

Future prospects for UNIMARC

Alan Hopkinson
Head of Library Systems and Bibliographic Services
Middlesex University
Hendon
The Burroughs
London NW4 4BT, UK
Email a.hopkinson@mdx.ac.uk

Abstract

UNIMARC is one of a rapidly decreasing number of bibliographic exchange formats. It facilitates the exchange of data. The cultural heritage field is one which is not awash with funding. The MARC formats are developing with time, taking into account the latest technologies, but the basic MARC formats are very much alive and facilitate exchange of data between systems even in the case of those that use low-cost software.

Keywords: bibliographic data; MARC; UNIMARC; cataloguing

1. Bibliographic exchange formats

Preamble

UNIMARC is one of a rapidly decreasing number of bibliographic exchange formats. These formats are conventions for the arrangement of bibliographic data in a computer file. By following these conventions (which are known as formats), developers of cataloguing systems software can provide an interface for the import of data from other systems and the export of data to other systems. Gredley and Hopkinson (1990) describe the development of MARC, beginning with Library of Congress MARC, followed by UK MARC, and the development of numerous dialects. MARC was supposed to be a means of exchanging bibliographic records but it quickly proved to be difficult to exchange between different national libraries because of changes that were made to the national format in each country. MARC (Machine Readable Cataloguing) was developed in 1966 at the Library of Congress by Henriette D Avram who had retired from the US government data processing service but could not rest and went to work on a Library of Congress project to develop automated cataloguing (she passed away in 2006 at a great age and it is good to take this opportunity to pause to remember the huge influence that she personally has had on libraries throughout the world). In 1968, the British National Bibliography (BNB), one of the forerunners of the British Library was invited to talks to develop a common format. Staff of the BNB made a few suggestions some of which were adopted. One suggestion was to replace punctuation in a catalogue record with codes to break up the data fields in a more systematic way. Library of Congress staff implemented subfields but not to the same extent as the BNB did. Library of Congress practice retained the punctuation which actually meant it was more difficult to provide the style of punctuation which the cataloguing rules demanded. To give two examples which illustrate this: in MARC formats the data is divided into fields which are identified by 3-digit tags, fields are divided into subfields: almost all fields begin with a subfield represented conventionally by **\$a**: 100 means personal name, main entry. In a catalogue, a personal name is represented as follows: Smith, John. Field 100 contains therefore **\$aSmith, John** in LC MARC. In UK MARC it is represented by **\$aSmith\$John**: the punctuation is replaced by a subfield. In the case of a title, data will be identified by title field 245: in LC MARC it is represented as **\$aUNIMARC :\$bthe future /\$cby Alan Hopkinson**, in UK MARC it is **\$aUNIMARC\$bthe future\$dby Alan Hopkinson**. In this second case, in LC MARC, the subfield has punctuation supplied as well as the data being identified by a subfield identifier. In general there is a small number of differences between the two formats but in order to take records from one format and import them into a database set up according to another format, a conversion program has to be run. However a principle was established by UK MARC that punctuation should wherever possible be produced by the computer program and not entered by the cataloguer into the MARC record. However, despite these differences outlined above, both formats use the same record structure. This was in the early 1970s adopted in the United States as a national ANSI standard and subsequently as an international standard, ISO 2709 (ISO, 2006). This was a standard based on the principles of exchange of records on magnetic tape which was the main method of transferring data between systems in the 1960s and 1970s. Subsequently but not until much later the standard was changed to recognise other transfer media but this was long after floppy disks and the internet were used.

From the outset then there were two formats though thankfully one data structure. Subsequently, Canada developed a MARC format, followed by other national libraries around the world which developed formats generally based on either LC or UK MARC. There were also formats developed for secondary services, using the same record structure. The UN agencies, Food and Agriculture Organization and International Atomic Energy Agency, used the AGRIS and INIS formats respectively and UNESCO developed as a medium of exchange between secondary services the UNISIST Reference Manual. Later, in an attempt to bring some order to the chaos of the different standards, UNESCO developed the Common Communication Format.

This situation meant that international exchange was not as easy as it should have been. IFLA then developed a format which was called the Universal MARC Format (UNIMARC), intended for international exchange of data. The idea was that each format would be converted into UNIMARC and so institutions wishing to take records from elsewhere would need to write only two computer programs, one into and one out of UNIMARC. UNIMARC was therefore intended to be a switching format and not a format for direct entry of data.

However, following the development of UNIMARC, a number of national libraries prepared national formats based on UNIMARC. In France and Belgium a format very close to UNIMARC called INTERMARC was and still is used. However, in France UNIMARC is used as the format outside the national library and the national library convert their data for export into UNIMARC. Portugal, South Korea, Japan, Taiwan, Italy, Greece and Croatia all began to use UNIMARC as their national formats. A cooperative based in Slovenia COBISS also used it making UNIMARC the format of what was then Yugoslavia. The former Soviet Union used a format derived from UNIMARC. However, after 1991 when many foreign donors established computer systems in eastern Europe, a plethora of library systems were acquired some of which were able to use only LC MARC.

Later, LC MARC became a national format and was called US MARC with input into its development from a group associated with the American Library Association known as MARBI. Moving on to the 21st century, US MARC was renamed MARC21, denoting it was international and not intended only for the US. Canada abandoned its format in favour of MARC21, as did South Africa and Australia. The National Library of Scotland had used US MARC for a number of years as they were a member of OCLC and were interested in downloading records in US MARC format from the OCLC database. Taiwan began to use MARC21 for Latin language materials.

Currently there are millions of MARC records around the world. They are transferred in the ISO 2709 format. Many library systems demand the use of this though XML has become a new transfer medium and currently a version of XML compatible with MARC has been proposed as an international standard, known as MarcXchange (ISO, 2006), which dispenses with the record structure but which allows the retention of the other features of the MARC formats, tags, subfields and indicators and indeed provides a standard way of representing these in XML. So MARC coding is still alive in XML!

2. Influence of the outside world

MARC was developed 40 years ago, but computers have advanced in many ways. So many systems have changed beyond recognition, but in our field MARC formats remain almost exactly as they were. As mentioned above, the structure of MARC is based on principles of tape exchange. Computer programmers coming to it now for the first time find it very quaint. There is pressure from the outside world to change this. However, the library community has always been different. For example, data processing in library systems has never used abbreviation and coded data to the same extent as other kinds of data processing used to. Any changes were resisted entirely until the advent of XML. Computer programmers might inform library managers that MARC using ISO 2709 is antiquated but on the other hand the system works. Having a format which requires special programming contributes an element of stability to the situation. Library management systems usually contain the tags in tables so that a system can implement MARC21, UK MARC or UNIMARC without special programming. So the influence of the outside world is minimal.

In this conference we are interested in libraries, archives and museums. These sectors also have their own practices for sharing data. However, libraries are different because every library's collection overlaps with every other's whereas archives and museums generally have unique items. There have

been attempts to develop MARC based formats for archives though not as far as I know for museums. In the 1970s, MARC for Archive and Manuscript Control was developed though it was never incorporated into the consolidated US MARC when MARC incorporated the different library materials such as serials and maps in 1991. Another difference between archives and museums on the one hand and libraries on the other concerns the entities in which the user is interested. Libraries traditionally have lent books and allowed people to read in the library journal articles. The book or the journal article are the sought items and are the items found in catalogues or bibliographies. Archives on the other hand are sorted and arranged into collections. The records of these are organised by archival level. A record can be produced for the entire collection and individual records produced only for important elements within the collection. Ideally these records are linked to each other. MARC can cope with linked records but it is not the ideal record structure for that purpose. In consequence MARC has not been used very much for archives and in general, where it has been so used, the archives are part of a library, such as a university archive under the administration of the library. That having been said, a number of extra fields have been added to the UNIMARC format for archival material. The Permanent UNIMARC Committee is going to publish guidelines for the use of UNIMARC for recording archives and manuscripts (Weitz, 2006).

3. Developments in MARC formats

At the beginning of this paper it was mentioned that the number of MARC formats is decreasing. Additionally, their total usage is decreasing but only marginally. The many national MARC formats which were based very closely on LC MARC or UK MARC have in many countries been replaced by MARC21. UNIMARC has not been replaced to the same extent. UNIMARC is a truly international format and additionally is maintained by the Permanent UNIMARC Committee which is ultimately responsible through the UNIMARC Core Activity (based at the National Library of Portugal) to IFLA. It has features that make it hospitable to multi-language records. It has a more sophisticated linking mechanism and incorporates a choice of methodologies of producing multi-level records. It would be difficult to convert catalogues of records with these features to MARC21, though it is possible to transfer individual records between the two formats particularly one by one when records can be edited manually.

The other way in which usage of MARC is declining is in the use of the ISO 2709 record structure for the exchanging of records. The US MARC users community developed MARCXML, a version of the record structure of MARC in XML format. This is now being voted on as an international standard. Unfortunately the methodology used is not hospitable to the UNIMARC embedded fields, one methodology of record linking which was developed for UNIMARC. UNIMARC records not using this technique will have no problem in implementing MARCXML. Additionally there is nothing to prevent UNIMARC users from developing their own methodology of representing UNIMARC records in XML.

Recently, the author of this paper was involved in an effort to download MARC21 records from the UK TALIS database. We discovered that TALIS users can obtain such records, but in an XML format (Barton, 2006). Users of other software need MARC in its native record structure (ISO 2709). This is not yet available. A number of years ago the ISO 2709 format would have come first and not the XML structured records. However, XML is open to different interpretations within the scope of the standard in the way that the MARC format is not.

However, publishers and booksellers are still predominantly using traditional MARC records for libraries to purchase alongside the books they procure. And national libraries are still distributing the traditional MARC records to the libraries in their countries. The British Library continue to distribute records in UK MARC as well as in the recently adopted MARC21. Records produced since July 2004 have been created in MARC21 but are converted to UK MARC for those who require them. .

Other national libraries are producing records in UNIMARC and making them available for use in customers' catalogues.

4. Promoting UNIMARC with low cost software

CDS/ISIS is a software package developed in 1985 which has progressed from the original DOS version to a Windows version. This software is basically an information retrieval package and it can be

used for library catalogues (Buxton and Hopkinson, 1994). It is eminently suitable for holding MARC records and a few years ago an implementation of CDS/ISIS called IsisMarc was designed which included two sample databases, MARC21 and UNIMARC (UNESCO, 2006). The package implements the fields and subfields in the MARC format and can produce records for the users in any kind of display desired. Although the package is free-of-charge, it is as sophisticated in its print formatting of MARC records as many commercial library automation systems. IsisMarc which has been developed by Ernesto Spinak of Montevideo is built on the concepts introduced in UNIBASE developed by the National Library of Portugal in 1990 for the DOS version of CDS/ISIS.

The cultural heritage field is generally an area where institutions do not have much funding so CDS/ISIS is an excellent software package for those that have expertise but no funds. In the 1980s CDS/ISIS was installed in a small museum near Brighton, where a large number of photographs were held in a museum which was based in a tower which was one of a chain of towers built along the coast of England to protect the country from an invasion by Napoleon. CDS/ISIS was then not a very easy to use package but the IT specialist was a retired lecturer in computing science. At the Tate Gallery, as Ray (1994) describes, CDS/ISIS was set up as an interim system while the library was making the choice of which commercial system it should purchase. This meant that when the new system was introduced there were already some thousands of records to import, records which had been created over the previous two years, in CDS/ISIS and from which catalogue cards had been printed so that the card catalogue continued to be maintained until the new system was ready to go live. So with the right computing expertise, one can implement CDS/ISIS, and now in its Windows version it is much easier, but the web version again requires computing knowledge. However, computer scientists may not know about cataloguing, so a ready-made format with the bibliographic data elements which are needed to provide a good catalogue record are all that is needed to get a museum's library up and running and transferring from catalogue cards to catalogue database. Standards serve two purposes; to enable institutions to set up systems in a way that will enable interoperability. But even if an institution has no desire to interact with the rest of the world, use of standards will prevent them from having to reinvent the wheel. MARC is a standard which takes the bibliographic data elements found in traditional cataloguing library cataloguing rules such as the Anglo-American Cataloguing Code and defines a place for them in a record structure. Archives have been placed in this same structure though usually this is in those archives with a close relationship to libraries. Museums have different requirements for their records though museums do not generally use MARC for their museum collections though they do use them for their library collections.

4. The future of MARC formats

MARC has been criticised as having an antiquated record structure but this is now overcome since records can be exchanged in an XML record structure. This does have its disadvantage: the common format is no longer necessarily the one that is supported first as we saw from the account of the instance with TALIS where the less standard XML transfer methodology was the one given priority. In the computer world anything over a decade old or even less is viewed with suspicion. However we live in a transient age when most objects are expected to last only five years. In the cultural heritage field we are tasked with counteracting this and giving permanence to objects which otherwise would disappear. We need to keep records of items which are more than five years old and we cannot change our systems as other industries do every five years without a huge amount of effort. In the UK after 35 years of using UK MARC most libraries have decided to adopt MARC21, a format with very few differences but different enough for us to have to spend six weeks of this summer along with our library system supplier (and at the cost of £7,000 which we were told was a bargain) converting our data! So in our field we want stability of systems in a way that other businesses do not. There are millions of records in catalogues all over the world and we do not want to make these obsolete but wish them to be alive for information retrieval. Tools like CDS/ISIS and other library management systems depend on these stable formats. The ability to convert them to XML structure has given them a new lease of life. Do not believe anyone who tells you that the MARC format is dead!

References

- Barton, L. (2006). *Talislist account management meeting* (internal document). London: Middlesex University
- Buxton, A. & Hopkinson, A. (1994). *CDS/ISIS handbook*. London: LAPL

- Gredley, E. & Hopkinson, A. (1990). *Exchanging bibliographic data: MARC and other international formats*. Chicago: ALA; London: LAPL
- ISO. (2006). *Information and documentation: Format for data interchange (ISO 2709)* First published 1973, as *Format for the exchange of bibliographic data*.
- ISO. (2007). *Information and documentation: MarcXchange (ISO 25577)*
- Ray, L. (1994). Evaluating library automation systems for archive management *On Line Information 94: 18th International Online Information Meeting. London. December 1994, Proceedings*. London, On-line Information, (pp.555-563)
- UNESCO (2006). *CDS/ISIS database software*. Available online: <http://www.unesco.org/isis>. Accessed Sept. 30, 2006
- Weitz, J. (2006). The 17th Meeting of the Permanent UNIMARC Committee, 2006 March 22-24, National Library of Portugal, Lisbon. *International Cataloguing and Bibliographic Control* 36 (3)