DESIDOC Bulletin of Information Technology, Vol. 17, No. 5, September 1997, pp. 17-22


# Bibliographic Databases and Exchange Formats

**Kamini S Mishra***

**Abstract**

Computers play an important role in the development of bibliographic databases. For generation and exchange of bibliographic data at defferent levels—international, national, regional and local—exchange formats are needed. This paper discusses about the formats available at national and international level for adoption by national, international agencies and, individuals/organisations as per their requirements.

## 1. INTRODUCTION

The advantage of computer is that it can rapidly and efficiently manipulate, retrieve any information/data which is stored in machine-readable form. Their use in creation and development of bibliographic databases has raised the hope of developing a universal bibliographic system through the cooperation of several national and international organisations. Generation and exchange of biblioraphic data takes place at different levels, i.e., international, national, regional and local. Thus there is need for exchange formats that are designed specifically for the transfer of machine-readable bibliographic data between systems.

## 2. BIBLIOGRAPHIC RECORD FORMAT

In the Unisist Reference Manual, a (machine-readable) bibliographic record is defined as a collection of information which pertains to a single document and which is stored in machine-readable form as a self-contained and unique logical structure.

*National Information System for Science & Technology (NISSAT), New Delhi

Format conveys the notion of a formalised framework or structure which will hold records of varying content according to certain set of rules or conventions controlling the representation of the data. These rules may be unique to a system, or shared with other systems.

Bibliographic record formats are used to describe the arrangement or structure of computer readable record of bibliographic items.

Formats can be of two types: internal/local format, and Exchange/communication/interchange format.

### 2.1 Internal Format

Internal formats are so called because they are internal/local to a software system. They can be changed specifically as per the needs of the local system and do not have to conform to any external standards.

### 2.2 Exchange/Communication Interchange Format

Exchange formats are also known as interchange/communication formats. Exchange formats are used for exchange of records between systems. Systems should be sufficiently

flexible to cope with the needs of many different software systems.

A major problem for designers of exchange format is the lack of international agreement on standards for constructing bibliographic records. Although ISBD is available, it is not accepted by many organisations.

## 2.3 Exchange Formats

Bibliographic data formats adopted for exchange of data consist of three basic components.

(a) A defined physical structure: rules for the arrangement (on a computer storage medium) of data to be exchanged.

(b) Content designators: codes to identify the different data element in the records (e.g., author, title, scale of map, starting date of journal, etc.)

(c) Content of the record: governed by rules for the formulation of the different data elements very closely tied up with content designators. The data elements separately identified by the codes in the exchange format are to be defined, not only in terms of content but also in form, if the records are to be suitable for use by another agency.

# 3. STANDARDS FOR RECORD FORMAT

Standardisation of the record format implies standardisation of the record structure, directory, content designator at national, regional and international levels. Design and implementation of a standard record format is uniformly acceptable to all bibliographic agencies involved in information transfer is very essential and urgent but it is very difficult to reach a consensus in this regard.

The International Conference of Cataloguing Principles (ICCP) held in Paris 1961 has set up standards for the headings of author and title fields in catalogues and bibliographies. The first standard developed in 1974 was meant for the description of monographs. It was followed by a series of specialised ISBDs for various forms of documents, called General International Standards Bibliographic Description ISBD(G).

ISBD(G) lists within its frame, all bibliographic elements which are required to describe and identify all types of materials which are likely to appear in library collection. It assigns an order to these elements and prescribes a distinct punctuation system to differentiate them from each other. It also serves as the basis for specialised ISBDs.

Development of ISBD may be termed as the greatest achievement contributing to the standardisation of bibliographic records for the following reasons:

- ☐ It facilitates records from various sources interchangable
- ☐ It assists in the interpretation of records across language barriers
- ☐ It assists in the conversion of bibliographic records to machine-readable form

## 3.1 ISO 2709–The International Standard Exchange Format

ISO 2709 is an international standard format for bibliographic information interchange on magnetic tape. It was developed for the exchange of bibliographic records on magnetic tape, originally in the Library of Congress (LC) MARC format.

## 3.2 Machine-Readable Record Format

MARC is an acronym for Machine Readable Catalogue or Cataloguing. This general description is misleading, implying that MARC is a kind of catalogue or method of cataloguing, whereas MARC is more accurately defined as a group of formats employing a particular set of conventions for the identification and arrangement of bibliographic data for handling by computer.

The original MARC format, from which current formats originated was developed at the LC in 1965-66. Since then more than 20 formats have appeared which are known as MARC. Their common characteristics are:

- ☐ Adherence to the ISO 2709 record structure, or its equivalent national standard including option which allows use of indicators and subfield identifiers in data fields.
- ☐ Most are national formats based on national library or national bibliographic agency and are designated communication formats for exchange of bibliographic recrords with other similar organisations.

MARC records have many other library and non-library related uses, though it is generally identified with the production of library catalogues & national bibliographies.

LC was the first to design and experiment on a MARC record format for the purpose of communicating bibliographic information to a large number of libraries. When MARC-I commenced as a pilot project in 1966 in LC, there were no established MARC formats available. Libraries had reached no consensus as to what all access points were required to take full advantage of an automated cataloguing system.

The MARC-II format was considerably improved in the light of experiences and opinions of important libraries and a specific survey was carried out for studying the requirements of the users. MARC-II format developed in 1968, was the result of Anglo-American co-operation. The new format was intended to be hospitable to all kinds of library materials, sufficiently flexible for a variety of applications in additions to catalogue production, and usable in a range of different computer systems. Despite Anglo-American cooperation, there were two versions of MARC i.e., LC MARC II and BNB MARC II.

## 3.3 USMARC Format

There are three USMARC communication formats:

(a) USMARC Format for Bibliographic Data (UFBD)

(b) USMARC Format for Authority Data (UFAD)

(c) USMARC Format for Holdings & Locations (UFHL)

All the three USMARC formats are implementations of ANSI Z39.2, American national standard for bibliographic information interchange on magnetic tape which conforms with the ISO 2709—the physical structure of the UKMARC record-although there are some differences in terminology and in definition of fields.

## 3.4 UKMARC Format

British National Bibliography has shown active interest in the possibilities of MARC from 1966. Cooperation with Aslib and OSTI within the country and with LC abroad, enabled it to develop a MARC format in 1968. There were no considerable changes in the format until the first edition of UKMARC Manual appeared in 1975. But later, the advent of BLAISE and the increased use of non-book materials in the educational field and publication of lack of standardisation in cataloguing as the major obstacle to a standard MARC format.

UKMARC is a single, unitary format designed to accommodate all types of material with some material specific fields. The physical structure of UKMARC is based on ISO 2709 and BS 4748. UKMARC format specifications and structure are set out in UKMARC Manual, Ed.2, 1980-87.

## 3.5 Other Formats

After the joint work of BNB and LC on the MARC format, other countries quickly started development of their national formats. They are Canada (CANMARC), Australia (AUSMARC), Germany (MAB1), Italy (ANNAMARC), Denmark (DENMARC), Spain (IBERMARC), Sweden (SWEMARC), etc.

Growing enthusiasm on the part of national bibliographic agencies to develop their own MARC records resulted in the wide-spread use of the MARC format. But the only area of standardisation in these national formats was the record structure effected through the adoption of ISO 2709. They still showed wide variations in respect of content and content designators. These differences necessitated suitable programmes to be written for one agency to use the records of another. Within

the LC MARC itself the variety of material added to the complexity of interchange of records.

International Federation of Library Asociations and Institutions (IFLA) Committee on Mechanisation sponsored the IFLA Working Group on Content Designators, and as a result an international standard for content designators was taken up and this led to the development of UNIMARC.

## 3.6 UNIMARC

IFLA Working Group on Content Designators recommended in 1973 a SUPERMARC which was based on ISBD. This was later called MARC International Format (MIF) from which the UNIMARC was developed. The final format was published in 1977. The group recognised the lack of standardisation in cataloguing as the major obstacle to a standard MARC format. UNIMARC is a communication format which necessitates writing and maintaining of only two conversion programs—one from the national format to the UNIMARC and the other from the UNIMARC to the national format. It was decided that each country can have its national format but, it should be the responsibility of the national bibliographic agency in a country to translate the records from the national format to the UNIMARC for purposes of interchange. ISBD was accepted as the basis of descriptive data elements within this format. The second edition of UNIMARC was published by IFLA International Office for UBC handbook with the intention of guiding the users in its application.

## 4.7 UNISIST Reference Manual

The UNISIST-ICSU/AB Working Group on bibliographic Description, set up in 1967 as part of the UNISIST programme, decided and developed the UNISIST Reference Manual for machine readable bibliographic descriptions.

The major feature of the format is that it gives equal prominence to bibliographic records whether they relate to analytics (i.e., journal article and contributions in journal, work published somewhere also), monographs or serial titles.

The format was designed to do this because it was developed from the secondary sources which give equal importance to the different bibliographical levels. The record contains no distinctive feature to permit a hierarchy to be indicated; instead, different tags are allocated to field at a particular level.

## 3.8 Common Communication Format

Although Unesco had developed the Reference Manual with the help of ICSU/AB, it had not been accepted by many organisations. These organisations continued to approach Unesco for assistance in developing bibliographic information system.

In April 1978, the Unesco General Information Programme (Unesco/PGI) sponsored an International Symposium on Bibliographic Exchange Formats which was held in Taormina, Sicily, organised by the UNISIST International Centre for Bibliographic Description (UNIBID) in cooperation with others to study the desirability and feasibility of etablishing maximum compatibility between existing bibliographic exchange formats. As a result of this symposium, a resolution passed at the symposium for Unesco to set the ad hoc group for the establishment of the Common Communication Format (CCF).

The objective of CCF was stated to provide a detailed and structured method for recording a number of mandatory and optional data elements in a computer readable bibliographic record for exchange purpose between two or more computerised systems. It is also useful to a single bibliographic agency engaged in structuring its own format and simultaneously keeping compatibility with the CCF. Non-computerised systems also can use CCF data elements because, it simplifies computerisation at a later state.

The first edition of CCF was published in 1984 and the second edition in 1988. Bibliographic agencies around the world dveloped national & local formats based on the CCF. The first CCF User's Group meeting was held at Geneva in 1989 and users recommended some minor changes in the later

editions. The CCF also contains mechanisms for providing three kinds of links between or within records and two different types of vertical/horizontal relationship. The record structure of the CCF constitutes a specific implementation of the international standard ISO 2709.

### *3.8.1 CCF/F: Common Communication Format (Factual)*

This is Common Communication Format for Factual Information, published in 1992. For this, there is separate set of data elements identified, as shown in the CCF(F) document:

### *3.8.2 Evaluation of CCF*

**Relationship with existing formats:** CCF is not meant for record of an institution for internal storage and processing purposes. Processing formats vary from institution to institution and also within the same institution. It is based on major exisiting international exchange formats and specially designed for transfer of records between systems which were already capable of providing output in these major exchange formats.

**CCF as an exchange format:** The CCF is intended as an exchange format and as such has to contain bibliographic data for exchanging between systems.

The CCF is intended for exchange of bibliographic records that were needed for the identification of a document in a catalogue or bibliography. It does not contain fields that would be required for library circulation systems or inter-library loan, etc. Any system that wants to exchange data elements other than those provided by CCF, is free to allocate unused tags to those data elements.

By using an exchange format like CCF, each system has to design only one conversion between its format and the common exchange format and back again.

## 4. INDIAN STANDARD

Standardisation of record format has not received due attention in Indian libraries. At national level, Indian Standards Institution (ISI), now renamed as Bureau of Indian Standards (BIS), had evolved a standard for bibliographical references in 1963 for use in non-computerised systems.

In July 1986, ISI published a standard IS: 11370-1985 titled 'Guide for Data Elements and Record Format for Computer based Bibliographic Description of Different Kinds of Documents.' During the late 1980's NISSAT organised a tripartite meeting (CALIBNET, DELNET, INFLIBNET) to sort out the difference in choice of formats: Comon Communication Format (CCF) *vis-a-vis* UNIMARC. Finally it was decided that a database producer can use either.

NISSAT also constituted a group to draft the INDIMARC guidelines based on the framework prescribed by the CCF. The progress made by the group is very slow.

Another effort by INSDOC, as SAARC national focal point on information, had produced a format for use by the participating countries of SAARC. Further progress on this is not known.

## 5. FUTURE TRENDS

During the last 25 years, a radical change has occured in the mechanism for transfer of bibliographic data. The current situation is both success and failure; success, because millions of records structured according to ISO 2709 are now available for exchange, and failure, because there is multiplicity of implimentation of ISO 2709 and the full potential of computing has not been harnessed to make the necessary conversion between each.

Among national formats, UNIMARC is a serious contender for the role of international format, but has been accepted only by library community. Some national formats, for example USMARC, have become virtually international. ISO 2709 will remain as a universally recognised standard for MARC. It is however a standard available in tape and the use of tape for bibliographic data exchange will decrease as time passes. Finally any change in record structure will lead to the change in conversion softwares from exchange format to systems internal format and vice-versa, but will be of little concern to MARC users where MARC is a

set of codes defining the data elements of a record in automated systems.

It is estimated that MARC and other exchange formats will continue for some time. As long as organisations wish to exchange record or derive bibliographic data from central agencies, and until computer technologists devise cost effective and relatively simple ways of transferring bibliographic data in different formats between systems, exchange formats remain necessary.

## 10. FURTHER READINGS


1. Gredley, Ellen and Hopkinson, Alan. Exchanging bibliographic data (MARC & other International formats). 1990.
2. Simmons, Peter and Hopkinson, Alan. The common communication format for bibliographic information. 1992.
3. Lahiri, Abhijit. An Indian model for database services. 1991.