
“The knowledge gained from learning information technology can be used to experiment with methods of transforming one metadata schema into another using various software solutions”

Violeta Ilik vivoweb.org

Hosted by ALCTS, Association for Library Collections and Technical Services
Why Catalogers?

- Survey conducted by Ma in 2007 revealed that the metadata qualifications and responsibilities required knowledge of MARC, crosswalks, XML, OAI …

- Analysis of cataloging position description performed by Park, Lu, and Marion reveal that advances in technology have created a new realm of desired skills, qualifications and responsibilities for catalogers.

- Terry Reese’s MarcEdit is a best friend to everyone in technical services

- Tosaka stresses the importance of metadata transformation to enable reuse

Violeta Ilik vivoweb.org

Hosted by ALCTS, Association for Library Collections and Technical Services
Recommendations:

- Cataloging departments need to be proactive in creation and maintenance of non-MARC metadata AND in the development of means for sharing the metadata.

- Catalogers need to participate in development and use of descriptive standards.

- Catalogers need to participate in development of semantic web technologies; creation of semantic web compliant data.
Outline:

- MARC data – challenges
- XML family of standards
  - XML & XSLT
- Metadata Map
- Metadata transformation (live demo)
- Resources
- Conclusion

Violeta Ilik vivoweb.org
Hosted by ALCTS, Association for Library Collections and Technical Services
MARC data – challenges
Common issues - metadata and crosswalks

The four categories of metadata problems according to Dushay and Hillman:

- missing data
- incorrect or erroneous data
- confusing or inconsistent data
- insufficient data


Violeta Ilik vivoweb.org
Hosted by ALCTS, Association for Library Collections and Technical Services
MARC data – challenges
The Challenges:

- ambivalent matches
- hybrid bibliographic records
- data mapping to multiple fields or combining into single fields during migration
- orphaned data parsed into incongruous fields
- mixed standards in original data
- MARC data loss during the migration
- flat structure versus hierarchical structures

MARC data – challenges

The Challenges (continued):

- reconciling metadata organization systems
- choice of unanalogous processes during metadata standards creation
- imprecise definitions or alternate naming choices that inhibit element to element mapping
- information being lost or combined during mapping
- unharmonious hierarchical structures

(Margaret St. Pierre and William P. LaPlant Jr., “Issues in Crosswalking Content Metadata Standards,” (white paper, National Information Standards Organization (NISO))
The XML family of standard

- Schema languages
- XSLT
- Xquery
- Xpath
- XSL-FO
- SVG
- Namespaces
- Schematron
- Xproc
- Regular expressions
- Xforms - An XML-oriented replacement for the forms used on the Web on commercial sites and elsewhere.


Violeta Ilik vivoweb.org

Hosted by ALCTS, Association for Library Collections and Technical Services
XML family of standards

Introduction to XML

- Ordered hierarchy of content objects (OHCO); three views of XML (containers, tree, serialization)

- Well-formedness: root element, matching tags, no overlapping elements, quoted attribute values, proper name characters, no reserved characters (ampersand, angle brackets)

- Document analysis (hierarchy; chunks and in-line elements)

- Elements and attributes

- Descriptive vs presentational markup

- Validity (vs well-formedness)


Violeta Ilik vivoweb.org

Hosted by ALCTS, Association for Library Collections and Technical Services
XML family of standards
Ordered hierarchy

- XML is a hierarchical tree

- Tags are not toggle switches that turn properties on and off.

- When you’re writing XML, insert the whole element, with both the start and end tags

XML family of standards

Rules of well-formedness:

- every start tag must have a closing tag <tag> </tag> Or <tag/>

- Tags must nest cleanly <creator><name>Birnbaum, David</name></creator>

- Attribute values must appear within quotation marks <page n="12"/>

- Tags are case sensitive and they must match

  <creator></creator> Or <CREATOR></CREATOR>

- Single root element

- The left angle bracket and ampersand are special characters


Violeta Ilik vivoweb.org

Hosted by ALCTS, Association for Library Collections and Technical Services
XML family of standards
Document analysis (hierarchy; chunks and in-line elements)

- The life cycle of a project starts by determining the structural hierarchy and the semantics
XML family of standards
Elements & attributes

- An element consists of a start tag, an end tag, and content, which is whatever occurs between the tags.
  - XML elements may have four types of content:
    - Element content
    - Text content
    - Mixed content
    - Empty element

- Attributes – provide supplementary information about an element
XML family of standards
Descriptive vs presentational markup

- Descriptive markup is usually mapped onto procedural markup. It is designed to support an open class of applications like information retrieval.

- Presentational markup is designed for reading as it clarifies the presentation of a document.


Violeta Ilik vivoweb.org
Hosted by ALCTS, Association for Library Collections and Technical Services
XML family of standards
Validity (vs well-formedness)

- This is a technical term that means that the document uses only certain elements, and that it uses them only in certain contexts.
XML family of standards
Overview: XSLT

- "XSLT (eXtensible Stylesheet Language Transformations) is one way to transform your document, manipulate the tree, and output the results as XML, HTML, SVG, or plain text."

- An XSLT stylesheet is an XML document and must be valid against the XSLT schema.

- The root element is `<xsl:stylesheet>` - elements inside the root are primarily `<xsl:template>` elements - template elements typically have a `@match` attribute

- XSLT is a declarative programming language
XML family of standards
Namespaces

- Input namespace
- Output namespace

<xsl:stylesheet xmlns:marc="http://www.loc.gov/MARC21/slim"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
    version="2.0" exclude-result-prefixes="marc">
    <xsl:import href="MARC21slimUtils.xsl"/>
    <xsl:output method="xml" indent="yes"/>
</xsl:stylesheet>
XML family of standards
Controlling the output with

\(<\text{xsl:output}>\)

- \(<\text{xsl:output method="xml" indent="yes"} />\>
- \(<\text{xsl:output method="text" indent="no"} />\>
The Library of Congress (LC) developed MARCXML architecture and MARCXML toolkit to standardize the exchange of MARC structured data in XML.
XML family of standards

Introduction to the use of eXtensible Stylesheet Language Transformation (XSLT) for repurposing, editing and reformatting metadata

Review:

- Mapping compares and analyzes two or more metadata schemas, while crosswalks are the product of the mapping process

- Each XSLT stylesheet describes how a set of XML documents (the source documents) should be converted to other documents (the result documents)

Difference between MARC and XML: XML uses beginning tags <> and ending tags ///</>

Violeta Ilik vivoweb.org

Hosted by ALCTS, Association for Library Collections and Technical Services
XML family of standards
Two opposing views on crosswalks:

- Crosswalks are a stopgap measure
- Crosswalks represent an attempt to identify interoperable elements among standards

Metadata can and should be reused, and libraries must ensure interoperability of their metadata for this very reason.

Violeta Ilik vivoweb.org
Hosted by ALCTS, Association for Library Collections and Technical Services
Metadata Map
15 base Dublin Core elements:

- title; creator; subject; description; publisher; contributor; date; type; format; identifier; source; language; relation; coverage; rights
## Metadata Map

<table>
<thead>
<tr>
<th>DC Field</th>
<th>MARC</th>
</tr>
</thead>
<tbody>
<tr>
<td>title</td>
<td>245</td>
</tr>
<tr>
<td>title.alternative</td>
<td>130, 210, 240, 242, 246, 730, 740</td>
</tr>
<tr>
<td>creator Or contributor</td>
<td>100, 110, 111, 700, 710, 711</td>
</tr>
<tr>
<td>date.original [xsitype=&quot;W3CDS&quot;]</td>
<td>260$c$g, 533$d</td>
</tr>
<tr>
<td>description</td>
<td>500-599, except 506, 530, 540, 546</td>
</tr>
<tr>
<td>subject (ex: xsitype=&quot;LCSH&quot;)</td>
<td>600, 610, 611, 630, 650, 653</td>
</tr>
<tr>
<td>type</td>
<td>Leader06, Leader07, 655</td>
</tr>
<tr>
<td>format</td>
<td>300</td>
</tr>
<tr>
<td>format.medium</td>
<td>340$a$, 856$q</td>
</tr>
<tr>
<td>format.extent</td>
<td>300$a$, 533$e</td>
</tr>
<tr>
<td>coverage [ex:scheme=&quot;TGN&quot;]</td>
<td>651, 662, 751, 752</td>
</tr>
<tr>
<td>publisher</td>
<td>260$a$b</td>
</tr>
<tr>
<td>rights</td>
<td>506, 540</td>
</tr>
<tr>
<td>language</td>
<td>008, 041 OR 546</td>
</tr>
</tbody>
</table>
## Metadata map [cont.]

<table>
<thead>
<tr>
<th>Metadata Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>identifier</td>
<td>020, 022, 024, 856$q</td>
</tr>
<tr>
<td>relation</td>
<td>530, 760-787$o$t</td>
</tr>
<tr>
<td>source</td>
<td>534$t, 786 0#$t</td>
</tr>
<tr>
<td>audience</td>
<td>521</td>
</tr>
<tr>
<td>provenance</td>
<td>561</td>
</tr>
</tbody>
</table>
How to get the MARCXML:

- Directly from OCLC
- via MarcEdit workflow (Export from OCLC the usual way >> convert to MARCXML via MarcEdit)
- Export from your ILS
- …
Under Options window select “Record Characteristics” and under Bibliographic Records section and “Record Standard” select MARCXML, under “Character Set” select UTF-8 Unicode:
Next steps:

- Export the selected bibliographic record. It will be stored in the new export destination file you selected.

- Open the file with OxygenXML Editor (or any other similar editor) and you are ready to manipulate the XML file with a XSLT.
Example MARCXML

<collection xmlns="http://www.loc.gov/MARC21/slim"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
<record xmlns="http://www.loc.gov/MARC21/slim">
<leader>00000ngm a2200000ka 4500</leader>
<controlfield tag="001">ocn271676443</controlfield>
<controlfield tag="008">081114s2008 xx 059 vleng d</controlfield>
<datafield tag="007" ind1=" " ind2=" ">
<subfield code="a">v</subfield>
<subfield code="b">f</subfield>
<subfield code="d">c</subfield>
<subfield code="e">z</subfield>
<subfield code="f">a</subfield>
<subfield code="g">h</subfield>
<subfield code="h">z</subfield>
</datafield>
<datafield tag="245" ind1="0" ind2="4">
<subfield code="a">The future of humanities scholarship in a digital world</subfield>
<subfield code="h">[videorecording].</subfield>
</datafield>
<datafield tag="260" ind1=" " ind2=" ">
<subfield code="c">2008.</subfield>
</datafield>
<datafield tag="300" ind1=" " ind2=" ">
<subfield code="a">1 videocassette (59 min.) :</subfield>
<subfield code="b">sd., col. ;</subfield>
<subfield code="c">1/5 in.</subfield>
</datafield>
<datafield tag="500" ind1=" " ind2=" ">
<subfield code="a">Title from container.</subfield>
</datafield>
<datafield tag="511" ind1=" " ind2=" ">
<subfield code="a">Laura Mandell.</subfield>
</datafield>
<datafield tag="518" ind1=" " ind2=" ">
<subfield code="a">Recorded on August 1, 2008, in 105 Dartmouth Hall, Dartmouth College.</subfield>
</datafield>
<datafield tag="538" ind1=" " ind2=" ">
<subfield code="a">DVCam.</subfield>
</datafield>
<datafield tag="655" ind1=" " ind2=" ">
<subfield code="a">College films and programs.</subfield>
<subfield code="2">mim</subfield>
</datafield>
<datafield tag="700" ind1="1" ind2="8">
<subfield code="a">Mandell, Laura.</subfield>
</datafield>
<datafield tag="710" ind1="2" ind2="8">
<subfield code="a">Dartmouth College.</subfield>
<subfield code="b">Media Production Group.</subfield>
</datafield>
<datafield tag="710" ind1="2" ind2="8">
<subfield code="a">Dartmouth College.</subfield>
<subfield code="b">Research Computing.</subfield>
</datafield>
<record>
</collection>
Using OxygenXML Editor

Violeta Ilik vivoweb.org

Hosted by ALCTS, Association for Library Collections and Technical Services
Resources
Cataloging Resources and tools:

- OCLC Bibliographic formats and standards (general MARC reference)
- US Library of Congress MARC 21 format for bibliographic data (general MARC reference)
- MARCXML home (reference and tools)
- MarcEdit (freeware Windows-only application to edit MARC records, including individual and batch transformation)
- Dublin Core Metadata Initiative (home)
- Metadata Object Description Schema (MODS) home (reference and tools)
- …

Violeta Ilik vivoweb.org

Hosted by ALCTS, Association for Library Collections and Technical Services
Resources
XSLT and related references:

*Online:*

- [FunctX XSLT](#) function library
- [Dave Pawson’s XSLT FAQ](#) page
- [w3schools](#) tutorials and references

*Books:*

- [Michael Kay’s XSLT 2.0 and XPath 2.0 programmer’s reference](#) (reference)
- [Jeni Tennison’s Beginning XSLT 2.0: From novice to professional](#) (tutorial)

Violeta Ilik [vivoweb.org](#)  
Hosted by ALCTS, Association for Library Collections and Technical Services
Resources
Files on Git:

- Transforming-MARCXML-with-XSLT

https://github.com/vioil/Transforming-MARCXML-with-XSLT

Violeta Ilik vivoweb.org

Hosted by ALCTS, Association for Library Collections and Technical Services
Conclusion

Catalogers have the potential to undertake metadata projects by active participation in metadata transformation.

Violeta Ilik [vivoweb.org](http://vivoweb.org)
My teachers and mentors

- Dr. David Birnbaum, Chair, Slavic Languages & Literature Department, University of Pittsburg

- Dr. Laura Mandell, Director, Initiative for Digital Humanities, Media, and Culture; Professor, Department of English, Texas A&M University

- Matthew Gibson, Director of Digital Initiatives; Editor, Encyclopedia Virginia; Virginia Foundation for the Humanities

- Cristine Ruotuolo, Digital Services Manager for Humanities and Social Sciences and Bibliographer for English Language and Literature at the University of Virginia Library

Violeta Ilik vivoweb.org

Hosted by ALCTS, Association for Library Collections and Technical Services
Thank you
Violeta Ilik
vivoweb.org

Hosted by ALCTS, Association for Library Collections and Technical Services