Universal Bibliographic Control and International MARC Core Programme

UNIMARC: An Introduction

Understanding the UNIMARC format

1. What is MARC ?

MARC is an acronym for *Ma*chine *R*eadable *C*atalogue or *C*ataloguing. This general description, however, is rather misleading as MARC is neither a kind of catalogue nor a method of cataloguing. In fact, MARC is a short and convenient term for assigning labels to each part of a catalogue record so that it can be handled by computers. While the MARC format was primarily designed to serve the needs of libraries, the concept has since been embraced by the wider information community as a convenient way of storing and exchanging bibliographic data.

The original MARC format was developed at the Library of Congress in 1965-6 leading to a pilot project, known as MARC I, which had the aim of investigating the feasibility of producing catalogue data in machine-readable form. Similar work was in progress in the United Kingdom where the Council of the British National Bibliography had set up the BNB MARC Project with the remit of examining the use of machine-readable data in producing the printed *British National Bibliography (BNB)*. These parallel developments led to Anglo-American cooperation on the MARC II project which was initiated in 1968. MARC II was to prove instrumental in defining the concept of MARC as a communication format.

MARC II established certain principles which have been followed consistently over the years. In general terms, the MARC communication format is intended to be:

- Hospitable to all kinds of library materials
- Sufficiently flexible for a variety of applications in addition to catalogue production
- Usable in a range of automated systems

Despite cooperation there emerged several versions, e.g. UKMARC, INTERMARC and USMARC, whose paths diverged owing to different national cataloguing practices and requirements. Since the early 1970s an extended family of more than 20 MARC formats has grown up. Differences in data content means that editing is required before records can be exchanged.

One solution to the problem of incompatibility was to create an international MARC format (UNIMARC) which would accept records created in any MARC format. So records in one MARC format could be converted into UNIMARC and then be converted into another MARC format. The intention was that each national agency would need to write only two programs - one to convert into

UNIMARC and one to convert from UNIMARC - instead of one program for each other MARC format, e.g. INTERMARC to UKMARC, USMARC to UKMARC etc.

So in 1977 the International Federation of Library Associations and Institutes (IFLA) published *UNIMARC : Universal MARC format*, stating that "The primary purpose of UNIMARC is to facilitate the international exchange of data in machine-readable form between national bibliographic agencies". This was followed by a second edition in 1980 and a *UNIMARC Handbook* in 1983. All focussed primarily on the cataloguing of monographs and serials and took advantage of international progress towards the standardisation of bibliographic information reflected in the International Standard Bibliographic Descriptions (ISBDs).

In the mid-1980s it was seen necessary to expand UNIMARC to cover documents other than monographs and serials. So a new description of the format - the *UNIMARC Manual* - was produced in 1987. By this time UNIMARC had been adopted by several bibliographic agencies as their in-house format. So the statement of purpose was amended to include "UNIMARC may also be used as a model for the development of new machine-readable bibliographic formats".

Developments did not stop there. Increasingly a new kind of format - an authorities format - was used. Previously agencies had entered an author's name into the bibliographic format as many times as there were documents associated with him or her. With the new system they created a single authoritative form of the name (with references) in the authorities file; the record control number for this name was the only item included in the bibliographic file. The user would still see the name in the bibliographic record, however, as the computer could import it from the authorities file at a convenient time.

So in 1991 *UNIMARC/Authorities* was published.

By that year users of UNIMARC realised that the occasional rewriting of manuals was not enough. What was needed was continuous maintenance. The Permanent UNIMARC Committee came into being that year, charged with regularly supervising the development of the format. In maintaining the format, care is taken to make changes upwardly compatible, i.e. no records created before a change would be invalid after it.

The latest development in the format has come about because of the requirement of European Community countries to produce unified specialised catalogues of their records. In order to produce such unified catalogues they had to adopt a common format for them - UNIMARC.

2. The UNIMARC format

The UNIMARC format, like any other version of MARC, involves three elements of the bibliographic record:

Record structure

The record structure is designed to control the representation of data by storing it in the form of strings of characters known as *fields*.

All data in the record must be stored using one or more character sets. Since computers can store and manipulate only numbers, each symbol, alphabetical character etc. is assigned a number following the rules of a particular character set. For example, one character set assigns the number '75' to 'K'. UNIMARC allows the use of certain character sets, approved by the International Organization for Standardization (ISO).

The record structure established by UNIMARC is an implementation of the relevant standard: *Format for bibliographic information interchange on magnetic tape (ISO 2709-1981)*. This structure utilises record labels and directories. As few users need concern themselves with such items, the description below covers the way a cataloguer sees the record.

Content designation

Certain conventions are followed in order to identify the data elements within records. Such elements which include author, title and subject access are further characterised where necessary. This supports the manipulation of the data for a variety of purposes:

- To provide multiple access points for searching,
- To allow the typography and layout to be varied,
- To permit certain elements of the record to be omitted where this is required.

For an example of such manipulation, see the "Displaying citations" section later in this document.

In addition, UNIMARC records may be formatted for visual display on a VDU, for output on CD-ROM or fiche and for printing out as hard copy.

In general, UNIMARC provides content designation only for data which is applicable to all copies of a work. However, information which applies only to some copies (or even a single copy) of a work may be of interest beyond the holding institution. In such cases UNIMARC assigns specific fields for such details. These fields are also available for cases where the information is for in-house purposes only.

Data content

The content is the data which is stored in the fields within the record. Data can be coded data or bibliographic data.

- Coded data is used to represent such items as control numbers, publication type, and main language of text. There is also provision for the characteristics of special types of items such as printed music.

- Bibliographic data is defined by reference to the International Standard Bibliographic Description for that type of material. In addition, each record can carry a class number and subject headings.

The purpose of UNIMARC, therefore, is to facilitate the description, retrieval and control of bibliographic items. This is achieved by providing a structure for recording bibliographic information which is input by reference to international standards.

3. The role of UNIMARC

Initially, UNIMARC was used for the exchange of records on magnetic tape but has since been adapted for use in a variety of exchange and processing environments.

The UNIMARC format is available to all agencies concerned with the exchange of bibliographic information. In practice, though, UNIMARC is orientated towards the requirements of libraries.

The fields, which are identified by three-character numeric tags, are arranged in functional blocks. These blocks organise the data according to its function in a traditional catalogue record. In the table below, fields 0-- - 1-- hold the coded data while fields 2-- - 8-- contain the bibliographic data:

Block	Example
0-- Identification block	010 International Standard Book Number
1-- Coded information block	101 Language of the work
2-- Descriptive information block	205 Edition statement
3-- Notes block	336 Type of computer file note
4-- Linking entry block	452 Edition in a different medium
5-- Related title block	516 Spine title
6-- Subject analysis block	676 Dewey Decimal Classification
7-- Intellectual responsibility block	700 Personal name - primary intellectual responsibility
8-- International use block	801 Originating source
9 - Reserved for local use	

In addition to the 9-- block any other tag containing a 9 is available for local implementation.

The fields defined by UNIMARC provide for different kinds and levels of information. This can be shown by looking at a typical record in the UNIMARC format.

4. Anatomy of a UNIMARC record

Example: Alain-Fournier's novel "Le Grand Meaulnes", translated into English as "The lost domain".

001 0192122622@
010##$a0-19-212262-2$d£12.95@
020##aUSb59-12784@
020##aGBbb5920618@
100##$a19590202d1959####|||y0engy0103####ba@
1011#$aeng$cfre@
102##aGBben@
105##$aac######000ay@
2001#$a{NSB}The {NSE}lost domain$fAlain-Fournier$gtranslated from the French by Frank Davison$gafterword by John Fowles$gillustrated by Ian Beck@
210##$aOxford$cOxford University Press$d1959@
215##$aix,298p,10 leaves of plates$cill, col.port$d23cm@
311##$aTranslation of: Le Grand Meaulnes. Paris : Emile-Paul, 1913@
454#1$1001db140203$150010$a{NSB}Le {NSE}Grand Meaulnes$1700#0$aAlain-Fournier$f1886- 1914$1210##$aParis$cEmile-Paul$d1913@
50010$a{NSB}Le {NSE}Grand Meaulnes$mEnglish@
606##$aFrench fiction$2lc@
676##$a843/.912$v19@
680##$aPQ2611.O85@
700#0$aAlain-Fournier,$f1886-1914@
702#1$aDavison,$bFrank@
801#0aUKbWE/N0A$c19590202$gAACR2@
98700$aNov.1959/209@

Before looking at the MARC fields in detail, it is important to understand how the coding defines the data content. This is done by means of *field enumerators* which are composed of the following elements:

- Tag: a three digit number, e.g. 700, which defines the type of bibliographic data.

- Indicators: two single digit numbers right after the tag, eg 700#0 , that either refine the field definition or show how the field should be treated for catalogue production, e.g. by signalling that a note should be made. Blanks are shown by the hash sign # to distinguish them from a space.

- Within each field, data is coded into one or more subfields, e.g. 700#0$a ... $b ..., etc., according to the kind or function of the information. The effect of the subfield coding is to refine further the definition of the data for computer processing. The subfield identifiers consist of a special character, represented by a $ in the examples, and a lower case alphabetic character or a number 0-9.

- Each field is followed by an end of field mark represented by the "at" sign @ in the examples.

Where necessary, # has been used to represent a blank.

The role of the field enumerators is explained with reference to the preceding record.

Details

001 0192122622@

001 (the record identifier) is a unique number or combination of letters and numbers that serves to identify the record in a file. It is almost the only field not to have indicators.

010##$a0-19-212262-2$d£12.95@
This field holds the ISBN ($a) and price ($d). If the item were also available as a paperback then a second 010 field would hold a second ISBN and price, with a $bPbk to show that the data referred to the paperback version. Since the indicators serve no special function, they are both blank.

020##aUSb59-12784@
020##aGBbb5920618@
These fields show that the item has been assigned a national bibliography number that is unique to that bibliography. The first (59-12784) has been assigned by the Library of Congress so the country code ($a) is US.

```
                1         2         3
        012345678901234567890123456789012345
100##   19590202d1959####|||y0engy0103####ba@
```

This is a fixed-length field where the meaning of a character is dependent on its position. Hence the transcription above is preceded by numbers showing the character positions (cp).

cp 0-7	show that the record was added to the file in 1959 February 2nd.
cp 8-12	show that the record is a monograph (d) published in 1959.
cp 13-16	are not used for monographs and so contain blanks.
cp 17-19	can include codes to show the intended audience, e.g. for children aged 9-14 (code d). Here the positions hold the fill character (|) showing that the information is not provided by the agency.
cp 20	indicates whether the item is a government publication. "y" means "no"
cp 21	indicates whether the record has been modified (1) or not (0). This means that changes have had to be made because of the character sets used cannot express certain special characters and transliteration has been used. For example, "where there is a heart symbol in the title - I [heart symbol] Paris

	- the title may have to be transcribed as "I [love] Paris".
cp 22-24	gives the language of cataloguing, in this case English.
cp 25	gives the transliteration code. "y" means that no transliteration has been used
cp 26-33	contain codes for the character sets used. 01 shows that the basic Latin set has been used, 03 covers the extended Latin set. The four blanks show that no additional sets have been used.
cp 34-35	gives the script of the title as given in the item (which may not be the same as that of the record's title). In this case it is Latin.

1011#$aeng$cfre@
This field gives details of the languages involved. The value of the first indicator (1) shows that the item is a translation. It is a translation into English ($a) from French ($c).

102##aGBben@
The country of publication field contains the ISO standard code for the country (Great Britain). The code for the locality within the country ("en" for England) is peculiar to the cataloguing agency, as there is no ISO standard for this.

105##$aac######000ay@
The coded data field for books and other monographic publications is almost as complex as field 100. It indicates what illustrations the item has, whether it is a biography etc.

2001#$a{NSB}The {NSE}lost domain$fAlain-Fournier$gtranslated from the French by Frank Davison$gafterword by John Fowles$gillustrated by Ian Beck@
The title field has first indicator "1", showing that the title is significant: in a browsable list - printed microform or electronic - there would be an added entry filing at "Lost domain". To avoid having the title file in the "T" part of an alphabetical listing, the "The[space]" is preceded and succeeded by a special character (represented here by {NSB}and {NSE}) to show where the non-sorting characters begin and where they end. These characters would not appear in any listing or on a reader's computer screen.
$f indicates the first statement of responsibility; subsequent statements are coded $g.

210##$aOxford$cOxford University Press$d1959@
Details of publication, distribution etc. can be quite complex. In this case only three subfields are needed: $a for place, $c for publisher and $d for date. Cases where more detail was needed would include items by a minor publisher whose address could not readily be obtained; the address would appear in subfield $b.

215##$aix,298p,10 leaves of plates$cill, col.port$d23cm@
This field holds the physical description.

311##$aTranslation of: Le Grand Meaulnes. Paris : Emile-Paul, 1913@

This is a Note pertaining to linking fields and is produced by the computer rather than input by the cataloguer. For details see field 454 below.

454#1$1001db140203$150010$a{NSB}Le {NSE}Grand Meaulnes$1700#0$aAlain-Fournier$f1886- 1914$1210##$aParis$cEmile-Paul$d1913@

This is a linking field, in this case pointing to the original of which the item is a translation. Each $1 (one) subfield holds the contents of a field: 001 Record identifier, 500 Uniform title, 700 author, 210 publication details. In sophisticated systems only the 001 would be needed: the reader could use it to call up the record for the original French item. For use in other systems and where the file does not contain a record for the original, enough detail is given to identify the item.
Second indicator "1" means "Make a note". At a convenient time the computer would produce 311 Translation of: [because this is what tag 454 means] and add the title and publication details.

50010$a{NSB}Le {NSE}Grand Meaulnes$mEnglish@

A uniform title. The first indicator serves the same function as that for the 200 field. The $m (language) subfield allows the catalogue to group together all English translations of this work.

606##$aFrench fiction$2lc

This field holds Topical name [i.e. Thing] as subject. The $2 code shows that the thesaurus used is the list of Library of Congress Subject Headings.

676##$a843/.912$v19@

This field holds a Dewey Decimal Classification number from the 19th edition of DDC (hence the $v). The "/" is a "prime mark": libraries with little French literature could drop it and everything beyond it - giving a class number of "843".

680##$aPQ2611.O85@

This is the Library of Congress class number.

700#0$aAlain-Fournier,$f1886-1914@

The tag means "Personal name - primary intellectual responsibility". The second indicator is 0 as this is a name entered under forename rather than under surname. The $f subfield holds the author dates of birth and death.

702#1$aDavison,$bFrank@

The tag means "Personal name - secondary intellectual responsibility". The second indicator is 1 as this is a name entered under surname. The forename is in the $b subfield

801#0aGBbWE/N0A$c19590202$gAACR2@

This "Originating source" field gives details of the creation of the record. This is especially useful for union catalogues, which by definition contain records from different agencies. Subfield $a holds the code for the country and $b the code for the agency creating the record. $c is the date of creation and $g holds details of the cataloguing code used - in this case the *Anglo-American cataloguing*

rules, 2nd edition.

98700$aNov.1959/209@
987 is a local field. In this case it contains the shelf-mark.

For an example of this record without the fields, subfields etc., see Displaying citations below.

5. Putting UNIMARC to work

Bibliographic records in the UNIMARC format are designed for use in automated library systems. Depending on the versatility of the system a range of related functions can be supported by manipulating the data. Two such functions are information retrieval and displaying citations.

Information retrieval

In the UNIMARC format each data element is identified for the purposes of information retrieval. Using computer software, it is possible to search on most of the MARC fields and subfields in the record. For example:

- Keywords (i.e. significant words)
- Subject headings
- Author
- Name, topical name, geographical name as subject
- Title and series title
- Standard numbers (ISBN, ISSN etc.) and numbers assigned by agencies (a national bibliographic agency, a government printing office etc.)
- Classification numbers
- Publisher
- Publication date and type
- Acronyms formed from name and title words
- Coded items. For example, FICTION would select the above record because field 105 character position 11 codes it as fiction.

While each record in the UNIMARC format is a discrete entity, a catalogue consisting of many such records becomes a database enhanced with the capacity to respond to highly specific or comprehensive search strategies. The range of search options will, of course, depend on the kind of software employed.

Displaying citations

UNIMARC offers a choice of formats for displaying records. Naturally, readers will not want to consult the full MARC record simply because the format is intended not for human perusal but for processing by computer. A sympathetic display for use by readers is the Catalogue card format:

843.912 (DC19)

```
Alain Fournier, 1886-1914
[Le Grand Meaulnes. English]. The lost domain / Alain-Fournier;
translated from the French

by
Frank
Davison; afterword by John Fowles; illustrated by Ian Beck. - Oxford:
Oxford University Press, 1959.
- ix,298p,10 leaves of plates; ill, col.port; 23cm
Translation of: Le Grand Meaulnes. Paris : Emile-Paul, 1913
ISBN 0-19-212262-2: £12.95
1.Ti 2.The lost domain 3.Davison, Frank 4.French fiction  B59-20618
Pressmark: Nov.1959/209
```

This citation represents a card in the classified sequence, which will be filed under 843.912. The second to last line shows the other headings under which the record will appear in a library catalogue, and the national bibliography number. The first tracing is an abbreviation for "Title". In this particular layout the author's name appears on a separate line above the title etc. With the exception of 7-- fields (which present problems and so need the cataloguer to put in the punctuation) most of the punctuation is supplied by the computer as it translates subfield codes into punctuation and typeface.
This is just one possible layout: local practice dictates how the entry will appear.

6. Maintaining UNIMARC

The interests of users of UNIMARC records are represented by the Permanent UNIMARC Committee (PUC), which plays an important role by acting as a focus for user views and reactions when amendments to UNIMARC are proposed. It does this on behalf of IFLA UBCIM, which is ultimately responsible for UNIMARC.

7. UNIMARC Authorities

The UNIMARC Authorities format is designed to allow an agency to hold in one place the authoritative form of name of an author, corporate body name etc., together with references from other forms of name. Such data is linked to a bibliographic record by subfield $3 (Authority record number) in fields in the 7-- block of the bibliographic format.
The data can be embodied in the bibliographic record either at the time of creation or when a user views that record.
There are three types of authority record, coded in the record label as "x" (authority entry record), "y" (reference entry record) and "z" (general explanatory entry record).

Structure of the UNIMARC Authorities format

0--	Identification block	(as in the UNIMARC bibliographic format)
1--	Coded information block	(as in the UNIMARC bibliographic format)
2--	Heading block	
3--	Information note block	(as in the UNIMARC bibliographic format)
4--	See reference tracing block	
5--	See Also reference tracing block	
6--	Classification number block	(as in the UNIMARC bibliographic format)
7--	Linking heading block	(as in 4-- of the UNIMARC bibliographic format)
8--	Source information block	(as in the UNIMARC bibliographic format)
9--	National use block	(as in the UNIMARC bibliographic format)

Anatomy of UNIMARC authorities records

The following are two typical examples of simple records:

Type of record:	"x" (authority entry record)
001	A369875@
100##	$a19810715aengy0103####ba@
152##	$aAACR2@
200#1	$aStewart,$bJ.I.M.@
500#1	$0For works written under his real name see$aInnes,$bMichael $3B329638@
801#0	aUKbBL$c198110629@
810##	$aWho's who@

Type of record:	"x" (authority entry record)
001	B329638@
100##	$a19810716aengy0103####ba@
152##	$aAACR2@
200#1	$aInnes,$bMichael@
500#1	$0For works written under his pseudonym see $aStewart,$bJ.I.M. $3A369875@

801#0	aUKbBL$c198110629@
810##	$aWho's who@

As both are similar, only the second will be explained:

001 B329638@

001 is the record identifier

100## $a19810716aengy0103####ba@

The general processing data field has the same sort of structure as the bibliographic 100 field. It gives the date entered on the file (16th July 1981). The record is "a" established (i.e. not provisional). The language of cataloguing is English. The code "y" shows that no transliteration system was used. In the eight-position character set part "0103" shows that the basic Latin and the extended Latin sets were used; the four blanks show that no others were used. The script of cataloguing is the Latin alphabet ("ba").

152## $aAACR2@

152 is the Rules field. The record follows the *Anglo-American cataloguing rules*, 2nd edition. Such information is held in 801 $g in the bibliographic format.

200#1 $aInnes,$bMichael@

Since the field is for a personal author, the indicators and subfield codes follow field 700 in the bibliographic format.

500#1 $0For works written under his pseudonym see $aStewart,$bJ.I.M. $3A369875@

This is a "See also" reference for a personal author; so the indicators and subfield codes follow field 700 in the bibliographic format. This includes the $3, which holds the record number for the J.I.M Stewart heading. There is the addition of $0 (zero) for "Instruction phrase".

801#0 aUKbBL$c19810629@

Like the same field in the bibliographic format, this gives the country, institution and date of latest transaction for an originating agency (second indicator 0).

810## $aWho's who@

This field gives the source in which the data was found - in this case a biographical dictionary.

Other Authorities format fields

The other equivalents of the bibliographic 7-- fields are 210 Corporate or Meeting Name (as 71-), 215 Territorial or Geographic Name (as 71-), 220 Family Name (as 72-).
Titles are covered by 230 Title, 240 Name and Title, 245 Name and Collective Title.
As the above do not use fields from the 6-- Classification number block or 7-- Linking heading block, examples are given below.

Field 676 contains the Dewey Decimal classification number, as in the bibliographic format, but with the addition of subfield $c Explanatory terms.

250## $aParsley@
676## $a583.48$cBotany$v20@
676## $a641.655$cCooking$v20@

250 is used for Topical subjects as headings (like 606 in the bibliographic format). When a document on the herb is about parsley as a plant, the class number should be 583.48; when it is about parsley as food, the class number should be 641.655.

The 7-- block is used to hold a form of name in a different language or script.

001	234566
100	$a character positions 9-11= ger
215##	$aSchweiz@
715##	$8fre$aSuisse$3A234567@
715##	$8ita$aSvizzera$3A234568@

001	234567
100	$a character positions 9-11= fre
215##	$aSuisse@
715##	$8ger$aSchweiz$3A234566@
715##	$8ita$aSvizzera$3A234568@

In a library's German language catalogue the authoritative form for the geographic name "Switzerland" is the German one (A234566). But this entry is linked to similar ones in French and Italian. A reader searching for books on "Svizzera" will be shown those where the subject is "Schweiz".
In the library's French language catalogue the authoritative form is the French one, and the authority record will be the A234567 one.

In a library where one language predominates, the authoritative form will be in that language, with "See" references (415 fields) from the name in other languages

8. Short bibliography

8 Short bibliography

ISBD(G) : General International Standard for Bibliographic Description .... - Revised ed. ; prepared by the ISBD Review Committee Working Group set up by the IFLA Committee on Cataloguing.- München, London, New York, Paris : K G Saur, 1992.

ISO 1001-1986. File structure and labelling of magnetic tape for information interchange.

ISO 2709-1981. Format for bibliographic information interchange on magnetic tape.

UNIBASE : UNIMARC Demonstration Database. - Frankfurt : IFLA Universal Bibliographic Control and International MARC Programme, 1994.

UNIMARC in Theory and Practice : Proceedings of the Workshop Held in Sydney, Australia, 1988. - London : IFLA Universal Bibliographic Control and International MARC Programme, 1989. (Available from K G Saur).

UNIMARC/Authorities. - München, London, New York, Paris : K G Saur, 1991.

UNIMARC/CCF : Proceedings of the Workshop Held in Florence, 5-7 June 1991. - München, London, New York, Paris : K G Saur, 1993.

UNIMARC and CDS/ISIS: Proceedings of the Workshops Held in Budapest, 21-22 June 1993 and Barcelona, 26 August 1993.- München, London, New Providence, Paris : K G Saur, 1994.

UNIMARC Manual : Bibliographic Format. - 2nd ed. - München, London, New Providence, Paris : K G Saur, 1994.

9. Glossary

IFLA	International Federation of Library Associations and Institutions
ISBD	International Standard Bibliographic Description
ISBN	International Standard Book Number
ISO	International Organization for Standardization
MARC	Machine-readable catalogue (or cataloguing) format
PUC	Permanent UNIMARC Committee
UBCIM	Universal Bibliographic Control and International MARC Programme
UNIMARC	Universal MARC format

Latest Revision: *March 3, 1999*