Representation of the UNIMARC bibliographic data format in Resource Description Framework

Gordon Dunsire, Mirna Willer, Predrag Perožić
Presented at DC-2013, Lisbon, Portugal, 5 September 2013
UNIMARC

• Universal Machine Readable Cataloguing
  – Maintained by the Permanent UNIMARC Committee (PUC) of the International Federation of Library Associations and Institutions (IFLA)
  – First published in 1977

• Specifies formats for encoding Authority, Bibliographic, Classification and Holdings data
  – Based on ISO 2709, library content standards, etc.
Project

• Representation of UNIMARC in RDF
  – Funded for first year by PUC
    • Will take more than 1 year ...
  – Focus on UNIMARC Bibliographic format

• To support production of datasets from UNIMARC catalogues
  – Used in Europe, North Africa, Russia, China, Japan

• To support linked data interoperability with related IFLA standards and beyond
Element sets

• “Bibliographic” format has same focus as International Standard Bibliographic Description (ISBD)
  – The entity [bibliographic] Resource ~ Manifestation

• Attributes => RDF properties

• Lossless data requires finest level of granularity
  – Qualified UNIMARC coded subfield
Value vocabularies

• Coded information stored in tag block 1xx
  – Code lists specify notation, term, description, and scope
• Represented as RDF/SKOS vocabularies
  – Italian and Portuguese translations – multilingual environment
  – Interoperability with vocabularies of other schema
• 12 published so far
  – For example: Target audience
<table>
<thead>
<tr>
<th>Preferred Label</th>
<th>URI</th>
<th>Status</th>
<th>Updated</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>adult, general</td>
<td>../ns/unimarc/terms/tac#m</td>
<td>Published</td>
<td>2013-03-28 9:37</td>
<td></td>
</tr>
<tr>
<td>adult, serious</td>
<td>../ns/unimarc/terms/tac#k</td>
<td>Published</td>
<td>2013-03-28 9:36</td>
<td></td>
</tr>
<tr>
<td>children, ages 9-14</td>
<td>../ns/unimarc/terms/tac#d</td>
<td>Published</td>
<td>2013-03-28 9:35</td>
<td></td>
</tr>
<tr>
<td>juvenile, general</td>
<td>../ns/unimarc/terms/tac#a</td>
<td>Published</td>
<td>2013-03-28 9:32</td>
<td></td>
</tr>
<tr>
<td>pre-primary, ages 0-5</td>
<td>../ns/unimarc/terms/tac#b</td>
<td>Published</td>
<td>2013-03-28 9:33</td>
<td></td>
</tr>
<tr>
<td>primary, ages 5-10</td>
<td>../ns/unimarc/terms/tac#c</td>
<td>Published</td>
<td>2013-03-28 9:34</td>
<td></td>
</tr>
<tr>
<td>unknown</td>
<td>../ns/unimarc/terms/tac#u</td>
<td>Published</td>
<td>2013-03-28 9:38</td>
<td></td>
</tr>
<tr>
<td>young adult, ages 14-20</td>
<td>../ns/unimarc/terms/tac#e</td>
<td>Published</td>
<td>2013-03-28 9:36</td>
<td></td>
</tr>
</tbody>
</table>
URI design templates

**Element set** granularity at subfield level with superstructure of fields (tags) and 2 qualifiers (indicators). Coded subfields refined by character position.

<table>
<thead>
<tr>
<th>Tag</th>
<th>Ind 1</th>
<th>Ind 2</th>
<th>Subfield</th>
<th>CharPos</th>
<th>URI</th>
<th>Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>1</td>
<td>_ [blank]</td>
<td>a</td>
<td></td>
<td>2001_a</td>
<td>Title proper</td>
</tr>
<tr>
<td>100</td>
<td>_</td>
<td>_</td>
<td>a</td>
<td>17</td>
<td>100__a17</td>
<td>Target audience code 1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vocabulary token</th>
<th>Code</th>
<th>URI</th>
<th>Vocabulary: Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>tac</td>
<td>m</td>
<td>tac#m</td>
<td>Target audience: adult, general</td>
</tr>
</tbody>
</table>
Target audience code

“applicable to records of materials in any media”

Subfield a, character positions 17-19, of tag 100 General processing data

3 instances of one-character code

100__a17

100__a18

100__a19

Order of position carries no significance in UNIMARC format

But content rules may assign significance
Mappings

• UNIMARC tags and subfields have corresponding ISBD “elements”
  – Now out-of-date after publication of ISBD consolidated edition
  – Category of alignment relationship to be determined
    • Equivalent or broader/narrower
  – To be used as basis for sub-property mappings

• Mappings from UNIMARC to other vocabularies being developed
Granularity

• Intellectual value of UNIMARC is preserved by a finest-grained semantic representation

• Data can always be dumbed-down to the level of coarseness required by applications
  – Processed with shared open maps
  – Including schema.org and dct!
    • And BIBFRAME too ...

• Data should be published without loss
  – For semantically rich applications

• Universal Bibliographic Control ~ Semantic Web
Thank you!