What's Wrong With MARC?

Thomas Meehan, Head of Current Cataloguing, University College London

In June 2011, the RDA Test Coordinating Committee recommended that the US national libraries should implement RDA but subject to certain conditions, one of which was that they should "Demonstrate credible progress towards a replacement for MARC[1]. The Committee found that those trying out RDA "felt any benefits of RDA would be largely unrealized in a MARC environment[2]. This is not the first time that MARC has been seen to be in need of replacement. Some twelve years ago, Roy Tennant wrote a now often quoted article in Library Journal entitled *Marc Must Die*:

> The problems with MARC are serious and extensive, which is why a number of us are increasingly convinced that MARC has outlived its usefulness.[3]

This article will seek to address some of these problems, including most of those mentioned by Tennant in his article. However, when talking about MARC it is worth bearing in mind Tennant's words on the difficulties of isolating MARC:

> When I refer to MARC in this column, I am conflating several interrelated things. There are the MARC syntax, the MARC data elements, and the Anglo-American Cataloguing Rules (AACR).

To this must be added RDA. This article will refer to MARC21, which is the dominant, if still not universal standard. Some problems, particularly those to do with punctuation, are less pronounced in UKMARC, for example. To many non-cataloguers, including those involved in systems work, MARC *is* cataloguing. The existence of the underlying content standards- AARC2 and RDA- is something of a mystery and, given the closed nature of them both, not a mystery that is easily solved. Indeed, it is the need for systems to reach beyond the cataloguing unit, the OPAC, and the library itself that generate much of the need to reassess MARC. Even before the Bibliographic Framework Initiative proposed a linked data approach the RDA Test Coordinating Committee already recognised that the problems with MARC stood in the way of translating RDA data to a linked data environment: "MARC may hinder the separation of elements and ability to use URIs in a linked data environment."[4]

Brief History

It is worth reflecting briefly on how MARC came to be as it is. Before MARC, catalogue records were generally recorded in print form- e.g. an index card or dictionary catalogue- or perhaps on a microfilm.

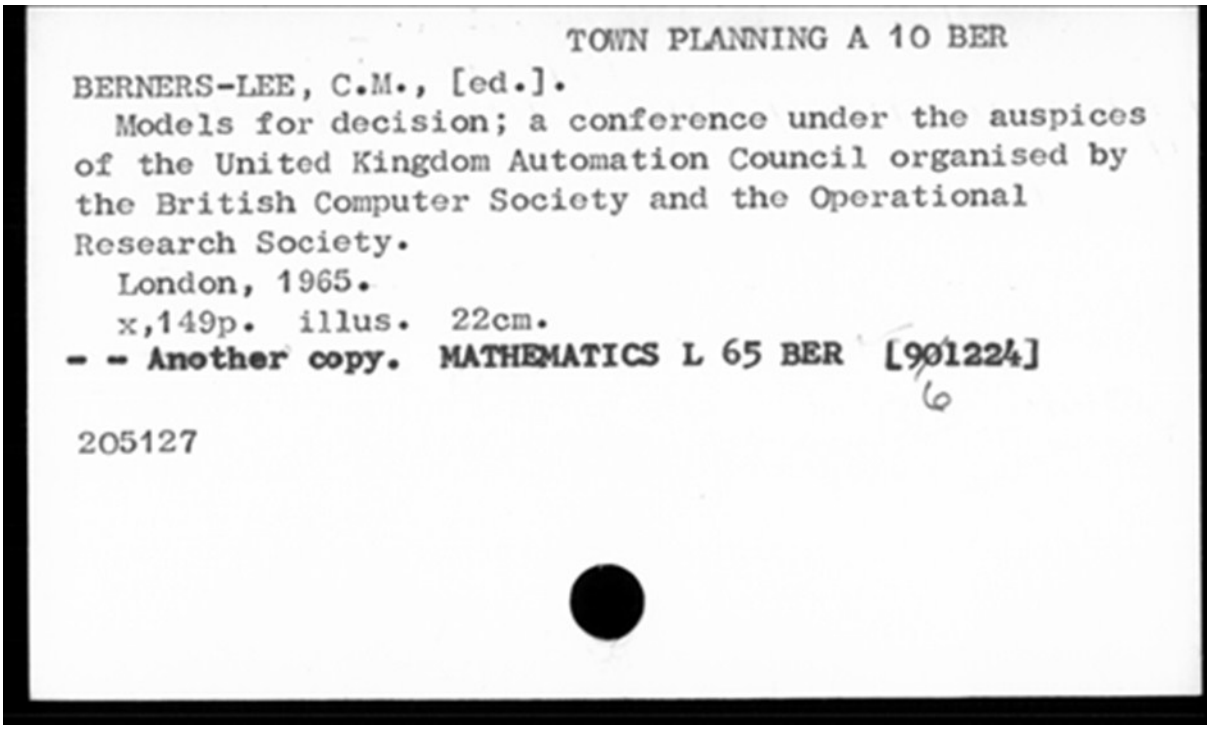
---

1 U.S. RDA Test Coordinating Committee. *Report and Recommendations of the U.S. RDA Test Coordinating Committee. Executive Summary*. 2011, P. 3. http://www.nlm.nih.gov/tsd/cataloging/RDA_report_executive_summary.pdf

2U.S. RDA Test Coordinating Committee. *Report and Recommendations of the U.S. RDA Test Coordinating Committee. Executive Summary*. 2011, P. 8. http://www.nlm.nih.gov/tsd/cataloging/RDA_report_executive_summary.pdf

3Tennant, Roy. *MARC Must Die*. 2002. http://lj.libraryjournal.com/2002/10/ljarchives/marc-must-die/

4U.S. RDA Test Coordinating Committee. *Report and Recommendations of the U.S. RDA Test Coordinating Committee. Executive Summary*. 2011, P. 8



5.

In either case, the point is that the record is designed to be read by eye by humans. The catalogue card is essentially text to be read in sequence and as a whole. The pre-AACR2 example above uses layout and some punctuation to separate out elements of data. Extracting out a line of data "London, 1965" only makes sense to us as we recognise "London" as a place and "1965" as a date. However it still needs the rest of the card for us to know which book was published in London in 1965.

MARC's basic function is to encode catalogue data, such as AACR2, onto a computer. By itself, a raw MARC record is essentially unreadable:

```
00788nam·a2200181·a·
450000100270000000500170002700800410004402400150008524502100010026000490031
300003200359504004100391650003300432700002300465710003900488710003000527710
04900557_UCL010000000000000000477125_20061112120300.0_850710s1965····enka·····
b····000·0·eng··_8·_ax280050495_00_aModels·for·decision·:_ba·conference·
under·the·auspices·of·the·United·Kingdom·Automation·Council·organised·by·the·
British·Computer·Society·and·the·Operational·Research·Society·/_cedited·by·
C.M.·Berners-Lee._··_aLondon·:_bEnglish·Universities·Press,_c1965._··_ax,·
149·p.·:_bill.·;_c23·cm._··_aIncludes·bibliographical·references._·
0_aDecision·making_vCongresses._1·_aBerners-Lee,·C.·M._2·_aUnited·Kingdom·
Automation·Council._2·_aBritish·Computer·Society._2·_aOperational·Research·
Society·(Great·Britain)__¶
```

[5] UCL Library Services. *Card Catalogue Online. Card Catalogue Single Card Display. Card number 887 of 931 from drawer: BEN-VENISTE to BERNHEIMER*. http://cardcat.ucl.ac.uk/cgi-bin/carddisplay.pl?card=887;drawer=13;max=931;ctype=C

In the above example, anything to do with the 245 field is marked in bold. You will notice that the 245 field tag is only mentioned in what is called the directory (the shorter bold sequence); the indicators and text of the field are carried separately. The subfield codes are marked with usually hidden characters, which have been changed to underscores above to make them visible. To extract any piece of data from this record, the whole record has to be disassembled, something that requires special software such as a Library Management System (LMS) or Marcedit.

A MARC display modified for human consumption, such as most of us are used to when editing MARC records or using MarcEdit, is more readable for cataloguers at least, although this is no longer a standard MARC record:

245 00 $a Models for decision :
$b a conference under the auspices of the United Kingdom Automation Council organised by the British Computer Society and the Operational Research Society /
$c edited by C.M. Berners-Lee.
260__ $a London :
$b English Universities Press,
$c 1965.
300__ $a x, 149 p. :
$b ill. ;
$c 23 cm.
504__ $a Includes bibliographical references.
700 1_ $a Berners-Lee, C. M.

The MARC record, in this case including all the ISBD punctuation, is ideal for carrying the index card's data for human consumption, whether on index cards, compiling a dictionary catalogue, or even a display screen for an OPAC, at least if you want to recreate the index card layout. The rest of this article will concentrate on some specific problems with MARC. In these cases, it helps to think of oneself as a computer or a computer programmer. One way to do this is to imagine writing out specific and foolproof instructions for finding specific pieces of information for someone who has no idea of the subtleties of MARC or cataloguing rules: how complex do they become and how many exceptions do you need to put in?

Finite Notation Problem

MARC uses field names that are three figure numbers, two numerical indicators, and single letter subfield codes. Like the Dewey Decimal (DDC) system, the notation itself can become constricting. For example, using the second indicator of the 650 field for the subject scheme…

650 _0 for LCSH
650 _1 for LC for Childrens
650 _2 for MeSH

…means that that there are only ten possible schemes that can be recorded. In the same way as DDC uses 0 to expand, MARC in this case resorts to a second indicator 7 and an additional subfield:

650 _7 Source specified in subfield $2

This solves the immediate problem while creating a degree of inelegance and forcing any system to look in two places. Another example that may be familiar is that of the 246 field for variant title. The first indicator controls the addition of a note or added entry, the second indicator details what kind of varying title is involved; if the title begins with an article, there is no way to show this as there is with the second indicator of the 240, 245, 730, or 740. There needs to be a third indicator and there is no way to add one:

246 184 $aThe title on the spine

Not that it hasn't been suggested!

Data in More than One Place

The solution to the limited number of subject schemes in the previous section forced data to be in more than one place. This is common in MARC. For instance, on an index card, this might be recorded in human-readable form in a textual note:

546 __ $a In English.

A computer readable form is provided in positions 35-37 of the MARC21 008 field where it would be recorded using a code: "eng". If the language of an item is more complicated- e.g. for a multilingual book- then these three characters are not sufficient and we run into the finite notation problem again, and the 040 field is used:

041 __ $a eng
$a ger
546 __ $a English and German parallel texts.

Furthermore, languages can appear in other places, such as part of uniform titles, this time not free text, but not using the same codes as the 008 or 041 either:

240 10 $l English

A computer programme trying to work out what language(s) a book was in or had been translated from could end up having to look in at least three MARC fields and understand two coding systems; the arguably richest field- the 546- would be effectively lost as it is too difficult to process English text (assuming the catalogue record is in English!) and the many possible wordings cataloguers may have used.

Double Encoding

One notable feature of MARC21, especially as compared with something like UKMARC, is that the data is effectively encoded twice: first using ISBD punctuation as mandated by AACR2, then again using MARC21 fields and subfields. RDA in theory only defines elements and is free from ISBD, but to enter RDA into MARC21 means putting the ISBD punctuation in. The following example uses ISBD punctuation only, i.e. no MARC:

Worcester : Edward Elgar, 1857-1934.

As pure text, this is confusing for the uninitiated and looks like biographical information about the composer Edward Elgar. In the context of a catalogue record display and in the right place, I know by the colon and the comma that "Worcester" is a place of publication, "Edward Elgar" is in fact a publisher, and "1857-1934" is the range of publication dates. Similarly, the coding alone of this 260 field makes these elements clear:

260 __ $a Worcester
$b Edward Elgar
$c 1857-1934.

Putting them both together is unnecessary:

260 __ $a Worcester :
$b Edward Elgar,
$c 1857-1934.

It makes public display on a traditional OPAC easier if the intention is to replicate the catalogue card, but this can cause significant problems if you want to isolate a piece of data.

Text, Not Data

For instance, what is the place of publication of the book in the foregoing example? In MARC, this should be a straightforward as the place of publication is recorded in the 260 field (at least in AACR2) subfield $a. This means the place of publication is:

> Worcester :

This is clearly not the case as the place is actually:

> Worcester

A computer programme wanting to display this information separately on a catalogue, or a list of places, or to index it, has to strip the space-colon away. If the record doesn't have a publisher or date there might even be commas or full-stops instead. If it's a cataloguer-supplied piece of data, there might also be square brackets to dispense with, so a computer programme now has to look out for any of the following to make sense of this information:

> Worcester :
> Worcester,
> Worcester.
> [Worcester] :
> [Worcester],
> [Worcester].

When that is done, the information is still not easy for a computer to understand in that "Worcester" is a human-readable, if standard English, string. However, under RDA, not counting spelling or transcription errors, this could also potentially be:

> Worcester, Worcestershire
> Worcester, Worcs
> Worcester, Worcestershire, England
> Worcester, England

(Not to mention the ones in South Africa, Massachusetts, New York, and elsewhere in the U.S.). Automatically establishing that these are the same thing is difficult, which is where a linked data approach is helpful. A URI such as *http://id.loc.gov/authorities/names/n81018214*
identifies the place as Worcester (in England) regardless of the transcribed form.

MARC stores a number of other elements that should be data as text instead. Sizes of books are stored with their units, and the units may vary; the 264 _4 subfield $c should store a copyright date but instead holds a string starting with a copyright symbol: to get the actual date, the copyright symbol needs to be removed.

Data Mixed Up

Different types of MARC data are frequently mixed up and take some disentangling. Consider the following:

245 00 $a Data on the web
$h [electronic resource] :
$b research and applications /
$c Antonis Bikakis, Adrian Giurca (eds.).

The GMD in the 245 subfield $h is not a part of the title except that AACR2 deals with it as such. If a library chose not to display the GMD- perhaps preferring icons instead- then removing the $h is not as simple as it could be as punctuation is removed at the same time:

245 00 $a Data on the web
$b research and applications /
$c Antonis Bikakis, Adrian Giurca (eds.).

The 245 subfield $c is defined in MARC as the "Statement of responsibility, etc. ". In the following example, the "etc." includes two parallel statements of responsibility, a title proper, and a subtitle:

245 10 $a Enduring resistance :
$b cultural theory after Derrida /
$c edited by Sjef Houppermans, Rico Sneller, Peter van Zilfhout. = La résistance persévère : la théorie de la culture (d')aprés Derrida / edité par Sjef Houppermans, Rico Sneller, Peter van Zilfhout.

It is still useful for displaying a catalogue record, but it is now hard to isolate any statements of responsibility, not to mention the parallel title, which is essentially lost. The 245 $c effectively loses any meaning.

Changing Text As Primary Key

Current authority control practice using MARC requires a match between the form of the heading in a bibliographic record and the corresponding authority record. This requires considerable maintenance work and potential for broken links. The heading for Oscar Niemeyer was previously:

Niemeyer, Oscar, 1907-

But changed on his death to:

Niemeyer, Oscar, 1907-2012.

This requires both the authority and bibliographical records to be changed. With record sharing and outsourcing of either type of record, this effectively needs changing on all systems at all libraries. When computers are doing the indexing and matching, even small details can break links between records:

Niemeyer, Oscar, 1907-2012.
Niemeyer, Oscar, 1907-2012

Using URIs such as *http://id.loc.gov/authorities/names/n82013357* for Niemeyer avoids such problems. Forms of name, death dates, or any other information can be added or changed without breaking the link itself.

Expressing Relationships

RDA has at least improved MARC's handling of relationships. In the following example, it is impossible to know, at least in isolation, what C.M. Berners-Lee's role is with regard to the item being catalogued:

700 1_ $a Berners-Lee, C. M.

He could be a second author, an editor, a translator, or perhaps even the subject of a Festschrift. Relationship designators will help to ease this:

700 1_ $a Berners-Lee, C. M.,
$e editor of compilation.

Although, note that the comma before the $e is another defiantly textual addition. Remove the $e as we did with the GMD above, and we are left with this:

700 1_ $a Berners-Lee, C. M.,

The comma now looks like part of the name!

Handling of RDA

Of course, what has really accelerated the proposed demise of MARC is the implementation of RDA. Some of these issues, such as the greater importance of relationships in RDA have already been mentioned. MARC is a record-based format. This makes it difficult to dissentangle data for the Manisfestation, Expression, and Work elements of an RDA description. Implementation of RDA along the lines of the Joint Steering Committee (JSC)'s Scenario 1 would require separate records for each WEMI entity[6]. The MARC record is based on the Manifestation. Expression-level data is found mixed in the MARC record. In the following simple RDA record, the Expression-level data is marked in bold:

245 00 $a Models for decision :
$b a conference under the auspices of the United Kingdom Automation Council organised by the British Computer Society and the Operational Research Society /
$c edited by C.M. Berners-Lee.
264 _1 $a London :
$b The English Universities Press Limited,
$c 1965.
300 __ $a x, 149 pages :
$b illustrations ;
$c 23 cm
336 __ $a text
$2 rdacontent
337 __ $a unmediated
$2 rdamedia
338 __ $a volume
$2 rdacarrier
504 __ $a Includes bibliographical references.
700 1_ $a Berners-Lee, C. M.,
$e editor of compilation.

[6] Joint Steering Committee for Revision of AACR2. *RDA Implementation Scenarios*. http://www.rda-jsc.org/docs/5editor2.pdf

It is difficult to express any relationships between the Expression and Manifestation here. Some data is even mixed up within MARC fields: in this case the illustrative material in 300 subfield $b. Some fields, such as the 700, are not even identifiable as any Work or Expression elements until they are followed by a relationship designator: “author” would make it a Work element, “editor of compilation” shows it is an Expression element.

Who uses MARC?

Only libraries use MARC. When MARC’s use is restricted to creating index cards or compiling a catalogue for internal use, this might be fine, but it imposes restrictions on how libraries can interact with other organisations. By continuing to use MARC, libraries are tied to library-specific systems and processes. For instance, it would be difficult to move beyond a library management system for metadata management and editing, whereas there are tools for dealing with standards like RDF which are not so restricted to libraries. The restriction of MARC to the library community also makes it harder for non-librarians to use and benefit from the rich data libraries hold and continue to create.

Even within the library, MARC is by no means universal. It is as noteworthy that repository databases are not using MARC as it is that AACR2 was not the obvious method of recording repository metadata. Discovery systems like Primo also bring in a great range of metadata to work alongside the catalogue’s MARC records; MARC records in Primo are also converted to an internal format for display and indexing.

Conclusion: the Purpose of MARC

If MARC is not used for display- and arguably it hasn’t for some time given the line-by-line display favoured by online catalogues over the card index display- then what is its purpose? Cataloguers interact with MARC most often through input interfaces. Why this happens, though, is not obvious. The rules we have used for years- AACR2 and to a greater extent RDA- have elements which are then encoded in MARC fields and subfields. For example, the title proper is entered into the 245 subfield $a; the publisher in 260 subfield $b. Why is the cataloguer faced with an extra layer of numbers and codes to learn, especially as they are then expected to impose ISBD punctuation as well? It is perfectly possible to create a system which asks the user to fill in these elements instead of MARC codes. We use systems like this all the time on the web in various web forms where the underlying structure of the database is hidden from us. A practical example is the software developed by the company The MARC of Quality called RIMMF (RDA in Many Metadata Formats)[7]. This allows the cataloguer to input RDA elements directly without using MARC or ISBD, as in the following screenshot:

[7]The MARC of Quality. *RIMMF2*. http://www.marcofquality.com/

Element Label	Text	RDA Rule	AAP
Manifestation			
RIMMF Identifier †	tmq-demo-5	no rule	
Title		2.3	
Title Proper	MARC21 for everyone	2.3.2	☑
Other Title Information	a practical guide	2.3.4	☐
Statement of Responsibility		2.4	
Statement of Responsibility Relating to Title Proper	Deborah A. Fritz, Richard J. Fritz	2.4.2	
Publication Statement		2.8	
Place of Publication	Chicago	2.8.2	☐
Publisher's Name	American Library Association	2.8.4	☑
Date of Publication	2003	2.8.6	☑
Copyright Date	©2003	2.11	☐
Mode of Issuance	single unit	2.13	☐
Identifier for the Manifestation	083890842X	2.15	
Media Type	unmediated	3.2	☐
Carrier Type	volume	3.3	☑
Number of units †	xvi, 188	no rule	☐
Extent	pages	3.4	☐
Dimensions	28 cm	3.5	☐
Composite Key †	**MARC21 for everyone. American Library Association. 2003. volume**		
Expression Manifested		17.1	
RIMMF Identifier †	**Fritz, Deborah A. MARC21 for everyone. Text ‹tmq-demo-4›**	17.4.2.1	

8

Here the cataloguer can concentrate on applying the rules and structure of the data. It is not necessary at the cataloguing stage to know how this is stored internally- in a relational database, raw MARC files, a linked data triplestore, or some other means. One reason why MARC input displays may have persisted is that, as Rob Styles once remarked, MARC has become the “lingua franca of library cataloguing”.

We have already discussed how MARC, especially as marked up with ISBD, can make manipulation of data difficult. For instance, indexing authors with relationship designators becomes more tricky if you have to worry about the comma that may be present (but isn’t always) before a subfield $e. Terry Reese’s MarcEdit software allows cataloguers to manipulate MARC records by converting them to a text format that is easy to deal with, i.e. that isn’t MARC. It converts them back so the library system can understand them again.

The real strength of MARC is as a medium of exchange and distribution. It is compact and all major library management systems understand it. There are systems and standards, such as z39.50, that make exchanging records straightforward. However, the system works mostly on the principle of search and delivery of records. Linked data offers the possibility of publishing data more openly for wider consumption using HTTP.

MARC has been a very successful, long-lasting, and efficient servant of library data, getting library catalogue data quickly onto computer systems for local use, distribution, and sharing. However, technology standards, and even cataloguers’ own rules, are moving beyond what MARC is capable of. Its narrow library focus, record-based structure, and reliance on text-based cataloguing are limiting what libraries can do with the rich and high quality bibliographic data cataloguers have been amassing for decades and continue to produce.

[8] Carty, Celine. *RIMMF, or, Cataloguing Without MARC. Slide 7*. http://www.cilip.org.uk/sites/default/files/documents/Celine%20Carty%20-%20RIMMF%2C%20or%2C%20Cataloguing%20Without%20MARC.pptx

References

Carty, Celine. *RIMMF, or, Cataloguing Without MARC. Slide 7*. http://www.cilip.org.uk/sites/default/files/documents/Celine%20Carty%20-%20RIMMF%2C%20or%2C%20Cataloguing%20Without%20MARC.pptx

Joint Steering Committee for Revision of AACR2. *RDA Implementation Scenarios*. http://www.rda-jsc.org/docs/5editor2.pdf

Library of Congress. Network Development and MARC Standards Office. *MARC Standards*. http://www.loc.gov/marc/

Tennant, Roy. *MARC Must Die*. 2002. http://lj.libraryjournal.com/2002/10/ljarchives/marc-must-die/

UCL Library Services. *Card Catalogue Online. Card Catalogue Single Card Display. Card number 887 of 931 from drawer: BENVENISTE to BERNHEIMER*. http://cardcat.ucl.ac.uk/cgi-bin/carddisplay.pl?card=887;drawer=13;max=931;ctype=C

U.S. RDA Test Coordinating Committee. *Report and Recommendations of the U.S. RDA Test Coordinating Committee. Executive Summary*. 2011. http://www.nlm.nih.gov/tsd/cataloging/RDA_report_executive_summary.pdf