

Harmonization of USMARC, CAN/MARC, and UKMARC

Sally H. McCallum

The Library of Congress, the National Library of Canada, and the British Library began discussing the harmonization of their respective MARC formats in 1994. The differences between USMARC and CAN/MARC were primarily in details rather than general specifications. Changes were made to CAN/MARC that eliminated many of the differences between CAN/MARC and the other two formats (USMARC and UKMARC). In addition, changes in USMARC that aligned USMARC and CAN/MARC were approved in 1997. The nature of the differences between UKMARC and CAN/MARC has necessitated a different process of harmonization. The differences between these two formats are many in extent, details, and approach to some requirements. Although total harmonization of USMARC-CAN/MARC with UKMARC is not feasible at this time, the British Library's program to add USMARC-CAN/MARC fields to UKMARC has increased the congruency of these formats. The National Library of Canada and the Library of Congress have begun to work on joint maintenance procedures and plan to have joint documentation.

Over the last four years the Library of Congress (LC) and the United States library community have been pursuing harmonization of the USMARC format with the formats used in Canada and the United Kingdom. At the outset of this project, harmonization of the USMARC, CAN/MARC, and UKMARC formats was regarded as highly possible because these formats were already similar in many respects. In the late 1960s, when USMARC was under development in the United States, under the direction of Henriette Avram at LC, consultation was carried out with colleagues in Canada and the United Kingdom. Among others, Ed Buchinski from Canada and Richard Coward from the United Kingdom visited LC and discussed the MARC Pilot Project and features of the emerging data exchange format. With similar cataloging traditions, Canadian and British librarians were very interested in these American developments.

This interaction had an impact on the final specification of the "MARC II" format that emerged at the end of the MARC Pilot Project, and colleagues in both Canada and the United Kingdom began follow-on efforts to develop versions of MARC in their own countries. Those formats today use many of the same tags for similar data: for example, IXX tags for main entries, tag 245 for title, tag 260 for imprint, tag 300 for collation, 5XX tags for notes, 6XX tags for subjects, 7XX tags for added entries, and tag 008 for coded data. In addition, developers in Canada made a commitment to minimize the divergence of CAN/MARC and USMARC, even at the detailed level.

With this degree of collaboration the logical question is, why did the formats diverge? Several general conditions "encouraged" differences in the formats. One such condition was national needs. For example, the National Library of Canada had to accommodate bilingualism from the outset in CAN/MARC and

the British Library centered its development around support for the British National Bibliography rather than a general catalog. Other factors were cataloging traditions and interrelationships that, while similar, differed in significant ways. It is difficult and disruptive for cataloging agencies to give up established practices; thus the format was adapted to their needs.

Another major factor that made format congruence less critical across country borders was the environment of the late 1960s and early 1970s. Because there were essentially no networks, files of records were moved from institution to institution on tapes sent through the mail. Tapes received could be put through a conversion program, which added very little extra time to the movement of a set of records from one agency to another. Also, there were no systems with large resources of records where libraries could obtain cataloging copy. Thus commonality of format was not an obvious critical need.

In the 1990s, a total transformation of the technical environment vastly increased the potential for international interchange and networking. Today there are enormous bibliographic record resources such as OCLC, RLIN, WLN, and AG Canada from which records are sought worldwide. Records are constantly retrieved and received from databases in other countries. The user's expectation and need is to view the record and be able to incorporate it immediately into a file and continue manipulations. Examples are the sets of records that come today via ftp from book vendors around the world, which might even arrive before the bibliographic items. LC's experience is that receipt of these records in USMARC saves resources and time in making them immediately available to acquisitions specialists. Another example is the Z39.50 Information Retrieval protocol, which supports responses in various record formats including MARC. MARC has been widely used in Z39.50 implementations because systems are often able to screen MARC records as if they came from the internal system, adding to the seamlessness of the retrieval. International connectivity has led to increased activity and expectations from international cooperative programs, making format conversions an obstacle to today's real time transfer needs.

Harmonization Meetings

With this setting, LC, the National Library of Canada, and the British Library began exploratory discussions in 1994 concerning harmonization of their respective formats in the present environment. During 1995 and 1996, staff from the three libraries held several meetings and discussed the costs, benefits, and possible impacts, and got down to a more detailed analysis of the changes that might be needed. Discussions of the benefits of harmonization raised the following points:

- cataloging copy would be easier to obtain and use, and more records would be available earlier;
- maintenance of costly conversion programs could be eliminated;
- earlier information from national bibliographies could benefit collection development;
- enhanced cross catalog searching would benefit reference and cataloging staff and users;
- the range of automated vendor cataloging and online catalog systems might increase, resulting in more choice in "size and shape" and competition that could stimulate innovation and provide price stability;
- system vendors would not have development costs associated with multifformat support; and
- format maintenance would be reduced when three separate formats were no longer maintained and documentation preparation shared.

The costs of changes associated with the harmonization effort were recognized as potentially serious. Any format changes cause the networks and utilities, which libraries depend on as cost-effective sources of records, to make costly system changes. Libraries that use the utilities ultimately pay for these system changes through their fees. Local systems, not just the large record stores, also need adjustment when the format changes, and libraries are dependent on system vendors to make those alterations. Some organizations might need retrospective record changes to their databases. The cost for retraining and new documentation for the changes, which will affect most format users, is often overlooked.

Thus it was clearly recognized by the three national libraries that the work on harmonization would require extensive consultation among all users of the three formats and careful procedures for making decisions and implementing any changes. Implementation would probably be a lengthy rather than brief process.

In 1995 the consultation process with constituencies began. The National Library of Canada held meetings with the Canadian Committee on MARC (CCM), looking at the differences between USMARC and CAN/MARC in detail. In the United States, the MARC Advisory Group and ALA's MARBI committee began investigating the possible changes and impacts. And a consultancy meeting and paper was prepared by the British Library for the UKMARC users to consider.

USMARC and CAN/MARC

While work proceeded simultaneously between USMARC and the other two formats, because USMARC and CAN/MARC had fewer differences, the project progressed

more quickly. The differences with CAN/MARC were primarily in details rather than general specifications. The National Library of Canada decided to make the first review and identified all the differences between the formats; it found more than 70 differences. The United States MARC Advisory Group viewed these differences in a June 1995 discussion paper, and the CCM held intensive meetings on them in June and October of 1995. After the CCM review, only 30 to 40 changes were presented to the United States and United Kingdom for consideration, as Canada had been able to eliminate the other differences. Another discussion paper went to the Machine Readable Bibliographic Information Committee (MARBI) in January 1996 with the reduced number of possible changes and a request that impact be carefully considered by USMARC format users. A formal proposal was made and considered in June 1996 and completed in January 1997.

Resolving Differences

An example of the types of differences and the process for resolution is illustrated by the proposal to align the Map 008 coded values for prime meridians. CAN/MARC had 38 values and used two character positions to code the meridians; USMARC had seven values and used only one character position. The overlap in values defined was five. Representatives of the Canadian map community and the United States community conferred on this sticky issue, and decided to take a fresh approach to the coding of prime meridian. They noted that actually any city can be a prime meridian for a map, so no list of values would be adequate. They also discussed other places in the record where the prime meridian is always recorded in textual form. The map specialists agreed that the coded prime meridian is not really needed and the formats could be harmonized on this point by making both obsolete. This characterized the "harmonious" give-and-take of the discussions over various differences.

There were several areas of differences between USMARC and CAN/MARC that required special attention. One dealt with archival material fields and the other with the Canadian bilingual requirements. The Canadian archival community developed new cataloging standards in the early 1990s. They are similar to, yet have some differences from, those used by the United States community. As a result of trying to accommodate the new rules, CAN/MARC had in 1995 adopted a number of new features for archivists that greatly increased the difference with USMARC. Archivists on both sides of the border took the initiative on this problem and organized a special meeting in Toronto of representatives from the Society of American Archivists' Committee on Archival Information and Exchange (SAA's CAIE) with Canadian counterparts to discuss these archive-related differences. As a result the differences were minimized and all

the remaining archival changes were eventually accepted in the harmonized format.

The bilingual requirements for Canada were a more difficult issue, consisting of two basic needs: (1) equivalent French and English headings were needed in bibliographic records and they had traditionally been carried in 9XX fields in CAN/MARC; and (2) a bilingual flag was required for authority records. The use of the 9XX fields for defined fields was especially problematic for USMARC users since reservation of the 9XX fields for local definition has long been a format principle. Most USMARC users have defined various 9XX fields for local needs. While CAN/MARC had the same philosophy on the 9XX fields, CAN/MARC had used them for equivalent headings partly because the equivalent heading fields were not needed in USMARC and they wanted to avoid any future clash with a non-9XX tag added to USMARC. After a great deal of discussion, an agreement was reached to include the CAN/MARC 9XX fields in an appendix of the harmonized format, clearly indicating that they are defined for a special situation and do not preclude other format users from having the same 9XX tags defined for local purposes in a different manner. In the future, when local tags are needed, this action will encourage users to consider using other tags than these, even if they do not expect any Canadian records. CAN/MARC users were satisfied with this solution.

The bilingual flag in authority records presented a different type of problem. Canada uses the bilingual flag to indicate whether a name or title authority is appropriate for use in a French language catalog, English language catalog, or can serve either. The United States, as well as USMARC users in other countries, might need to treat other languages as the Canadians do French and English. For example, in Switzerland, catalogs may be needed for French, English, Italian, or German, and in Finland, Finnish and Swedish are both official languages. In the United States, Spanish is more likely to be a language for which alternative language headings are needed. The 008 fixed field position used to indicate language(s) of catalogs for this information was very limited for accommodating additional languages. The United States community felt that if language were going to be accommodated, then a more universal technique was needed. Fortunately, it was determined that there already existed a subfield in the USMARC-CAN/MARC field 040 (Cataloging Source) subfield \$b (Language of Cataloging) where multiple languages could be recorded to indicate the languages for which the heading and cross reference structure might be valid. In addition, the character position in the 008, traditionally used by Canada for the English/French information, was also approved for the harmonized format, since it has been used in all Canadian authority records.

The changes to USMARC that aligned the two formats, USMARC and CAN/MARC, were finally approved in

January 1997, and the July 1997 update to USMARC included those changes. CAN/MARC has also distributed an update document that brings the two specifications into a separate-but-equal status. The changes to USMARC may be summarized as follows:

- There were minor changes to coded data—some elements were made obsolete and others enhanced.
- Field 016 for any National Library Record Identifier was added.
- Special changes were made to several fields to accommodate archival needs.
- Multilingual accommodation was improved.

USMARC and UKMARC

The process between UKMARC and USMARC has been quite different from that with CAN/MARC, primarily because of the nature of the differences between the formats. The foundations for harmonization with UKMARC are, of course, the common roots and the many similarities between the formats. USMARC and UKMARC have the same ISO 2709 (or ANSI/NISO Z39.2) structure, which governs the general layout such as the record leader, directory system, fields, indicators, and subfield codes. In places where choices could be made, such as the number of indicators per field, USMARC and UKMARC are certainly aligned. Also, as was noted earlier, many of the tags are the same or nearly so, and especially the 008 is used for the coded data in UKMARC.

The differences are many, however, in extent, detail, and approach to some requirements. UKMARC supports only one published format, for bibliographic data, and an internal British Library format for authority data, whereas USMARC-CAN/MARC has five related formats: bibliographic, holdings, community information, authority, and classification. Prior to the beginning of the harmonization process, UKMARC had never been filled out with a full complement of fields to fully accommodate nontextual material. In many fields UKMARC has different subfielding and data recording specifications from that of USMARC-CAN/MARC, and unfortunately they are the key fields that appear in all records, such as title, author, imprint, and collation. Another major difference is the way multilevel items are handled in UKMARC.

Before the harmonization process began, UKMARC had only a single 008 field with limited coded data for non-print material and many specialized fields were missing among the variable fields. The British Library therefore viewed harmonization as an opportunity to add specialized fields they needed without going over the same ground as the USMARC and CAN/MARC users had in order to develop the fields. This has been a successful area of collaboration. UKMARC was able to adopt a number of

USMARC-CAN/MARC coded and textual fields, adjusting them to fit the needs of the British Library systems.

Subfield Differences

The "ISBD area" of the record was a special problem on further investigation of the differences between USMARC and UKMARC. This includes the "transcription fields" such as title (245), edition (250), imprint (260), series statement (490), and collation (300). In those fields, UKMARC focuses on support of ISBD formulated data, providing subfield markers for each piece of ISBD punctuation and then omitting the punctuation from the field. The UKMARC approach creates the need for many more subfields in those fields—seven additional ones in the title 245 field, six in the imprint 260 field, eight in the collation 300 field—leading the British Library itself to take some liberties with subfield use in recent years.

In USMARC, rules other than ISBD/AACR have always needed to be accommodated, therefore subfield markers are generally used only to identify and delineate access points. Whatever punctuation an agency wants is included with the data. In the title-imprint-collation area, the information is not usually formulated for access, but is recorded as it appears on the item. A common convention (endorsed by AACR) is to repeat important names and places in the record in other fields in controlled access point form. An example is the author statement that appears in the \$c subfield in field 245, transcribed with other statements about joint authors, illustrators, etc. The 1XX and 7XX fields are used to give the names of these creators from the 245 in inverted and controlled form for access, but the title and subtitle are separate subfields as they are not repeated but access occurs from the 245 field. The differences in approach are illustrated in figure 1.

Adopting the UKMARC approach to the transcription fields would be a very costly change to USMARC-CAN/MARC users, with enormous impact. It would also be costly for the British Library to adopt the practices of USMARC-CAN/MARC, so it was agreed that the ISBD area content designation would not be useful to pursue for harmonization at the present time.

Multilevel Techniques

Another area that contained major differences in approach was multilevel records and linking. In USMARC-CAN/MARC each record has a target item that is described in the record, and whose title is in the 245 field. All hierarchically related items are linked to that item using the 4XX/8XX series fields or the 7XX fields. The USMARC-CAN/MARC format supports and AACR2 cataloging rules allow use of a single record for multiple volumes of a multi-volume work where the individual volumes have "weak"

| USMARC 245 Title field | UKMARC 245 Title field | UKMARC British Library 245 Title field |
|--|---|---|
| \$aWildlife of the world /\$cBent Jorgensen ; illustrated by Gabriele Pozzi ; translated from the German by Anthea Bell | \$aWildlife of the world\$oBent Jorgensen\$rillustrated by Gabriele Pozzi\$mtranslated from the German by Anthea Bell | \$aWildlife of the world\$rBent Jorgensen\$rillustrated by Gabriele Pozzi\$rtranslated from the German by Anthea Bell |
| USMARC subfields only the title for access/indexing and display; a 1XX field provides access by the author's name and 7XX fields provide access for the illustrator and translator, if desired by the cataloging agency. | UKMARC subfields for punctuation; as with USMARC, separate 1XX and 7XX access points are provided for the creators, if desired. | The British Library reduces subfields, thus ISBD punctuation is supplied by tracking the occurrences of the \$r subfield. |

Figure 1. Sample 245 Fields in USMARC and UKMARC

titles. In this case, the higher level (and stronger) title is in the 245 title field with the weak volume titles recorded in the 505 contents field.

In UKMARC whether the title of the individual item is "weak" or strong, the title of the item will be carried in a 248 field and titles higher in the hierarchy for the item title will be in other 248 fields with the highest in the 245 field. There can be other information pertaining to any of the title levels in the record, which is linked to the appropriate title through an indicator and subrecord technique.

The UKMARC technique is complex and not possible in USMARC-CAN/MARC without a great deal of change to basic areas of the format, affecting large numbers of records. So again, it was agreed that this area was not ready for harmonization at this time. There are several other significant differences between USMARC-CAN/MARC and UKMARC that make full alignment difficult.

Harmonization

Despite these problems, advances in harmonization have been made between USMARC-CAN/MARC and UKMARC. One seemingly minor but important change was to make the multiple surname indicator in personal name fields obsolete in both formats. Application of the indicator followed different rules in the two formats, resulting in a significant barrier to name authority cooperation between our countries. It was determined that very few organizations claimed to have indexing or sorting routines that used the indicator.

In addition, as noted above, the British Library has had a program to add a significant number of USMARC-CAN/MARC fields to UKMARC, going through its consultative process for several groups of fields over the last few years. This has increased the congruency of the two formats.

Next Steps with UKMARC

The major differences cited above and several other significant differences have, after careful analysis, made total har-

monization of USMARC-CAN/MARC with UKMARC not feasible at this time. Alignment has been boosted, however, by the adoption of USMARC-CAN/MARC fields that the British Library has undertaken in the last few years. With closer consultation on change proposals at the technical level, it is planned that the formats can perhaps move closer and certainly not continue to diverge in significant ways. Continuing work on joint programs, such as the successful NACO-sponsored authority cooperation, will be a catalyst for keeping the formats as close as possible. There is agreement that, if opportunities arise, specific fields and areas will be aligned in the future.

Next Steps with CAN/MARC

Maintenance Procedures

Now that full alignment has been achieved, the National Library of Canada and LC are looking at the next steps to sustain this progress. Already work has begun on joint maintenance procedures. For USMARC the preharmonization maintenance procedure was the following. Proposals and discussion papers originated from any users. They were developed into MARBI documents by the Network Development and MARC Standards Office at LC with review by LC staff. These proposals and discussion papers were then published on the Web for wider discussion and taken to a meeting of the ALA MARBI committee and the USMARC Advisory Group, held at each ALA meeting. Following the meeting LC made the final decision on the change, usually agreeing with the consensus of the MARC Advisory Group and vote of the MARBI committee. The National Library of Canada has a similar process between the National Library's MARC office and the Canadian Committee on MARC.

In the future, and as has been done for the last two USMARC Advisory Group meetings, LC, and the National Library of Canada will consult and both will conduct internal reviews while the proposals are being developed. They will begin their broader reviews with the advisory committees

and MARC users simultaneously when the proposals are posted on the Web. After separate national consultation processes, LC and the National Library of Canada will together make the final decision. Past experience has indicated that there would seldom be differences in consensus and certainly not critical ones, but in some cases a change might go back to the consultation stage if agreement cannot be reached or the National Library of Canada and LC see other reasons for further consideration of a change.

Because the format is used worldwide, international consultation will continue to be supported as it has been through the MARC listserv, which has more than 900 members, primarily from North America, and includes participants from over 35 other countries. An announcement goes to the list when the proposals and discussion papers are posted and available for electronic access. Anyone can comment over the listserv either before the Advisory Group meeting or by attending the open MARC Advisory Group/MARBI meetings during ALA.

Joint Documentation

LC and the National Library of Canada are planning to have joint documentation published in a variety of forms as USMARC is now: full version in print (and later on the Web), concise version on the Web and in print, and field list on the Web. The joint bibliographic format and the joint authority format were published, with the other three formats issued in 1999. For both CAN/MARC and USMARC, these will just be new editions of the same specification with a new name, as all changes to the bibliographic and authority formats needed to align USMARC and CAN/MARC have already been issued in a previous update.

The Network Development and MARC Standards Office will take advantage of the Web for some material previously found in the formats' printed appendixes. Dynamic appendixes would be better accessed from the Web than print. Also the descriptions in the format are being careful-

ly reviewed for United States-specific and Canada-specific language, so the new editions will reflect the harmonization. LC has agreed to take primary responsibility for the English edition of the format, while the National Library of Canada will take responsibility for a French version. This provides a convenient division of the work.

Implementation

LC and the National Library of Canada were not in a position to implement all of the format changes until LC completed installation of its new automated system in mid-1999 and the National Library of Canada updated its still relatively new AMICUS systems. Thus implementation was planned for 1999 to give vendors of systems and networks or utilities ample time to make local and network changes and carry out any data conversions.

Conclusion

The harmonization process has been difficult but successful in a number of ways. With Canada, we have achieved full harmonization. There are many cooperative programs between Canadian and American institutions that will benefit from a common format, making procedures simpler and more cost effective. The Library of Congress can cease to maintain its CAN/MARC to USMARC conversion programs. And the National Library of Canada and the Library of Congress MARC offices are looking forward to sharing documentation responsibilities. Although full harmonization was not possible with UKMARC, the opportunity has been taken to move the USMARC-CAN/MARC and UKMARC formats closer together. The differences between the formats are also better understood, which will enable the MARC offices in the three countries to identify opportunities to increase compatibility and minimize divergence.