Gordon Dunsire, Independent Consultant
Edinburgh, United Kingdom
gordon@gordondunsire.com

Ontology in Practice

Ontologija u praksi

UDK/UDC: 004.822
026/027:004.738.5
Pregledni znanstveni rad – Scientific review
Primljeno / Received on: 04. 12. 2021.
Prihvaćeno / Accepted on: 13. 12. 2021.

Abstract

This paper discusses issues that arise from implementing the IFLA Library Reference Model (LRM) and its impact on bibliographic practices and services. The LRM has been fully implemented in RDA: resource description and access (RDA), and is in the process of being applied to ISBD (International Standard Bibliographic Description) and UNIMARC. These standards were developed before the LRM, and a major issue is the need to ensure that existing entities and elements are conformant with those given in the model ontology. This involves checking semantics, amending definitions, and resolving gaps and inconsistencies. The LRM is a conceptual model and provides only broad relationship and attribute elements that need to be augmented by finer elements to support operational practices and services. The precision of bibliographic description and retrieval services is driven by the granularity of elements: the finer the

Arhivi, knjižnice, muzeji:
mogućnosti suradnje u okruženju globalne informacijske infrastrukture
24 (2021), 89-104
ISSN 1849-4080

granularity, the greater the precision, so the level of granularity in an implementation of the LRM requires analysis of services to inform decisions. The paper also describes ontology, metadata content guidance, and metadata encoding format as the principal components of a system for universal bibliographic control that treats national cultural agencies as the meeting point of the local and the global.

Keywords: Ontologies; LRM; RDA; bibliographic control; Semantic Web; application profiles

Sažetak

U radu se raspravlja o pitanjima povezanima s implementacijom *IFLA-ina knjižničnog referentnog modela* (*IFLA Library Reference Model: LRM*), te njegovu utjecaju na bibliografske prakse i usluge. LRM je u potpunosti implementiran u *RDA: resource description and access*, a u postupku je njegova implementacija u ISBD (*Međunarodni standardni bibliografski opis*) i UNIMARC. Ti su standardi razvijeni prije objavljivanja LRM-a, pa je glavni problem potreba da se osigura da su postojeći entiteti i elementi usklađeni onima u ontologiji modela. To uključuje provjeru semantike, doradu definicija i rješavanje razlika i nedosljednosti. LRM je konceptualni model i kao takav pruža samo općenite odnose i elemente atributa koje je potrebno proširiti granuliranjem elemenata kako bi se podržale stvarne prakse i usluge. Točnost bibliografskoga opisa i usluga za pronalaženje uvjetuju granulirani elementi: veće granuliranje, veća točnost, pa prema tome razina granuliranja u implementaciji LRM-a zahtjeva analizu usluga pri donošenju odluka. U radu se također opisuju ontologija, smjernice za sadržaj metapodataka i enkodirajući format za metapodatke, koji čine glavne dijelove sustava za univerzalni bibliografski nadzor, prema kojemu su nacionalne ustanove u kulturi središnje mjesto za lokalno i globalno.

Ključne riječi: ontologije, LRM, RDA, bibliografski nadzor, semantički *web*, aplikacijski profili

Introduction

This paper discusses some of the issues that are encountered when deriving an operational ontology from a conceptual ontology to form part of an integrated set of tools for managing library and cultural heritage metadata in practical applications.

Effective operational metadata requires three main aids or tools. Any form of metadata that is packaged for re-use needs tools to determine the structure of the package, the encoding format for the structure and its contents, and guidance on formulating the content of the metadata. The diagram in Figure 1 shows the functional drivers that determine the "species" of bibliographic metadata that inhabits a specific cultural heritage environment.

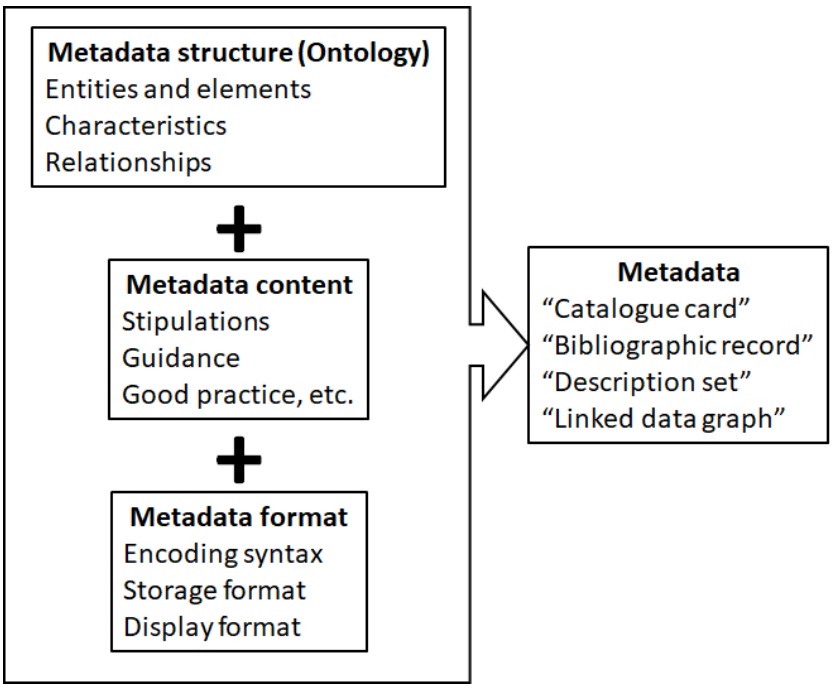

Figure 1. The bibliographic metadata ecosystem

Metadata may be packaged in a variety of ways, from catalogue card to linked data graph. The structure of the package is supported by an ontology that specifies the classes, characteristics, and relationships of things that are described. The content that describes an individual member of the species is supported by instructions and guidance on describing the things specified in the ontology. The format of the package is supported by a syntax for encoding the content so that it can be stored, displayed, and exchanged between specific applications. There is no standard definition of the term "ontology". It is used with different meanings in philosophy, where it origi-

nated, and in its application in fields of research.[1] This paper is concerned with ontology in bibliographic metadata and information retrieval systems, and uses the definition given in RDA: "A formal naming and definition of the types, properties, and interrelationships of the entities that exist for a particular universe of discourse".[2]

"IFLA Library Reference Model (LRM) is a high-level conceptual reference model developed within an entity-relationship modelling framework … that covers all aspects of bibliographic data … in linked data environments."[3] It is one of several bibliographic standards maintained by IFLA (International Federation of Library Associations and Institutions). The LRM is "comprehensive at the conceptual level, but only indicative in terms of the attributes and relationships that are defined." The LRM provides detailed definitions, scope notes, and constraints for its entities and the broad relationships between them, and for those attributes and specific relationships that are indicative of a fuller implementation of the model. These collectively form a high-level ontology for the "bibliographic universe".

The model is intended to be refined and extended with additional entity and element subtypes before it is implemented in a practical application. This involves several distinct stages:

- Choose which entities to implement from the model
- Fit additional entities to the model entities
- Choose which attributes to implement from the model
- Fit additional attributes to the entities
- Choose specific relationships from the model
- Fit additional relationships to the entities.

The distinction between an attribute and a relationship element is not absolute, and different implementations may treat the same element as either an attribute or a relationship. A relationship element has a target entity, so for example if an element for "colour" takes a value of a Colour

[1] Löfgren, Kent. What is ontology? A general introduction to the word and the concept [accessed: 2021-11-29]. Available at: https://www.youtube.com/watch?v=XTsaZWzVJ4c.

[2] RDA Registry. Ontology [accessed: 2021-11-29]. Available at: http://www.rdaregistry.info/termList/RDATerms/#1185.

[3] Riva, Pat; Patrick Le Bœuf; Maja Žumer. FLA Library Reference Model: a conceptual model for bibliographic information. As amended and corrected through December 2017 [accessed: 2021-11-29]. Available at: https://repository.ifla.org/bitstream/123456789/40/1/ifla-lrm-august-2017_rev201712.pdf.

entity it is a relationship, but if it takes a value that is just the name of a colour it is an attribute. It depends on what additional entities are fitted to the model entities. The LRM itself made similar choices when accommodating elements from the previous models on which it was based.

The result of refining and extending the LRM is an ontology for practical application that is derived or implemented from the model ontology. The derived ontology provides only structure and semantics, and needs to be complemented by tools for choosing and assigning values to the elements that describe and relate the individuals, members, or instances of the entities that are to be described by a metadata application. The full set of tools and stages of implementation is shown in Figure 2.

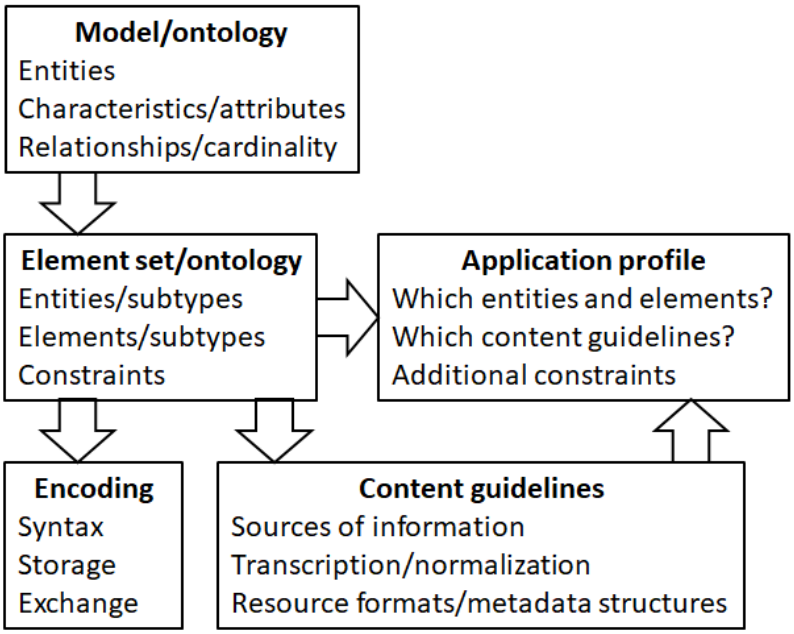

Figure 2. Component tools for a practical implementation of an ontology

Figure 2 indicates the dependences between the component tools.

A model ontology specifies the core entities and their basic characteristics. It also indicates the basic relationships between the entities and any cardinality restrictions on the number of entities that can be related. For example, the LRM states that an expression realizes one and only one work, but a work is realized by at least one expression. Most metadata applications

require a more granular set of entities and elements to improve precision. An operational ontology refines and extends the model ontology by specifying additional entity and element subtypes, and additional cardinality restrictions or other structural constraints. The ontology may also specify an encoding system for its entities and elements.

Guidelines for assigning values to create metadata content are based on the elements that are used to describe a specific individual thing or instance of an entity. Each element needs to be included. Guidelines typically cover the source of information for a value and how to transcribe a value. These may be different for specific categories of cultural heritage resource.

An encoding format for elements and content specifies the syntax used for data storage and exchange. For example, the format may represent each element with a code, and specify the character set used to represent element values.

If the operational ontology is intended to cover a range of applications, it is likely that any one application will not use every element, and that the application may require additional restrictions to be imposed on the use of appropriate elements. The relevant elements and constraints are specified by an application profile. The profile may also specify which of the content guidelines to apply to an element if there is a choice. An application profile is a practical guide for the cataloguer of which values to assign to which elements to create effective metadata for what is being described. It can be used to generate or select data input or retrieval templates and display content guidelines in context, presenting only what is relevant to the application.

The LRM ontology has been implemented as an operational ontology in RDA, and is in the process of being implemented in ISBD. "RDA: Resource Description and Access is a package of data elements, guidelines, and instructions for creating library and cultural heritage metadata …".[4] Alignment with IFLA's bibliographic conceptual models is a key component of RDA, so the RDA Steering Committee agreed at its 2016 annual meeting to adopt the LRM, which was then near to its final draft, within the forthcoming RDA Toolkit Restructure and Redesign (3R) Project.[5] The redesigned Toolkit beca-

[4] RDA Toolkit. Introduction to RDA. July 2021 [accessed: 2021-11-29]. Available at: https://access.rdatoolkit.org/en-US_ala-d3b460ae-1818-3459-8c9c-9dd5fda8edf8/p_k1m_zty_mdb (subscription required).

[5] RDA Steering Committee. 3R Project final report. 2021 [accessed: 2021-11-29]. Available at: http://www.rda-rsc.org/sites/all/files/3R%20Project%20Final%20Report.pdf.

me the official version of RDA with its December 2020 release. "The International Standard Bibliographic Description (ISBD) is intended to serve as a principal standard to promote universal bibliographic control … the standard that determines the data elements to be recorded or transcribed in a specific sequence as the basis of the description of the resource being catalogued …".[6] ISBD is an IFLA standard, and in 2017 a working group for the analysis of the alignment and impact of IFLA LRM on ISBD reported that "ISBD may be developed as an implementation of LRM, and that more ISBD elements may be aligned to relationships in the model".[7] This is being carried out as a two-stage process; the first stage will implement ISBD as an operational ontology and content standard for the LRM entity Manifestation, and the second stage will extend coverage to other LRM entities, including Expression and Work. The development of RDA and ISBD is coordinated via a protocol established between the RDA Steering Committee and the ISBD Review Group.[8]

Entities

The Agent entity in the LRM is in the middle of a hierarchy of entities, from the universal entity Res to the finer granularity of Collective Agent. This can be shown as a linked data graph of triples or entity-relationship diagram, as in Figure 3. The Agent entity in RDA is also in the middle of a hierarchy, but with top and bottom entities that are not in the LRM: RDA Entity, and Family.

---

[6] ISBD International standard bibliographic description. Consolidated edition. 2011 [accessed: 2021-11-29]. Available at: https://repository.ifla.org/bitstream/123456789/786/1/ifla-isbd-international-standard-bibliographic-description-2011.pdf.

[7] ISBD Review Group. Minutes. 2017 [accessed: 2021-11-29]. Available at: https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/isbdrg/isbdrg_meeting_2017.pdf.

[8] 8. RDA Steering Committee. Protocol between the RSC and the ISBD Review Group. 28 January 2016 [accessed: 2021-11-29]. Available at: http://www.rda-rsc.org/sites/all/files/RSC-Chair-5.pdf.

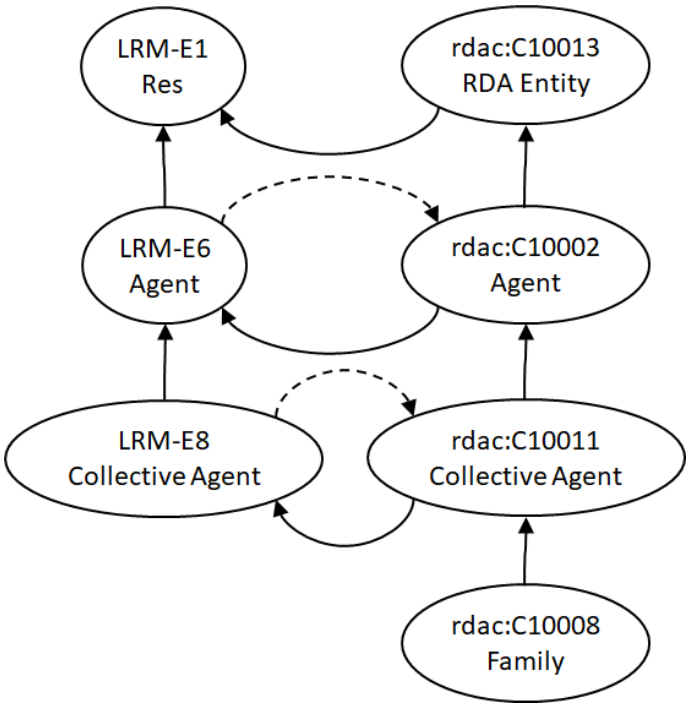

Figure 3. Entity-relationship diagram of partial map from RDA to LRM agent entities

The hierarchies can be aligned using the same subtype relationship. All of the relationships in Figure 3 are rdfs:subclassOf, an element from the Resource Description Framework Schema (RDFS) that is used to describe ontologies for data processing.[9] Its verbalized label is "is sub-class of". It is the same element as the "isA" relationship mentioned in the LRM. The subclass element implies that every thing that is a member of the subclass is also a member of the superclass. The subclass relationship element is used to describe the hierarchy of entities within an ontology. Figure 3 shows both the LRM and RDA hierarchies that include the Collective Agent entity. It also shows that the subclass element can be used to map two hierarchies together at their points of intersection. For example, the RDA Collective Agent entity is mapped as a subclass of the LRM Collective Agent entity, as shown in Figure 3. This states that every RDA family is also an RDA

[9] W3C. RDF Schema 1.1: W3C recommendation 25 February 2014 [accessed: 2021-11-29]. Available at: https://www.w3.org/TR/rdf-schema/.

collective agent, and every RDA collective agent is an LRM collective agent. The relationship is transitive: every instance of an RDA family is an instance of LRM collective agent. The RDA Registry provides an RDFS map from the RDA entities to the LRM entities.[10]

What this means in practice is that RDA is certain of the implication that every individual RDA collective agent is also an individual LRM collective agent, but is not certain that every individual LRM collective agent is also an individual RDA collective agent. The LRM community can resolve the uncertainty by creating the inverse mapping "LRM-E8 rdfs:subclassOf rdac:C10011", indicated by the dashed lines in Figure 3. If two entities are each a subclass of the other, they are effectively the same entity. The practical utility of applying this approach to mapping is that the RDA community is assured that its metadata is compatible with the LRM without having to wait for confirmation from the LRM community.

Note that the LRM's Res and RDA's RDA Entity are not at all equivalent. The RDA entity is clearly a subclass of the LRM entity because Res is the superclass of all entities in the bibliographic universe. The RDA entity, on the other hand, is the superclass of all entities in the RDA universe, a subset of the bibliographic universe. For example, the RDA universe does not cover entities for metadata for administrative or rights management support. It would be incorrect if the LRM community mapped Res as a subclass of RDA Entity. Figure 3 shows how an entity that is not in the LRM, RDA's Family, is made compatible with the LRM by using the RDA hierarchy and crossing over to the LRM via a superclass, thus refining the LRM in practice.

Elements

LRM provides a universal relationship element between two universal Res entities; LRM-R1 associates Res with Res. The relationship works the same in both directions, so it is an inverse of itself; this is known as a symmetric relationship. The LRM entity hierarchy cascades this relationship to any two LRM entities, for example Manifestation and Agent. Any specific relationship between Manifestation and Agent is therefore a subtype of the universal relationship. The subtype relationship between the relationship

[10] RDA Registry. Map from RDA classes to IFLA LRM. 22 December 2020 [accessed: 2021-11-29]. Available at: http://www.rdaregistry.info/Maps/mapRDAEntity2LRM.html.

elements is rdfs:subpropertyOf or "is sub-property of", from narrow to broad, and is also taken from RDF Schema. An example of the LRM approach is:

- LRM-E1 "Res" LRM-R1 "is associated with" LRM-E1 "Res"
- Implies: LRM-E4 "Manifestation" LRM-R1 "is associated with" LRM-E6 "Agent"
    - Implied subtype: LRM-E4 "Manifestation" "LRM-R7 was created by" LRM-E6 "Agent"

This device provides a complete coverage of high-level pair-wise relationships between LRM entities.

RDA provides a specified matrix of high-level association relationships between all of the RDA entities, but uses distinct relationship elements for each pair. The corresponding RDA relationship element to LRM-R7 is rdam:P30329 "has creator agent of manifestation". It is assumed again to be a subtype of the LRM element and is mapped accordingly. Again, the RDA Registry provides an RDFS map from the RDA elements to the LRM elements.[11]

The relationship element hierarchies are usually refined still further within an operational implementation of the LRM. The RDA element itself is subtyped for specific Agent subtypes such as Person. RDA splits the "publication" role between mechanically produced (publisher) and hand produced (producer) manifestations. These are both subtypes of the high-level creator role. For legacy reasons, RDA also refines the publisher relationship for audio-visual manifestations that are accessed by end-users via broadcast technologies. This results in the RDA hierarchy:

- rdam:P30363 "has creator person of manifestation"
    - rdam:P30360 "has producer person of unpublished manifestation"
    - rdam:P30362 "has publisher person"
        - rdam:P30347 "has broadcaster person"

Another implementation of the LRM may require even more granular relationships, for example to described podcast manifestations. This might

[11] RDA Registry. Map from RDA properties to IFLA LRM. 22 December 2020 [accessed: 2021-11-29]. Available at: http://www.rdaregistry.info/Maps/mapRDA2LRM.html.

be a role that is deemed to be narrower than "broadcaster". In principle, any element (relationship and attribute) can be refined to any level of precisions that is required. In practice, this may become unwieldy and better approaches may be available.

Shortcuts

The complexity of an ontology can be simplified by replacing a chain of specific relationship elements with a single element that relates the first entity directly with the last entity. Such simplification comes at a price; no information about the intermediate entities in the chain is recorded, resulting in 'dumber', lower quality metadata. However, if a shortcut is latent in a legacy element, then the price has already been accepted, and the legacy element can be retained as a specified shortcut within the operational ontology.

The LRM defines a "shortcut" as "A single relationship which serves to represent a more developed path consisting of two or more relationships". A shortcut element is an element that collapses a chain of multiple relationships into a single relationship that ignores the detail of how the chain is constructed and preserves only the source and target entities. A model of a simple shortcut is:

- Entity A has shortcut relationship with entity Z

This is equivalent to:

- Entity A has relationship 1 with entity B + entity B has relationship 2 with entity Z

The specification of shortcut elements is an implementation decision. The LRM itself specifies only 2 shortcut elements: LRM-R15 "is equivalent to"; and LRM-R25 "was aggregated by". They provide a simplification of the full model in the complex areas of appellations and aggregates respectively. Shortcut elements are a useful tool for aligning existing relationship elements to a new model. The example of a shortcut given in Figure 4 is the most basic one found in RDA. RDA has a legacy "primary" element that relates a manifestation to a work that it embodies. There is no equivalent element in the LRM. The RDA element is a shortcut for the model chain of primary relationship elements between a Manifestation and Work via the Expression that is the content that is embodied.

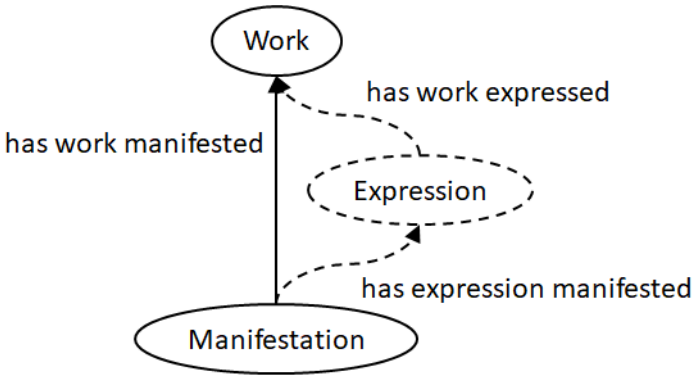

Figure 4. Expansion of a shortcut relationship element

The embodied expression is not recorded, so the shortcut ignores language, content type, and other Expression attributes. This is useful for some applications, for example in a monolingual environment, but is not useful for multilingual environments.

The LRM also notes that relationship elements that have a source or target entity of Agent can be extended implicitly to Agent subtypes such as Person by chaining them with the ontological "isA" subclass relationship. The 3R Project chose to provide explicit relationship elements for the RDA Agent entity subtypes. These are arranged in hierarchies based on the hierarchy of Agent subtypes. For example, rdam:P30362 "has publisher person" is an element subtype of rdam:P30083 "has publisher agent". These "agent hierarchies" intersect with the semantic hierarchies of relationships to form polyhierarchies where most agent relationships have two, rather than one, parent or broader relationship. The "publisher" relationship is more granular than the 'creator of manifestation' relationship, so rdam:P30362 "has publisher person" is an element subtype of rdam:30363 "creator person of manifestation". The 'Related elements' section of the instructions for "publisher person" therefore states[12]:

> "For broader elements, see:
>
> Manifestation: creator person of manifestation
>
> Manifestation: publisher agent"

[12] RDA Toolkit. Publisher person. July 2021 [accessed: 2021-11-29]. Available at: https://access.rdatoolkit.org/en-US_ala-e4e22111-816f-300a-a3df-7c6b1ba74f76/4c2f63a4-4528-434c-a131-9446cfddaa92 (subscription required).

This illustrates the net-like, not tree-like, structure of relationships in an ontology that uses entity subtypes. The underlying semantics can be represented in different ways by different implementations.

National agencies

LRM is a global model that covers the bibliographic universe, and ISBD, RDA, and other implementations are intended for international use. Local applications of such standards use application profiles that select the appropriate tools from the global toolset and stipulate how they should be used.

Application profiles can be structured in layers, with each layer being more precise and detailed in its formulation. This can be used to coordinate and interoperate local specialized applications within a wider framework, typically at national or supranational level. A common core profile might cover the broad requirements of cultural, economic, and social communities, and be refined or extended for more specific applications within the broad ecosystem. This can be applied in principle to multiple levels of profile. The broader profiles are a suitable activity for national cultural heritage agencies that can specify the basic requirements for a national bibliographic ecosystem, and coordinate and distribute more local requirements that are sectoral or special. For example, sources of authorized access points for entities are not prescribed at the global level because there is no agreement about which ones to use. Sources include authority files and the rules for construction which vary according to cataloguing tradition and social, cultural, and political factors. Even aggregator services such as VIAF, the virtual international authority file, do not have universal coverage. National agencies are able to take these factors into account and coordinate the maintenance of national sources of access points that can then be prescribed in an application profile.

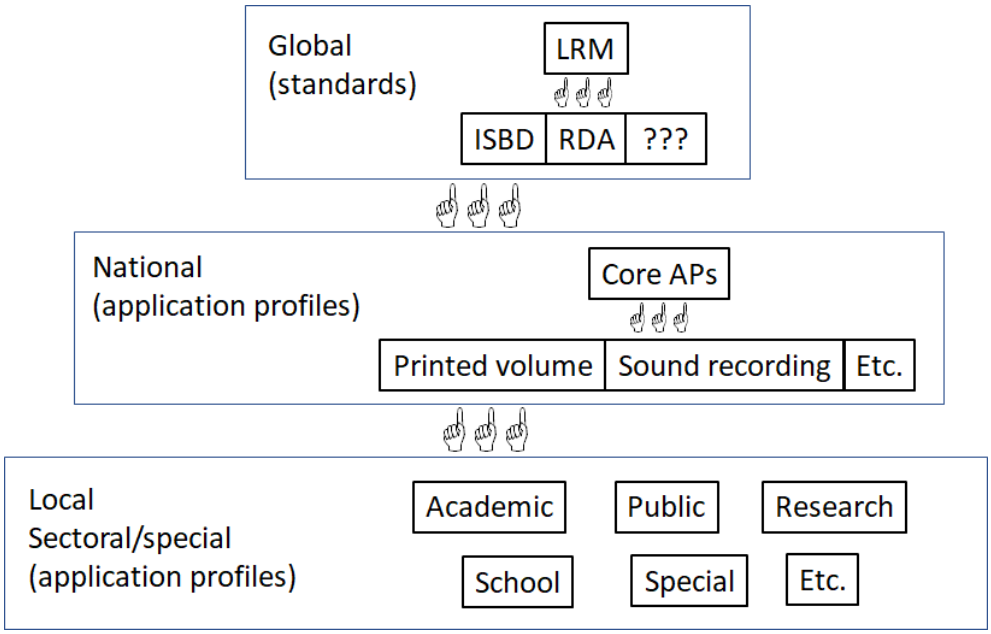

Figure 5. Application profiles as a bridge between the local and the global

Figure 5 is a schematic diagram showing the three main levels of application: global, national, and local. Local application profiles for metadata for specific kinds of libraries can be derived from national application profiles for global metadata standards. National profiles may be based around a common core set of elements and content that is augmented by elements required for specific kinds of resource.

The arrangement is flexible; if national agencies are based on kinds of library service, then the sectoral profiles may be maintained at national level, with local refinement based on kinds of resource. This is a powerful tool for coordinating metadata from the bottom-up.

Conclusion

A model ontology usually requires refinement and extension to provide the entities and elements required for practical application. An operational ontology reflects a view of the world from a specific community that is compatible with a view shared by all communities.

Legacy entities and elements can be accommodated within an operational ontology by subtyping and shortcuts. Subtyping an entity or element increases the precision of an ontology, but using a shortcut decreases precision. The integrity of the model ontology should be respected and not simplified to an extent that renders it ineffective.

Application profiles are required to select entities and elements for specific practical applications from an operational ontology with wide coverage. A local metadata service is unlikely to require every entity or an element from every element hierarchy, or the most granular element in a hierarchy, so a profile can filter out unwanted "noise" in metadata creation and maintenance processes. A profile also supports interoperability between similar applications and within broader cultural heritage applications.

National agencies have an intermediation role between global and local metadata environments in creating and coordinating application profiles. Such agencies are well-placed to apply cultural and social norms to the selection of entities, elements, content standards, and encoding format from universal tools, and to coordinate refinement of national tools for specific local applications.

Literature

ISBD International standard bibliographic description. Consolidated edition. 2011 [accessed: 2021-11-29]. Available at: https://repository.ifla.org/bitstream/123456789/786/1/ifla-isbd-international-standard-bibliographic-description-2011.pdf.

ISBD Review Group. Minutes. 2017 [accessed: 2021-11-29]. Available at: https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/isbdrg/isbdrg_meeting_2017.pdf.

Löfgren, Kent. What is ontology? A general introduction to the word and the concept [accessed: 2021-11-29]. Available at: https://www.youtube.com/watch?v=XTsaZWzVJ4c.

RDA Registry. Map from RDA classes to IFLA LRM. 22 December 2020 [accessed: 2021-11-29]. Available at: http://www.rdaregistry.info/Maps/mapRDAEntity2LRM.html.

RDA Registry. Map from RDA properties to IFLA LRM. 22 December 2020 [accessed: 2021-11-29]. Available at: http://www.rdaregistry.info/Maps/mapRDA2LRM.html.

RDA Registry. Ontology [accessed: 2021-11-29]. Available at: http://www.rdaregistry.info/termList/RDATerms/#1185.

RDA Steering Committee. Protocol between the RSC and the ISBD Review Group. 28 January 2016 [accessed: 2021-11-29]. Available at: http://www.rda-rsc.org/sites/all/files/RSC-Chair-5.pdf.

RDA Steering Committee. 3R Project final report. 2021 [accessed: 2021-11-29]. Available at: http://www.rda-rsc.org/sites/all/files/3R%20Project%20Final%20Report.pdf.

RDA Toolkit. Introduction to RDA. July 2021 [accessed: 2021-11-29]. Available at: https://access.rdatoolkit.org/en-US_ala-d3b460ae-1818-3459-8c9c-9dd5fda8edf8/p_k1m_zty_mdb (subscription required)

RDA Toolkit. Publisher person. July 2021 [accessed: 2021-11-29]. Available at: https://access.rdatoolkit.org/en-US_ala-e4e22111-816f-300a-a3df-7c6b1ba74f76/4c2f63a4-4528-434c-a131-9446cfddaa92 (subscription required)

Riva, Pat; Patrick Le Bœuf; Maja Žumer. FLA Library Reference Model: a conceptual model for bibliographic information. As amended and corrected through December 2017 [accessed: 2021-11-29]. Available at: https://repository.ifla.org/bitstream/123456789/40/1/ifla-lrm-august-2017_rev201712.pdf

W3C. RDF Schema 1.1: W3C recommendation 25 February 2014 [accessed: 2021-11-29]. Available at: https://www.w3.org/TR/rdf-schema/.