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

Vocabulary: Show detail for UNIMARC: Target audience

| Detail | Concepts | History | Versions | Maintainers |

Preferred Label	URI	Status	Updated	Actions
✔ adult, general	.../ns/unimarc/terms/tac#m	Published	2013-03-28 9:37	
✔ adult, serious	.../ns/unimarc/terms/tac/#k	Published	2013-03-28 9:36	
✔ children, ages 9-14	.../ns/unimarc/terms/tac#d	Published	2013-03-28 9:35	
✔ juvenile, general	.../ns/unimarc/terms/tac#a	Published	2013-03-28 9:32	
✔ pre-primary, ages 0-5	.../ns/unimarc/terms/tac#b	Published	2013-03-28 9:33	
✔ primary, ages 5-10	.../ns/unimarc/terms/tac#c	Published	2013-03-28 9:34	
✔ unknown	.../ns/unimarc/terms/tac#u	Published	2013-03-28 9:38	
✔ young adult, ages 14-20	.../ns/unimarc/terms/tac#e	Published	2013-03-28 9:36	

8 results

http://metadataregistry.org/concept/list/vocabulary_id/322.html

URI design templates

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

Value vocabulary granularity at code level. Hash URIs used if code list is small, or self-referential ("other", etc.)

Tag	Ind 1	Ind 2	Subfield	CharPos	URI	Attribute
200	1	_ [blank]	a		2001_a	Title proper
100	_	_	a	17	100__a17	Target audience code 1

Vocabulary token	Code	URI	Vocabulary: Term
tac	m	tac#m	Target audience: adult, general

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

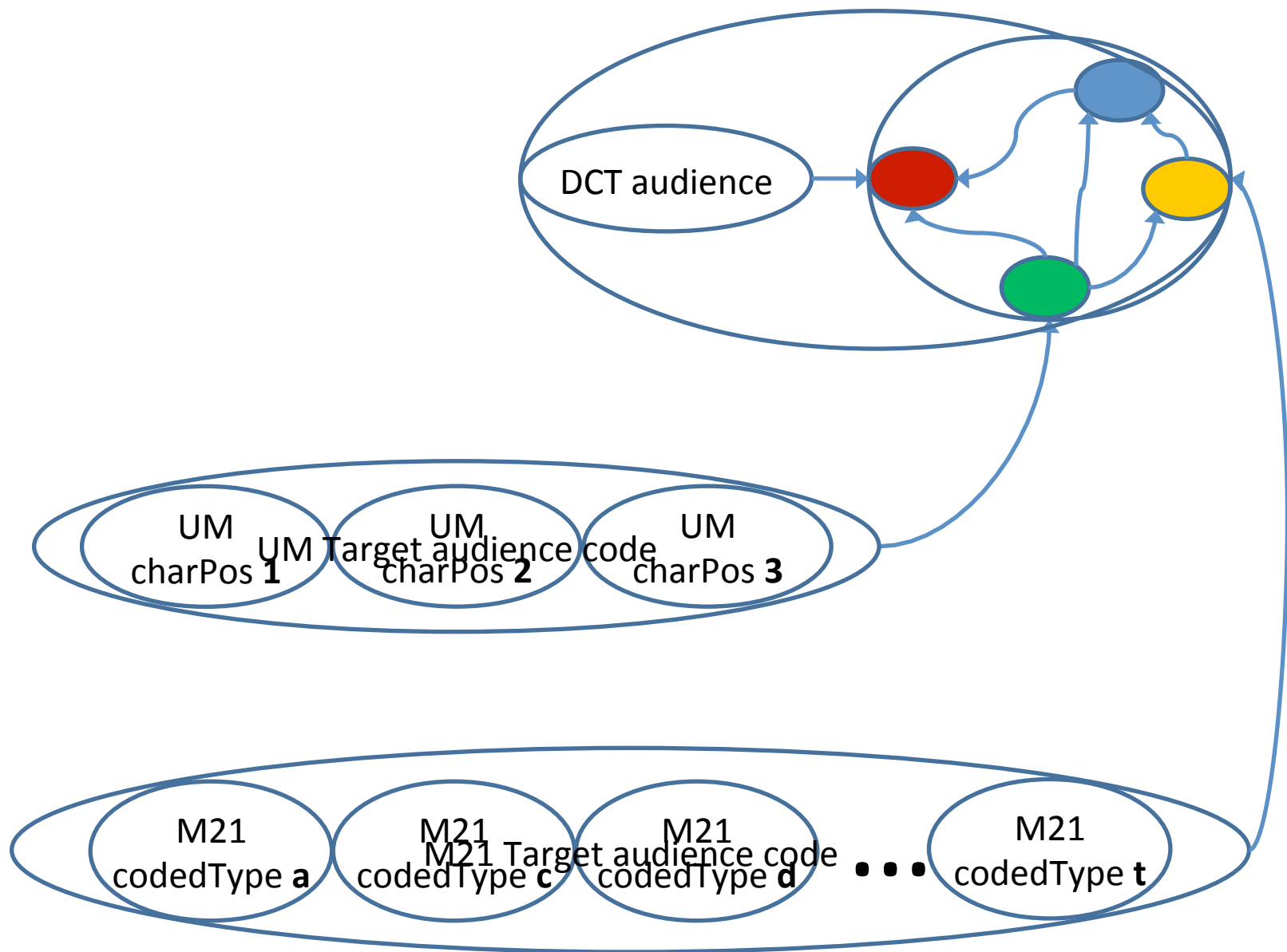**100__a17-19**

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

DCT audience

UM Target audience code

UM charPos **1**

UM charPos **2**

UM charPos **3**

M21 Target audience code

M21 codedType **a**

M21 codedType **c**

M21 codedType **d**

M21 codedType **t**

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!