Adding XSLT to the cataloger's toolbox

Efficiencies for transforming and analyzing bibliographic data

Presented by Annie Glerum for the Technical Services Workflows Efficiency Interest Group meeting at ALA Midwinter, January 27, 2014

Adding XSLT for efficiency

- Metadata transformations for MARC21

- Quality control of vendor batch files

- Example XSLT coding snippets

- Resources for learning

Metadata transformation for MARC21

- Repurposing ProQuest Metadata for Batch Ingesting ETDs into an Institutional Repository / Shawn Averkamp and Joanna Lee
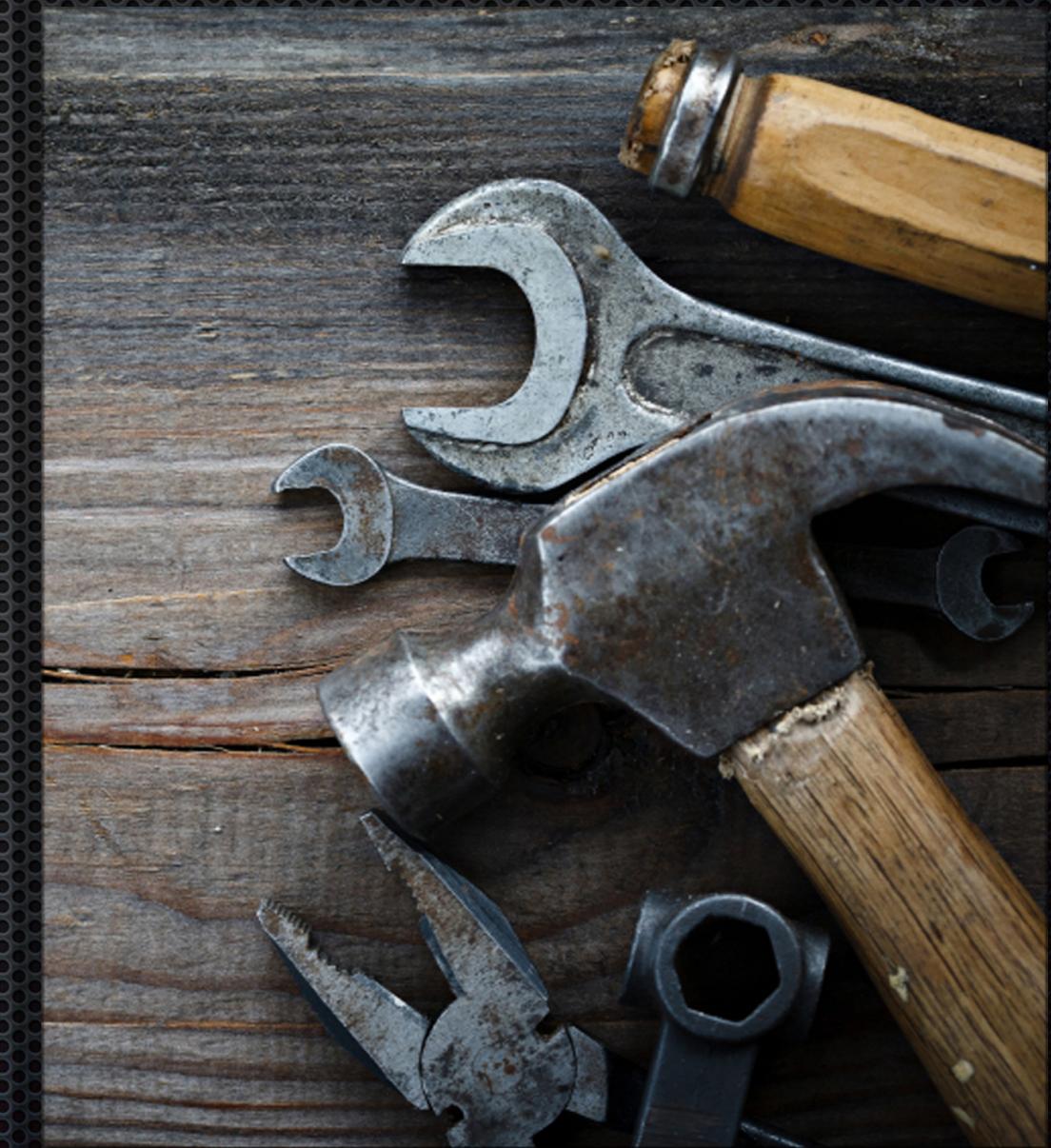
- http://journal.code4lib.org/articles/1647

Sources from "Repurposing ProQuest Metadata for Batch…"

XSLT files

- ETD-bepress2MARC21XML.xsl

- ETD-bepressXML2Txt.xsl

- ETD-ProQuestXML2bepressXML.xsl

Bepress schema

- http://www.bepress.com/document-import.xsd
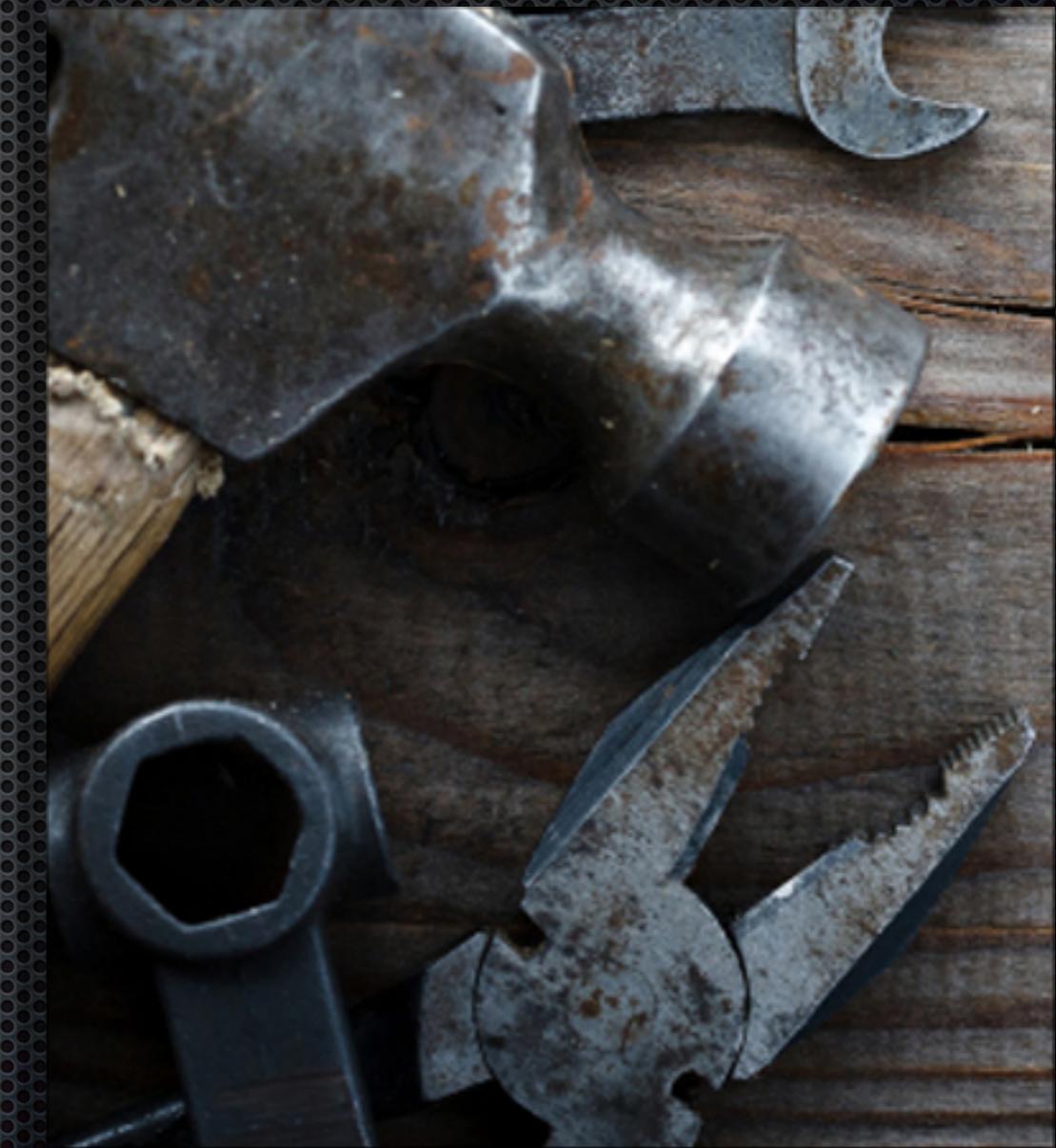
Refinements to ETD-bepress2MARC21XML.xsl

- Proper title case

- Personal name suffixes

- Parsed dissertation note

- Local committee and department notes

- Local access points

Quality control of vendor batch files

- CatQC and Shelf-Ready Material: Speeding Collections to Users While Preserving Data Quality / Michael Jay, Betsy Simpson, and Doug Smith

- Cleanup of Netlibrary Cataloging Records: A Methodical Front-End Process / Elaine Sanchez, Leslie Fatout, and Aleene Howser

- Outsourcing Monograph Cataloging at the UNLV Libraries / Kenneth J. Bierman and Judith A. Carter

See Bibliography for full citations

Quality control of vendor batch files

Fast, cheap, and good: Balancing timeliness, efficiency, and quality for outsourced bibliographic records.

Similar to UF's CatQC program, eXtensible Stylesheet Transformations (XSLT) would be applied to files to generate staff-friendly reports for elements requiring human review. Yet instead of using MarcEdit to analyze and correct MARC records, XSLT would perform analysis of the data as well as perform transformative or augmentative global editing.

Elements for quality control report and review

Coding

- *Less-then-full Elvl; Non-English cataloging agency*
- *Cont=b but no 504; x11 but Conf=0; 504 or 500 containing 'ind*' and Indx=0; 546 but no 041; 651 or 6xx $z but no 043*
- *ISBN for e-books in 020 $a*
- *Dates match in fixed field, call number, and publication*

Access

- *Complete call numbers*
- *At least one LCSH 6xx*
- *Contents notes containing "/" not coded 505 00*

Identification

- *CIP without pagination*

Elements for quality control report and review

Checklists to scan for:

- *Possible set records*
- *Possible non-print records*
- *Possible miscoded initial articles in 245*
- *Possible miscoded 246*
- *No 1xx field, may need one*
- *No 7xx fields, may need one*
- *490 Ind1=0 to change to 490_1/830 pair*

- *440 to convert to 490/830*
- *Nonstandard coding in 856*
- *8xx fields to verify authority and local practices*
- *Call number ranges that need holdings location edits*
- *Oversize books that need holdings location edits*
- *300 $e for accompanying material*

Example XSLT code snippets

Simple variable

```
<xsl:variable name="catLang"
   select="marc:datafield[@tag='040']/marc:subfield[@code='b']"/>
```

Complex variable

```
<xsl:variable name="oclc">
   <xsl:choose>
      <xsl:when test="contains(marc:controlfield[@tag='001'], 'ocn')">
         <xsl:value-of select="substring-
            after(marc:controlfield[@tag='001'], 'ocn')"/>
      </xsl:when>
      <xsl:when test="contains(marc:controlfield[@tag='001'], 'ocm')">
         <xsl:value-of select="substring-
            after(marc:controlfield[@tag='001'], 'ocm')"/>
      </xsl:when>
      <xsl:otherwise/>
   </xsl:choose>
</xsl:variable>
```

Example XSLT code snippets

Simple list

```
<xsl:for-each select="xml/DISS_submission">
  <title>
    <xsl:value-of select="DISS_description/DISS_title"/>
  </title>
</xsl:for-each>
```

Complex list

```
<xsl:for-each select="marc:collection/marc:record">
  <barcode>
    <xsl:value-of select="marc:datafield[@tag='049']/
      marc:subfield[@code='l']"/>
    <CouttsNo>
      <xsl:value-of select="marc:datafield[@tag='909']/
        marc:subfield[@code='a']"/>
    </CouttsNo>
  </barcode>
</xsl:for-each>
```

Example XSLT code snippets

Simple conditionals

Is a 440 field present?

```
<xsl:when test="marc:datafield[@tag='440']">
```

Is the language of the cataloging agency English?

```
<xsl:when test="$catLang='eng'">
```

Is the 1st indicator of the 490 field zero?

```
<xsl:when test="marc:datafield[@tag='490']/@ind1='0'">
```

Is the encoding level less-than-full?

```
<xsl:when test="substring(marc:leader,18,1)='2'
or $encoding='3'
or $encoding='5'
or $encoding='7'
or $encoding='E'
or $encoding='J'
or $encoding='K'
or $encoding='M'">
```

Example XSLT code snippets

Complex conditionals

Is the encoding level 8 *and* are there no digits present in the 300 $a subfield?

```
<xsl:when test="$encoding='8'
        and not(matches(marc:datafield[@tag='300']/
                         marc:subfield[@code='a'], '[0-9]'))">
```

Is the 2nd indicator of the 505 field not zero *and* does the field contains " / " characters?

```
<xsl:when test="marc:datafield[@tag='505'][@ind1]!='0'
        and contains($contentsNote, ' / ')">
```

(Does the 504 field contain the substring "ind" *or* a "Includes index" 500 field) *and* (is the Indx fixed field 0)?

```
<xsl:when test="(contains(marc:datafield[@tag='504']/
  marc:subfield[@code='a'],'ind') or marc:datafield[@tag='500']/
  marc:subfield[@code='a']='Includes index')
        and (substring(marc:controlfield[@tag='008'],32,1)='0')">
```

Example XSLT code snippets

Output if the conditional is true

```
<rows>
<!-- CIP: ELvl 8 records without numbers in extent-->
   <xsl:choose>
      <xsl:when test="$encoding='8' and not(matches(marc:datafield
           [@tag='300']/marc:subfield[@code='a'], '[0-9]'))">
         <column><xsl:value-of select="$batch"/></column>
         <column><xsl:value-of select="'CIP'"/></column>
         <column><xsl:value-of select="$oclc"/></column>
         <column><xsl:value-of select="$title"/></column>
         <column><xsl:value-of select="
              concat('Extent:', $extent)"/></column>
      </xsl:when>
      <xsl:otherwise/>
   </xsl:choose>
</rows>
```

Example quality control report

Excel Report

Batch	Flag	OCLC	Title	Data
20130913	CIP	796750185	Cultivating cosmopolitanism for intercultural communication : communicating as [snip]	Extent: xiv pages : 23 cm.
20130913	CIP	231580825	Fact and relevance : essays on historical method / by M.M. Postan.	Extent: p. cm.
20130913	CIP	841898362	Neuroscience for social work : current research and practice / [edited by] Holly C. Matto, [snip]	Extent: pages ; cm
20130913	CIP	818414664	Growing up Canadian : Muslims, Hindus, Buddhists / edited by Peter Beyer and [snip]	Extent: p. cm.
20130913	CIP	828776017	Peacebuilding in practice : local experience in two Bosnian towns / Adam Moore.	Extent: pages cm
20130913	CIP	832706687	Romantic intimacy / Nancy Yousef.	Extent: pages cm
20130913	CIP	842499932	Unfinished business : screening the Italian Mafia in the new millennium / Dana Renga.	Extent: pages cm.
20130913	CIP	838760712	Who is an indian? : race, place, and the politics of indigeneity in the Americas / edited by [snip]	Extent: p. cm.

Resources for learning XSLT

Online

- http://www.w3schools.com
- http://www.lynda.com
- http://stackoverflow.com/

Texts

- Beginning XSLT and XPath: Transforming XML Documents and Data / Ian Williams

- Learning XSLT / Michael James Fitzgerald

- XSLT Cookbook: Solutions and Examples for XML and XSLT Developers, 2nd Edition / Sal Mangano

- XSLT 2.0 and XPath 2.0 Programmer's Reference / Michael Kay

Any questions?

Annie Glerum

Head of Complex Cataloging

Department of Description and Cataloging

Florida State University Libraries

aglerum@fsu.edu

Resources consulted

- *Bierman, Kenneth J., and Judith A. Carter. "Outsourcing Monograph Cataloging at the UNLV Libraries." Technical Services Quarterly 25.3 (2008): 49-65. Library Literature & Information Science Full Text (H.W. Wilson). Web. 23 July 2013.*

- *Fitzgerald, Michael James. Learning XSLT. Sebastobol, California: O'Reilly Media,Inc., 2003.*

- *Jay, Michael, Betsy Simpson, and Doug Smith. "CatQC and Shelf-Ready Material: Speeding Collections to Users While Preserving Data Quality." Information Technology & Libraries 28.1 (2009): 41-48. Library Literature & Information Science Full Text (H.W. Wilson). Web. 23 July 2013.*

- *Kay, Michael. XSLT 2.0 and XPath 2.0 Programmer's Reference. 4th edition. Indianapolis, Indiana: Wiley Publishing, Inc., 2008.*

- *Mangano, Sal. XSLT Cookbook: Solutions and Examples for XML and XSLT Developers. 2nd Edition. Sebastobol, California: O'Reilly Media,Inc., 2005.*

- *Sanchez, Elaine, Leslie Fatout, and Aleene Howser. "Cleanup of Netlibrary Cataloging Records: A Methodical Front-End Process." Technical Services Quarterly 23.4 (2006): 51-71. Library Literature & Information Science Full Text (H.W. Wilson). Web. 23 July 2013.*

- *Williams, Ian. Beginning XSLT and XPath: Transforming XML Documents and Data. Indianapolis, Indiana: Wiley Publishing, Inc., 2009.*