DCPAPERS
*Proc. Int'l Conf. on Dublin Core and Metadata Applications 2013*

# Representation of the UNIMARC Bibliographic Data Format in Resource Description Framework

Gordon Dunsire
Consultant, Scotland
gordon@gordondunsire.com

Mirna Willer
University of Zadar, Croatia
mwiller@unizd.hr

Predrag Perožić
University of Zadar, Croatia
pperozic@gmail.com

## Abstract

This paper describes the history and role of the UNIMARC bibliographic data formats, as background to a discussion of preliminary outcomes of a project to represent the formats in Resource Description Framework (RDF) and map them to related standards, including the International Standard Bibliographic Description (ISBD). These include the testing of the strategy for namespace and URI design and the methodology for populating them with content, identification of alignment inconsistencies, and preliminary mappings to Dublin Core, RDA: resource description and access, and MARC 21. The paper discusses the relevance of these standards to the aims of Universal Bibliographic Control (UBC) and user-focused applications in the Semantic Web.

**Keywords:** semantic maps; mapping; ontologies; libraries; library metadata; dumb-down; Semantic Web; Linked Data; Linked Open Data; LOD; RDF; RDFS; OWL; UBC; FRBR; ISBD; International Standard Bibliographic Description; cataloguing; UNIMARC; RDA: resource description and access; MARC 21; Dublin Core Metadata Initiative; DCMI.

## 1. Introduction and Background

The UNIMARC: UNIversal MAchine Readable Cataloguing format is a standard supported by the International Federation of Library Associations and Institutions (IFLA) with a primary function to facilitate the international exchange of bibliographic data in machine readable form. It is intended to be a carrier format for interchange purposes, and therefore does not stipulate the form, content or record structure of the data as it is held within individual systems. It may also be used as a model for the development of new machine-readable bibliographic formats.

UNIMARC is maintained by the IFLA Permanent UNIMARC Committee (PUC) formed in 1991; the secretariat is the UNIMARC Core Activity (UCA), one of the six IFLA Strategic Programmes, and has been hosted by the National Library of Portugal in Lisbon since 2003 (IFLA, 2013-1). UNIMARC comprises a set of four formats covering bibliographic description, name and subject authority control, classification, and collection holdings data, with correspondence and interrelation based on the same underlying logical and physical record structure, together with a system of content designators for content data. The first UNIMARC format, which contained specifications for bibliographic data, was published in 1977 as "UNIMARC Universal MARC Format", followed by "UNIMARC/Authorities Format" in 1991, "UNIMARC Classification Format" in 1996, and "UNIMARC Holdings Format" in 2004. All four formats have been maintained and updated in regular intervals required by the stability of the formats, development of the standards on which they are based, and user requirements (IFLA, 2013-2).

The UNIMARC formats are built on content standards, including other IFLA standards relevant for bibliographic information organization such as ISBD: International Standard Bibliographic Description, GARR: Guidelines for Authority and Reference Records, classification schemes such as Universal Decimal Classification (UDC), Dewey Decimal Classification (DDC), and ISO 10324: 1997 – Information and documentation – Holdings





https://doi.org/10.23106/dcmi.952136192

statements – Summary level for holdings data. Logical and physical levels are based on ISO 2709 which specifies the structure of records containing bibliographic data for exchange purposes; adherence to this standard allows interoperability at structural level with other MARC formats such as MARC 21. Other ISO standards used include ISO 464 for control functions and graphic characters such as record label, directory, indicators, subfield identifiers and code values; ISO 6630 for filing, sorting, permuting, etc., and ISO 5426 (extended Latin set), ISO 5427 (extended Cyrillic set), ISO 6438 (African coded character set), ISO 10646 Level 3 (Unicode, UTF-8), etc. for alphabet and/or script of contents. The formats also use a system of codes for some data elements, developed specifically for UNIMARC; some of these have close correspondence to MARC 21 codes.

The PUC liaises with other related MARC maintenance agencies such as Comité français UNIMARC, LC MARC Standards Office, Canadian Committee on MARC, OCLC Bibliographic Formats and Standards, as well as ISSN Network and other relevant bodies. Several mappings between UNIMARC formats and MARC 21, Dublin Core and EAD (Electronic Archival Description) have been published for various needs and purposes. UNIMARC and the other IFLA bibliographic standards such as ISBD and Functional Requirements for Bibliographic Records (FRBR) form the basis of the Universal Bibliographic Control (UBC) core programme. UBC is based on the goal of "promotion of a world-wide system for control and exchange of bibliographic information. The purpose of the system is to make universally and promptly available, in a form which is internationally acceptable, basic bibliographic data on all publications in all countries" (Anderson, 1974, 11). The continuing functioning of the system is assured by IFLA, as the international bibliographic standards body, and a network of national bibliographic agencies.

A 2008 survey (Cordeiro, 2008) based on 80 responses from 33 countries showed that the UNIMARC format is used as the internal, national format by 23 countries, and as an exchange-only format by 10. Five UNIMARC-based formats were reported. The users are located in hot-spots: European countries such as Portugal, Italy, France, Croatia, Slovenia, Serbia, Greece, etc., Maghreb countries such as Algeria, Morocco, and Tunisia, Russia, and China and Japan. In recent years, however, some of the countries using UNIMARC have moved to the MARC 21 formats, including the Czech Republic, Slovakia and India. UNIMARC formats are translated into the languages of all the countries in which it is used.

## 2. Project on representing UNIMARC in RDF

The project on representing UNIMARC in RDF has its genesis in 2010 when the PUC discussed and formally approved the proposal made by Dunsire and Willer, in their roles as the chair of the IFLA Namespaces Task Group, and ISBD Review Chair and consultant to the PUC, respectively, who argued the case that UNIMARC should join the existing namespaces created for such as ISBD and FR (Functional Requirements) family of models to complete the suite of IFLA standards (Weitz, 2010). The next step was made by Dunsire and Willer when they presented their research based on experience with representing bibliographic standards in RDF at the UNIMARC Session during the IFLA Conference in 2011 with a paper subsequently updated and published in the IFLA Journal. (Dunsire and Willer, 2011) analyses in detail issues in the design of namespaces for UNIMARC elements and vocabularies, identifies possible solutions in the namespaces of related standards such as ISBD and the FR family of conceptual models, draws attention so the need to develop one or more DC Application Profiles, and makes recommendations to the PUC for further discussion and approval. These recommendations cover strategies and patterns for identifying namespaces for the component element sets and value vocabularies and RDF classes, concepts, and properties, as well as research topics for developing semantic maps between UNIMARC vocabularies and with related standards.

The PUC fully committed itself to the project during the IFLA conference in 2012, and made a proposal for funding to the IFLA Professional Committee. This was rejected because of budget



https://doi.org/10.23106/dcmi.952136192

constraints, so the UNIMARC Core Activity decided to finance it as a development project out of its own, limited funds. The two-year project commenced in November 2012 with a meeting between the consultants who are carrying out the work: the authors of this paper. The basic methodology proposed by Dunsire and Willer remains unchanged, although the scheduling and scope will be more fluid. The general aim is to concentrate on the UNIMARC bibliographic format in the first year, and the UNIMARC authorities format and pilot mappings in the second. The experience gained from the initiative in 2011 by Metadata Management Associates (MMA) to develop namespaces for MARC 21 has been also taken into account (Metadata Management Associates, 2011). The current focus of the project is getting the semantic content of the UNIMARC bibliographic format into an RDF representation so that further issues can be identified and the vocabularies published as soon as possible. The work covers a mix of element sets and value vocabularies, both internal to the format and external, such as ISBD.

### 2.1. UNIMARC namespaces and their management

An early decision of the project was to favour representing all UNIMARC Bibliographic (UNIMARC/B) elements in a dedicated namespace using the Open Metadata Registry (OMR) for vocabulary management, and link to existing ISBD classes and properties as appropriate, rather than re-use them. This was based on the lack of a formal protocol between the PUC and the ISBD Review Group for ensuring the semantic stability of a common set of RDF elements. The project subsequently provided evidence to justify the additional effort required by the decision; an example is discussed in the Mappings section of this paper.

The project also approved the pattern for namespaces for UNIMARC/B elements and vocabularies proposed by Dunsire and Willer. This is based on the element encoding format specified by ISO 2709 and UNIMARC usage of indicator and subfield options. Thus the finest semantic granularity is specified by subfields, encoded by a single alphanumeric character and the delimiter "$", while the coarsest level is specified by fields, encoded with a tag composed of three numeric characters. Two indicators, each a single character, are used to qualify the tag, both syntactically and semantically. The need to separate the semantics within the encoding had resulted in the MMA MARC 21 namespace creating a separate RDF element for each subfield for every combination of indicators, so the UNIMARC/B element set takes the same approach. Each subfield is represented by an RDF property with the local part of its URI corresponding to the encoding of tag + indicator 1 + indicator 2 + subfield code. Character positions in coded data fields are added when required. For example, the *UNIMARC/B tag 100 General processing data* has a single subfield $a that contains a string of codes. The *Target Audience Code* can be used up to three times in character positions 17-19; the local part of the URI of the corresponding property for target audience is therefore ***100__a17-19***. The underscore is used to indicate a blank value of an indicator, following the practice adopted for MARC 21. This is a super-property of properties corresponding to the individual character positions, for example ***100__a18***. This approach was necessary for MARC 21, but is redundant in many, if not all, coded elements of UNIMARC because the encoding has a clearer semantic structure. However, the project will accept the redundancy in case local variations of practice, for example attaching significance to the order of code positions, require it.

The UNIMARC/B value vocabularies are represented in SKOS with the local part of the concept URI corresponding to the UNIMARC/B code. The value vocabulary for *UNIMARC: Target Audience Code* (UNIMARC audience vocabulary, 2013) has the URI ***http://iflastandards.info/ns/unimarc/terms/tac***; the URI for the code value "c" for the category "primary, ages 5-10" (UNIMARC audience primary, 2013) is ***http://iflastandards.info/ns/unimarc/terms/tac#c***. RDF labels, descriptions, definitions, and scope notes are based on the text of the format's documentation, using exact transcription where possible to reduce subsequent synchronization costs and support string matching between the namespace and the source documentation.



https://doi.org/10.23106/dcmi.952136192

## 2.2. Data conversion

This encoding data and other UNIMARC content are not available as a structured database or spreadsheet, so the sources used by the project are confined to the Word version of the 2008 edition (Hopkinson, 2008) and unpublished PDF versions of the 2012 updates. These have a structured layout, so it is easy to identify the relevant data. The data are copied from the source using standard document interface functions and pasted into a spreadsheet. Using a spreadsheet allows the data to be parsed using formulas into standard design patterns for URIs and annotations. The spreadsheet also allows blocks of rows of data, one row for each subfield, to be copied and amended to change the indicator values. When data for one set of subfields has been pasted, the block of rows is copied for each differing value of the first indicator and the new value painted into the appropriate column. When each value for the first indicator has been completed, the entire set of blocks of rows is copied for each differing value of the second indicator. The result covers every combination of indicator values at the granularity of the subfield.

Data for the field name, indicator value captions, and subfield name are also pasted into the spreadsheet. These are combined using a spreadsheet formula to create a unique label for the RDF property. This design pattern for the formula will be similar to that used by MMA for MARC 21: subfield name + field name + first indicator caption + second indicator caption, delimited by suitable conjunctions and punctuation. A MARC 21 example is the label "Title in Title Statement (Added entry) (No nonfiling characters)" of the URI ***http://marc21rdf.info/elements/2XX/M24510a***: the subfield/field delimiter is the conjunction "in" and the indicator captions are delimited by round brackets. Similarly, data for code captions and descriptions are pasted into the spreadsheet as the basis of value vocabulary definitions and scope notes. The project has identified a number of issues requiring human interpretation of the UNIMARC/B documentation to provide complete and consistent definitions and notes. These include the separation of scope notes from the definition, removal or resolution of cross-references to other elements, removal of usage and formatting information, and correction of typographical and other, sometimes more serious, errors. Similar issues have with other IFLA standards have been identified by Dunsire (2011).

## 2.3. Intended utility

The base domain of the UNIMARC/B namespace is based on a standard template used by other IFLA namespaces: ***http://iflastandards.info/ns/unimarc/***. As indicated above, this is further divided to give separate namespace domains for subdivisions of the format, such as ***http://iflastandards.info/ns/unimarc/terms*** for value vocabularies. Using a regular design pattern for the local part of the element and concept URIs based on the UNIMARC encoding format has several benefits. It allows data triples to be derived automatically from an encoded record. The subject of the data triple is the resource described, the predicate is the RDF property corresponding to the subfield, and the object is the value of the subfield content. All of the data is normally present in the record: the record number as the local part of the subject URI, and the format encoding as the local part of the predicate URI. This reduces the resources required to process the data before publication in RDF. The availability of properties for the finest level of semantics and syntax, the subfield and character position, allows UNIMARC records to be published as linked data without loss of information. An additional benefit of this particular design pattern is that the property URI is essentially opaque because it does not depend on human language. It meets the multilingual requirements of IFLA by providing equal treatment, and of UNIMARC users through familiarity. In the authors' experience, cataloguers and systems librarians refer to UNIMARC elements by the encoding rather than caption, "Two hundred first indicator one dollar a" rather than "Title proper of Title and statement of responsibility (Title is significant)" or any coherent contraction such as "Significant title proper", so the semantic interpretation of ***unimarcb:U2001_a*** should be easy and accurate for such expert UNIMARC users. The open availability of UNIMARC ontologies combined from the element sets and internal mappings is intended to assist linked data applications without access to such expertise to



https://doi.org/10.23106/dcmi.952136192

use UNIMARC data correctly and effectively. The development of a framework of internal mappings to a smaller set of broader level properties at the tag or code position range level to conceal unnecessary complexity and syntactical redundancy from external applications is an essential part of the methodology used by the project. It is part of a wider approach to improving interoperability between library standards and the wider bibliographic and cultural heritage linked data environment, for example the ongoing work with ISBD, ONIX, and RDA (ISBD Review Group, 2012-1, 2012-2).

A survey on the use of namespaces by national libraries and union catalogues conducted at the end of 2012 and the beginning 2013 for a forthcoming book (Willer and Dunsire, 2013) shows that those libraries that use UNIMARC or MARC 21 formats do not publish their bibliographic linked open data using element sets from native MARC namespaces, such as the openly available MMA version of MARC 21 or any private representations. Instead, all of them use elements from other schema such as ISBD and the FR family of models, FOAF, schema.org, etc., or develop local namespaces representing coarse-grained tag-level properties. In some of the cases, several schemas are used in a mix and match methodology, for example the RDF output by the British Library of MARC 21 records of the British National Bibliography (British Library, 2012). This indicates that none of the available trusted schema namespaces cover all of the specific needs for the expression of data granularity developed within the MARC formats, and still required by the functionality of descriptive and retrieval systems of the library community. The project aims to improve this situation for UNIMARC users by providing clearer associations between the bibliographic encoding format and the rules governing usage and content adopted to meet national policies, as well as providing feedback to the PUC to improve the semantic quality of the format.

## 3. Mappings

### 3.1. UNIMARC bibliographic format and ISBD

In defining the purpose of the ISBD as providing the stipulations for compatible descriptive cataloguing worldwide, the standard aims, among other things, to assist in the conversion of bibliographic records to electronic form (ISBD, 2011). The UNIMARC/B format was designed to fulfill that aim, so that each ISBD element has a corresponding UNIMARC content designator or tag defined. Thus, *UNIMARC/B 2-- Descriptive Information Block* encompasses the first six ISBD areas and *UNIMARC/B 3-- Notes Block* contains *ISBD 7 Note area*, while *ISBD 8 Resource identifier and terms of availability area* is covered by fields in *UNIMARC/B 0-- Identification Block* and by the field *UNIMARC/B 856 Electronic Location and Access*.

Therefore each UNIMARC field and subfield is defined in relation to its correspondence to an ISBD area and element: Figure 1 shows the correspondence between field *UNIMARC/B 200 Title and Statement of Responsibility* and *ISBD 1 Title and statement of responsibility area*; the strikeouts identify amendments to the published third edition of UNIMARC/B. ISBD uses prescribed punctuation as a syntactical device to delimit elements in a language-independent, though human-readable, record display.



https://doi.org/10.23106/dcmi.952136192

| UNIMARC subfield | Element name | ISBD ~~(G)~~ section | Preceding punctuation |
|---|---|---|---|
| | | | |
| $a | Title proper | 1.1 | New area |
| $a (repeated) | Title proper by the same author | 1.1.4.4 | ; |
| $b | General material designation | 1.2 | [ ] |
| $c | Title proper by another author | 1.5.5.11.2 | . |
| $d | Parallel title proper | 1.3 | = |
| $e | Other title information | 1.4 | : |
| $f | First statement of responsibility | 1.5 | / |
| $g | Subsequent statement of responsibility | 1.5 | ; |
| $h | Number of a part | 1.1.5.3 ~~ISBD(S)~~ | . |
| $i | Name of a part | 1.1.5.3 ~~ISBD(S)~~ | , if after $h, else . |

The above table illustrates the preceding punctuation which occurs in most situations but not necessarily when parallel data is present. Further information is found in the ISBD.

Figure 1. UNIMARC/B field 200 and ISBD area 1 correspondence (Hopkinson, 2008)

This alignment, however, precedes the consolidated edition of ISBD issued in 2011, and thus has some impact on defining the UNIMARC namespaces and mappings. Namely, the *General material designation* (GMD) element was removed from ISBD area 1 and replaced by the new *ISBD 0 Content form and media type area* in the consolidated edition. That also renumbered the elements so the ISBD element numbered 1.2 changed from being the GMD and now refers to the *Parallel title* element, and so on. The UNIMARC 2012 Update does not reflect this renumbering, and retains subfield *UNIMARC/B 200 $b* for the GMD. The point is that the format cannot delete a defined element in the expectation that it is used in legacy records; it can only make it obsolete. The PUC has chosen not to take action on this element. The issue also shows that the ISBD Review Group did not fully appreciate the importance of deprecating, rather than deleting, the element, although the consolidated edition refocuses on the element name as the primary label and reduces the impact of renumbering.

Other differences were identified in this example. The repeated subfield *UNIMARC/B 200 $a Title proper by the same author* is aligned with ISBD stipulation 1.1.4.4 for "Resources with two or more works without a collective title", but according to the consolidated ISBD it should be 1.1.5.2 for "Resources without a collective title". There is a significant semantic difference between these ISBD stipulations. In subfield *UNIMARC/B 200 $c*, the reference to ISBD stipulation 1.5.5.11.2 seems to be a definite error, as the correspondence should be to stipulation 1.4.5.11.2. Finally, subfields *UNIMARC/B 200 $h* and *UNIMARC/B 200 $i* do not correspond to any of the stipulations given under 1.1.5.3. The project will recommend to PUC that the UNIMARC to ISBD correspondence tables should be further updated to match the consolidated ISBD edition, and that the use of ISBD numbering should be replaced with element names or URIs. The project will recommend to the ISBD Review Group that elements should be deprecated rather than removed, to take into account the consequences for existing data and related formats.

The next phase of the analysis of the correspondences between UNIMARC and ISBD took into account the ISBD namespaces. The elements named, with corresponding stipulation numbers, in the outline of the ISBD have been represented as RDF properties. However, the namespace contains additional properties which are not explicitly listed in ISBD, and these will be required for RDF mappings between the UNIMARC/B and ISBD element sets. Table 1 shows the correspondence between UNIMARC elements and the ISBD namespaces.



https://doi.org/10.23106/dcmi.952136192

Table 1: Alignment of UNIMARC/B field 200 and ISBD RDF properties

| **UNIMARC/B** | | | **ISBD** | |
|---|---|---|---|---|
| **Field** | **Subfield** | **Name** | **Property** | **Label** |
| 2-- | | Description information block | | |
| 200 | | Title and statement of responsibility | | has title and statement of responsibility area |
| | | | P1170 | has title statement |
| | | | P1012 | has title |
| 200 | $a | Title proper | P1004 | has title proper |
| | (rep) $a | | P1117 | has title of individual work by same author |
| | $b | General material designation | | |
| | $c | Title proper by another author | P1118 | has title of individual work by different author |
| | $d | Parallel title proper | P1005 | has parallel title |
| | | | P1182 | has common title of parallel title |
| | | | P1183 | has dependent title of parallel title |
| | | | P1184 | has dependent title designation of parallel title |
| | $e | Other title information | P1006 | has other title information |
| | $f | First statement of responsibility | P1007 | has statement of responsibility relating to title |
| | $g | Subsequent statement of responsibility | P1007 | has statement of responsibility relating to title |
| | | | P1141 | has parallel statement of responsibility relating to title |
| | | | P1137 | has common title of title proper |
| | $h | Number of part | P1139 | has dependent title designation of title proper |
| | $i | Name of part | P1138 | has dependent title of title proper |
| | $v | Volume designation | | |
| | $z | Language of parallel title proper | | |

These alignments cover all simple types of semantic relationship: equality (***unimarcb:200__a*** and ***isbd:P1004***); different granularity (***unimarcb:200__f*** and ***isbd:P1007***) and aggregation (***unimarcb:200__d*** and ***isbd:P1182***). There are also more complex relationships, not fully expressed in Table 1, which include conditional relationships between UNIMARC subfields.

### 3.2. Outside the box: the sub-property ladder

The use of OWL or an application profile will be required to represent the complex alignments, while RDFS and OWL will be used for the simple ones, such as ***owl:equivalentProperty*** and ***rdfs:subPropertyOf***. In particular, the project is investigating the use of what Dunsire (2012) calls the "sub-property ladder" and Powell, Nilsson and Naeve (2003) refer to as "intelligent dumb-down". An example of such a chain of sub-property relationships is ***unimarcb:2001_a*** <> ***isbd:P1004*** < ***uncrda:titleProper*** < ***uncrda:title*** < ***dct:title*** < ***dc:title*** (where "uncrda" is the set of RDA properties unconstrained by FRBR). This is taken from a map of a partial ontology for the bibliographic *title* attribute from Dunsire, Hillmann and Phipps (2012). The strategy is to develop internal maps for aggregating fine-grained properties and relating the coarse-grained properties to dumber, coarser-grained properties in related element



https://doi.org/10.23106/dcmi.952136192

sets while avoiding duplicating existing pathways in related semantic maps, using the approach discussed in Dunsire, Hillmann, Phipps, and Coyle (2011).

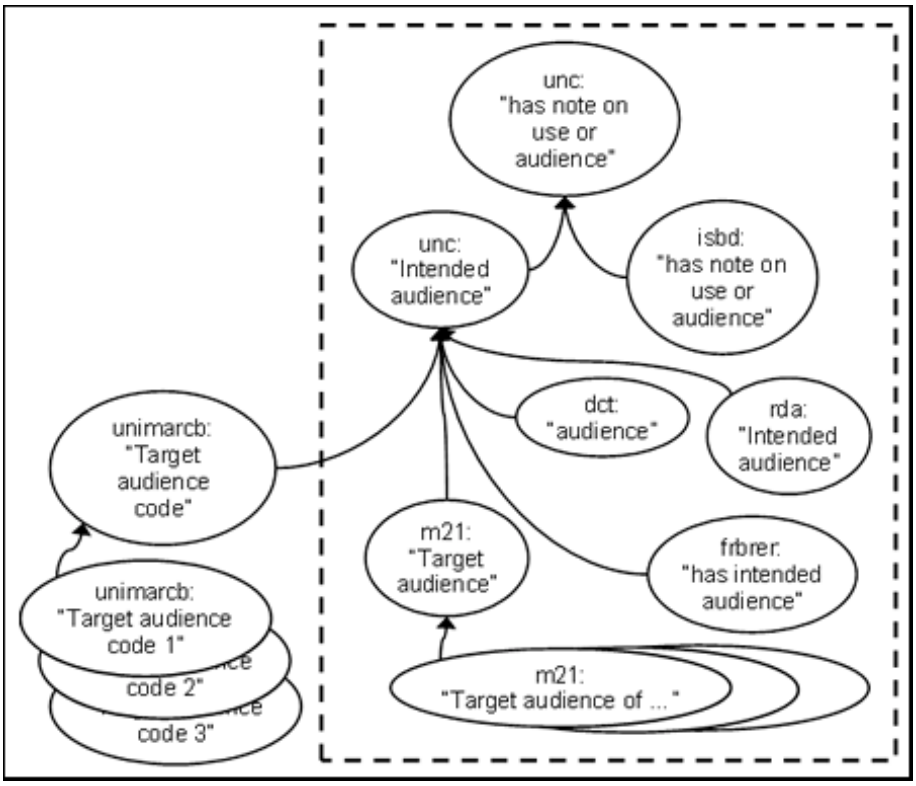


Figure 2: RDF graph of bibliographic elements for "audience"; all properties/predicates are *rdfs:subPropertyOf*

Figure 2 shows a possible mapping from the Target audience code element ***unimarcb:U100__a17-19*** to an ontology proposed by Willer and Dunsire (2013) for similar elements for the intended audience of a bibliographic resource taken from MARC 21 (m21), FRBR (frbrer), RDA: resource description and access (rda), Dublin Core terms (dct), and ISBD (isbd). The elements are identified with a QName and English label for clarity; the only RDF property used for mapping is ***rdfs:subPropertyOf***. The elements with the QName "unc" are unconstrained versions of the RDA and ISBD sub-properties, that is, with no domain or range. The external ontology is marked by the dashed lines. The figure also includes an internal UNIMARC/B map connecting the properties for the individual character positions. As already noted, the map is redundant in the standard UNIMARC/B semantics, because no significance is assigned to the order of codes or which codes are used. However, it serves to conceal this from the external ontology by restricting the connection from the internal map to the external ontology to a single link, and allows the internal map to be refined in application profiles for local implementations of the format.

Figure 3 is a serialization of the audience ontology of Figure 2 in terse triple format. The ontology entails "dumber" data triples from the fine grained output of a UNIMARC or MARC 21 record directly mapped to the "level 0" properties. A record with three audience codes produces three direct data triples, each using a different positional property, and three entailed triples using the same non-positional property. There is no effective dumb-down because only syntactical information, the code position, is lost. The entailed triples are simpler to use by external applications because only one UNIMARC property is involved, and retain all of the semantic data of the source record.



https://doi.org/10.23106/dcmi.952136192

This example demonstrates some of the benefits of a bottom-up approach to mapping relatively local schema, "universal" though UNIMARC is to a significant number of libraries world-wide. As Dunsire, Harper, Hillmann and Phipps (2012) observe, the local schema can preserve, and make available, the semantics of its context and application, while external, global applications can make the semantic granularity of the local data as coarse as they require.

```
# Audience sub-property ladder.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix dct: <http://purl.org/dc/terms/>.
@prefix frbrer: <http://iflastandards.info/ns/fr/frbr/frbrer/>.
@prefix isbd: <http://iflastandards.info/ns/isbd/elements/>.
@prefix m2100: <http://marc21rdf.info/elements/00/>.
@prefix m2100X: <http://marc21rdf.info/elements/00X/>.
@prefix rdaGroup1: <http://rdvocab.info/Elements/>.
@prefix unc: <>.
@prefix unimarcb: <>.
# UNIMARC level 0 to higher aggregate UNIMARC property (unpublished).
unimarcb:U100__a17 rdfs:subPropertyOf unimarcb:U100__a17-19 .
unimarcb:U100__a18 rdfs:subPropertyOf unimarcb:U100__a17-19 .
unimarcb:U100__a19 rdfs:subPropertyOf unimarcb:U100__a17-19 .
# MARC 21 level 0 to higher aggregate MARC 21 property (unpublished).
m2100X:M006c05 rdfs:subPropertyOf m2100:M00Aud .
m2100X:M006g05 rdfs:subPropertyOf m2100:M00Aud .
m2100X:M006i05 rdfs:subPropertyOf m2100:M00Aud .
m2100X:M006j05 rdfs:subPropertyOf m2100:M00Aud .
m2100X:M006k05 rdfs:subPropertyOf m2100:M00Aud .
m2100X:M006o05 rdfs:subPropertyOf m2100:M00Aud .
m2100X:M006r05 rdfs:subPropertyOf m2100:M00Aud .
m2100X:M006t05 rdfs:subPropertyOf m2100:M00Aud .
m2100X:M008BK22 rdfs:subPropertyOf m2100:M00Aud .
m2100X:M008CF22 rdfs:subPropertyOf m2100:M00Aud .
m2100X:M008MU22 rdfs:subPropertyOf m2100:M00Aud .
m2100X:M008VM22 rdfs:subPropertyOf m2100:M00Aud .
# Aggregate UNIMARC property to unconstrained/RDA property
unimarcb:U100__a17-19 rdfs:subPropertyOf unc:intendedAudience .
# Aggregate MARC 21 property to unconstrained/RDA property
m2100:M00Aud rdfs:subPropertyOf unc:intendedAudience .
# RDA property to unconstrained/RDA property
rdaGroup1:intendedAudience rdfs:subPropertyOf unc:intendedAudience .
# FRBRer to unconstrained/RDA property
frbrer:P3006 rdfs:subPropertyOf unc:intendedAudience .
# DC terms to unconstrained/RDA property
dct:audience rdfs:subPropertyOf unc:intendedAudience .
# ISBD property to unconstrained/ISBD property (unpublished)
isbd:P1091 rdfs:subPropertyOf unc:P1091 .
# Unconstrained/RDA property to unconstrained/ISBD property (unpublished)
unc:intendedAudience rdfs:subPropertyOf unc:P1091 .
```

Figure 3: TTL serialization of graph of bibliographic elements for "audience".





https://doi.org/10.23106/dcmi.952136192

## Conclusion

The representation of the UNIMARC data formats in RDF is an opportunity to preserve the intellectual value of UNIMARC in the form of semantic data structures. That value extends well beyond the large quantities of high-quality and increasingly linked metadata created to meet library requirements in the global digital environment, or the rigour and dedication brought to the development of the standard. The formats are a core component of UBC and its relationship to the current conceptual bases of the international professional cataloguing community expressed in International Cataloguing Principles (ICP) (Tillett and Cristán, 2009). The purpose of UBC is just as relevant in the digital environment as it is in the print. It is not much different from schema.org's intent to "make it easier for users to find relevant information on the web" (Schema.org, 2011).

It is important that these data formats, and the metadata that uses them, are properly represented in the current development environment of the Semantic Web. Commercial interests are interacting with the public availability strategies of the national, educational, and public library communities. Libraries know from experience that there is much to be gained from such interaction, but also much that might be lost from the quality of the user experience. All metadata standards face challenges in adapting to linked data technologies. The shift in focus from the bibliographic record to the triple, in the much wider field of view of the Semantic Web, presents particular problems to data encoding formats, as this paper has described. The intellectual and ethical inheritance of the "universal" library community should be available to meet such challenges and inform solutions. Yet the current response has been disjoint. The Library of Congress' Bibliographic Framework Transition Initiative (BIBFRAME) aims "to determine a transition path for the MARC 21 exchange format" (Library of Congress, 2011). The first documentation issued by the initiative (Library of Congress, 2012) states "the BIBFRAME model is the library community's formal entry point for becoming part of a much larger web of data", yet it does not mention UNIMARC or ICP, and there has been no direct consultation with IFLA about collaboration.

This is unlikely to lead to technical problems, as the UNIMARC elements can be readily dumbed-down to whatever is required by BIBFRAME or, indeed, any other entry point into the linked data environment. The door swings both ways, allowing UNIMARC data values to be indexed by broad-based services which can then drill-down or connect to more specific services based on the UNIMARC standard. The initial BIBFRAME model is coarse-grained, as is schema.org. The availability of much finer semantic granularity from UNIMARC should help to ensure that end-users continue to benefit from high-quality as well as high-quantity information access services in the future Semantic Web.

## References

Anderson, Dorothy. (1974). Universal Bibliographic Control: A long term policy – A plan for action. Pullach/München: Verlag Dokumentation.

British Library. (2012). British Library Data Model – Book. V.1.4. Available at: http://www.bl.uk/bibliographic/pdfs/bldatamodelbook.pdf

Cordeiro, Maria Inês. (2008). The UNIMARC Landscape: Updated Overview. Available at: http://archive.ifla.org/VI/8/unimarc-survey-2008_results.pdf

Dunsire, Gordon. (2011). Interoperability and semantics in RDF representations of FRBR, FRAD and FRSAD. Concepts in context: proceedings of the Cologne Conference on Interoperability and Semantics in Knowledge Organization July 19th-20th, 2010. Würzburg: ERGON. Preprint available at: http://www.gordondunsire.com/pubs/docs/InteropSemanticsRDFFRBRPre.pdf

Dunsire, Gordon. (2012). Using the sub-property ladder. Metadata Matters blog. Available at: http://managemetadata.com/blog/2012/05/12/using-the-sub-property-ladder/

Dunsire, Gordon, and Mirna Willer. (2011) UNIMARC and Linked Data. IFLA Journal, 37, 4 (December 2011) (pp314-326). Available at: http://www.ifla.org/files/hq/publications/ifla-journal/ifla-journal-37-4_2011.pdf





https://doi.org/10.23106/dcmi.952136192

Dunsire, Gordon, Diane Hillmann, and Jon Phipps. (2012). Reconsidering Universal Bibliographic Control in Light of the Semantic Web. Journal of Library Metadata, Vol. 12, Issue 2-3 (pp. 164-176).

Dunsire, Gordon, Corey Harper, Diane Hillmann, and Jon Phipps. (2012). Linked data vocabulary management: infrastructure support, data integration, and interoperability. Information Standards Quarterly, v.24, no. 2/3 (Spring/Summer 2012). Available at: http://www.niso.org/apps/group_public/download.php/9411/FE_Dunsire-etal_VocabMgmt_isqv24no2-3.pdf

Dunsire, Gordon, Diane Hillmann, Jon Phipps, and Karen Coyle. (2011). A Reconsideration of Mapping in a Semantic World. Proc. Int'l Conf. on Dublin Core and Metadata Applications 2011. Available at: http://dcevents.dublincore.org/index.php/IntConf/dc-2011/paper/view/52/6

Hopkinson, Alan (ed.) (2008). UNIMARC Manual: Bibliographic Format. 3rd. ed. München: K. G. Saur.

IFLA. (2013-1). About the UNIMARC Core Activity. Available at: http://www.ifla.org/about-unimarc

IFLA. (2013-2). UNIMARC formats and related documentation. Available at: http://www.ifla.org/publications/unimarc-formats-and-related-documentation

ISBD. (2011). ISBD : International standard bibliographic description. Consolidated ed. Berlin ; München : De Gruyter Saur.

ISBD Review Group. (2012-1). Alignment of the ISBD element set with RDA element set – RDA, Appendix D.1. Available at: http://www.rda-jsc.org/docs/6JSC-ISBD-Discussion-1.pdf

ISBD Review Group. (2012-2). Mapping of ISBD Area 0 vocabularies to RDA/ONIX Framework vocabularies as part of the ISBD/RDA Available at: http://www.rda-jsc.org/docs/6JSC-ISBD-Discussion-2.pdf

Library of Congress. (2011). Bibliographic Framework Transition Initiative. Available at: http://www.loc.gov/marc/transition/

Library of Congress. (2012). Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services. Available at: http://www.loc.gov/marc/transition/pdf/marcld-report-11-21-2012.pdf

Metadata Management Associates. (2011). The MARC21 Vocabularies from Metadata Management Associates. Available at: http://marc21rdf.info/

Open Metadata Registry. Available at: http://metadataregistry.org

Powell, Andy, Mikael Nilsson, and Ambjrn Naeve. (2003). DCMI Abstract Model. Available at: http://www.ukoln.ac.uk/metadata/dcmi/abstract-model/2003-12-18/

Permanent UNIMARC Committee. Available at: http://www.ifla.org/unimarc/puc

Schema.org. (2011). What is schema.org? Available at: http://schema.org/

Tillett, Barbara, and Ana Lupe Cristán. (2009). IFLA Cataloguing Principles: Statement of International Cataloguing Principles (ICP) and its Glossary. München: K.G. Saur. Available at: http://www.ifla.org/files/assets/cataloguing/icp/icp_2009-en.pdf

UNIMARC audience vocabulary. (2013). Vocabulary: Show detail for UNIMARC: Target Audience Code. Available at: http://metadataregistry.org/vocabulary/show/id/322.html

UNIMARC audience primary. (2013). Vocabulary: UNIMARC: Target Audience Code, Concepts: primary, ages 5-10. Available at: http://metadataregistry.org/concept/show/id/6342.html

Weitz, Jay. (2010). Informal Meeting of the Permanent UNIMARC Committee, 13-14 August, 2010, IFLA Congress, Gothenburg, Sweden. International cataloguing and bibliographic control, Vol. 39, no. 4 (pp. 82-83).

Willer, Mirna, and Gordon Dunsire. (2013). Bibliographic Information Organization in the Semantic Web. In press.





https://doi.org/10.23106/dcmi.952136192