



## EVOLUTION OF BIBLIOGRAPHIC STANDARDS AND THEIR COMPATIBILITY ISSUES WITH BIBLIOGRAPHIC DATABASE TOOLS

**Dr. Nanaji Shewale**

Librarian, Gokhale Institute of Politics and Economics (GIPE),  
B M C C Road, Deccan Gymkhana, Pune (India)

### ABSTRACT

This article maps the historical developments of the cataloguing systems from print catalogues to the development of the Machine Readable Catalogues (MARC). Theoretically, the very purpose of MARCs is the data exchange or resource sharing where one should follow the standards in terms of a) Cataloguing of documents b) Machine Readable Cataloguing c) Bibliographic Database Management Software. Author has made an effort to study various rules of cataloguing of documents, variety of MARCs and their data structures and tried to use the MARC structure with the database management system WINISIS and noted down some of the observations.

**KEYWORDS:** Bibliographic Standards, MARCs, Machine Readable Cataloguing (MARC), Library of Congress, USMARC, British Library, UKMARC, MARC21, WINISIS, AACR2, Bibliographic Database Management Systems, B-DBMS.

### 1. INTRODUCTION

The library catalogue whether printed or machine readable plays one of the vital roles in a library and information centre. The emergence of computer coupled with database management systems (DBMS) ushered in an era of Machine Readable Catalogues (MARC), and Online Public Access Catalogue (OPAC). The networking technology and the Internet accelerated the resource sharing activities across the libraries of the world by forging standards like Z39.50 and its offshoot like SRU/SRW, OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting) and Dublin Core. The substratum of all these activities remains and will remain to be cataloguing. In its new forms, whether it is Dublin Core or MARC21/MARCXML, cataloguing has greater role in managing the World Wide Web. Hence, the growing importance of Semantic Web.

For an automated library system, the current printed card catalogue, the shelf list, and authority files, etc. need to be converted in a form that a computer can read. When these files are put in a standardized manner, they become Machine Readable. Machine Readable Cataloguing (MARC) and this is the first step towards library automation, resource sharing and networking.

---

One of the outcomes of information technology in the field of library and information science is networking. The basic idea behind networking is resource sharing, be it, our good old co-operative cataloguing, centralized cataloguing, data exchange, etc. Data sharing or data exchange at national and international level are more popular and practical concepts. This expects that all the participating libraries follow some kind of standards. For successful data exchange or resource sharing, one should follow the standards in terms of a) Cataloguing of documents b) Machine Readable Cataloguing c) Bibliographic Database Management Software.

### **1.1 cataloguing of documents**

Catalogue is 'a list of library materials contained in a collection, a library, or a group of libraries, arranged according to some definite plan'<sup>1</sup>. The term '*definite plan*' in this definition tells the cataloguer what kind of description the catalogue card should give? How to present and distinguish such information? This calls for the cataloguing rules that are essential in terms of the

1. Choice of descriptive data elements for documents - like monographs, technical reports, journal articles, conference proceedings, etc.
2. Representation of the data in each of the descriptive elements
3. Order of the data with necessary punctuation marks.

One of the reasons for choosing AACR2 for the current study is its worldwide popularity among the librarians, which may be because of its regular updates and long-term existence.

### **1.2 Machine Readable Cataloguing (MARC)**

The importance of MARC standards cannot be overstated. If libraries and information centers are to share their cataloguing efforts, and take advantage of the available MARC databases, it is essential to adhere to MARC standards. The creation of MARC records would allow electronic access to bibliographic records, thus facilitating resource sharing using Z39.50 client/server. Some of its advantages are:

1. MARC prevents duplication of work and allows the libraries better sharing of bibliographic resources.
2. MARC enables libraries to acquire cataloguing data that is predictable and reliable.
3. MARC enables libraries to adopt commercial or open source library automation systems.
4. MARC allows libraries to replace one system (hardware or software) with another with the assurance that their data will still be compatible and reusable.

One of the major characteristics of MARC is to assign unique numeric codes called the content designators or simply 'tags' to each and every descriptive data element identified by the cataloguing code while describing the document. The advantage of using numeric codes is that of overcomes the language barrier in the data exchange. While designing a database, instead of using the field names in a particular language, the data elements like author, title, publisher, etc., are assigned numeric codes. The MARC record contains directories before each piece of bibliographic information to tell the cataloguer and the computer what data is to be found in a specific field. Normally the directory structure, also called record structure is

represented in ISO 2709 format, though using XML for representing MARC records is gaining more popularity. Perhaps, the ISO 2709 is nearing its extinction.

In 1950, Library of Congress planned for MARC and brought out the first MARC pilot project in 1963. Since then till 1995 various MARCs are designed and implemented. Some of the popular MARCs are USMARC, UKMARC, CAN/MARC, CATMARC, UNIMARC, etc. The Library of Congress and the National Library of Canada harmonized USMARC and CAN/MARC formats in a single edition in early 1999 under a new name: MARC21. The name points to the future as we moved into the *21st Century*. Many national and local libraries which were following various MARCs are willing to join the MARC21 bandwagon. In this work, MARC21 is chosen as it is used internationally and hopefully will continue to be there. Since the present study attempts to study the compatibility of MARC21 with AACR2R and WINISIS, it is proposed to cover only the tags related to bibliographic data elements leaving out the database management related tags.

### **1.3 Database Management Software - DBMS**

DBMS software packages are quite popular in the field of library and information science for the past few decades. These have been designed not only to store massive information but also to meet the different access needs of the users. DBMS have become very vital as they are more versatile and flexible than the traditional catalogue. Whether printed or digitally stored, the library catalogue has always been a database of records, i.e. database of catalogue entries. In its new format with MARC and online access, the card catalogue becomes an OPAC.

Many commercial and open source library automation software use Relational Database Management Systems (RDBMS), e.g. Virtua uses Oracle, Koha uses MYSQL. Although many of commercial library software packages use RDBMS, bibliographic software like CDS/ISIS is more suited to bibliographic data. The main characteristics of bibliographic data are a) they are repeatable and b) many fields are divided into subfields, to keep the logical unity of the data.

Some of the RDBMS packages offer arrays as a basic data and allow repeatable subfield to be stored. However, the performance drastically goes down while searching and retrieving data from array. In addition, RDBMS packages allow only one unit columns (fields) and subfields cannot be created.

As of today, there are many DBMS software, but WINISIS has been chosen for this study, as this database system is designed keeping in view the bibliographic data, its structure, etc. The WINISIS has most of the features that are required for any bibliographic database system like its print formatting language and Pascal interface, enabling one to print the catalogue. It allows use of tag codes instead of field names with variable field length, variable records length, repeatable fields, etc. It allows to export or import data in ISO 2709 format. In addition, WINISIS has been widely used by many library and information centers across the world. One of the features behind its popularity could be the cost. It is available free of cost from UNESCO.

## **2. Development of Bibliographic Standards and Machine Readable Catalogues**

### **2.1 USMARC**

During 1950s, Library of Congress (LC) came up with an idea of producing catalog data in machine-readable form. MARC – Machine Readable Cataloguing, was thus originally developed by the Library of Congress as a means of transforming the information on catalogue cards that

can be read, stored and processed by a computer. Library of Congress MARC was initially known as the 'LC MARC' format, later it was, referred to as USMARC. Both USMARC and LC MARC are often used interchangeably.

Library of Congress received a grant from Council on Library Resources (CLR) to study the feasibility of applying automated techniques to the operations of Library of Congress. The study published in 1963 by Gilbert<sup>2</sup> recommended developing the design and implementation procedures required to automate the cataloguing, searching, indexing, and document retrieval functions. Another study entitled: *The recording of Library of Congress Bibliographical Data in Machine-Readable Form; A Report* prepared for the Council of Library Resources, Inc. became the subject of a conference held at Library of Congress in January 1965.

From November 1966 - June 1967, 16 libraries received bibliographical data on magnetic tapes for experimentation. According to Crawford<sup>3</sup>, the pilot project format designed by Library of Congress, and subsequently called MARC I to distinguish it from its successor, was restricted to book materials, owing to the limited time for the analysis and design. The single fixed field of 108 characters (catering to Library of Congress card number, language, date of publication and other data useful for machine searching) and only 20 variable fields indicated that the format had been designed for monographs. Gredley<sup>4</sup> mentions that, the character set was also limited and the project recommended further work on an extended character set that use roman-alphabet languages.

Soon after LC MARC was launched, various countries started designing their own MARCs to automate the publication of their national bibliographies. These MARCs differed because of:

1. Different cataloging standards that determine the data content of a MARC record;
2. Different subject control and classification systems;
3. Different official languages sometimes spoken even within a single country; and
4. Different character sets and codes developed to cater for non-western languages and scripts such as, Indian, Chinese and Japanese.

## 2.2 UKMARC:

British National Bibliography (BNB) for UKMARC Pilot Project, and many other librarians, exerted a strong influence on the design of the MARC II format. The purpose was the standard communications format for interchange of bibliographic data. MARC II structure was capable of containing bibliographic information for books, serials, maps, music, journal articles, etc. MARC II format, therefore, reflected the consensus of a large number of librarians and systems personnel.

In December 1967, there was a discussion on MARC II format and proposed the character set. In 1968, a project report was published on experiences of Library of Congress and participating organizations, discussing the expanded character set, and the MARC II format itself. The expanded character set was an extended ASCII (American Standard Code for Information Interchange) included 46 additional character positions.

On the issue of varied languages, the Unicode<sup>5</sup> character encoding of 16 bits, provides a sufficient number of unique codes for the world's scripts and technical symbols in common use. In December 1993 MARBI (an American Library Association interdivisional committee on Representation in MACHine Readable Form of Bibliographic Information), distributed a

discussion paper that presented a mapping of characters in USMARC Latin character sets to the Universal Character Set (UCS) <sup>6</sup>.

In 1969, USASI (United States of America Standards Institute, later, ANSI, American National Standards Institute) Z39 Standards Committee dealing with machine input records, revised draft of LCMARC format for Bibliographic Information Interchange on Magnetic Tape, was approved by the membership of the Z39 Committee on 21 March 1969. In July 1970, ANSI approved the proposed format for Bibliographic Information Interchange on Magnetic Tape on May 13, 1970. This Standard was published in mid-September 1971.

According to Campos<sup>7</sup>, the National formats LCMARC and BNB MARC's efforts to achieve compatibility and harmonization carried out in 1970 gave birth to ISO 2709 standard.

Cayless<sup>8</sup> says, UK introduced MARC in Britain from January 1966 with the supply of forthcoming title entries of British National Bibliography (BNB) to Library of Congress as part of a Shared Cataloguing Programme. According to Gredley,<sup>4</sup> during the MARC Pilot Project, BNB commissioned a feasibility study to use machine-readable data for the production of the printed BNB and other library-related purposes. Thus the BNB MARC Project was launched in January 1966. MARC I pilot project, BNB and Library of Congress staff developed the MARC II format.

Library of Congress's MARC II<sup>9</sup> format was formally published in January 1968 and a preliminary version of BNB's MARC II was issued in June 1968, as MARC Record Service Proposals. Later, in March 1969, BNB MARC II Specifications document was published.

MAB<sup>7</sup> Maschinelles Austausch format fur Bibliotheken, was developed in Germany in 1973 and was revised in 1990. This had the features for record linking between different bibliographic levels which influenced later MARC developments such as UNIMARC.

Based on USMARC, in 1974, CANMARC – Canadian MARC Communication format was designed with some fields adopted from UKMARC and revised in 1979. IBERMARC, the first national format for Spain was developed in 1976 which was based on USMARC. ANNAMARC was designed in 1978 for National Library of Florence in Italy and switched over to UNIMARC in 1985. CATMARC, published in 1987, is very close to UKMARC, was developed for National Bibliography of Catalonia.

While several national formats were being advocated, efforts were already being carried out by IFLA in order to establish an international MARC format, INTERMARC appeared as an international format for the national libraries of France and Belgium. Thus, based on UKMARC, INTERMARC (M) was published in 1975. INTERMARC never reached the international status, as during the mean time UNIMARC was adopted by most of the public and university libraries.

### **2.3 UNIMARC**

The main reason behind the design of UNIMARC was the creation of an international MARC format that would accept records created in any MARC format and act as common format in terms of conversion. UNIMARC (Universal MARC) was published in 1977 by the International Federation of Library Association (IFLA) with primary purpose of facilitating international exchange of data in machine-readable form between national bibliographic agencies. A second edition of UNIMARC published in 1980 and followed by a handbook in 1983. It was restricted to monographs and serials but demonstrated the advantages of flexibility and perfect compatibility with ISO standards and ISBDs. The necessity to expand UNIMARC to

accommodate different sorts of documents envisaged resulted in a new *UNIMARC Manual* (published in 1987). UNIMARC aims not only to be the carrier, or universal MARC for conversion purpose but also as a model for the development of new machine readable bibliographic formats. Till 1990s, UNIMARC got good support for international bibliographic standardization. A seminar in Florence<sup>10</sup> in June 1991 highlighted the importance of UNIMARC format.

In 1992, IFLA identified 30 national MARC formats throughout the world. These MARCs were based on 3 major MARCs, i.e. USMARC, UKMARC, and later UNIMARC. Mortaza<sup>9</sup> had categorised them as:

1. USMARC based - CANMARC in Canada, INTERMARC in France, IBERMARC in Spain, and INDOMARC in Indonesia.
2. UKMARC based - AUSMARC in Australia, THAIMARC in Thailand, SINGMARC in Singapore and ANNAMARC in Italy.
3. UNIMARC based - SAMARC in South Africa, Chinese MARC in Taiwan, JAPANMARC in Japan, and the Croat adaptation of UNIMARC in former Yugoslavia.

Permanent UNIMARC Committee (PUC), worked from 1991-1994 on a new edition of the UNIMARC Manual which was published as loose-leaf format in 1994 with updates on microforms, old printed books and component parts. After this, it concerned more on block 4XX, i.e. the linking block. Till 1997, the interest in UNIMARC and its use had increased and the efforts to keep up good maintenance and updating the relevant issues were ongoing.

Even after using common machine format, identical field tagging systems and a very high level of identity in data content, both LC MARC and UK MARC were not the same, because

1. Differing texts of Anglo-American Cataloguing Rules (AACR), also led to some divergence between the two formats.
2. Library of Congress's format was used to produce catalog data on Library of Congress catalog cards in machine-readable form, while UK format was developed for the production of the printed BNB and other library-related purposes.
3. There were differences of terminology, like: 'leader' in USMARC format while 'record label' in UKMARC format, or 'bibliographical level' in US format and 'class of record' in UK format.
4. Both USMARC and UKMARC also differed in, including requirements for punctuation, treatment and labeling of subfields, and use of filing information.

Some conversion tools and programs like Pygmalion<sup>13</sup> were introduced and updated from time to time depending on modifications in each national MARC format. Some of the national bibliographic agencies had to have several conversion programs to be able to communicate bibliographic data to other national agencies. The cost, time and efforts required to maintain these conversion programs as well as national formats were major problems caused by the proliferation of national MARC formats.

## 2.4 MARC21:

The internationalization of MARC took place considering the large amounts of money and effort spent on duplicating tasks of formatting, keyboarding, proofing, and maintenance of machine-readable records, especially for cataloging, if the practice of developing local formats were to be continued. Representatives of Library of Congress and British Library (BL)<sup>11</sup> met on 21 June 1994 to review principles for harmonization and simplification of USMARC and UKMARC. The objective of the discussion was the possibility of increasing compatibility between the two formats with the ultimate goal of achieving a single MARC format.

In 1999 the Library of Congress and the NLC (National Library of Canada), merged the USMARC and CANMARC formats to create MARC21<sup>12</sup>. The name points to 21st century. The MARC21 bibliographic format, as well as all official MARC21 documentation, is maintained by the Library of Congress. It is published as *MARC21 Format for Bibliographic Data*. As on today MARC21 is available in 11 Languages.

In 2000, the British Library, many library and information centers in UK preferred for a move from UKMARC to MARC21. Recent survey has revealed that only 7 percent of the respondents want to retain UKMARC, while 30 percent preferred partial harmonization and 57 percent want full conversion to MARC21. It is expected that the British Library will implement MARC21 in the summer of 2004, after which bibliographic data will be originated in MARC21.

## 3. Some preliminary observations

In spite of all the efforts of internationalization and trying to bring all the MARCs under MARC21 umbrella, some of the problems went un-noticed. The basic definition of MARC is machine readable catalogue but the very machine philosophy is undermined. After studying the MARC21 design, it is found that it inherits more the rules of AACR2R cataloguing and ISBD punctuation marks than the principles of DBMS. e.g. entering the punctuation marks in the data. Similarly, MARC21 design imposes the AACR2 rules of card cataloguing to generate the catalogue records that look like a printed catalogue. This resulted in the introduction of fields like main and added entries which are data redundant and not required in DBMS. Also grouping of many data elements from the 'title and statement of responsibility' area under one field is not convenient in a database environment. In the present study an attempt is made to see if these problems can be resolved by an alternate approach.

As of now, except the ISIS family (CDS/ISIS, WINISIS, etc.), there are not many known Bibliographic DBMS available either commercially or freely. Generally, various commercial library software use their own DBMS to store the bibliographic data. WINISIS being one of the recent versions in the ISIS family and is a freely available BDBMS. Hence, a detailed study is desired to check out the compatibility between newly designed MARC21 and WINISIS. The ultimate objective of a library catalogue is to display the records according to a standard catalogue code. AACR2 is the most widely recognized cataloguing code and MARC21 claims full compatibility with AACR2. So a study is required to find out the compatibility of AACR2 with MARC21. There is also a need to suggest alternatives where discrepancies are worked.

#### **4. Focus on compatibility issues**

Some of the issues of the compatibility of cataloguing rules of AACR2, bibliographic standard MARC21 and the bibliographic database Management Software B-DBMS are

1. Issues of compatibility of AACR2R, MARC21 and WINISIS as these standards are developed at different times, by different organizations and for different purposes.
2. Design and develop a bibliographic database in WINISIS using AACR2R, following the bibliographic tag codes of MARC21.
3. A study in structural design of the system that would accommodate AACR2R rules, amenable for data exchange and can be implemented in a bibliographic database environment.
4. Generation of automated catalogue fully compatible with AACR2R and which is amenable to data exchange in a network environment using MARC21.

As mentioned in the objectives, the present study deals with the issues of compatibility of AACR2R, MARC21 and WINISIS, it deals with all the three aspects in detail.

#### **4.1 AACR2R:**

The present work is a compatibility study and not meant to be a critique of AACR2R. As such it does not make any attempt to evaluate the validity of the rules of AACR2R and any controversies surrounding it. For the purpose of study, it will be verified whether the rules of AACR2R are fully amenable to computational environment. During the study, it is assumed that AACR2R is an ideal standard. If at all any changes are required then it is recommended that this burden should be taken care by MARC21. Accordingly, alternative approaches are suggested for MARC21 implementation.

#### **4.2 MARC21:**

Library of Congress has introduced different MARC formats like bibliographic, authority, holdings, classification and the community. Each of these formats has different tag codes for varied purposes. Even within each format there are various tag codes like a) bibliographic format i.e. descriptive elements tags b) tags for the database management i.e. linking fields and c) other tags. Since the present study deals mainly with the bibliographic tags and their compatibility in bibliographic cataloguing, it is restricted to bibliographic data elements only. Similarly, MARC21 fields 600-65X meant for subject headings or access terms to a bibliographic record through a heading or term that is constructed according to established subject cataloging or thesaurus-building principles and guidelines. Hence, these fields are not covered in this study.

#### **4.3 WINISIS:**

It is a bibliographic DBMS and used as a tool for the data entry and displaying or printing the records. An attempt is made to find out whether WINISIS is accommodative of AACR2R rules with regard to the choice, presentation, and order of descriptive elements and the use of punctuation marks. An effort is made to find out whether or not WINISIS accommodates all the subfields that are recommended by MARC21 that follows AACR2R in major part.



### 5. Compatibility issues of bibliographic standards with bibliographic database tools

The importance of standards on library and information services cannot be overemphasized. Accordingly, libraries adopted several standards for classification, indexing and cataloguing. In modern scenario, the library tools such as catalogues have to manifest as computerized systems or even more desirable are web based services. It then becomes imperative to study how amenable are the standards such as bibliographic standards like AACR2R rules or MARC21 amenable to be directly implemented in bibliographic database management system. In close observation of such system the following compatibility issues are observed:

#### CONCLUSION

AACR2R, MARC21, WINISIS are developed at different times and by different organizations, so in practice they may not be fully compatible with each other.

- Considering the data redundancy, it looks like the database system's computational principles and advantages are undermined.
- While entering data according to MARC21, it recommends entering the punctuation marks as part of data, which may lead to serious problems while indexing and globally editing the punctuation marks.
- Though MARC21 is designed in conformance to AACR2 and its display format, the vision of making it to MARC of new millennium has broadened its coverage to Authority, Holdings, Classification, and Community formats also which resulted in generating many fields and subfields beyond the capacity of library and information centers.
- Since AACR was designed for printed card catalogue, some of its rigid rules may not be compatible or relevant in the context of automation.
- Even though WINISIS is a very good bibliographic DBMS, it may not be fully amenable for bibliographic data; for example, the WINISIS cannot handle repeatable subfields and there is an unnoticed limit on the number of subfields, etc.

In view of the discussion above, one should take a systematic approach to compatibility study of the standard and technology tools that are used to implement them.

#### REFERENCES

1. Anglo-American Cataloguing Rules / Prepared under the direction of the Joint Steering Committee for the Revision of AACR; edited by Michael Gorman and Paul W. Winkler. 2nd. ed. 1988 revision. Chicago: American Library Association, 1988. p. 616
2. UNESCO. CDS/ISIS for Windows <http://www.unesco.org/isis> Browsed on March 2003.
3. Crawford, W., MARC for Library Use: Understanding Integrated USMARC, 2nd, G. K. Hall & Co., Boston, 1989. pp. 337-41.
4. Gredley, E. and Hopkinson, A. Exchanging Bibliographic Data: MARC and other International Formats, Canadian Library Association, Ottawa, 1990. pp. 72-77.
5. Unicode Consortium, The. The Unicode Standard: Worldwide Character Encoding, Addison-Wesley, Reading, MA, 1991.
6. UCS and USMARC mapping: MARBI Discussion Paper No. 73, Library of Congress, Washington, 1993.

7. Campos, E M; Lopes, M I and Galvao, R M. MARC formats and their use: an overview. Program, Vol. 29, no.4, Oct 1995. pp. 445-459.
8. Cayless, C. The development of MARC in the United Kingdom, 1966-71, Nelson, J.R., Proceedings of the MARC II Seminar; 18-20 August 1971, The School of Librarianship, The University of New South Wales, Sydney, 1973. p 14.
9. Mortaza, Kokabi. The internationalization of MARC. Part I: the emergence and divergence of MARC. In: OCLC Systems & Services Volume 12 Number 1 1996 pp. 21-31
10. Die Deutsch Bibliothek. Study to establish the feasibility of using UNIMARC amongst EC national libraries, bibliographic utilities and book-trade based upon their present computer facilities: final report. Luxemburg : CEC, 1992
11. Reser, D.W., Towards a common MARC format (Press release), Message to Multiple Recipients of List AUTOCAT (AUTOCAT@UBVM.cc.buffalo.edu, 21 July, 1994).
12. Network Development and MARC Standards Office. Library of Congress. MARC21: Harmonized USMARC and CAN/MARC. Oct. 1998.  
<http://www.loc.gov/marc/annmarc21.html> Browsed on March 2003.
13. Prasad, A R D. Pygmallion. In: Workshop on Application of Optical Character Recognition Technology to Library and Information Work. Bangalore : DRTC, 23rd - 24th March, 1999.