

An XML representation of the UNIMARC manual: a working prototype

Joaquim Carvalho

BookMARC, Ida
jrc@bookmarc.pt

Summary

- Why a XML representation of the UNIMARC manual is needed?
- Main structure: Leader, Fields and Subfields.
- Small details: positional fields, examples cross references.
- How it works: documentation, validation, decoding of records.
- Open issues: internationalization, conditional; rules, other formats

Why?

- The UNIMARC manual contains information relevant for the validation and interpretation of specific records and needs to be readily available in human and machine readable formats.
- The information in the manual must be represented in computer systems that process UNIMARC data.
- Providing the manual in XML greatly simplifies the full implementation of the format in computer systems.

Goal (from UNIMARC roadmap)

- Produce a formal representation of the UNIMARC format, so that the standard can be incorporated in software systems in a transparent way.
 - 1) An XML representation of the UNIMARC manual, which can be processed automatically by applications that need to enforce the format rules, provide help information, or lists of legal values.
 - 2) A set of services definitions related to validation or introspection of records.

The Schema

- Overview
 - The Schema FORMAT.xsd
 - Graphical overview

Sample:Leader

```
<?xml version="1.0" encoding="UTF-8" ?>
- <FORMAT name="UNIMARC" domain="Bibliographic" version="1" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
- <LEADER mandatory="y" repeatable="n">
- <RECORD-STATUS name="Record Status" start="6" length="1">
  <OPTION name="Corrected Record" value="c" />
  <OPTION name="Deleted Record" value="d" />
  <OPTION name="New Record" value="n" />
  <OPTION name="Previously Issued Higher Level Record" value="o" />
  <OPTION name="Previously Issued as an Incomplete, Pre-publication Record" value="p" />
</RECORD-STATUS>

- <TYPE-OF-RECORD name="Type of Record" start="7" length="1">
  <OPTION name="Language Materials, Printed" value="a" />
  <OPTION name="Language Materials, Manuscript" value="b" />
  <OPTION name="Music Scores. Printed" value="c" />
  <OPTION name="Music Scores, Manuscript" value="d" />
  <OPTION name="Cartographic Materials, Printed" value="e" />
  <OPTION name="Cartographic Materials, Manuscript" value="f" />
  <OPTION name="Projected and Video Material (Motion Pictures, Filmstrips, Slides, Transparencies)" value="g" />
  <OPTION name="Sound Recordings, Non-musical Performance" value="i" />
  <OPTION name="Sound Recordings, Musical Performance" value="j" />
  <OPTION name="Two-dimensional Graphics (Pictures, Designs etc.)" value="k" />
  <OPTION name="Electronic resources" value="l" />
  <OPTION name="Multimedia" value="m" />
  <OPTION name="Three-dimensional Artefacts and Realia" value="n" />
</TYPE-OF-RECORD>
```

Sample: control field

- <FIELD tag="001" name="RECORD IDENTIFIER" mandatory="y" repeatable="n">
<DESCRIPTION>This field contains characters uniquely associated with the record, i.e. the control field.</DESCRIPTION>
- <IND1>
<DESCRIPTION>In conformance with ISO 2709 this field does not have indicators.</DESCRIPTION>
</IND1>
- <IND2>
<DESCRIPTION>In conformance with ISO 2709 this field has no indicators.</DESCRIPTION>
</IND2>
- <SUBFIELD tag="*" name="" mandatory="n" repeatable="n">
<DESCRIPTION>In conformance with ISO 2709 this field does not contain subfields.</DESCRIPTION>
</SUBFIELD>
- <NOTES>There are no restrictions on the form of the record identifier. When the record identifier is used to identify a publication, this information is nevertheless to be entered in the field specified for that data in a separate field.</NOTES>
- <RELATED tag="*" name="">
<DESCRIPTION>An agency may be using other separately identified numbers such as ISBN or National Library of Medicine Number.</DESCRIPTION>
</RELATED>
- <EXAMPLES>
- <EX n="1">
001 2002-00000000011
<P />
A number in the Joint Catalogue file of the South African State Library.
</EX>

Sample: data field

- <FIELD tag="071" name="PUBLISHER'S NUMBER (SOUND RECORDINGS AND MUSIC)" mandatory=<DESCRIPTION>This field contains publisher's numbers used for sound recordings and publicatio
- <IND1 name="Type of Publisher's Number"><DESCRIPTION>This indicator specifies the type of publisher's number contained in the field. It r<OPTION value="0" name="Issue number" /><OPTION value="1" name="Matrix number" /><OPTION value="2" name="Plate number" /><OPTION value="3" name="Other publisher's number (printed music)" /><SEEEX N="2" /><SEEEX N="4" /></IND1>
- <IND2 name="Note Indicator"><DESCRIPTION>This indicator specifies whether the agency supplying the record would use the (ould be set to 0.</DESCRIPTION><OPTION value="0" name="Do not make a note" /><OPTION value="1" name="Make a note" /></IND2>
- <SUBFIELD tag="a" name="Publisher's Number" mandatory="n" repeatable="n"><DESCRIPTION>The number as assigned by the agency. Not repeatable.</DESCRIPTION></SUBFIELD>
- <SUBFIELD tag="b" name="Source" mandatory="n" repeatable="n"><DESCRIPTION>The agency which assigned the number. Not repeatable.</DESCRIPTION></SUBFIELD>

Sample: coded field

- `<FIELD tag="100" name="GENERAL PROCESSING DATA" mandatory="y" repeatable="n">`
 `<DESCRIPTION>This field contains fixed-length coded data applicable to records of materials in a`
 `<IND1 name="blank (not defined)" />`
 `<IND2 name="blank (not defined)" />`
- `<SUBFIELD tag="a" name="General Processing Data" mandatory="y" repeatable="n">`
 `<DESCRIPTION>All data entered in $a is identified by character position within the subfield. Con`
- `<PSUBFIELD n="1" name="Date Entered on File" mandatory="y" start="0" end="7">`
 `<DESCRIPTION>Eight numeric characters in ISO standard form (ISO 8601-1988) for dates: YYYY`
 date when the record was originally created and entered into machine-readable form, to give sor
 retain its original date.`</DESCRIPTION>`
- `<EXAMPLES>`
 `<EX n="1">Character positions 0-7: 19671005 The record was originally entered into a machine`
 `</EXAMPLES>`
 `</PSUBFIELD>`
- `<PSUBFIELD n="2" name="Type of Publication Date" start="8" mandatory="n" end="8">`
 `<DESCRIPTION>A one-character code indicates the type of date according to eleven categories w`
 ns 9-12) and Publication Date 2 (character position 13-16). These dates are always entered as a 1
 been supplied as an approximation in a corresponding field.`</DESCRIPTION>`
- `<VOCABULARY type="local" name="Type of Publication Date">`
- `<ITEM code="a" name="currently published serial">`
 `<DESCRIPTION>Publication date 1 contains the beginning year of publication or coverage if cove`
 RIPTION>

Detail: examples, notes and relations

- <FIELD tag="208" name="MATERIAL SPECIFIC AREA: PRINTED MUSIC SPECIFIC STATEMENT" |
<DESCRIPTION>This field contains a statement, transcribed from the item, relating to the part
<IND1 name="blank (not defined)" />
<IND2 name="blank (not defined)" />
- <SUBFIELD tag="a" name="Printed Music Specific Statement" mandatory="n" repeatable="n">
<DESCRIPTION>The statement related to the particular format of the item in hand, taken from
<SEEEX N="1" />
<SEEEX N="2" />
</SUBFIELD>
- <SUBFIELD tag="d" name="Parallel Printed Music Specific Statement(s)" mandatory="n" repeata
<DESCRIPTION>Any additional statements in a different language and/or script, taken from th
</SUBFIELD>
- <NOTES>
Parallel data: Parallel data appearing on the item, i.e. data repeated in another language or s
ata whenever required by ISBD, preceding and in the same subfield as the data to which it rela
<SEEEX N="2" />
</NOTES>
- <RELATED tag="125" name="CODED DATA: SOUND RECORDINGS AND MUSICAL SCORES">
<DESCRIPTION />
</RELATED>
- <EXAMPLES>
<EX n="1">208 ##\$aMiniature score</EX>
<EX n="2">208 ##\$aOrchestr-Partitur\$d= Full score</EX>
</EXAMPLES>
</FIELD>

Detail: conditional rules

```
- <FIELD tag="206" name="MATERIAL SPECIFIC AREA: CARTOGRAPHIC MATERIALS -  
MATHEMATICAL DATA" mandatory="c" repeatable="y">  
  <DESCRIPTION>The statements of scale, projections, coordinates and equinox of a  
  cartographic item.</DESCRIPTION>  
  <IND1 name="blank (not defined)" />  
  <IND2 name="blank (not defined)" />  
  
- <MANDATORY-CONDITION>  
  <RULE>substring($leader, 7, 1) = 'e' or substring($leader, 7, 1) = 'f'</RULE>  
  <NOTES>This field is mandatory for cartographic material.</NOTES>  
</MANDATORY-CONDITION>
```

How it works

- The XML representation of the manual was designed to be convenient to use together with standard XML transformations.
- By applying transformations to the manual and/or UNIMARC records in XML many useful tasks can be accomplished.

Sample transformations

Producing HTML view of the manual

```
./bin/transform.sh src/xml/manual/Leader.xml  
src/stylesheets/FORMATtoHTML.xsl leader.html
```

```
./bin/transform.sh src/xml/manual/Unimarc2.xml  
src/stylesheets/FORMATtoHTML.xsl unimarc2.html
```

Explaining records

```
./bin/transform.sh src/xml/Record.xml  
src/stylesheets/HTMLFormater.xsl explained.html
```

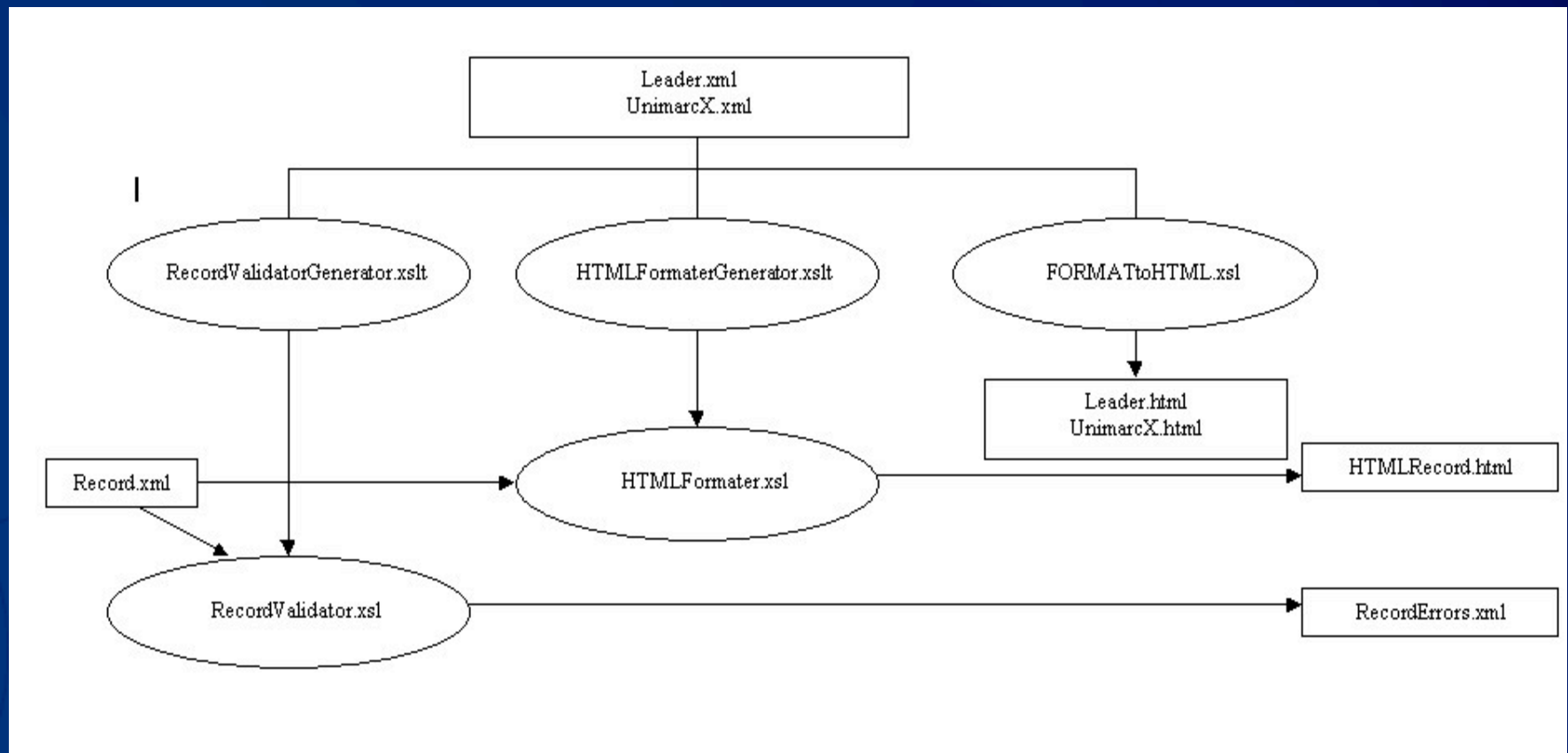
Validating records

```
./bin/transform.sh src/xml/Record.xml  
src/stylesheets/RecordValidator.xsl validated.xml
```

More transformations

- Generating an updated validation stylesheet
- `./bin/transform.sh src/xml/FORMATDescription.xml
src/stylesheet/RecordValidatorGenerator.xslt
validator.xsl`

Functionality of the kit



Command line examples

- **Generating documentation in HTML**
- `./bin/transform.sh src/xml/manual/Leader.xml
src/stylesheets/FORMATtoHTML.xsl LeaderDoc.html`
- **Validating Records**
- `./bin/transform.sh src/xml/Record.xml
src/stylesheets/RecordValidator.xsl isitvalid.xml`
- **Decoding records**
- `./bin/transform.sh src/xml/Record.xml
src/stylesheets/HTMLFormater.xsl myrecord.html`

Possibilities

- Automatic updating of software systems from the standard source.
- Generic library systems read XML format rules for UNIMARC or MARC21 or variants.
- Lower cost of developing sophisticated functionality like complex conditional validation, on line help etc.

Open Issues

- Internationalization
 - What would be the best way to deal with different language versions of the Manual?
 - Different XML documents?
 - Single Document with language variants?
- Are all the conditional rules expressed in terms of the “record”?
- Should the Manual assume a given standard for the representation of bibliographic records, or should it be neutral?
- Should an effort be made to express MARC21 in the same schema?
- What about authorities, classification, ...?
- Licensing and copyright issues.

More information

- Joaquim Carvalho
– Jrc@bookmarc.pt