

## Sharing of Inhouse Databases in India: Need for Standardization

RAM KUMAR\*

AND

DR ASHOK KUMAR\*

The paper discusses in brief the features and importance of inhouse databases being developed by various institutions in India. Emphasizes the need for standardization in record 'format' and 'structure' in order to exchange/merge records effectively and economically. Various exchanging formats developed and accepted by international community have been discussed with their merits and demerits.

### 0 INTRODUCTION

Use of data processing techniques on a large scale for processing, organising and disseminating information in libraries and information centres in India has led to the emergence of various bibliographic databases on specific subjects. These databases differ in their level of coverage of information sources, hence, they may be local, national or international in scope. Apart from few national databases developed by NIS-SAT sectoral centres in India, much efforts in database building have been made on individual basis by various institutions. As a result a sizeable number of 'inhouse' databases exist in India. These 'inhouse' databases cover a specific discipline, usually in depth, and serve specific group of users on institutional level. Hence, such databases

are named as "local" or "corporate" databases.

Despite a number of limitations in their creation and use, inhouse databases can play vital role in development of our country by fulfilling information needs of users at local as well as national level. We can generate our own database of national level in all relevant fields by cooperation and coordination among various institutions and other bodies engaged in database building. More specifically, inhouse databases, so generated, could be made more useful for local and national users by making them more comprehensive in their coverage. This can be achieved by exchange/merging records from databases on a similar subject created by different institutions. However, this exchanging/merging of records among vari-

---

\* Structural Engineering Research Centre, CSIR, Post Bag No 10, Kamla Nehru Nagar, Ghaziabad (UP) - 201 001.

ous databases is possible only if they have similarity in their record 'format' and 'structure'. Uniformity in record 'format' and 'structure' facilitates the exchanging/merging of records among databases and their consequent use across institutional, national and international boundaries by using various networks.

## 1 INHOUSE DATABASE

Inhouse database comes from within an organisation and developed by library staff and/or others by using infrastructures available with library or parent body. An inhouse database contains references to information sources available with that library/institution or available with other libraries of that locality. Inhouse databases are locally oriented, collection-specific and generated to serve a specific group of users who are limited in number at institutional level. An inhouse database is designed to cover a specified discipline and the coverage can be general or limited to some specific and definable aspects. An inhouse database may be defined as "a collection of related information organised so that a particular group of users can find what they want by searching a corporate or internal database on a computer in the organisation rather than only being able to search commercial of external database using telecommunications."<sup>1</sup>

Inhouse database is a very important means of information organisation and its subsequent dissemination in specific discipline. It can provide important services to its specific users by providing information generated and/or organised locally. Such local information may not be covered by other national and international databases. However, such local information may be very useful, manytimes, dealing with special aspects of a locality. Moreover, such local

information might be lost if not covered by inhouse databases. Thus, inhouse databases could serve following main purposes—(a) to generate indigenous capabilities for database creation and utilization (b) complementary and supplementary to the national and international databases (c) recording of information generated locally (d) to encourage the users to make use of national databases rather than costly commercial databases and (e) immediate availability of information sources referred to.

## 2 STANDARDIZATION

Inhouse databases developed in India have a number of anomalies in their creation and use. They have been developed in complete isolation at institutional level by record 'format' and 'structure'. The record format and structure, thus followed, is non-compatible with other exchanging formats created and accepted by global community. Consequently, exchanging/merging of records among databases created by various institutions in different formats has become a serious problem. They need a number "intermediary" of "conversion" programmes for exchanging/merging the records from different databases. Thirty different conversion programmes, for instance, are required for exchanging/merging records from the databases developed by adopting six different formats. Similarly, twenty different record formats followed by different institutions require one hundred eighty different conversion programmes to enable a free flow of data among them.<sup>2</sup> Interestingly, such intermediary programmes are not easily available in India and if available they are not fully effective in their purpose. Moreover, they are very costly. Thus, making it uneconomical to share records from different databases. It is wise, therefore, to use similar, standardized exchanging

format in order to exchanging/merging the records efficiently and economically. Standardized format of a database is not only useful for identification of original documents referred to but also helpful for users to 'judge' the immediate relevancy of document from its bibliographic description.

### 3 BIBLIOGRAPHIC RECORD AND FORMAT

A bibliographic record comprises many fields showing various data elements (such as author, title, edition, name of publisher etc.) and/or sub-elements in sub-fields (such as surname, middlename and forename of an author) or other characteristics of a document through which it could be identified. These data elements are arranged in various fields of a record in a set sequence for enabling a user to identify his/her documents referred to.

A bibliographic record should be complete in itself containing sufficient data elements necessary for identification of original documents; sub-elements useful for local needs. The identification and representation of data elements/sub-elements to be included in a record constitute the "format" of that record; while the logical and physical methods to apply the format constitute the "structure" of concerned record or database.<sup>3</sup>

The data elements as well as sub-elements identified for inclusion in a record are necessarily represented by unique term/codes/content designators. These codes are necessary to identify and subsequent recalling of the corresponding data elements in a computerized bibliographic record. This process of assigning unique tags/content designators is called as "tagging" or "codification". The tags/content designators are basically symbolic codes (numericals and/

or alphabets) or names representing the different data elements in unique way and allways preceding them for their self identification in a machine readable record. A distinct set of such codes/tags for identification and representation of data elements pertaining to a particular type of documents is called a "format".<sup>4</sup>

Apparently, a different format (better to say standard) contains a distinct set of tags/content designators to represent various data elements identified. Thus, each standard adopts a distinct system of 'tagging' to identify and represent the data elements in a record. It is important to remind here, therefore, that the 'tagging' system (codification) of data elements adopted for creation of database by a library must be *compatible* with other systems of 'tagging' adopted by other libraries to facilitate the free flow of data among concerned libraries.

### 4 NATIONAL AND INTERNATIONAL EXCHANGING FORMATS

A number of exchanging formats are available which have been formulated by various national and international bodies:

#### 41 LIBRARY OF CONGRESS

The Library of Congress has developed MARC<sup>5</sup> format in 1966. The MARC format was originally prepared for automatic production of printed catalogues as it adheres to AACR2. MARC has now been adopted very well in many countries, still it is not a sole standard for exchanging the bibliographic data in machine-readable form on international level. It is partly due to the lack of unique record identifier, thus, exhibiting inability to match records at the time of searching. Moreover, MARC has a 'flat' (non-hierarchical) records structure, contains only the 'root' elements, thus, unable to link logically a record with other records.



As a result of such inabilities, many modified versions of MARC have been derived by individual library and bibliographic agency adopted to suit their local needs. There is need to have simplified version of MARC format to make it more suitable for online searching, especially, for non-professionals. It should have capability of full-scale record display necessary for the purpose of bibliographic verification. MARC should have a 'hierarchical' record structure to handle multirecord entries in which a record could logically be contained in another record and could itself contain a record. Thus, it would link various data elements included in a record and/or with other records in a file.

#### 42 INTERNATIONAL STANDARDS ORGANISATION (ISO)

ISO: 2709-1981<sup>6</sup> has been developed by ISO/TC/46 in 1981 for exchange of bibliographic records. It does not include the standard content designators and specifies only the record structure. Thus, it presents a generalized communication format without assigning any meaning to tags/identifiers/indicators. This standard, therefore, can be used to formulate specific format by individual country to suit its local needs.

#### 43 INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS (IFLA)

There are two sets of standard developed by IFLA for this purpose:

(A) International Standard Bibliographic Description (ISBD)<sup>7</sup> was formulated under UBC (Universal Bibliographic Control) programme, one of the six major programmes of IFLA. It provides the guidelines for the selection and identification of various data elements of a bibliographic entity; order of data elements identified; and the punctuation marks by which these data elements are to be demarcated.

(B) UNIMARC—another major IFLA standardization effort was the development of international MARC format called as UNIMARC<sup>8</sup>, under the International MARC Programme (IMP). The UNIMARC format was built on ISO and the ISBD standards designed to follow AACR2 rules for catalogue like output and, thus, biased towards a catalogue code.

Despite many merits of the format, UNIMARC has some complexities in its structure. It has many superfluous, reduplicated or replicated data fields. For instance, the fields '430 through 436' are defined to relate a serial with its predecessor, while the fields '440 through 448' are used to relate a serial with its successor. Only one field or tag is sufficient to indicate bibliographical relation or link of a serial with its successor or predecessor. There also exist many such redundant fields in the format. The field '520: former title of serial', for instance, is an example of a reduplicate field.

#### 44 UNESCO

It has developed following standards for the exchange of data:

(A) UNISIST Reference Manual—UNISIST Reference Manual<sup>9</sup> was published in 1974 and maintained by the UNISIST International Centre for Bibliographic Description (UNIBID). As the first edition of the manual was biased towards the requirements of indexing and abstracting agencies, therefore second edition appeared in 1981 to serve as a 'standardized communication format' for the exchange of bibliographic records.

The reference manual basically lists various data elements with their descriptions. It does not specify any cataloguing

rules for rendering and inputting the data for fixed form of output. The manual presents guidelines for machine-readable exchange of data.

(B) Common Communication Format (CCF)—The first edition of CCF<sup>10</sup> was published in 1984 in English and French as an exchanging format for all kinds of libraries and related agencies. CCF has a generalised structure and does not prescribe rules for description of data elements. It provides, therefore, autonomy up to some extent to the individual library to formulate their own descriptive practices for data elements of a bibliographic/non-bibliographic entity by providing mandatory and optional data elements, such practice is useful for individual library to fulfil its local needs.

CCF has been designed for use of all kinds of libraries or related agencies to produce records in similar format. This facilitates not only the exchange of the data easily and quickly but also acts as a bridge among databases produced in different internal formats to convert data contained into converseable format.

CCF provides autonomy to the individual library to get output in desired form. It is due to the non-adherence of CCF to any cataloguing code for fixed type of output. Moreover, it identifies and defines relationships between the data elements and their corresponding codes irrespective of kinds of bibliographic entity. This unique feature of CCF has led to the Information Specialists to apply and use it for creation of databases of entities other than bibliographic nature. This can be possible by making slight changes or modifications in the existing fields which can be used to represent same attributes but possessed by different entities. Importantly, there are a number of unused tags available in CCF which can be used to add or interpo-

late additional data elements useful for individual library to serve as local needs of its users. All these features of CCF have led the information community to use it as a national and international exchanging format.

#### 45 AMERICAN NATIONAL STANDARDS INSTITUTE (ANSI)

ANSI has developed ANSI: Z39-1979<sup>11</sup> for exchange of bibliographic records in machine-readable form. This standard has been used as base for devising MARC format by Library of Congress and, thus, MARC format is a particular implementation of Z39.2 standard. This standard has a hierarchical record structure like ISO: 2709 to handle multiple entry records in a single reference. Though accepted very well, yet some limitations are there. Lack of unique record identification field (*ie*, a key required to identify each record and to place it in a fixed location in the file having the same length *ie*, number of characters), for instance, makes relatively difficult the processing of records.

#### 46 INDIAN STANDARDS

The Bureau of Indian Standards (BIS) has formulated IS: 11370-1985<sup>12</sup> for exchanging the bibliographic records in machine-readable form. IS: 11370 defines the standard structure of the communication format and presents guidelines for exchange of bibliographic records of any kind of documents. The Indian standard assigns meaning to data fields, specify 'tags' or 'field indicators' for fixed portion of the fields and contents for the variable fields. This standard has been formulated as a national communication format and is being used widely. However, it is felt now to revise it by incorporating some of CCF like features keeping in view changing technology and thus needs of the Indian libraries<sup>13</sup>.

## 5 CONCLUSION

The merging/exchanging of records among databases developed by different institutions on specific subjects may bring forth the emergence of comprehensive and useful databases in India. These databases, then, may provide useful services to the users at national level. Moreover, such databases will represent the pool of literature published in India in various fields and, thus, they may best be utilized commercially at international level. This can be achieved by making them comprehensive, compatible as well as by formulating appropriate marketing strategy. However, this is only possible when all these databases are generated with identical records by using standardized exchanging format. More efforts are expected in our country from national organizations such as NISSAT and BIS with the co-operation and co-ordination with international organisations to develop a national communication format. However, it is imperative to use CCF as it provides many autonomies useful for local needs.

## REFERENCES

- 1 EDDISON (Betty). Tools to start with: consultants and their words. *Database*. 11, April; 1988; 88-90.
- 2 VISWANATHAN (T). Emerging standards for library networks. *Annals of Library Science and Documentation*. 38 (3); 1991; 79-84.
- 3 MUDHOL (M V) and Rout (R K). 'Standardization with special reference to ISBN: an implication for computer education in the library systems'. *Lucknow Librarian*. 21, 4; 1989, 131-136.
- 4 ATTING (John C). 'The concept of a MARC format'. *Information Technology and Libraries*. March, 1983, 7-17.
- 5 LIBRARY OF CONGRESS, AUTOMATED SYSTEMS OFFICE "MARC format for bibliographic data". Washington, D C: LC. 1980.
- 6 INTERNATIONAL STANDARDS ORGANISATION (ISO). "Documentation: format for bibliographic interchange on magnetic tape (ISO: 2709-1981)". 2nd ed. 1981. GENEVA: ISO.
- 7 IFLA. International standard bibliographic description (for single volume and multi-volume monographic publications. London: IFLA Committee on Cataloguing. 1971.
- 8 IFLA, Working Group on Content Designators. "UNIMARC: Universal MARC format". 2nd revised ed. 1977. IFLA, International Office for UBC, London.
- 9 UNISIST/ICSU-AB Working Group on Bibliographic Description. "UNISIST reference manual for machine readable bibliographic description". 1974, UNESCO, Paris.
- 10 CCF: The common communication format". Ed by Peter Simmons and Alan Hopkinson. Paris: General Information Programme on UNISIST, Unesco. 1988.
- 11 AMERICAN STANDARDS NATIONAL INSTITUTE (ANSI). "American national standard format for bibliographic information interchange on magnetic tape (ANSI: Z39.2-1979)". New York: ANSI. 1979.
- 12 IS: 11370-1985. "Guide for the data elements and record format for computer based bibliographic databases for bibliographic description of different kinds of documents". New Delhi: Indian Standards Institute. 1985.
- 13 KASHYAP (M M). "Standard bibliographic data designating system and communication format for Indian libraries and information organisations". *ILA Bull* 25 (3); October-December 1989. 11-23.



**ILA**

ISSN 0970-4728

# BULLETIN

---

Vol XXVIII

No 3-4

October-March 1993

---



---

**INDIAN LIBRARY ASSOCIATION**

---