MARC and MARC XML

What is MARC XML (MARCXML), and why should we care? The short answer is that MARC XML is MARC encoded in XML. The longer answer is that MARC XML moves MARC 21 catalog data from the proprietary data format traditionally used by libraries to XML, the open-standard data format used by just about everybody else. By joining the rest of the world, libraries can take advantage of the wealth of knowledge and tools available for XML data.

First, let us discuss MARC’s origins and development. Back in the sixties, when computers still used punch-cards and magnetic tape, the Library of Congress (LC) developed a machine-readable data format for cataloging, which caught on as libraries around the world realized the need for interoperable, computerized catalog records. In essence, the old card catalog data was broken apart into separate data fields and tagged in a way that identified the type of data. This process is well-understood by computer scientists, and the MARC format is a typical example of a data format created for the specific needs of an industry.

However, MARC is a proprietary format not commonly used outside of libraries, and today there are better options. XML was introduced in the late nineties, not long after the introduction of the web, as a generalized replacement for HTML, the data format used to represent web documents. Via XML, data may be kept in tagged fields and displayed according to the needs of the user, including as human-readable documents. Many of us have seen this idea before in ‘mail merge’, a handy trick for turning databases into address labels. The implications
of this simple idea are staggering, and people are still developing new ways to harness the power of XML.

I believe I am required by tradition to point out that XML is derived from an older standard called SGML, the Standard Generalized Markup Language. This standard is extraordinarily complicated, and hardly anyone uses it. Now you know everything the typical person needs to know about the relationship between XML and SGML.

From the XML 1.0 (Third Edition) W3C Recommendation,¹ the design goals for XML are:

1. XML shall be straightforwardly usable over the Internet.
2. XML shall support a wide variety of applications.
3. XML shall be compatible with SGML.
4. It shall be easy to write programs which process XML documents.
5. The number of optional features in XML is to be kept to the absolute minimum, ideally zero.
6. XML documents should be human-legible and reasonably clear.
7. The XML design should be prepared quickly.
8. The design of XML shall be formal and concise.
9. XML documents shall be easy to create.
10. Terseness in XML markup is of minimal importance.

As you can see, XML provides an updated framework compatible with MARC, but with the benefit of broad support. As a result, the Library of Congress, the originator of the MARC format, has chosen to update MARC to the XML format; thus, MARC XML. MARC XML holds the same data as MARC using the same fields and indicators, and data in either format can be easily converted to the other. As an example from the LC website:²

```xml
<datafield tag="100" ind1="1" ind2=" ">
    <subfield code="a">Sandburg, Carl,</subfield>
    <subfield code="d">1878-1967.</subfield>
</datafield>
```

¹ [http://www.w3.org/TR/REC-xml/](http://www.w3.org/TR/REC-xml/)
² [http://www.loc.gov/standards/marcxml/Sandburg/sandburg.xml](http://www.loc.gov/standards/marcxml/Sandburg/sandburg.xml)
As you can see, the data encoded in MARC XML is the same as what you would find in a MARC record. An additional benefit is that XML does not require any magical software to convert the data into MARC. However, should you want software to assist with data-entry, a wide variety of tools exist for writing XML data, some of which specifically supports MARC XML.

Saying MARC XML is essentially the same as MARC is wonderful, but why is it better? The best answer is that XML can be directly interpreted by modern web browsers, greatly reducing the need for ‘solutions’ like Innopac and other Online Public Access Catalogs (OPACs). Browsers already exist for particular populations, including the visually impaired, and the underlying MARC data is immediately available for use without the inevitable garbling likely in current OPAC solutions. Perhaps even more importantly, XML can be understood by ‘machines’ such as Google, allowing library holdings to be searchable from outside the library environment. This ability is still being explored, but promises to give libraries a greater presence on the web.
than is currently the case.

Related to MARC is MARC Lite, a subset of MARC created by the Library of Congress to serve as a general purpose cataloging data standard. It is important because the XML version of MARC Lite is MODS, an increasingly important metadata standard. Just as MARC and MARC XML can be transformed into each other, MARC Lite and MODS can be similarly transformed. The primary benefit to using MODS over MARC XML is the simpler language used. As an example:

```xml
<titleInfo>
  <title>Sound and fury :</title>
  <subTitle>the making of the punditocracy</subTitle>
</titleInfo>

<name type="personal">
  <namePart>Alterman, Eric</namePart>
  <role>
    <roleTerm type="text">creator</roleTerm>
  </role>
</name>

<typeOfResource>text</typeOfResource>
<genre authority="marc">bibliography</genre>
<originInfo>
  <place>
    <placeTerm authority="marccountry" type="code">nyu</placeTerm>
  </place>
  <place>
    <placeTerm type="text">Ithaca, N.Y</placeTerm>
  </place>
  <publisher>Cornell University Press</publisher>
  <dateIssued>c1999</dateIssued>
  <dateIssued encoding="marc">1999</dateIssued>
  <issuance>monographic</issuance>
</originInfo>

<language>
  <languageTerm authority="iso639-2b" type="code">eng</languageTerm>
</language>

<physicalDescription>
  http://www.loc.gov/standards/mods/v3/mods99042030.xml
```
As you can see, all of the information and formatting you would expect to find in a MARC record are in the MODS record, but with more accessible tags. Because MODS is in the MARC family of data formats, the catalog record can be transformed into MARC XML, MARC Lite, or MARC, as necessary, with a minimal time investment.

What do these developments mean for libraries? Better service for a minimal investment of time. A great amount of time and money has gone into making MARC the ubiquitous standard it is today, and MARC is not likely to change drastically or to be abandoned any time soon. However, the world outside of libraries has found its own solution to cataloging, and that solution involves XML. With surprising foresight, the Library of Congress has made the transition to XML easy and reversible, with the added benefit of a compatible data format for generalists' use. I would say the future of MARC cataloging is looking well.
Bibliography


