RDA and the eXtensible Catalog

By
Jennifer Bowen and David Lindahl
eXtensible Catalog Organization

The eXtensible Catalog (XC) is a set of open-source software tools and metadata schema designed to facilitate library metadata management and resource discovery.¹ XC is the result of a multi-million dollar, multi-year software development project funded by The Andrew W. Mellon Foundation’s Scholarly Communications Program, with additional support from The University of Rochester, The Consortium of Academic and Research Libraries in Illinois (CARLI) and other XC partner institutions.² XC software is supported and maintained by the not-for-profit eXtensible Catalog Organization (XCO).

XC software represents the first live implementation of a subset of RDA in a FRBR-based, non-MARC environment. XC’s implementation of RDA has been led by individuals who have participated in the development of both the RDA Toolkit and the RDA vocabulary registry. XC’s use of RDA has also been informed by the real-world requirements of actual working software, as well as through a user research process conducted at four ARL libraries.³

RDA in the XC Schema

XC uses a metadata schema, called the XC Schema, to facilitate the functionality of the XC discovery environment and take full advantage of metadata created by libraries. The XC Schema uses the concept of a Dublin Core Application Profile to use metadata elements from various schemas within a single environment. The schema currently contains twenty two RDA elements and eleven RDA role designators, as well as all Dublin Core “dcterms” data properties and a few other data elements defined specifically to enable XC’s system functionality. To be eligible for inclusion within a Dublin Core Application Profile, schema terms or elements must be defined on the basis of RDF (Resource Description Framework).⁴ Metadata elements created using AACR2 and MARC do not fit this criterion and therefore cannot be used within an Application Profile. Fortunately, RDA elements and vocabularies DO fit this criterion because of their development and maintenance, in parallel with the development of the RDA Toolkit, in the Open Metadata Registry.⁵ RDA elements can therefore interact easily with elements from other metadata schemas, making RDA a much more flexible standard than other standards currently in use within the library community, and therefore attractive for use within XC.

³ Nancy Fried Foster et al., eds., Scholarly Practice, Participatory Design and the eXtensible Catalog (Association of College and Research Libraries, forthcoming), http://hdl.handle.net/1802/12375
For the first version of the XC Schema, we selected RDA elements that enable us to retain the granularity of bibliographic data currently coded within MARC records. RDA elements used within the XC Schema include serial numbering, scale for cartographic materials, plate number for musical scores and other very specific elements that are not represented in more general schemas such as Dublin Core. Using these particular RDA elements within the XC Schema enables us to map several MARC data elements directly to RDA properties. In developing the XC Schema, we have been fortunate to have access to elements from a standard such as RDA that has been being developed within the library community, and which therefore aligns closely with defined elements in existing library catalog data. The developers of RDA have wisely created a standard that can function both within the existing MARC environment as well as in an RDF-based, linked data environment. We have found that RDA thus serves as an important “bridge” between present library systems and emerging applications such as XC.

Preparing MARC data to be reused in an open system environment has required XC software developers to build a robust metadata processing platform to analyze, clean up, and repurpose MARC data. The resulting XC Schema metadata records enable the functionality of XC’s next-generation user interface, and can be potentially converted into RDF-based linked data, to make data about library resources available for use as part of the Semantic Web. While we have demonstrated that existing MARC records can be reused successfully, we have also confirmed that a significant amount of AACR2/MARC data cannot be reused without considerable programming or manual record editing. One AACR2/MARC bibliographic record may contain references to multiple resources, but lack identifying information for the related entities. The process of converting data from AACR2/MARC records to linked data is complicated because the data has not originated in RDF-compatible structures and the definitions do not always map correctly. We believe that the library community can derive benefit from current MARC data in future systems. However, to knowingly continue to create metadata that cannot be reused effectively in other systems is potentially a waste of current library resources, especially when using a more forward-looking standard (RDA) will begin to address these problems.

Implementing RDA in a FRBR-based Environment

One of XC’s goals is to enable legacy MARC metadata to be reused within a new system architecture that is not based upon MARC. The XC Schema is FRBR based, and makes use of separate but linked records for the FRBR Group 1 entities, in an architecture that approaches an RDA “Scenario 1” implementation. XC software enables the processing of these FRBR-based records in an end-to-end system, managing the relationships between these records even as records are updated and deleted and as new records are added to a source repository. XC’s current abilities to create, parse, and manage FRBRized data records demonstrate that basing a system upon a FRBRized data structure is indeed feasible, and show how such a system can work alongside current MARC-based Integrated Library Systems. We envision a scenario where XC will harvest both RDA records in MARC from an ILS, plus records in some FRBR-based RDA carrier or schema (yet to be defined) from another source, and enable

---


both types of metadata to interact within the same discovery interface. With both sources of metadata using the same RDA elements and vocabularies, the amount of metadata manipulation necessary will be much reduced, and a greater percentage of the metadata will be usable by other applications. Using RDA in a non-MARC environment is not something that is far off in the future – it is possible now.

We have used RDA-defined associations between a data element and a particular FRBR or FRAD entity as the basis for such associations within the XC Schema, and XC’s mapping of MARC fields to the XC Schema is informed by the MARC to RDA mappings presented in the RDA Toolkit. By basing as much of our work as possible upon RDA, XC has set the stage for additional transformations of metadata into a more complete FRBR-based implementation of RDA. This could occur either through the addition of more RDA elements and roles to the XC Schema, or through the development of another carrier or schema defined specifically for RDA. In either scenario, XC software’s ability to parse MARC data into a hierarchical FRBR-based structure can ease the transition of the library community into an RDA-based world that goes beyond an initial implementation of RDA in MARC. XC’s transformation of MARC data to the XC Schema can be thought of as an interim step that will pave the way for a full system implementation of RDA in the future.

Given the current lack of a metadata carrier or schema for RDA other than MARC, some have speculated that it is premature to implement RDA, considering the seemingly meager benefits of implementing RDA in MARC. We do not see this as an issue. As we have demonstrated with the XC Schema, RDA elements and properties can be used in combination with other schema elements within an XML-based Application Profile. The use of RDA elements, even within a MARC-based structure, will help XC’s metadata cleanup and transformation programs work more effectively than does AACR2 data. We look forward to experimenting with RDA elements expressed as RDF statements, which may be a more promising alternative than developing a new XML record-based carrier for RDA data. These scenarios can be investigated using XC software as soon as RDA is implemented more widely and libraries begin sharing larger sets of records that have been cataloged using RDA.

**Benefits of RDA in XC**

XC’s implementation of RDA elements within a FRBR-based XML schema positions the XC Schema as the most promising way forward for implementing RDA in XML in the near term. XC’s potential use for this purpose was apparently evident to the developers of the RDA Toolkit, who included a link to the eXtensible Catalog website as a resource within the RDA Toolkit itself. XC not only enables the conversion of MARC fields to RDA data elements using bulk processing services, but also successfully makes use of that same RDA data within a working discovery system. XC Schema is a foundation for a solid RDA implementation that is usable in real systems, addresses real use scenarios, and works with existing Integrated Library Systems and web content management systems.

A community-wide implementation of RDA within the library world will benefit not only users of the eXtensible Catalog, but also developers and users of other applications that make information about library collections accessible via the open web. One of the strengths of the library community has always been its adoption of community-wide standards such as AACR2 and MARC, which encourage other communities to interact with our metadata. A widespread adoption of RDA will continue this tradition of library leadership in metadata standards, and provide a clearer vision for the development of future library systems.