Convergence of UNIMARC and RDA: an update

Presented and discussed at EURIG Satellite Meeting, Sept 15th 2020

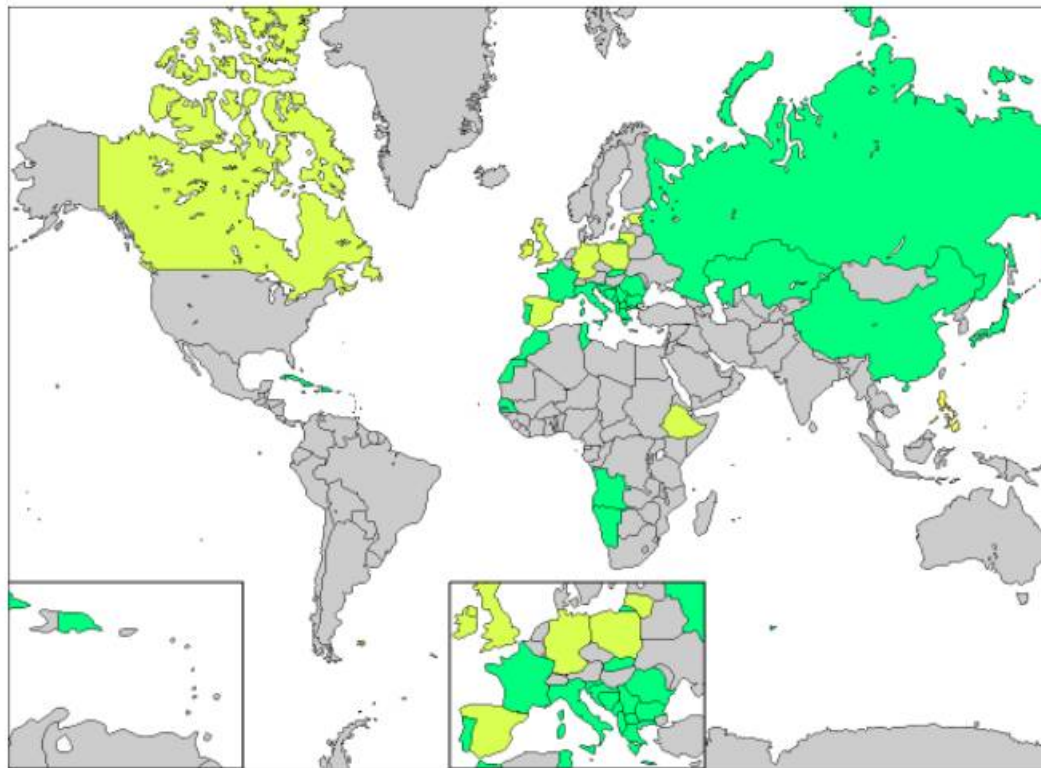Prepared by Héloïse Lecomte (Abes) and Françoise Leresche (BnF)

Program

- ▶ UNIMARC Users & LRM/RDA

- ▶ PUC Action plan 2019-2021 & RSC Liaison

- ▶ UNIMARC updates 2012-….

- ▶ Last updates (2020)

- ▶ Expressing LRM/RDA in UNIMARC: assets and challenges

❷ The importance of the users' community

CURRENT LANDSCAPE

UNIMARC as internal format
UNIMARC as exchange format only

Countries
Albania
Angola
Bosnia and Herzegovina (COMARC)
Bulgaria (COMARC)
China (CNMARC)
Cuba
Cyprus
France
Greece
Iran (IranMARC)
Italy
Japan (JAPANMARC)
Lithuania
Macedonia (COMARC)
Montenegro (COMARC)
Morroco
Namibia
Portugal
Romenia
Russia (RUSMARC)
Senegal
Serbia (COMARC)
Slovakia
Slovenia
Tunisia

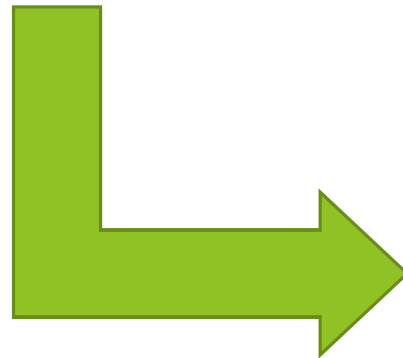| UNIMARC Publications |
|---|
| |
| English |
| Portuguese (1989, 2008) |
| French (2007) |
| Spanish (Cuba, 2003, Spain, 1996) |
| Greek (1993-1996) |
| Persian (Iran,2002) |
| Italian (2000) |
| Russian |
| Slovene |
| Lithuanian |
| Croatian |
| Serbian |
| Slovak (1994-1996) |
| Chinese (2004) |
| Japanese (Authorities, 1991) |
| Romanian |

![IFLA logo] International Federation of
Library Associations and Institutions

Permanent UNIMARC Committee

Members

* **Chair:** Gordana Mazić, Slovenia
* **Vice Chair:** Jay Weitz, USA
* Rosa Galvão, Portugal
* Olga Zhlobinskaya, Russia
* Héloïse Lecomte, France
* Genovaite Rutkauskiene, Lithuania
* Saeedeh Akbari-Daryan, Iran
* **Honorary member and special consultant:** Mirna Willer, Croatia
* **Corresponding members:** Flavia Bruni, Italy; Stanislaw Golubtsov, Russia; Branka Badovinac, Slovenia

Implementing **LRM-structured data** :
- Croatia
- France
- Slovenia
- Lithuania
- ?

Working on **RDA-based cataloguing code** :
- France (RDA-FR)
- Portugal (RDA)
- ?

+ Italy (REICAT, based on FRBR, developed simultaneously to RDA)

Action plan

Name of Professional Unit: Permanent UNIMARC Committee

Focus Area 2
UNIMARC Harmonization with IFLA LRM and new cataloguing codes

IFLA Strategic Direction
2.3 Develop standards, guidelines, and other materials that foster the best professional practice

Key Initiatives	
1	Analyze IFLA LRM, cataloguing standards that accommodate LRM (like RDA, REICAT, RDA-FR) and UNIMARC formats
2	Mapping UNIMARC to LRM
3	Coordinate UNIMARC National Committees using RDA and LRM like description rules to prepare new proposals for UNIMARC

Funding Needed	**Projects/Activities**	**Tasks & Responsibility**	**Timeframe**
☐	Develop a mapping tool for the aligning model – cataloguing code – format elements	Te PUC WG, CfU (Héloïse Lecomte)	Jan. 2020 to March. 2020
☒	Mapping table LRM – UNIMARC	PUC WG, CfU (Héloïse Lecomte)	March. 2020 March. 2021
☐	Prepare UNIMARC proposals for missing fields	PUC WG, CfU (Héloïse Lecomte)	December 2021

How will you communicate your activities and results?
New and updated fields will be integrated into UNIMARC text documentation; Existing UNIMARC Namespaces will be updated or a new one will be registered.

How will you measure the impact of your activities?
Feedback information from Nacional UNIMARC committees and their representatives in PUC and through the regular questionnaire to UNIMARC users in national libraries.

Identify other Unit(s) which may be interested in this Focus Area, or with which you could collaborate on projects/activities.
BCM Review Group; ISBD Review Group; Cataloguing section.

To: RDA Steering Committee

From: Gordon Dunsire, Chair, RSC

Subject: Protocol between the RSC and the Permanent UNIMARC Committee

Background

"RDA is a package of data elements, guidelines, and instructions for creating library and cultural heritage resource metadata that are well-formed according to international models for user-focussed linked data applications" – http://rda-rsc.org

The RDA Steering Committee (RSC) maintains and develops vocabulary encoding schemes (VESs) for several RDA recorded elements. The vocabularies are amended following established RSC policies and procedures.

The RSC provides linked data versions of the RDA vocabularies in the Open Metadata Registry (OMR) and RDA Registry.

The RSC maintains, jointly with the IFLA ISBD Review Group, a mapping between ISBD and RDA elements.

UNIMARC (Universal MARC format) is a format for the representation and communication of bibliographic, authority, classification and holdings information in machine readable form, originally created by IFLA. The IFLA Permanent UNIMARC Committee (PUC) within the IFLA UNIMARC Strategic Programme is responsible for maintaining the UNIMARC format and for the editorial development of the UNIMARC documentation.

The PUC updates the UNIMARC format following established policies and procedures, and maintains and develops a number of VESs as UNIMARC code lists.

Purpose

The purpose of this protocol is to support communication between the editorial processes of the RSC and the PUC in order to support functional and semantic interoperability between data compatible with RDA and data compatible with the UNIMARC format.

The result of the editorial processes is RDA/UNIMARC mappings added to the RDA Toolkit.

Methodology

The protocol is intended to be light-weight to avoid significant changes to the current practices, priorities, and workflows of the RSC and PUC.

The usual channel of communication between the two groups are the RSC consulting liaison to the PUC and the PUC consulting liaison to the RSC, who are responsible for routine communication and reviewing each group's activity for potential issues.

Duties of the RSC

The RSC will:

- Inform PUC at the earliest opportunity of impending changes to RDA that may impact on UNIMARC.
- Liaise with PUC on the development and maintenance of linked data mappings between RDA and UNIMARC value vocabularies.
- Liaise with PUC on the development and maintenance of linked data mappings between RDA and UNIMARC element sets.
- Nominate a consulting liaison to the PUC.

Duties of the PUC

The PUC will:

- Inform the RSC at the earliest opportunity of impending changes to UNIMARC vocabularies that may impact on RDA.
- Liaise with the RSC on the development and maintenance of linked data mappings between RDA and UNIMARC value vocabularies.
- Liaise with the RSC on the development and maintenance of linked data mappings between RDA and UNIMARC element sets.
- Nominate a consulting liaison with the RSC.

RSC-PUC Agreement,
31 December 2018

RDA-FR/UNIMARC mapping

▶ CfU (French UNIMARC Committee) working on it, entity by entity

  ▶ Works and Expressions: done > see proposed updates 2020

  ▶ Manifestations, Items, etc.: to do

Example

Section RDA-FR	Numéro d'instruction RDA-FR	Nom d'élément RDA-FR (prédicat)	Entité sujet	Entité objet/attribut	Etiquette de zone UNIMARC B-FR	Code de sous-zone UNIMARC B-FR	Etiquette de zone UNIMARC A-FR	Code de sous-zone UNIMARC A-FR	Demande de création ou d'évolution à faire au PUC
2	6.20	Public destinataire de l'expression	Expression	Attribut	B/100	pos.17-19	à créer		
2	6.20	Public destinataire de l'expression	Expression	Attribut			A/333	a	
2	6.21	Information sur l'expression	Expression	Attribut			A/300		
2	6.22	Circonstances de la captation	Expression	Attribut			A/640	ind. 1	à créer nouvelle valeur d'ind.
2	6.22.2	Lieu de captation	Expression	Attribut			A/640	a, b, c, d, e	
2	6.22.3	Date de captation	Expression	Attribut			A/640	f, g, i	
2	6.22.4	Occasion de la captation	Expression	Attribut			A/640	h	
2	6.22.5	Type de captation	Expression	Attribut	B/620		à créer		zone ? Sous-zone ?
2	6.23	Forme de la notation	Expression	Attribut			à créer		Proposition de création d'un nouvelle zone A/ pour les formes de notation, avec des sous-zones spécifiques pour chaque type de document
2	6.23.2	Écriture du contenu linguistique	Expression	Attribut	B/100		à créer		
2	6.23.3	Forme de la notation musicale	Expression	Attribut	B/125	a pos.0	à créer		
2	6.23.4	Forme de la notation tactile	Expression	Attribut	B/106	a pos.0 val f	à créer		
2	6.23.5	Forme de la notation du mouvement	Expression	Attribut			à créer		
2	6.24	Caractéristique de couleur	Expression	Attribut	B/115	$a/4	à créer		

OK

UNIMARC/A update necessary

UNIMARC & FRBR/LRM/RDA : 2012-2019

- Development of UNIMARC Authorities Format:
  - New record types: works, expressions
  - New authorised access points for works and expressions (231, 232, 241, 242) and other access points (Variant Access Points, Parallel Access Points, Subject Analysis)
  - New Related Access Points (5-- block) in order to link works and expressions to other works and expressions, to agents, etc.
- New Related Titles in the UNIMARC Bibliographic Format (5-- block) to allow linking of manifestations (or classic bibliographic records) to works and expressions
- RDA Analysis for Production, Publication, Distribution, Manufacture Statements > field B214
- RDA Content Type and RDA Media Type > fields B181/182
- Lots of UNM/B fields deprecated in the context of LRM/RDA cataloguing

2020: Works and expressions attributes and relationships (UNM/A) – main updates underway

- ▶ Sources for the update propositions:
  - ▶ 1/ LRM attributes and relationships
  - ▶ 2/ RDA attributes and relationships
  - ▶ 3/ RDA-FR
- ▶ Works (UNM/A):
  - ▶ New Coded Data: form and category of work (140)
  - ▶ New Note Fields: award note (334), abstract (378), note on work (370)
  - ▶ Link to Form or Genre Entity (608)
- ▶ Expressions (UNM/A):
  - ▶ New Coded Data: language (101), form of notation (105), intended audience (125), circumstances of capture (127), category of expression (145), medium of performance (146), colour and sound content (147)
  - ▶ New Note Fields: note on expression (371), abstract (378)
  - ▶ Link to Form or Genre Entity (608)

Works: indicative list of fields (UNM/A)

UNDER DISCUSSION

- **001 Record identifier**
- **003 Persistent Record identifier**
- 005 Version identifier
- 033 Other System Persistent Record Identifier
- 035 Other system control numbers
- **050/051/052 ISTC/ISWC/ISAN**
- **100 General Processing Data**
- 101 Language of the Entity [language of the representative expression]
- 102 Nationality of Entity
- **104 Coded Data Field: Main date of Entity**
- 122 Coded Data Field: Time Period of work Content
- 125 Coded Data Field: Intended audience of Work or Expression
- 128 Coded Data Field: Form of Musical Work and key or mode
- **140 Coded Data Field: Category and form of work**
- **152 Rules**
- **154 Coded Data Field : Title**
- **231/41 Authorised Access Point for Works**
- 333 Users/Intended Audience Note

- 334 Awards Note (Work or Expression)
- 370 Note on Work
- 378 Abstract of work or expression [NB: Expression attribute in RDA]
- 431/41 Variant Access Point for Works
- 501/511/521 Related Access Point – Personal Name, Corporate Body, Family Name with Responsibility for the Work
- 515 Related Access Point – Place
- 531/541 Related Access Point – Work
- 532/542 Related Access Point – Expression
- 6XX Subject Access points (including 631, 41, 32, 42)
- 608 Form or Genre of Entity
- 640 Place(s) and Date(s) associated with the entity
- 731/41 Authorized Access Point in Other Language and/or Script for Works
- 801 Originating Source
- 810 Source Data Found
- 815 Source Data Not Found
- 820 Usage or Scope Information
- 830 General Cataloguer's Note

10

Legend:
Mandatory elements
Proposals 2020 (new or change)

Expressions: indicative list of fields (UNM/A)

UNDER DISCUSSION

- **001 Record identifier**
- **003 Persistent Record identifier**
- 005 Version identifier
- 033 Other System Persistent Record Identifier
- 035 Other system control numbers
- **050/051/052 ISTC/ISWC/ISAN ?**
- **061 ISRC**
- **100 General Processing Data**
- 101 Language of the Entity
- 102 Nationality of Entity
- **104 Coded Data Field: Main date of Entity**
- 105 Coded Data Field: Form of notation of expression
- 125 Coded Data Field: Intended audience of Work or Expression
- 127 Coded Data Field: Duration and Type of Sound Recordings, Visual Projections, Musical Compositions and Performances
- 128 Coded data field: Form of musical work and key or mode
- **145 Coded Data Field: Content Type**
- 146 Coded Data Field: Medium of Performance
- 147 Coded Data Field: Colour and Sound Content
- **152 Rules**

- **154 Coded Data Field : Title**
- **232/42 Authorised Access Point for Expressions**
- 300 Information Note [RDA-FR : Information sur l'expression]
- 333 Users/Intended Audience Note
- 334 Awards Note (Work or Expression)
- 371 Note on expression [RDA-FR : caractéristiques de couleur, sonores, précisions sur la forme de la notation, etc.]
- 432/42 Variant Access Point for Expressions
- 502/512/522 Related Access Point – Personal Name, Corporate Body, Family Name with Responsibility for the Expression
- 531/541 Related Access Point – Work
- 532/542 Related Access Point – Expression
- 640 Place(s) and Date(s) associated with the entity
- 732/42 Authorized Access Point in Other Language and/or Script for Expressions
- 801 Originating Source
- 810 Source Data Found
- 815 Source Data Not Found
- 820 Usage or Scope Information
- 830 General Cataloguer's Note

11

Legend:
Mandatory elements
Proposals 2020 (new or change)

Expressing LRM/RDA in the UNIMARC format

Assets of UNIMARC:

- A large community, a widely used format for production and exchange

- Millions of records encoded in UNIMARC (challenge ?)

- Widespread use of the Authorities Format

- Lots of links between records: relationships between works, between expressions, between all entities and agents, places, dates, etc.

- Good granularity (very structured format)

- Use of controlled vocabularies extracted from well-known international standards: multilingual format (code>text)

Challenges:

- Not *directly* for the semantic web

- Local implementations: many « flavours » of UNIMARC > interoperability ??

- The « nomen » entity

- The treatment of « persons » and bibliographic identities

- …

Data structured in UNIMARC is well prepared for a conversion to web-specific formalism

→ UNIMARC as a transition tool towards… something else ?