diagram



Selective Harvesting

- ❑ **Define in MarcEdit**
 - ❑ by identifier (e.g. oai:soar.wichita.edu:10057/255)
 - ❑ by set (e.g. hdl_10057_351)
 - ❑ by date (e.g. from=2007-01-01&until=2008-01-01)

- ❑ Or, http://soar.wichita.edu/dspace-oai/request?verb=ListRecords&metadataPrefix=oai_dc&from=2007-01-01&until=2008-01-01

- ❑ **How do we define harvesting these only?**
 - ❑ **Define by set**
(<http://soar.wichita.edu/dspace-oai/request?verb=ListSets>)
 - ❑ **Sets by schools and departments**
 - ❑ AE Theses and Dissertations (hdl_10057_313)
 - ❑ ANTH Theses (hdl_10057_233)
 - ❑ BIO Theses (hdl_10057_389)
 - ❑ CE Theses and Dissertations
 - ❑ ...
 - ❑ **Or sets in two categories**
 - ❑ Master's These (hdl_10057_351)
 - ❑ Dissertations (hdl_10057_352)

The screenshot shows the 'Metadata Harvester' application window. The title bar reads 'Metadata Harvester'. The main window is titled 'Harvest OAI Data'. It contains several input fields and checkboxes:

- Server Address:** A dropdown menu with the value 'http://soar.wichita.edu:8080/dspace-oai/request'.
- Set Name:** A text input field containing 'hdl_10057_253'.
- Metadata Type:** A dropdown menu with the value 'Dublin Core'.
- Crosswalk Path:** A text input field containing 'C:\Program Files\MarcEdit 5.0\xml\OAIIDCtoMARCXML.xml' with a folder icon to its right.
- Advanced Settings:** A section with a blue header containing:
 - GetRecord:** An empty text input field.
 - ResumptionToken:** An empty text input field.
 - Start:** An empty text input field.
 - End:** An empty text input field.
 - Translate to MARC-8
 - Timeout:** A text input field containing '100' followed by 'secs.'.
 - Harvest Raw Data (save OAI data to local file system)

At the bottom of the window are two buttons: 'OK' and 'Close'.

Alternatively, Define These Sets in XSLT

□ Dublin Core to MARCXML Stylesheet

```
<?xml version="1.0" encoding="UTF-8" ?>
- <xsl:stylesheet version="1.0" xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/1.1"
  xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
  http://www.openarchives.org/OAI/2.0/oai_dc.xsd" xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  xmlns="http://www.loc.gov/MARC21/slim" exclude-result-prefixes="dc dcterms oai_dc">
<xsl:import href="MARC21slimUtils.xsl" />
<xsl:output method="xml" encoding="UTF-8" indent="yes" />
- <xsl:template match="/">
- <collection xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.loc.gov/MARC21/slim
  http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
  <xsl:apply-templates />
</collection>
</xsl:template>
- <xsl:template name="OAI-PMH">
- <xsl:for-each select="ListRecords/record/metadata/oai_dc:dc">
  <xsl:apply-templates />
</xsl:for-each>
- <xsl:for-each select="GetRecord/record/metadata/oai_dc:dc">
  <xsl:apply-templates />
</xsl:for-each>
</xsl:template>
```

XSLT Customization: Transform and Display Theses and Dissertations Only

```
<record>
  <xsl:variable name="myType" select="dc:type" />
  - <xsl:choose>
    - <xsl:when test="($myType!=) and (($myType='Thesis') or ($myType='Dissertation'))">
      - <xsl:element name="leader">
        <xsl:variable name="type" select="dc:type" />
        - <xsl:variable name="leader06">
          - <xsl:choose>
            <xsl:when test="$type='collection'">p</xsl:when>
            <xsl:when test="$type='dataset'">m</xsl:when>
            <xsl:when test="$type='event'">r</xsl:when>
            <xsl:when test="$type='image'">k</xsl:when>
            <xsl:when test="$type='interactive resource'">m</xsl:when>
            <xsl:when test="$type='service'">m</xsl:when>
            <xsl:when test="$type='software'">m</xsl:when>
            <xsl:when test="$type='sound'">i</xsl:when>
            <xsl:when test="$type='text'">a</xsl:when>
            <xsl:when test="($type='Thesis') or ($type='Dissertation')">t</xsl:when>
            <xsl:otherwise>a</xsl:otherwise>
          </xsl:choose>
        </xsl:variable>
      </xsl:element>
    </xsl:when>
    ...
  </xsl:choose>
```

Sample Result Exported to OCLC

Books		Rec stat	Entered	20080602	Replaced	20080602122029.45
Type	ELvl	Src	Audn	Ctrl	Lang	
BLvl	Form	Conf	Biog	MRec	Ctry	
	Cont	GPub	LitF	Indx		
Desc	Ills	Fest	DtSt	Dates		
040						KSW #c KSW
042						dc
090						#b
049						KSWA
100	1	0				Niles, Rae #e author
245	0	0				A study of the application of emerging technology: teacher and student perceptions of the impact of one-to-one laptop computer access
260						#c 2006-12-24T14:56:10Z.
500						d06005
500						http://hdl.handle.net/10057/373
520						Thesis (Ed.D.)—Wichita State University, College of Education.
520						"May 2006."
520						Includes bibliographic references (leaves 129-145).
520						The purpose of this study was to describe and identify Sedgwick High School's teacher and student perceptions of the impact of one-to-one laptop computer access using an appreciative inquiry theoretical research perspective and the theoretical frameworks of change and paradigm shift. An appreciative inquiry theoretical research perspective was used to structure a qualitative, embedded descriptive case study design. An embedded case study design was used to describe the perceptions of high school teachers and their students who were involved in a one-to-one laptop computer wireless environment on student learning and how teachers teach. Data were collected through teacher and student focus groups, as well as administration of the Left-Hand Right-Hand Column Case Method. Data were analyzed using the comparative analysis matrix method (Miles & Huberman, 1994). The analyzed data revealed six salient findings: (1) Students functioned in the capacity of teacher, (2) technology changed the way teachers and students communicated, (3) the culture of the classroom dynamics between teacher and student changed, (4) technology made learning enjoyable for students, (5) teachers and students believed immersion in a technology-rich learning environment created advantages for student success after high school graduation, and (6) teachers believed that access to ubiquitous technology created new challenges for maintaining student engagement in the learning process. Five of the six findings suggested that technology had changed teaching and learning, and helped to create a paradigm shift in the teacher and student roles. Additionally, those findings also identified the positive core of Sedgwick High School, serving to describe the life giving forces within the organization. One of the six findings revealed challenges associated with the application of emerging technology in the classroom. The findings from this study have the potential to contribute to areas of study that focus on the use of technology in schools. Moreover, research from this study has the potential to help serve as a foundation for other school leaders who are seeking opportunities that prepare students for life in the technology-rich 21st Century through one-to-one laptop computer access.
540						Access restricted to WSU students, faculty and staff
540						Copyright Rae Niles, 2006. All rights reserved.
546						en_US
690						Educational technology
690						Education—Data processing
690						Electronic dissertations
655	7					Dissertation #2 local
856	4	1				#u 71805797 #z Connect to this object online.

Mapping Problems and Error Reports (for Variable Fields)

- ❑ 100 occurrence 1, indicator 2 - invalid code
- ❑ 520 occurrence 4, \$a occurrence 1, position 76 - invalid character - data must be ALA characters
- ❑ 655 occurrence 1, indicator 1 - invalid code
- ❑ 655 occurrence 1, indicator 2 - invalid code
- ❑ 655 occurrence 1, \$2 - invalid relationship - when element is present, then 655 indicator 2 must equal 7
- ❑ ...
- ❑ Need customization to meet our needs.

Mapping Test Results Using OAIDCtoMARCXML.xsl (in MarcEdit)

- DSpace (version 1.4 or below) only responds with simple Dublin Core xml file (to be transformed to MarcXML using xslt).

<input type="checkbox"/> Fields in DSpace	Transformed fields in OCLC	Correction and Customization Needed
<input type="checkbox"/> dc.contributor.author →	100 1 0 Niles, Rae †e author	(Delete †e author.)
<input type="checkbox"/> dc.date.accessioned →		
<input type="checkbox"/> dc.date.available →		
<input type="checkbox"/> dc.date.copyright →		
<input type="checkbox"/> dc.date.issued →	260 †c 2006-05	(Only keep 2006)
<input type="checkbox"/> dc.identifier.other →	500 d06005	(Change to 099)
<input type="checkbox"/> dc.identifier.uri →	500 http://hdl.handle.net/10057/373	(Change to 856 4 0)
<input type="checkbox"/> dc.description →	520 Thesis (Ed.D.)--Wichita State University, College of Education.	(Change to 502)
<input type="checkbox"/> dc.description →	520 "May 2006."	(Change to 500)
<input type="checkbox"/> dc.description →	520 Includes bibliographic references (leaves 129-145).	(Change to 504)
<input type="checkbox"/> dc.description.abstract →	520 The purpose of this study was to describe and identify Sedgwick High School's teacher and student perceptions of the impact of one-to-one laptop computer access using an appreciative inquiry theoretical research perspective and the theoretical frameworks of change and paradigm shift...	(Change to 520 3)
<input type="checkbox"/> dc.format.extent →		
<input type="checkbox"/> dc.format.extent →		
<input type="checkbox"/> dc.format.mimetype →		
<input type="checkbox"/> dc.language.iso →	546 en_US	(delete)
<input type="checkbox"/> dc.rights →	540 Access restricted to WSU students, faculty and staff	(delete)
<input type="checkbox"/> dc.subject.lcsh →	690 Educational technology	(Change to 650 _0)
<input type="checkbox"/> dc.subject.lcsh →	690 Education--Data processing	
<input type="checkbox"/> dc.subject.lcsh →	690 Electronic dissertations	
<input type="checkbox"/> dc.title →	245 0 0 A study of the application of emerging technology: teacher and student perceptions of the impact of one-to-one laptop computer access	(if 100 exists, use 245 1_; or else use 245 0_)
<input type="checkbox"/> dc.type →	655 7 _ Dissertation †2 local	(Change to 655 _7)
<input type="checkbox"/> dc.thesis.advisor →		(Add 700 1 2 ... †e advisor.)
<input type="checkbox"/> dc.identifier.oclc →	856 4 1 †u 71805797 †z Connect to this object online.	(replace †u with value from dc.identifier.uri)
<input type="checkbox"/> Appears in Collections: →		

Customized Mapping in XSLT

❑ Resolving data ambiguity

❑ Same DC fields to different MARC fields:

- ❑ description → 502(Dissertation)
500(General Note)
504 (Bibliography)

❑ Qualified DC element:

- ❑ description.abstract → 520(Summary)

❑ Solution: element positioning

```
<xsl:for-each select="dc:description[1]">  
  - <datafield tag="502" ind1="" ind2="">  
    - <subfield code="a">  
      <xsl:value-of select="normalize-space(.)" />  
    </subfield>  
  </datafield>  
</xsl:for-each>  
<xsl:for-each select="dc:description[2]">  
  - <datafield tag="500" ind1="" ind2="">  
    - <subfield code="a">  
      <xsl:value-of select="normalize-space(.)" />  
    </subfield>  
  </datafield>  
</xsl:for-each>
```

...

Customized Mapping in XSLT

□ Authorized data processing

□ Primary entries vs. added entries: title and personal names processing

□ Template to deal with personal names (in MarcEdit)

- E.g. `<dc:creator>Webb, Kyle M.</dc:creator>`
`<dc:creator>Webb, Kyle M., 1977 -</dc:creator>`
transformed to
`=100 1$aWebb, Kyle M.`
`=100 1$aWebb, Kyle M., $d1977-`

□ Identify field relationship and correct indicators

100, 245 (author, title) relationship: if 100 exists, 245 1 _
or else, 245 0 _

- **Local element: dc.thesis.advisor** transformed to 700 1_
(If more than one dc.thesis exists, positioning is needed.)

Customized Mapping in XSLT

❑ Processing of non-filing characters in title

❑ 245 (title) 2nd indicator: a, an, the... (2, 3, 4)

```
<xsl:for-each select="dc:title[1]">
- <xsl:choose>
- <xsl:when test="$exist100!="">
- <xsl:choose>
- <xsl:when test="substring(., 1, 2)='A '">
- <datafield tag="245" ind1="1" ind2="2">
- <xsl:choose>
- <xsl:when test="contains(., ':')">
- <subfield code="a">
  <xsl:value-of select="concat(substring-before(., ':'), ' : ')" />
</subfield>
- <subfield code="b">
  <xsl:value-of select="concat(substring-after(., ':'), ' / ')" />
</subfield>
</xsl:when>
...

```

❑ Alternatively, it can be defined in the title template.

Customized Mapping in XSLT

❑ Subjects vs. Keywords

- ❑ Only kept common subject in the test (when keywords and subjects mixed inconsistently)

```
- <xsl:for-each select="dc:subject">  
- <xsl:if test=".='Electronic dissertations'">  
- <datafield tag="650" ind1="" ind2="0">  
- <subfield code="a">  
  <xsl:value-of select="." />  
</subfield>
```

...

❑ Subject template (OSU solution)

```
❑ <dc:subject>ocean wave energy</dc:subject>  
   <dc:subject>direct-drive</dc:subject>  
   <dc:subject>fluid-structure interaction</dc:subject>  
   <dc:subject>Ocean wave power</dc:subject>  
   <dc:subject>Fluid-structure interaction</dc:subject>
```

Transformed to

```
=650 \0$aOcean wave power.  
=650 \0$aFluid-structure interaction.  
=690 \\\$aocean wave energy.  
=690 \\\$adirect-drive.  
=690 \\\$afluid-structure interaction.
```

Customized Mapping in XSLT

❑ String Processing

❑ Functions

- ❑ normalize-space()
- ❑ translate()
- ❑ substring()...

❑ Example: Extract partial value from DC element

- ❑ 260 (Date): only extract year from the issuing date in DC

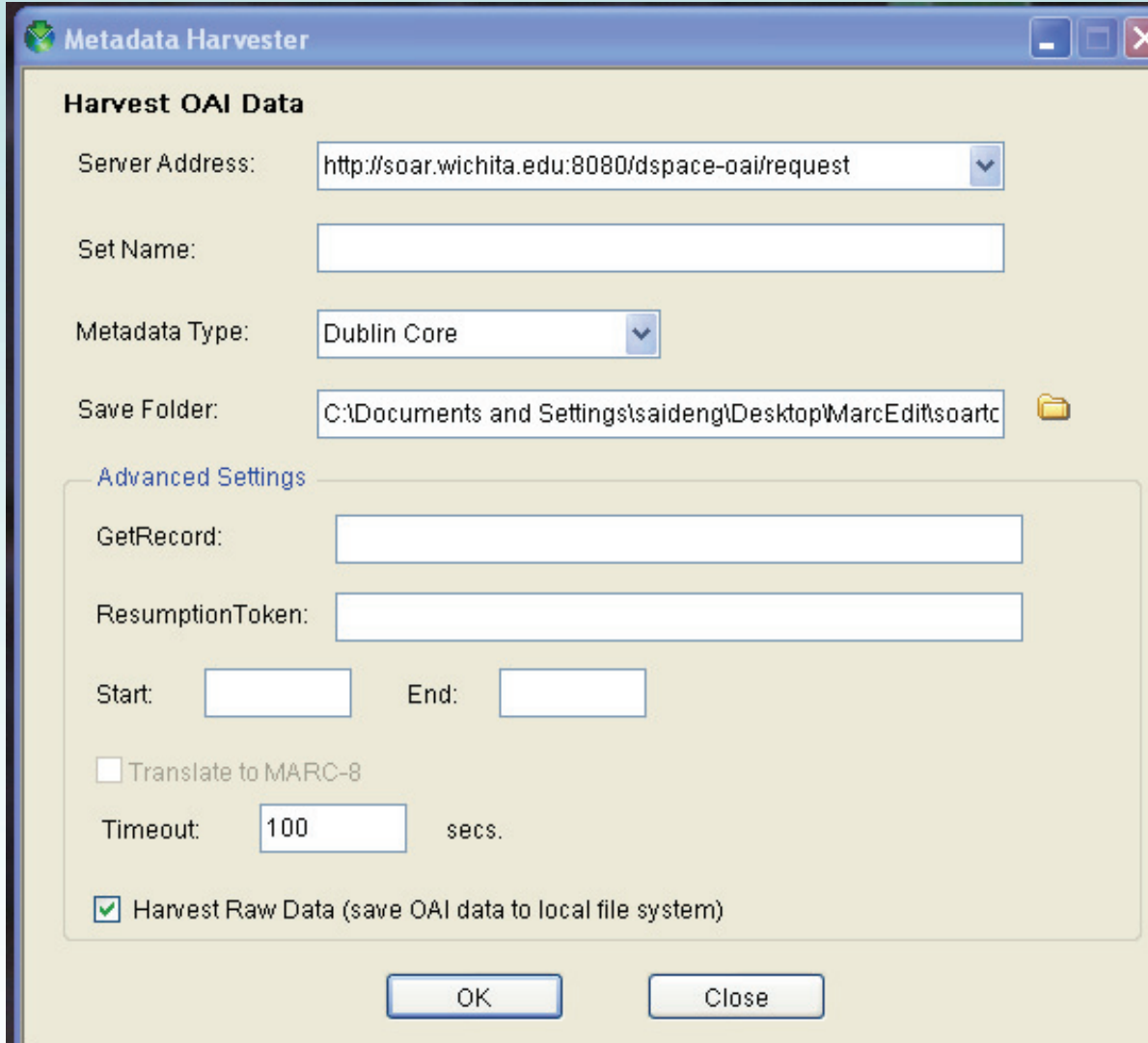
```
- <xsl:for-each select="dc:date[4]">
- <xsl:if test="!="">
- <datafield tag="260" ind1="" ind2="">
- <subfield code="c">
  <xsl:value-of select="substring(.,1,4)" />
  .
</subfield>
</datafield>
</xsl:if>
</xsl:for-each>
```

Customized Mapping in XSLT

- ❑ **Leaders:** fixed fields that comprise the first 24 character positions (00-23) of each MARC record. They provide information for the processing of the record.
- ❑ **008 field (Fixed-Length Data Elements)**
 - ❑ Type (t, manuscript language material) BLvl (m, Encoding level is monograph)
 - ❑ Desc (a) ELvl (l, encoding level is full level) Form (s, form of item is electronic)
 - ❑ Cont (b, m, content is theses with bibliographies) Ills (a, illustration included)
 - ❑ Srce (d, cataloging source) Conf (0, not a conference publication)
 - ❑ Fest (0, not a festschrift) LitF (0, not fiction) DtSt (s, single date)
 - ❑ Indx (0, no index) Lang (eng, language is English) Ctry (xx)
- ❑ **Ways to handle:**
 - ❑ Scripting and adding all fixed fields (leader and 008 fields) in OAIDCtoMARCXML.xsl;
 - ❑ Or, Adding 008 in MarcEditor after record export;
 - ❑ Or, applying fixed field template after records being exported to OCLC.

Harvesting Using the Revised XSLT Crosswalk

Harvest Raw Data

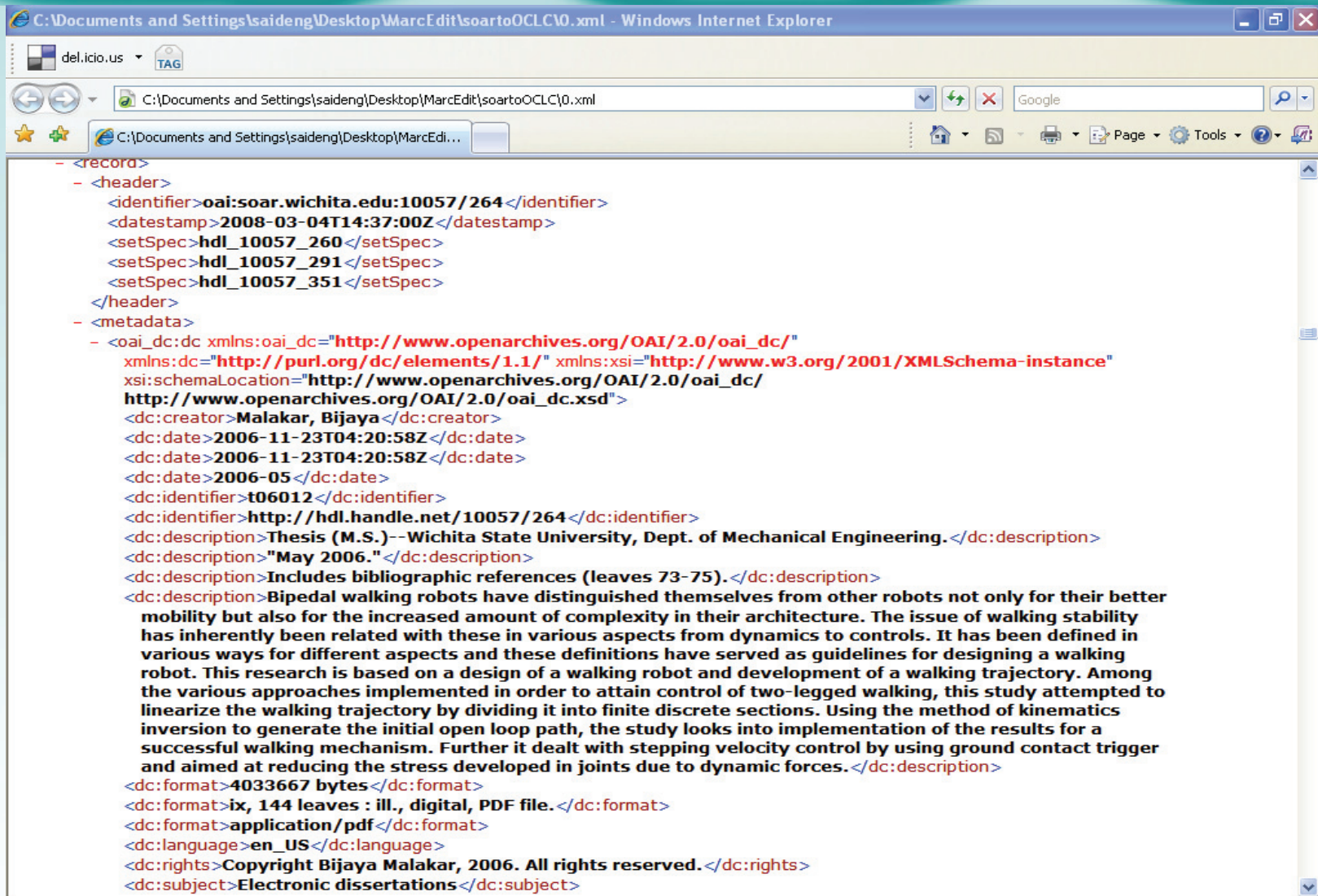


The screenshot shows the 'Metadata Harvester' application window. The title bar reads 'Metadata Harvester'. The main area is titled 'Harvest OAI Data' and contains the following fields and options:

- Server Address:** A dropdown menu with the value 'http://soar.wichita.edu:8080/dspace-oai/request'.
- Set Name:** An empty text input field.
- Metadata Type:** A dropdown menu with the value 'Dublin Core'.
- Save Folder:** A text input field containing 'C:\Documents and Settings\saideng\Desktop\MarcEdit\soarc' with a folder icon to its right.
- Advanced Settings:** A section enclosed in a rounded rectangle containing:
 - GetRecord:** An empty text input field.
 - ResumptionToken:** An empty text input field.
 - Start:** An empty text input field.
 - End:** An empty text input field.
 - Translate to MARC-8
 - Timeout:** A text input field with '100' and the label 'secs.' to its right.
 - Harvest Raw Data (save OAI data to local file system)

At the bottom of the window are two buttons: 'OK' and 'Close'.

Raw DC XML (Harvest oai Data to Local File)



```
-- <record>
- <header>
  <identifier>oai:soar.wichita.edu:10057/264</identifier>
  <datestamp>2008-03-04T14:37:00Z</datestamp>
  <setSpec>hdl_10057_260</setSpec>
  <setSpec>hdl_10057_291</setSpec>
  <setSpec>hdl_10057_351</setSpec>
</header>
- <metadata>
- <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
  xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
  http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <dc:creator>Malakar, Bijaya</dc:creator>
  <dc:date>2006-11-23T04:20:58Z</dc:date>
  <dc:date>2006-11-23T04:20:58Z</dc:date>
  <dc:date>2006-05</dc:date>
  <dc:identifier>t06012</dc:identifier>
  <dc:identifier>http://hdl.handle.net/10057/264</dc:identifier>
  <dc:description>Thesis (M.S.)-- Wichita State University, Dept. of Mechanical Engineering.</dc:description>
  <dc:description>"May 2006."</dc:description>
  <dc:description>Includes bibliographic references (leaves 73-75).</dc:description>
  <dc:description>Bipedal walking robots have distinguished themselves from other robots not only for their better
  mobility but also for the increased amount of complexity in their architecture. The issue of walking stability
  has inherently been related with these in various aspects from dynamics to controls. It has been defined in
  various ways for different aspects and these definitions have served as guidelines for designing a walking
  robot. This research is based on a design of a walking robot and development of a walking trajectory. Among
  the various approaches implemented in order to attain control of two-legged walking, this study attempted to
  linearize the walking trajectory by dividing it into finite discrete sections. Using the method of kinematics
  inversion to generate the initial open loop path, the study looks into implementation of the results for a
  successful walking mechanism. Further it dealt with stepping velocity control by using ground contact trigger
  and aimed at reducing the stress developed in joints due to dynamic forces.</dc:description>
  <dc:format>4033667 bytes</dc:format>
  <dc:format>ix, 144 leaves : ill., digital, PDF file.</dc:format>
  <dc:format>application/pdf</dc:format>
  <dc:language>en_US</dc:language>
  <dc:rights>Copyright Bijaya Malakar, 2006. All rights reserved.</dc:rights>
  <dc:subject>Electronic dissertations</dc:subject>
```

Harvest and Transform DC to MarcXML

Metadata Harvester

Harvest OAI Data

Server Address:

Set Name:


Metadata Type:

Crosswalk Path:

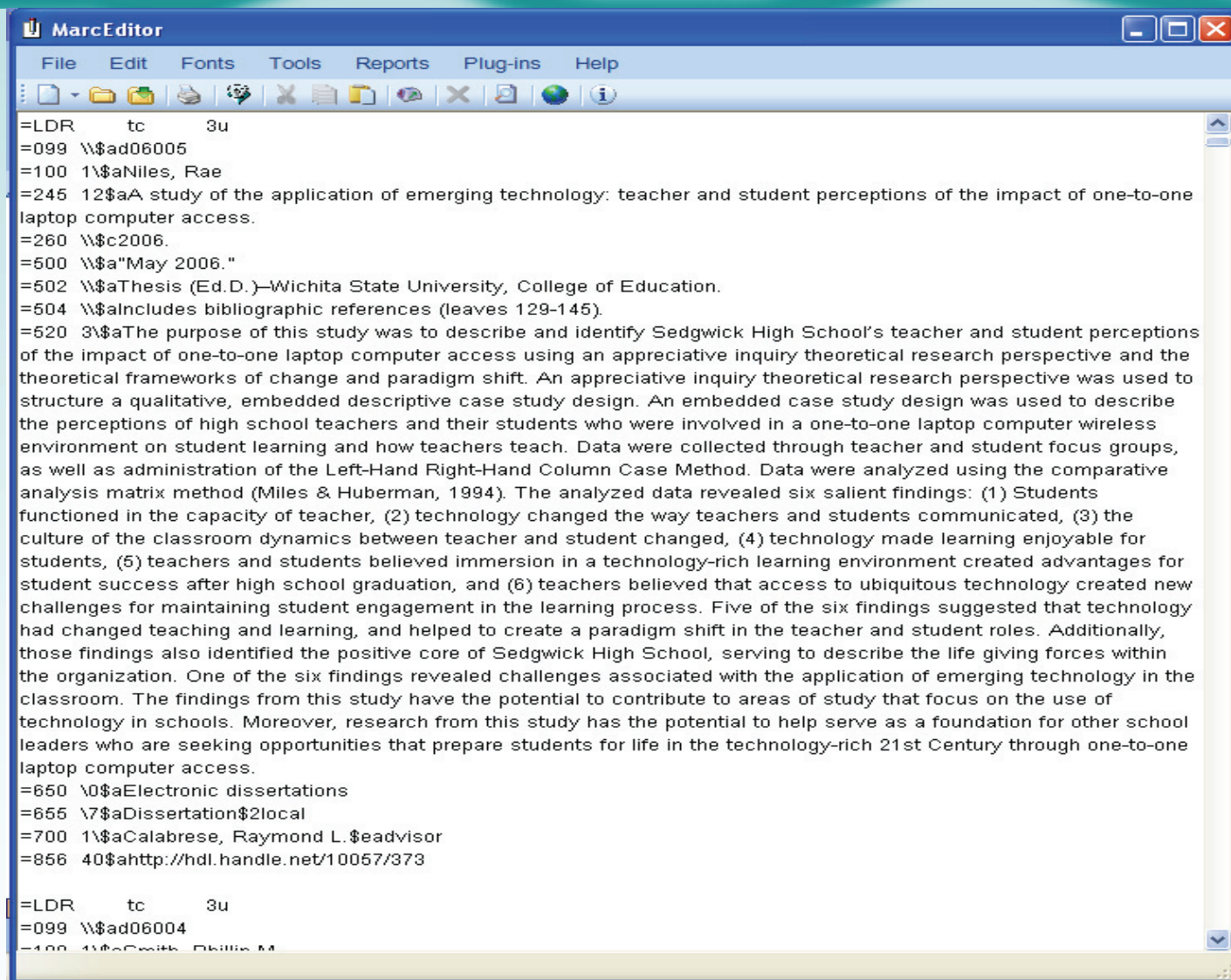
[Advanced Settings](#)

Results

Total Records Processed: 1036
ResumptionToken's Processed: 10
Last ResumptionToken Processed: 0001-01-01T00:00:00Z/9999-12-31T23:59:59Z//oai_dc/1000



Records will be Dumped to MarcEdit- MarcEditor



MarcEditor

Edit harvested theses in MarcEditor

- Batch edit fields, subfields, indicators (if needed)

 - E.g.: add 008 field for all records

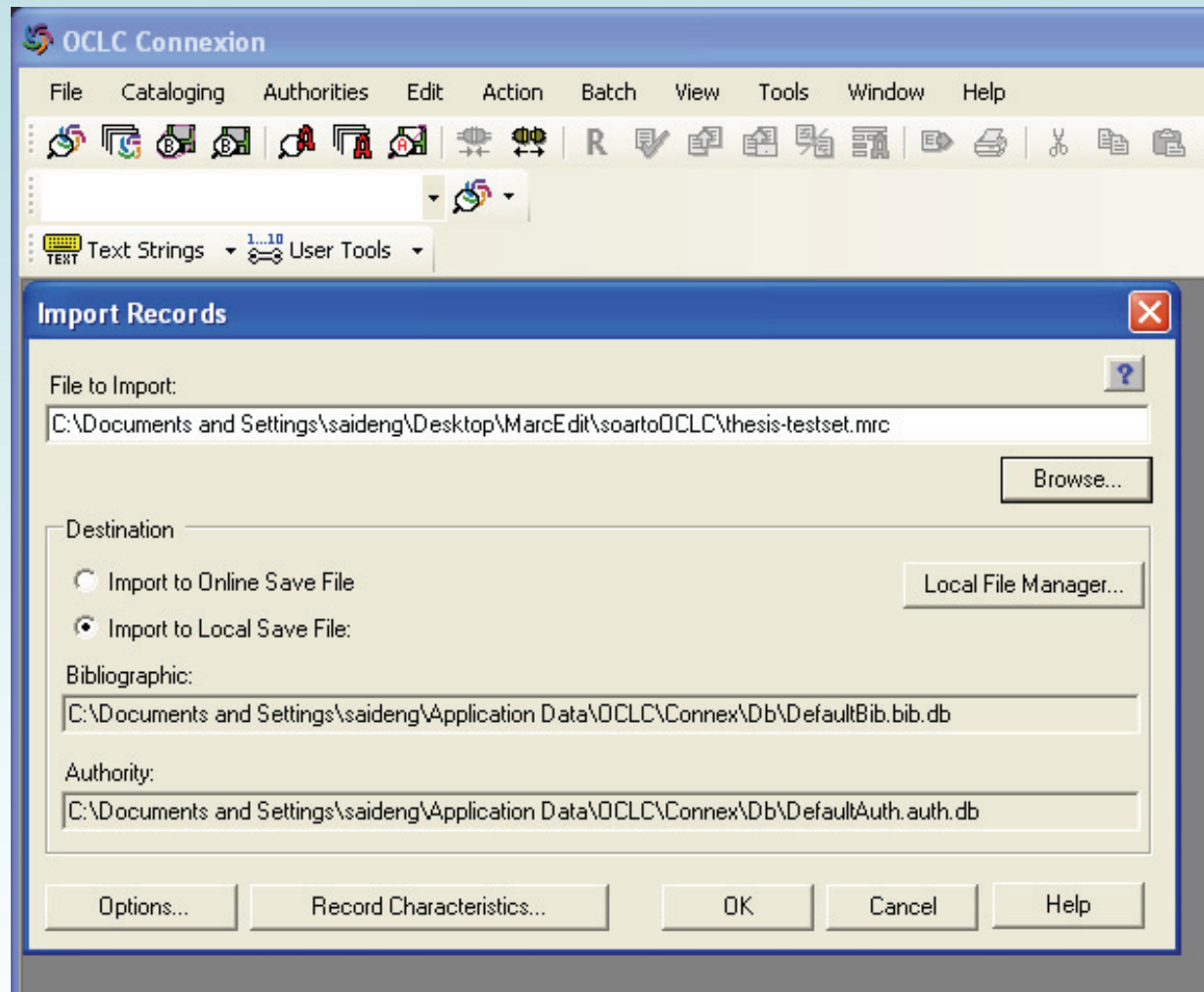
- .mrk (MARC text file) → Compile to .mrc (MARC)

Or

- Save as .mrk8 (MARC UTF8 text file) → Compile to .mrc (MARC)

Import Records to OCLC

- ❑ Click “File-Import Records...”
- ❑ Select “Import to Local Save File”



Import Records to OCLC

OCLC Connexion - [Local Bibliographic Save File List (DefaultBib.bib.db)]

File Cataloging Authorities Edit Action Batch View Tools Window Help


Text Strings User Tools

Save #	Contributor/Creator	Title	Control #	Date	Call Number	Date/Time Added	Held	Code	Fo
11	Goldbarth, Albert.	Title /	NEW	1981		5/22/2008 10:26:51 AM		KSWA	
12	Shaikh, Audrey Dawn author	The effect of line length and passage type on reading...	NEW			5/22/2008 2:30:04 PM		KSWA	
13	Sriririraju, Subhadrakumari K...	Signal processing of acoustic reflectometry in estimati...	NEW			5/22/2008 2:30:04 PM		KSWA	
14	Potter unknown author	Santo Domingo bowl	NEW			5/22/2008 2:30:05 PM		KSWA	
15	Nuest, Brian Dale	Where did I park? Connecting lower-level and higher-...	NEW		d06001	6/13/2008 4:13:10 PM		KSWA	
16	Niles, Rae	Study of the application of emerging technology: teac...	NEW		d06005	6/13/2008 4:13:10 PM		KSWA	
17	Smith, Phillip M.	Studies on electron and energy transfer in porphyrin a...	NEW		d06004	6/13/2008 4:13:11 PM		KSWA	
18	Miller, Russell K.	Impact of educational technology on learner interactio...	NEW		d06006	6/13/2008 4:13:11 PM		KSWA	
19	Gadde, Suresh	Supramolecular porphyrin-fullerene conjugates: desig...	NEW		d06008	6/13/2008 4:13:11 PM		KSWA	
20	Mosack, Victoria	Student ratings of university teaching: similarities and...	NEW		d06009	6/13/2008 4:13:11 PM		KSWA	
21	Kim, Young-Man	Robust and reduced order h-infinity filtering via LMI ap...	NEW		d06010	6/13/2008 4:13:11 PM		KSWA	
22	Hyun, Inha	Stochastic control of unified decentralized singularly p...	NEW		d06011	6/13/2008 4:13:11 PM		KSWA	
23	Kumbhar, Sachin Shivaji	Development of a finite element model and analysis of...	NEW		t06081	6/13/2008 4:13:11 PM		KSWA	
24	Summers, Bryce	Effects of family structure and parenting style on scho...	NEW		t06018	6/13/2008 4:13:11 PM		KSWA	
25	Swartzendruber, Rachel D.	Discovering voices among peculiar quietness: an anal...	NEW		t06061	6/13/2008 4:13:11 PM		KSWA	
26	Binns, Rebecca Kay	"On the cover of a Rolling Stone": a content analysis o...	NEW		t06063	6/13/2008 4:13:12 PM		KSWA	
27	Ozgun, Recep	Design and timing analysis of wave pipelined circuits.	NEW		t06064	6/13/2008 4:13:12 PM		KSWA	
28	Briggs, Roger T.	Dubliners and the Joycean epiphany.	NEW		t06065	6/13/2008 4:13:12 PM		KSWA	
29	Ott, Sara	Paradox and philosophical anticipation in Melville's M...	NEW		t06069	6/13/2008 4:13:12 PM		KSWA	
30	Shenoy, Sudhir Shivaraya	Energy absorption of a car roof reinforced with a grid...	NEW		t06073	6/13/2008 4:13:12 PM		KSWA	
31	Gowda, Supriya Srinivasa	Effect of large neutral amino acids on maternal phenyl...	NEW		t06074	6/13/2008 4:13:12 PM		KSWA	
32	Davis, Thomas Arthur	Evolution of literary theory: towards a bio-cultural appr...	NEW		t06075	6/13/2008 4:13:12 PM		KSWA	
33	Shelite, Thomas Robert	West Nile virus and wild bird populations.	NEW		t06076	6/13/2008 4:13:12 PM		KSWA	
34	Emami, Tooran	Unified procedure for continuous-time and discrete-ti...	NEW		t06077	6/13/2008 4:13:12 PM		KSWA	
35	Bajracharya, Bijay	Effect of variations of riveting process on the quality of...	NEW		(KSW)1238911	6/13/2008 4:13:12 PM		KSWA	

After Being Exported to OCLC...

❑ In OCLC Connexion client:

Open each file, do some review/editing as needed, attach KSW holding and apply fixed field template of ETD (if needed) in OCLC.



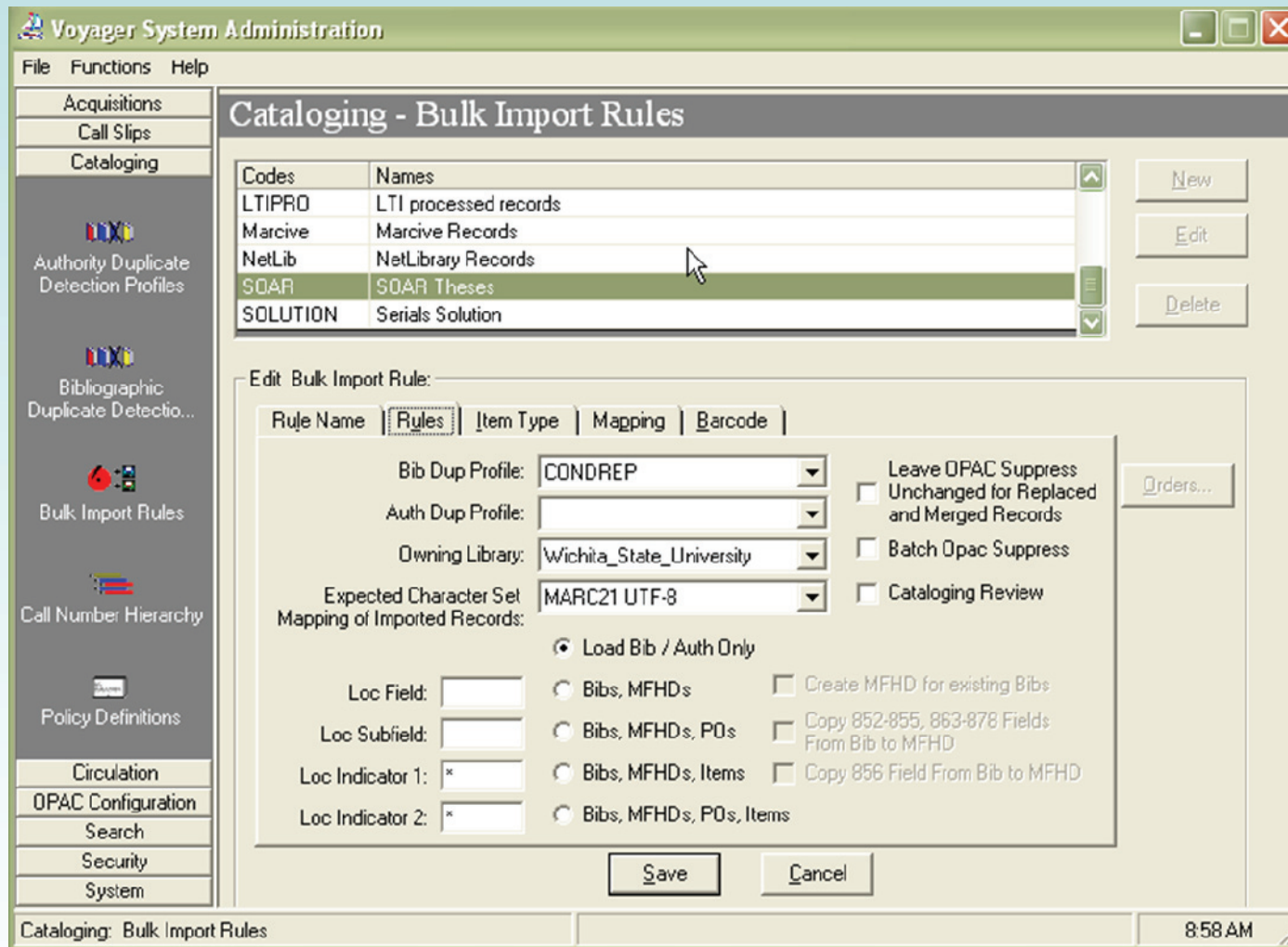
The screenshot shows the OCLC Connexion software interface. The title bar reads "OCLC Connexion - [Online Bibliographic Constant Data List: in: {"KSW"}]". The menu bar includes File, Cataloging, Authorities, Edit, Action, Batch, View, Tools, Window, and Help. The toolbar contains various icons for file operations and editing. Below the toolbar, there are dropdown menus for "Text Strings" and "User Tools". The main window displays a dialog box titled "Choose Fields to Apply" with three radio buttons: "Fixed" (selected), "Variable", and "Both". An "Apply Selected" button is also present. Below the dialog box, a table shows the "Default Record" with the following data:

Name	My Status	
ETD	Susan	

Alternatively, records exported to Voyager directly

- This part is performed by Gemma Blackburn.
- Send .mrc file to the Voyager server.
- Create a Bulk Import rule in Voyager System Administration module.
 - Go to: Cataloging → Bulk Import Rules → New
 - Name the rule
 - Choose (or create a new) Bib De-Duplication Rule
 - Modify mapping as needed
 - Save the rule

Voyager System Administration Bulk import rules screenshot



The screenshot shows the Voyager System Administration interface for Bulk Import Rules. The window title is "Voyager System Administration" and the main title is "Cataloging - Bulk Import Rules".

Left Navigation Panel:

- Acquisitions
- Call Slips
- Cataloging
- Authority Duplicate Detection Profiles
- Bibliographic Duplicate Detecto...
- Bulk Import Rules
- Call Number Hierarchy
- Policy Definitions
- Circulation
- OPAC Configuration
- Search
- Security
- System

Main Content Area:

Cataloging - Bulk Import Rules

Codes	Names
LTIPRO	LTI processed records
Marcive	Marcive Records
NetLib	NetLibrary Records
SOAR	SOAR Theses
SOLUTION	Serials Solution

Buttons: New, Edit, Delete

Edit Bulk Import Rule:

Rule Name: [Rules] | Item Type | Mapping | Barcode

Bib Dup Profile: CONDREP
Auth Dup Profile: []
Owning Library: Wichita_State_University
Expected Character Set: MARC21 UTF-8
Mapping of Imported Records: Load Bib / Auth Only

Loc Field: []
Loc Subfield: []
Loc Indicator 1: *
Loc Indicator 2: *

Leave OPAC Suppress Unchanged for Replaced and Merged Records
 Batch Opac Suppress
 Cataloging Review

Create MFHD for existing Bibs
 Copy 852-855, 863-878 Fields From Bib to MFHD
 Copy 856 Field From Bib to MFHD

Buttons: Save, Cancel, Orders...

Bottom Status Bar: Cataloging: Bulk Import Rules | 8:58 AM

Bulk import to Voyager

- ❑ Bulk Import the records using the Bulk Import rule
 - ❑ On your Voyager server, go to:
.../voyager/xxxdb/sbin/
 - ❑ Write the command for Bulk Import to run:
`Pbulkimport -ftheses-sample.mrc -iSOAR -b1 -e3`
 - ❑ `-f` and the file name (required)
 - ❑ `-i` and the Bulk Import rule name (required)
 - ❑ `-o` and your name (not required, but will let people know who ran the bulk import)
 - ❑ `-b` and a number. This will define the beginning record in the file that you want to import if you prefer to import a select set at a time (not required)
 - ❑ `-e` and a number. This will define the end record in a set to import (not required)
 - ❑ There are several other options. Check the Technical User's Guide

A real case

- ❑ Transformation of ETDs of 2007
 - ❑ Ph.D. Dissertations (Summer, Fall 2007): 23
 - ❑ Master's Theses (Summer, Fall 2007): 55

- ❑ Some adjustment in the transformation:
 - ❑ Transfer dc.format.extent[1] to physical description (Marc 300)
 - ❑ E.g. ix, 53 leaves, ill. → 300 \$a ix, 53 leaves : \$b ill.

 - ❑ Keep 3 description fields
 - ❑ description [1] → 500(General Note)
 - description [2] → 502(Dissertation)
 - description.abstract → 520(Summary)

 - ❑ 008 field values added in MarcEditor rather than applied in OCLC
 - ❑ E.g. =008 ...s2008\\x\\s\\sbm\\000\\0\\eng\\d

Discussion and Conclusion

- ❑ The customized mapping and metadata transfer can eliminate the need of double entry in DSpace and OCLC/Voyager and significantly improve our ETD work flow.
- ❑ Metadata management
 - ❑ One single crosswalk and style sheet will not meet all needs;
 - ❑ Needs to be based on standard practice but add local variations;
 - ❑ Application-specific mapping is needed for special projects;
 - ❑ Coordination in metadata repurposing is important.
- ❑ Data mapping, manipulation and transformation
 - ❑ Using qualified DC instead of element positioning in XSLT;
 - ❑ DSpace 1.5 enables qualified DC crosswalk for OAI-PMH;
 - ❑ Handling of MARC fixed fields and 008 field.
- ❑ Other technical issues
 - ❑ Using other tools for harvesting besides MarcEdit;
 - ❑ Using DSpace Item Importer and Exporter instead of Metadata Harvester.

Project team and Acknowledgement

- ❑ Sai Deng, Metadata mapping and transformation
- ❑ Susan Matveyeva, ETD cataloging and mapping
- ❑ Tse-Min Wang, Programming assistance
- ❑ Sandy Oswald, Manoj Gogoi, ETD cataloging assistance

- ❑ Terry Reese, Consultant
- ❑ Nancy Deyoe, Administrative Support
- ❑ Connie, Basquez, Voyager support
- ❑ Gemma Blackburn, Voyager support



Thank you!