

From Excel ETD Metadata to MARC Bib and NACO Records in ~~4-8~~ 12 Easy Steps!

Steven W. Holloway
Metadata Librarian
James Madison University

- Challenge: ETD metadata in Excel format. Really?
- Challenge: Extract NACO dissertant NARs from ETD metadata.
- Challenge: Assign globally unique identifiers to ETD records.
- Challenge: Store and edit ETD metadata as MARCXML.
- Challenge: Distribute NARs through LCNAF, set holdings in OCLC.

Our method of processing ETDs at James Madison University:

- train staff in XML;
- perform multiple XSL transformations;
- store XML files in eXist-db;
- edit using oXygen XML editor;
- upload and perform final edits in OCLC Connexion.



- Challenge 1: Staff need XML literacy.
- Solution 1: Enroll them in Library Juice Academy
Introduction to XML.



- Challenge 2: Extract NACO-level dissertant data from ETD submissions.
- Solution 2: Work with PCC; bribe the Digital Collections Librarian into adding the requisite form fields with explanations.



From Excel ETD Metadata to MARC Bib and NACO Records in 4-8 12 Easy Steps!

Scholarly Commons upload metadata field, preferably mandatory: Preferred Name ¶

Description: The form of your name that you anticipate publishing under, or that you choose to be known by. Example: Holloway, Steven W. (not: Holloway, Steven Winford) ¶

Scholarly Commons upload metadata field, optional: ORCID Identifier ¶

Description: ORCID (Open Researcher Community ID) provides a persistent, digital identifier that distinguishes you from every other researcher or author. Through integration in key workflows such as academic manuscript and grant submission, your ORCID supports automated linkages between you and your professional activities, ensuring that your scholarship is correctly attributed to you. Increasing numbers of publishers, academic organizations and funding agencies require you to use an ORCID. By creating an ORCID and associating it with your work in JMU Scholarly Commons, you are adding insurance that your name, and no other, will be unambiguously linked to the work you have chosen to publish here. ¶

Creating an ORCID is simple and free of charge! You can register for an ORCID ID at: <https://orcid.org/register>. Example ORCID: 0000-0002-7139-2625 ¶



- Challenge 3: Add globally unique identifiers to ETD bib records.
- Solution 3: Become an ARK registrant; use NOID to mint ARKs; insert ARKs into Excel data.

arks
ark:/27361/brzdzt.pdf
ark:/27361/brzdzm7.pdf
ark:/27361/brzdznn.pdf
ark:/27361/brzdzp2.pdf
ark:/27361/brzdzqg.pdf
ark:/27361/brzdzrw.pdf
ark:/27361/brzdzs9.pdf
ark:/27361/brzdzttq.pdf
ark:/27361/brzdzv4.pdf
ark:/27361/brzdzwj.pdf
ark:/27361/brzdzxz.pdf
ark:/27361/brdzcc.pdf

- Challenge 4: Excel XLSX is the Nightmare Before & After Christmas.
- Solution 4: Perform initial damage control in Excel, then save as Excel 2003 XML Spreadsheet.

Excel 2003 XML. Sigh.

```

201 <Row>
202 <Cell ss:StyleID="s68"><Data ss:Type="String">Cataloging Completed</Data></Cell>
203 <Cell ss:StyleID="s64"><Data ss:Type="String">title</Data><Comment ss:Author=""><Data><Font
204     html:Face="Tahoma" html:Size="8" xmlns="http://www.w3.org/TR/REC-html40">Title: Thesis Title</Font>
205 <Cell ss:StyleID="s65"><Data ss:Type="String">preferred_name</Data><Comment
206     ss:Author=""><Data><Font html:Face="Tahoma" html:Size="8"
207     xmlns="http://www.w3.org/TR/REC-html40">Preferred Name</Font></Data></Comment></Cell>
208 <Cell ss:StyleID="s65"><Data ss:Type="String">orcid</Data><Comment ss:Author=""><Data><Font
209     html:Face="Tahoma" html:Size="8" xmlns="http://www.w3.org/TR/REC-html40">ORCID</Font></Data></Comm
210 <Cell ss:StyleID="s64"><Data ss:Type="String">publication_date</Data><Comment
211     ss:Author=""><Data><Font html:Face="Tahoma" html:Size="8"
212     xmlns="http://www.w3.org/TR/REC-html40">Date of Award: Date Degree Awarded</Font></Data></Comment>
213 <Cell ss:StyleID="s65"><Data ss:Type="String">season</Data><Comment
214     ss:Author=""><Data><Font html:Face="Tahoma" html:Size="8"
215     xmlns="http://www.w3.org/TR/REC-html40">Enter season of publication: Winter, Spring, Summer, or Fa
216 <Cell ss:StyleID="s64"><Data ss:Type="String">embargo_date</Data><Comment
217     ss:Author=""><Data><Font html:Face="Tahoma" html:Size="8"
218     xmlns="http://www.w3.org/TR/REC-html40">Embargo Period: Embargo Expiration</Font></Data></Comment>
219 <Cell ss:StyleID="s64"><Data ss:Type="String">document_type</Data><Comment
220     ss:Author=""><Data><Font html:Face="Tahoma" html:Size="8"
221     xmlns="http://www.w3.org/TR/REC-html40">Document Type: Document Type</Font></Data></Comment></Cell>
222 <Cell ss:StyleID="s65"><Data ss:Type="String">arks</Data><Comment ss:Author=""><Data><Font
223     html:Face="Tahoma" html:Size="8" xmlns="http://www.w3.org/TR/REC-html40">ARKs</Font></Data></Comm
224 <Cell ss:StyleID="s64"><Data ss:Type="String">degree_name</Data><Comment
225     ss:Author=""><Data><Font html:Face="Tahoma" html:Size="8"
226     xmlns="http://www.w3.org/TR/REC-html40">Degree Name: Degree Name</Font></Data></Comment></Cell>
227 <Cell ss:StyleID="s64"><Data ss:Type="String">department</Data><Comment
228     ss:Author=""><Data><Font html:Face="Tahoma" html:Size="8"
  
```

- Challenge 5: Turn Excel 2003 XML file into a format that is a reasonable candidate for XSL transformation.
- Solution 5: Generate “Named Tag” XML files using Excel2NamedXML.xsl, a beautifully recursive script coded by Andy Schmidt.



“Named Tag” XML

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<row number="1">
```

```
<title>Suicidal behavior among elementary school students and current needs in  
prevention practices: a survey of Virginia school counselors</title>
```

```
<preferred_name>Mandy K.D. VanDyke</preferred_name>
```

```
<orcid/>
```

```
<publication_date>2015-01-01T00:00:00.000</publication_date>
```

```
<embargo_date>2015-04-15T00:00:00.000</embargo_date>
```

```
<arks>ark:/27361/brzdzkt.pdf</arks>
```

```
<degree_name>Educational Specialist (EdS)</degree_name>
```

```
<department>Department of Graduate Psychology</department>
```

```
<advisor1>Deborah Kipps-Vaughan</advisor1>
```

```
<keywords>suicide, school psychology, school counseling, children, collaboration  
</keywords>
```

```
<abstract>Most of the research on suicidal behavior ....</abstract>
```

```
<author1_fname>Maranda</author1_fname>
```

```
<author1_mname>K.D.</author1_mname>
```

```
<author1_lname>VanDyke</author1_lname>
```




- Challenge 6: Store XML files in a network environment that supports WebDAV operations.
- Solution 6: Create an eXist-db instance as a web service.




From Excel ETD Metadata to MARC Bib and NACO Records in 4-8 12 Easy Steps!

← → ↻ xdb.lib.jmu.edu:8080/exist/apps/dashboard/index.html


admin ▾




Backup




betterFORM Demo and Reference XForms




Collections




eXide - XQuery IDE




eXist-db Demo Apps




eXist-db Documentation




Java Admin Client



JMU_ETDs



Markdown Parser in XQuery



Monitoring and Profiling for eXist

- Challenge 7: Edit XML files using an enterprise-level XML editor and transformation engine.
- Solution 7: Edit the eXist-db files via WebDAV with oXygen XML editor.

From Excel ETD Metadata to MARC Bib and NACO Records in 4-8 12 Easy Steps!

The screenshot displays the XSLT Editor interface, which is used for transforming XML data. The main window shows the XSLT code for the file **Edspec2014-2015.xml**. The code is as follows:

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <table><!--Transformed by Excel2NamedXML, © 2014, Andy Schmidt, ASchmidt@Anamera
3   <row number="1">
4     <title>Suicidal behavior among elementary school students and current ne
5     <preferred_name>Mandy K.D. VanDyke</preferred_name>
6     <orcid/>
7     <publication_date>2015-01-01T00:00:00.000</publication_date>
8     <season>Summer</season>
9     <embargo_date>2015-04-15T00:00:00.000</embargo_date>
10    <document_type/>
11    <arks>ark:/27361/brzdzt.pdf</arks>
12    <degree_name>Educational Specialist (EdS)</degree_name>
13    <department>Department of Graduate Psychology</department>
14    <archive_dept/>
15    <advisor1>Deborah Kipps-Vaughan</advisor1>
16    <advisor2/>
17    <keywords>suicide, school psychology, school counseling, children, colla
18    <disciplines>Counseling Psychology; School Psychology</disciplines>
19    <abstract>Most of the research on suicidal behavior has focused on the m
20    <fulltext_url/>
21    <author1_fname>Maranda</author1_fname>
22    <author1_mname>K.D.</author1_mname>
23    <author1_lname>VanDyke</author1_lname>
24    <author1_suffix/>
25    <author1_email>vandykmk@dukes.jmu.edu</author1_email>
26    <author1_institution>James Madison University</author1_institution>
27    <author2_fname/>
28    <author2_mname/>
29    <author2_lname/>
30    <author2_suffix/>
```

The left pane shows the **Data Source Explorer** with a tree structure of connections. The **Connections** list includes:

- eXist_WEBDAV
 - apps
 - betterform
 - dashboard
 - demo
 - doc
 - eXide
 - fundocs
 - JMU_ETDs
 - data
 - Doc_2014-2015.xml
 - Edspec2014-2015.xml
 - ETD_AUTH_20014.xml
 - ETD_BIB_2014_Fall.xml
 - Fall 2014 Grad ETD
 - Masters_2014-2015.xml
 - MX_Auth_Doc2015.xml
 - MX_Auth_Edspec2015.xml
 - MX_Auth_Masters2015.xml
 - MX_Bib_Doc2015.xml
 - MX_Bib_Edspec2015.xml
 - MX_Bib_Masters2015.xml

The right pane shows the **Attributes** table, which is currently empty. Below it, the **Transformation Scenario** pane is visible, showing a list of scenarios under the **Project (18)** filter. The scenarios include:

- Authority2Textfil
- ETD_Authority
- ETD_Excel2MARC
- EXCEL_Numbere
- Excel_Tagged2M
- NamedXML2Aut
- NamedXML2MAF
- NamedXML2MAF

The bottom of the interface features a search bar and a set of buttons for navigating between different views: **Text**, **Grid**, and **Author**. The **Text** view is currently selected.

- Challenge 8: Turn ETD metadata into MARC bib records.
- Solution 8: Perform initial damage control in Excel; generate MARCXML bib records using multiple XSLs.
 - Receive ETD metadata as Excel XLSX
 - => save as Excel 2003 XML Spreadsheet format
 - => generate "Named Tag" XML using XSL
 - => generate MARCXML bib using XSL



MARCXML from “Named Tag”

```
<marc:datafield tag="024" ind1="8" ind2=" ">
  <marc:subfield code="a">ark:/27361/brzdzkt</marc:subfield>
</marc:datafield>
<marc:datafield tag="040" ind1=" " ind2=" ">
  <marc:subfield code="a">VMC</marc:subfield>
  <marc:subfield code="b">eng</marc:subfield>
  <marc:subfield code="e">rda</marc:subfield>
  <marc:subfield code="c">VMC</marc:subfield>
</marc:datafield>
<marc:datafield tag="049" ind1=" " ind2=" ">
  <marc:subfield code="a">VMCI</marc:subfield>
</marc:datafield>
<marc:datafield tag="100" ind1="1" ind2=" ">
  <marc:subfield code="a">VanDyke, Mandy K.D. ,</marc:subfield>
  <marc:subfield code="e">dissertant</marc:subfield>
</marc:datafield>
<marc:datafield tag="245" ind1="1" ind2="0">
  <marc:subfield code="a">Suicidal behavior among elementary school students and
    current needs in prevention practices:</marc:subfield>
  <marc:subfield code="b">a survey of Virginia school counselors /</marc:subfield>
  <marc:subfield code="c">Mandy K.D. VanDyke</marc:subfield>
```


- Challenge 9: Extract NACO dissertant NARs from ETD metadata.
- Solution 9: Check for pre-existing NARs, generate MARCXML authority records from “Named Tag” XML.
 - Receive ETD metadata as Excel XLSX
 - => save as Excel 2003 XML Spreadsheet format
 - => generate “Named Tag” XML using XSL
 - => generate MARCXML authority using XSL



MARCXML NAR from “Named Tag”

```
<marc:leader>  nz a22  n 4500</marc:leader>
<marc:controlfield tag="005">20150914132333.5</marc:controlfield>
<marc:controlfield tag="008">150914n| azannaabn      a aaa  c</marc:controlfield>
<marc:datafield tag="040" ind1=" " ind2=" ">
  <marc:subfield code="a">ViHarT</marc:subfield>
  <marc:subfield code="b">eng</marc:subfield>
  <marc:subfield code="e">rda</marc:subfield>
  <marc:subfield code="c">ViHarT</marc:subfield>
</marc:datafield>
<marc:datafield tag="100" ind1="1" ind2=" ">
  <marc:subfield code="a">VanDyke, Mandy K.D.</marc:subfield>
</marc:datafield>
<marc:datafield tag="370" ind1=" " ind2=" ">
  <marc:subfield code="e">Harrisonburg (Va.)</marc:subfield>
  <marc:subfield code="2">naf</marc:subfield>
</marc:datafield>
<marc:datafield tag="373" ind1=" " ind2=" ">
  <marc:subfield code="a">James Madison University. Department of Graduate Psychology
</marc:subfield>
  <marc:subfield code="2">naf</marc:subfield>
```




MARCXML NAR from “Named Tag”

```
<marc:datafield tag="375" ind1=" " ind2=" ">
  <marc:subfield code="a">[DELETE IF UNSURE]</marc:subfield>
</marc:datafield>
<marc:datafield tag="377" ind1=" " ind2=" ">
  <marc:subfield code="a">eng</marc:subfield>
</marc:datafield>
<marc:datafield tag="400" ind1="1" ind2=" ">
  <marc:subfield code="a">VanDyke, Maranda K.D.</marc:subfield>
</marc:datafield>
<marc:datafield tag="510" ind1="2" ind2=" ">
  <marc:subfield code="i">Corporate body:</marc:subfield>
  <marc:subfield code="a">James Madison University</marc:subfield>
  <marc:subfield code="w">r</marc:subfield>
</marc:datafield>
<marc:datafield tag="670" ind1=" " ind2=" ">
  <marc:subfield code="a">Suicidal behavior among elementary school students and
    current needs in prevention practices: a survey of Virginia
    school counselors, James Madison University Ed.S. thesis,
    2015:</marc:subfield>
  <marc:subfield code="b">ETD online submission form (Maranda K.D. VanDyke, Mandy
    K.D. VanDyke)</marc:subfield>
```




- Challenge 10: Import MARCXML bib and authority records into OCLC.
- Solution 10: Use MarcEdit to transform MARCXML into MARC binary files, then import as bib & authority collections into OCLC.





- Challenge 11: Final QA needed for bib and authority records, including assignment of subject headings, set holdings, and distribute NACO NARs.
- Solution 11: Use OCLC Connexion.





- Challenge 12: Make ETD bib and authority records available to James Madison University community.
- Solution 12: Import MARC binary records into Sierra ILS.



- All in 12 easy steps!



- Excel “Named Tag”2MARCXML bib and authority XSLs available on Github under a CC-BY-NC 4.0 license:

<https://github.com/Stevenholloway/JMU-metadev.Excel2MARCXML>