

MARC in XML

Description and Application

Sally H. McCallum
Library of Congress

Content

❖ XML

❖ MARC in XML

- MARCXML
- Tool Kit

❖ Samples of applications

XML – eXtensible Markup Language

❖ Why interested in XML

- XML is flexible, thus suitable for MARC data
- Has powerful (and easy to use) transformation language, XSLT
- Has combining characteristics through namespaces
- Embraced by the open source movement – computer community popularity
- Many electronic resources are XML
- New generation systems support XML
- Extensive tool creation taking place
- Used for other new metadata formats

XML basics

- ❖ XML is not a programming language – similar to ISO 2709 (structure for MARC)
- ❖ XML is a set of elements with tags and rules that can be used to markup data – capable of extensive hierarchy
- ❖ The tags are well-defined by, for example, XML Schema
- ❖ Developers can define their own tags and schema – tagging freedom

XML basics

- ❖ **Element tags**

- `<name>`

- ❖ **Subelement tags**

- `<name><namePart>...<date>`

- ❖ **Elements can have *attributes***

- `<name type="personal">`

- ❖ **All tags close**

- `<name>...</name>`

- ❖ **Example:**

- `<name type="personal"><namePart>Smith,
John</namePart><date>1930-</date></name>`

One example of XML

❖ MARC documentation is marked up in XML

- Using one XML file, can produce:
 - pdf for printed full and concise formats
 - Online concise
 - Online full
 - Online lite format
 - Online field list
- Other XML files are maintained for
 - MARC code lists
 - MARC online character set listings

MARC 21 in XML requirements

❖ Need to take advantage of emerging tools and systems that use XML

- SRU (next generation of Z39.50 search protocol)
- OAI (metadata harvesting protocol)
- METS (Metadata Encoding & Transmission Schema)
- *Establish standard MARC 21 in an XML structure*

❖ Need interoperability with other new XML schemas

- DC (use data from Dublin Core in MARC environment)
- ONIX (use data from ONIX in the MARC environment)
- *Assemble coordinated set of tools*

MARC 21 in XML requirements

- ❖ **Must have easy interchange with current data and systems**
 - **Pathway from MARC 21 “classic” to MARCXML and other metadata formats**
 - ***Provide flexible transition options***

Early experimentation for MARC

❖ SGML DTD developed ~1995

- Standard Generalized Markup Language (SGML) – Document Type Definition (DTD)
- Bibliographic DTD
- Authority DTD
- Defined element tag for each MARC subfield and character position
 - Enabled detailed validation
 - Enabled element use out of context
 - But, DTD is very large – difficult to use

Establish standard MARC 21 in XML

New approach - MARCXML

- ❖ **Simple “slim” schema, no change needed when MARC 21 changes**
- ❖ **All the elements of MARC 21 in an XML structure**
- ❖ **Lossless roundtrip conversion to/from MARC 21 – all tags, indicators, and data convert**
- ❖ **MARC tag numbers used**

Establish standard MARC 21 in XML

MARXML tags

- ❖ <leader>
- ❖ MARC directory not relevant to MARXML
- ❖ <controlfield> (MARC21 tags 001-009)
- ❖ <datafield> (MARC21 tags 010-)
 - <datafield><subfield>
 - With attributes for tags, indicators, subfield codes
 - ❖ <datafield tag="xxx" ind1="x" ind2="x">
 - ❖ <subfield code="x">

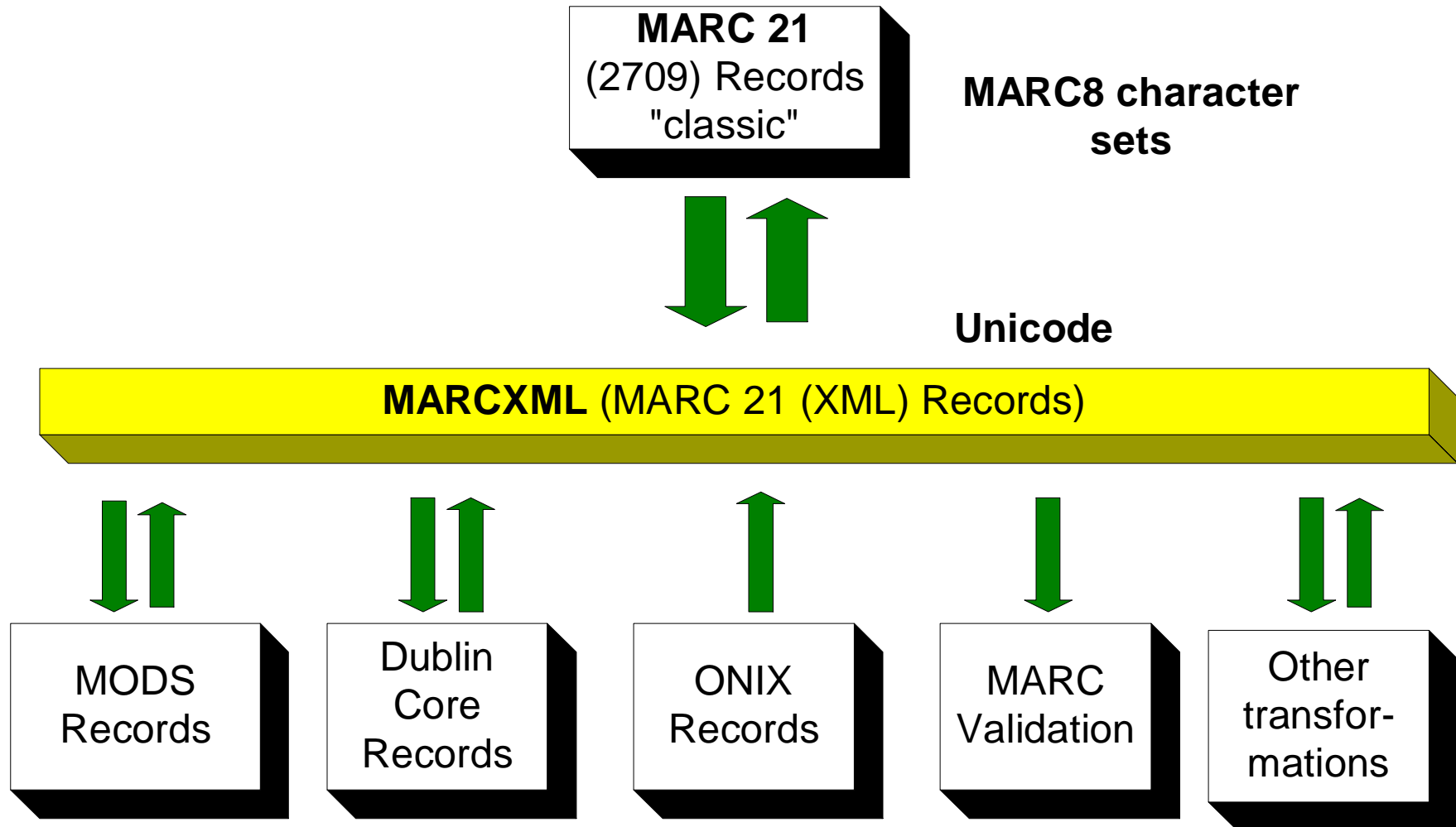
Snip of MARCXML data

```
<leader>01295cam a22003134a 4500</leader>
<controlfield tag="001">2004004615</controlfield>
...
<datafield tag="100" ind1="1" ind2=" " >
  <subfield code="a">Kent, Neil,</subfield>
</datafield>
<datafield tag="245" ind1="1" ind2="0">
  <subfield code="a">Helsinki :</subfield>
  <subfield code="b">a cultural and literary history /</subfield>
  <subfield code="c">Neil Kent</subfield>
</datafield>
<datafield tag="260" ind1=" " ind2=" " >
  <subfield code="a">New York :</subfield>
  <subfield code="b">Interlink Books,</subfield>
  <subfield code="c">2005.</subfield>
</datafield>
```

Assemble tools

MARC tool kit

(arrows indicate transformations downloadable from MARC website)



Tool kit transformation

MARC 21 → MARCXML → DC

<title>**Helsinki: a cultural and literary history**</title>

<creator>**Kent, Neil**</creator>

<type>**text**</type>

<publisher>**New York : Interlink Books,**</publisher>

<date>**2005.**</date>

<language>**eng**</language>

<description>**Includes bibliographical references (p. 237-238) and indexes.**</description>

<coverage>**Helsinki (Finland)—Intellectual life.**</coverage>

<coverage>**Helsinki (Finland)--Description and travel.**</coverage>

<identifier>**URN:ISBN:1566565448 (pbk.)**</identifier>

Sample applications of
MARCXML

Metadata switch

❖ Terminology Project of the OCLC Office of Research

- Switching service for vocabularies, e.g., DDC, LCC, LCSH, MeSH, GSAFD, ERIC, NGL
- Receive XML, html, MARC 21, etc. from thesaurus source
- Normalizing format – MARCXML
 - ❖ Utilizes rich detail of MARC 21
 - ❖ Utilizes flexibility of XML and XSLT style sheets

“Vendor-neutral” format

- ❖ **Los Alamos National Labs needed vendor-neutral format**
 - required a format for 87,000,000 metadata records from a variety of sources
 - Evaluated several different formats, MARC was best at accommodating a wide variety of data elements
 - Transform all incoming data into MARCXML from native format
 - Needed XML data for working with other parts of system
- ❖ **Selected MARCXML based on:**
 - XML
 - granularity, versatility, extensibility, hierarchy support
 - crosswalks available, tools available
 - cooperative and stable management, and widespread use.

MARC open source tool

❖ MarcEdit utility

- <http://oregonstate.edu/~reese/marcedit/html>
- **Editors**
 - MARC 21 to MARCXML – then variety of tools
 - Integration with other software
- **Crosswalks via MARCXML**
 - EAD to MARC 21
 - Geospatial to MARC 21
 - DC to MARC 21
 - ❖ Ex. Conversion of Dspace's Dublin Core records to MARC21 for loading into a catalog

Record maintenance at New York University

❖ Records transformed to MARCXML for change processing

- **New batches of MARC 21 records are converted to MARCXML and adjusted prior to load**
 - Change URLs and create MARC 21 holdings records
 - Create reproduction notes from data in record and system supplied data
- **“Global update”**
 - Subject heading changes
- **Identify special subsets of records**
 - Match publisher numbers, insert URIs for digitized material
 - Extract records for cooperative projects

XML-based protocols

❖ OAI-PMH – XML required for records

- Open Archives Initiative-Protocol for Metadata Harvesting (OAI-PMH)
- MARCXML became recommendation for MARC records in 2002
- Standard format a great help for harvesters

❖ SRU – XML required for records

- Search and Retrieve via URL (SRU)

❖ Virtual International Authority File (IFLA initiative)

- MARCXML records to be accessible via SRU (for persons) and OAI (for machines)

Library of Congress distribution

- ❖ **OPAC bibliographic records accessible via SRU, with records retrieved sent back in choice of MARCXML, MODS and DC**
- ❖ **Provide records for LC digital projects for OAI harvesting in choice of MARCXML, MODS, DC – conversion from MARC 21 “on-the-fly” using tool kit transformations**
- ❖ **Bibliographic and authority MARC records distributed by the LC Cataloging Distribution Service are available in MARCXML**

Summing up

- ❖ MARCXML provides the basis for evolution of MARC to the XML environment
- ❖ Access to XML tools is essential for the expanding ability to change records
- ❖ Downloadable transformations help to keep us standard
- ❖ Should MARCXML take on XML features that will not translate to MARC 21?
- ❖ Visit MARCXML at www.loc.gov/marcxml
- ❖ Questions?