Highlights on Library Standards, Protocols, Formats and Model

Kanchan Kamila

Highlights the different standards, protocols, formats and model for libraries which standardized the library processing work as well as search & retrieval service. It brings uniformity in the work and service and helps in international information interchange.

**Keywords:** Standard, ISO 2709, ISBD, MARC, CCF, FRBR

1. **Introduction**

Library practices are based on standardized methods of description based on the data elements and the provision of access points. The use of the computer has made standardization even more important than before. Human nature being what it is, however, there is never only one accepted way of doing anything that can involve different purposes, processes, costs, or historical antecedents. Using a bibliographic file that departs from recognized standards occasions neither physical disaster nor, in a stable person, emotional trauma. A catalogue card measuring 3-by-5 inches is functional despite its slight departure from the standard measurement of 7.5-by-12.5 centimetres. It is not hard to recognize a series statement that precedes an imprint in a citation even if library cataloguing rules would locate it after the collation. One can write a programme in a number of ways for an independent computer-based system and still achieve the desired result. However, if common basic practices are applied in producing different files or in using different systems to maintain and search them, then

1. Records can be interchanged among them;
2. A person consulting more than one such file becomes familiar with the pattern and can more easily find and interpret the relevant data; and
3. The same programs can be used to process records coming from various sources.

Reasonable people are persuaded that there are more advantages than disadvantages to standardization and eventually one particular practice gains overwhelming favour while its competitors wither from disuse. When consensus is well enough advanced in a particular area, the prevailing practice is usually codified formally by a representative committee of practitioners who ensure that it is stated as clearly and comprehensively as possible. This codification is a standard, whether it is formally ratified by an officially recognized body or only informally but widely implemented in practice. In this paper, some library standards, search protocols, formats and models are discussed.
2. **What is Standard**

A standard should provide (1) a unified structure, (2) a statement of minimum expectations, and (3) guidelines to determine when absolute uniformity in execution is essential and when it is not.

2.1 **Reasons for Ignoring Standard**

The most frequent reason for ignoring a standard is the perception that it is too difficult or costly to adopt or that it is unnecessary.

2.2 **Qualities of Good Standard**

It is therefore desirable for a standard, particularly one ratified at the international level, to be as flexible as its purposes permit. The larger and the smaller, the richer and the poorer, the general and the specialized institutions of the world can then reasonably aspire to adhere to it instead of simply ignoring it. What is needed is compatibility, not necessarily uniformity of a product down to the last comma.

2.3 **National and International Standards**

Being relatively rich, libraries throughout the English-speaking world have been able to devote much professional staff time to the formal codification of bibliographic practices. Active Anglo-American cooperation is almost a century old although interrupted by periods of apathy and by the inactivity imposed by two World Wars. This cooperation formed the basis for broader-based international cooperation after the Second World War when existing Anglo-American initiatives were expanded, notably under the aegis of IFLA (particularly its programme named the Universal Bibliographic Control and International MARC programme) and by UNESCO (particularly its General Information Programme (PGI) and its role in the United Nations Information System in Science and Technology (UNSIST). The projects of IFLA in particular are forwarded largely through the volunteer work of committee members from many countries who have the vision, political acumen, and fund raising ability needed (UNESCO often provides the funding for IFLA projects). Even where a particular standard is developed under other auspices, its approval by one or another of these bodies can help foster widespread adoption.

The ultimate international standard-approving organization is the International Organization for Standardization (ISO). Both the original initiative for a new standard and the ultimate responsibility for announcing it and persuading people to implement it rest with ISO’s component national bodies, for example, the British Standards Institute (BSI), Bureau of Indian Standards (BIS), Canadian Standards Association (CSA), and the American National Standards Institute (ANSI). International MARC’s national
Library Standards

ISO 2709

In 1973, ISO published its standard 2709. “Documentation – Format for Bibliographic Information Interchange on Magnetic Tape,” which must be used in conjunction with other ISO standards relating to character sets and tape labelling. This is the same as the American ANSI Z39.2 and BS: 4748. It regulates only the essential structural features of MARC as a variable field format and is not even limited to bibliographic records. Librarians use this standard also to format records in name and subject authority files, including all links. Archivists use a format known as MARC (AMC) [AMC=Archival and Manuscript Control]. ISO 2709 based formats can be, and are being, used to process any kind of information requiring variable-length fields, bibliographic or other.

The delimiters, content designators, record directory, and label are all part of ISO 2709 specifications. Only the following features are regulated in absolute terms:

1. The record and field delimiters are each described as a particular type in the character set
2. Every tag must be three bytes long and the label must be twenty-four bytes long, with specified data in fixed locations therein (some bytes in the label are left underlined; a local use may be defined).

ISBD

The International Standard Bibliographic Descriptions (ISBD) date back to 1969, when the IFLA Committee on Cataloguing sponsored an International Meeting of Cataloguing Experts (IMCE). This meeting produced a resolution that proposed creation of standards to regularise the form and content of bibliographic descriptions.
The ISBD specifies the requirements for the description and identification of the most common types of published resources that are likely to appear in library collections. The ISBD also assigns an order to the elements of the description and specifies a system of punctuation for the description.

<table>
<thead>
<tr>
<th>ISBD for different types of library materials with year are furnished below:</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISBD(M) – 1971 (1973 in several other languages) [M = Monograph]</td>
</tr>
<tr>
<td>ISBD(S) – 1974 [S=Serial]</td>
</tr>
<tr>
<td>ISBD(M) revised to bring it into line with ISBD(G) and first revised edition was published in 1978.</td>
</tr>
<tr>
<td>ISBD(CM),ISBD(NBM) and revised ISBD(S) – 1977. [CM = Cartographic materials]</td>
</tr>
<tr>
<td>ISBD(A) for older monographic publications (antiquarian) and ISBD(PM) – 1980. [PM = Printed Music]</td>
</tr>
<tr>
<td>ISBD(M), ISBD(CM) and ISBD(NBM) republication – 1987 [NBM = Non Book Material]</td>
</tr>
<tr>
<td>ISBD(S) republication – 1988</td>
</tr>
<tr>
<td>ISBD(CF) – 1990 [CF = Computer Files]</td>
</tr>
<tr>
<td>ISBD(A) and ISBD(PM) – 1991</td>
</tr>
<tr>
<td>ISBD(G) republication – 1992 [G = General]</td>
</tr>
<tr>
<td>ISBD(CF) became ISBD(ER) – 1997 [ER = Electronic Resources]</td>
</tr>
</tbody>
</table>

To date, in this general revision project, ISBD(S) was revised to ISBD (CR) for serials (S) and other continuing resources (CR), and was published in 2002. A revised ISBD (M) was also published in 2002, and a revised ISBD (G) in 2004. ISBD (CM) and ISBD (ER) underwent the world-wide review process and were revised following that process, but were not finished at that time because work was begun on a consolidated ISBD, resulting from a decision by the Review Group at the Berlin IFLA Conference in 2003 to form the Study Group on Future Directions of the ISBDs.

The terms of reference of this Study Group were set out as follows:

- To consider the issues and values of an ISBD that combines into a single document provisions for the entire Family, with different chapters for the information specific to a particular type of material;
- To improve consistency of terminology and content throughout the ISBDs;
- To consider administrative issues related to the Review Group's growing workload; and
- To assign priorities to new and ongoing projects to ensure timely, balanced completion of Review Group's agenda.
- There was quick agreement on the feasibility and usefulness of producing a Consolidated ISBD. The Study Group was charged by the Review Group with preparing a definitive text. Its work has been guided by the following Objectives and Principles:
Objectives

- To prepare a consolidated, updated ISBD from the specialized ISBDs in order to meet the needs of cataloguers and other users of bibliographic information;

- To provide consistent stipulations for description of all types of resources, to the extent that uniformity is possible, and specific stipulations for specific types of resources as required to describe those resources.

Principles

- The primary purpose of the ISBD is to provide the stipulations for compatible descriptive cataloguing worldwide in order to aid the international exchange of bibliographic records between national bibliographic agencies, national bibliographic agencies and throughout the international library and information community (e.g. including producers and publishers).

- Different levels of cataloguing will be accommodated, including those needed by national bibliographic agencies, national bibliographies, universities and other research collections.

- The descriptive elements needed to identify and select a resource must be specified.

- The set of elements of information rather than the display or use of those elements in a specific automated system will provide the focus.

- Cost effective practices must be considered in developing the stipulations.

ISBD Specified Eight Areas for Bibliographic Data

1. Title and Statement of Responsibility Area
2. Edition Area
3. Material or Type of Resource Specific Area
4. Publication, Production, Distribution, etc., Area
5. Physical Description Area
6. Series Area
7. Note Area
8. Resource Identifier and Terms of Availability Area (for older monographic resources)

4. Searching Protocols/ Standards

4.1 Z39.50

There are several technologies available to implement real time research, like Z39.50, SRU (Search/Retrieve via URL)/SRW (Search/Retrieve via Web), CQL (Common Query Language) etc. These are basically
kinds of protocol on which application services can be developed. The 'http' (Hyper Text Transfer Protocol), is the protocol of the WWW that is most widely used for hosting different services including various web transactions and mail. However 'http' is not without drawbacks, when it comes to accessing more than one database using a single interface. For example:

1. HTTP does not support the concept of ‘session’
2. As it deals with understand data, it results in poor indexing and noise in the retrieval

The Z39.50 species a client/server-based protocol for searching and retrieving information from remote databases. In other words Z39.50 is a protocol which specifies data structures and interchange rules that allow a client machine (called an ‘origin’ in the standard) to search databases on a server machine (called a ‘target’ in the standard) and retrieve records that are identified as a result of such a search. This specification describes the application service definition and the protocol specifications for real time searching. “Z39.50” refers to the International Standard ISO 23950: “Information Retrieval (Z39.50): Application Service Definition and Protocol Specification),” and to ANSI/NISO Z39.50 – 1995. It is maintained by International Standards and Maintenance Agency, Library of Congress.

Basically, Z39.50 is designed to enable communication by specifying both a general framework for transmitting and managing queries and results, and syntax for formulating queries between computers, typically those containing huge bibliographic data, like library catalogues.

**Apprehensions about Z39.50**

It is still under development for different extensions as well for basic element mapping

1. Not widely used
2. It is too complex to implement
3. It is often deemed that is not required any more as we have web
4. It sometimes does not work due to complications in implementations

But,

1. It is fairly matured standard
2. Fairly widely implemented for LIS work
3. Organizations like museums, art galleries, archives have started using it. Latest version supports non-bibliographic information
4. It is still useful in web environment. In fact, Web provides access to more than one Z39.50 enabled backend databases
5. Supports maintenance of centralized union catalogues
4.2 Meta Data

The task of the information architecture is to create web sites where users can actually find the information they are looking for. Controlled vocabularies produce the metadata that are used to organize Web sites in two ways: (1) by providing structure for the navigational scheme of the page or site, “which should use unambiguous labels and where the primary organization is usually hierarchical,” and (2) by ensuring optimal performance of the searching system, “where search terms are selected and organized for tagging content and searching for them.” A controlled vocabulary in a Web information architecture context is primarily a mechanism for implementing the effective use of metadata.

The standard definition of metadata is “data about data”, but in practice it usually means “information about objects’... that is, information about a document, an image, a reusable content module, and so on... In general, metadata is best understood as ‘any statement about an information resource,’ regardless of what it is being used for, which metadata vocabulary is being used, and how the metadata is represented.” Taylor examined competing definitions of metadata and found that “what they all have in common is the notion that metadata is structured information that describes the attributes of information packages for the purposes of identification, discovery, and sometimes management.”

Taylor distinguishes three major categories of metadata: (1) administrative metadata, (2) structural metadata, and (3) descriptive metadata. Administrative metadata are most closely associated with the document as property and may include such features as the source of the document, document ownership and rights, legal information related to the document, locational information, document use statistics, and information regarding retention and preservation. Structural metadata have to do with the document as physical entity and may address such factors as technical documentation related to the document, species of the physical nature of the document, creation or revision date, version or edition control, information about the software or browser processes required to use the document, compression or resolution information, encryption or password access information, and search protocol identification. Descriptive metadata are those that identify the intellectual and conceptual characteristics of a document and include information directly of value to the user of the content of the document, such as title, responsibility, date of creation, institutional affiliation, authority, links to other documents, and access to document content through a taxonomy, keywords, or other items intended to facilitate retrieval and use.

Metadata Schema

Although standard markup languages such as HTML and XML provide for flexible use of metadata to describe Web content, the flexibility that is a positive characteristic of general purpose markup languages translates into a negative due to the absence of metadata standardization in markup languages. The solution to this is generally thought to be found in metadata schema. “A metadata schema consists of a set of elements designed for a specific purpose, such as describing a particular type of information
resource. Many competing metadata schemata have been developed, some designed to achieve general purpose use and some for highly specialized purposes. Metadata schema are intended to do for the organization and representation of information on the Web what bibliographic standards such as the Anglo-American Cataloguing Rules did for libraries.

Any metadata schema is a form of artificial language, but metadata schema differ from controlled vocabulary in that a metadata schema makes no attempt at being a complete listing of usable terminology. Metadata schema possess three essential characteristics: (1) syntax, (2) semantics (content), and (3) structure. Any of these characteristics may be implemented at a variety of levels ranging from fundamentally simple to extremely complex.

**Purposes of Metadata**

1. Resource description
2. Information retrieval
3. Management of information
4. Rights management, ownership and authenticity
5. Interoperability and e-commerce

**Metadata Elements**

1. Elements related mainly to the Content of the resource
2. Elements related mainly to the resource when viewed as Intellectual Property
3. Elements related mainly to the Instantiation

**1. Content**

- Title
- Subject
- Description
- Source
- Language
- Relation
- Coverage

- **Title of the Resource**
- **Subject, Keyword**
- **Annotation, Abstract etc.**
- **Resource (physical, digital) from which the current resource was derived, digitized etc.**
- **Language of the Resource**
- **Relationship to other Works**
- **Geographic or Temporal Coverage**

**2. Creator (Author, Creator)**

- Publisher
- Contributor
- Rights

- **Publisher (Person or Institution)**
- **Contributing Person or Institution**
- **Rights Management Statement (or Link to), Copyright**
OAI-PMH

Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH), is a protocol developed by Open Archive Initiative for harvesting metadata from the repositories who expose their metadata. This harvested metadata from various repositories is further stored to build services for providing search facility. It uses XML (eXtensible Markup Language) over HTTP. The main purpose behind the development of this protocol was to bring application-independent interoperability and extensibility. One of the simplest forms of interoperability among individuals Digital Library systems is the harvesting of metadata.

![Figure 1: Metadata Creation Outflow](image1)

![Figure 2: Quality Assurance Loop for Metadata](image2)
5 Formats

5.1 MARC

MARC is an acronym that stands for Machine-Readable Cataloguing. The MARC standards consist of the MARC formats, which are standards for the representation and communication of bibliographic and related information in machine-readable form, and related documentation. It defines a bibliographic data format that was developed by Henriette Avram at the Library of Congress beginning in the 1960s. It provides the protocol by which computers exchange, use, and interpret bibliographic information. Its data elements make up the foundation of most library catalogues used today.

The record structure of MARC is an implementation of ISO 2709, also known as ANSI/NISO Z39.2. The permitted tag numbers for this part of the record run from 010 through 999. MARC records are composed of three elements: the record structure, the content designation, and the data content of the record. The record structure implements national and international standards (e.g., Z39.2, ISO2709). The content designation is “the codes and conventions established to identify explicitly and characterize ... data elements within a record” and support their manipulation. The content of data elements in MARC records is defined by standards outside the formats such as AACR2, Library of Congress Subject Headings, and MeSH. (The latest format of MARC is MARC21 (MARC for 21st Century).

5.2 CCF

Genesis: As a direct result of the International Symposium on Bibliographic Exchange Formats with the initiation by the UNESCO General Information Programme in April 1978 in Taormina, Sicily organised by UNISIST International Centre for Bibliographic Description in collaboration with International Council of Scientific Unions Abstracting Board, IFLA and ISO, an Adhoc Group was constituted for developing the Common Communication Format (CCF). After prolonged deliberations and discussions by the experts, the Group decided:

1. That the structure of the new format would be in accordance with the ISO 2709
2. That the core record would consist of a small number of mandatory data elements essential to bibliographic description, identified in a standard manner.
3. That the core record would be larger in number by adding optional data elements, identified in a standard manner
4. That a standard technique would be developed for accommodating levels, relationships, and links between bibliographic entities
Need for CCF

The need of CCF is of paramount importance of two or more wish to examine records with one another. It cannot be achieved unless exchange purposes. If there is a single national standard exchange format, information interchange within that country is possible so also will be greatly facilitated both technically and economically. But on the other hand, if each nation's standard format is different from all others, then it will be more problematic and complex to have international information interchange among national bibliographic agencies because of the number of computer programs that must be written to accommodate the translation of records from one format to another.

The other crucial factor is to go for a common and compatible format is to achieve uniformity, standardisation and consistency among different kinds of agencies creating bibliographic records. The goals of some of these differ greatly from the goals of others. Abstracting and indexing agencies tend to operate differently from most libraries and must work within different limits and constraints. Hence, various kinds of rules for bibliographic description have come into common use which ultimately gives rise to the creation and distribution of widely varying and mostly incompatible bibliographic records contained within equally varied and incompatible formats.

Scope and Uses

The Common Communication Format (CCF) is designed to provide a standard format for three major purposes:

(i) To permit the exchange of bibliographic records between groups of libraries and abstracting and indexing services
(ii) To permit a bibliographic agency to use a single set of computer programs to manipulate bibliographic records received from both libraries and abstracting and indexing services.
(iii) To serve as the basis of a format for an agency’s own bibliographic database, by providing a list of useful data elements. To assist the development of individual systems, other UNESCO documentation will provide implementation notes for the CCF, and a guide for AACR2 cataloguers who use the CCF.

Structures

The record structure of the Common Communication Format constitutes a specific implementation of the international standard ISO 2709. Each CCF record consists of four major parts:
When building a physical volume (magnetic type or disk) of bibliographic records, consideration must be given to the structure of the volume.


The above two publications replace the second edition of the CCF published in 1988. The CCF provides detailed and structured method for recording data elements in a computer readable record for exchange purpose between two or more computer based systems. Though it is a good format but it is now less used due to non-revision.

6. Model

6.1 FRBR

The cataloguing environment today is global. The use of integrating search interfaces on the Internet and new techniques of record discovery and record import have given more realism to the utopian goal that a bibliographic resource shall only be described once.

Another important factor is the emergence of electronic publishing on the Internet during the 1990s. This area presents a whole raft of problems which cataloguers have not had to face before, most of them connected with the dynamic and volatile nature of digital publishing, which probably cannot be solved to any satisfaction until there is a general and - at least in principle - working equivalent of legal deposit for this publishing environment. On the other hand, electronic, or digital, publishing also offers new opportunities to work with the producers to generate bibliographic data directly from the full text of documents, and we witness today a better understanding from the producers' side of the importance of providing basic bibliographic data.
It is logical, under such circumstances, that cataloguing codes are analysed and revised, and this situation offers an opportunity to investigate whether harmonizing of different codes is achievable. In this context we find a third influencing factor, the existence of the report Functional Requirements for Bibliographic Records (FRBR), which was presented at IFLA in Copenhagen in 1997, and since then has inspired both a theoretical analysis of existing cataloguing codes, especially the AACR2, recently extended even to the MARC21 format, and experiments with database structures, in order to arrive at more user-friendly solutions.

From 1992-1995 the IFLA Study Group on Functional Requirements for Bibliographic Records (FRBR) developed an entity relationship model as a generalized view of the bibliographic universe, intended to be independent of any cataloguing code (e.g. AACR2, the German RAK [Regeln für die alphabetische Katalogisierung] and RICA [Regole Italiane di Catalogazione per Autore] or implementation. One immediate consequence of this development was the decision to suspend most revision work on the ISBDs while the FRBR Study Group pursued its charge to “recommend a basic level of functionality and basic data requirements for records created by national bibliographic agencies.” In 1998, the FBBR Study Group published its Final Report after its recommendations were approved by the IFLA Section on Cataloguing's Standing Committee. At that time the ISBD Review Group was reconstituted to resume its traditional work. As expected, the IFLA Section on Cataloguing's Standing Committee asked the ISBD Review Group to initiate a full-scale review of the ISBDs. The objective of this “second general review project” was to ensure conformity between the provisions of the ISBDs and FRBR’s data requirements for the “basic level national bibliographic record.”

The ISBD Review Group felt that it was essentially to clarify the relationship between the ISBDs and the FRBR model. The Review Group believed that development of a table to detail the relationship of each of the elements specified in the ISBDs to its corresponding entity-attribute or relationship as defined in the FRBR model would satisfy the need to make clear that the ISBDs and FRBR themselves enjoy a harmonious relationship. The document that develops the mapping entitled “Mapping ISBD Elements to FRBR Entity Attributes and Relationships” was approved by the Cataloguing Section’s Standing Committee on July 9, 2004.

The FRBR report itself includes a description of the conceptual model (the entities, relationships, and attributes or metadata as we’d call them today), a proposed national level bibliographic record for all types of materials, and user tasks associated with the bibliographic resources described in catalogues, bibliographies, and other bibliographic tools.
IFLA continues to monitor the application of FRBR and promotes its use and evolution.

The IFLA Cataloguing Section's Working Group on FRBR, chaired by Patrick LeBoeuf, FRBR offers us a fresh perspective on the structure and relationships of bibliographic and authority records, and also a more precise vocabulary to help future cataloguing rule makers and system designers in meeting user needs. Before FRBR our cataloguing rules tended to be very unclear about using the words “work,” “edition,” or “item.” Even in everyday language, we tend to say a “book” when we may actually mean several things. For example, when we say “book” to describe a physical object that has paper pages and a binding and can sometimes be used to prop open a door or hold up a table leg, FRBR calls this an “item.”

When we say “book” we also may mean a “publication” as when we go to a bookstore to purchase a book. We may know its ISBN but the particular copy does not matter as long as it’s in good condition and not missing pages. FRBR calls this a “manifestation.” When we say “book” as in ‘who translated that book,’ we may have a particular text in mind and a specific language. FRBR calls this an “expression.”

When we say “book” as in ‘who wrote that book,’ we could mean a higher level of abstraction, the conceptual content that underlies all of the linguistic versions, the story being told in the book, the ideas in a person’s head for the book. FRBR calls this a “work.”

The emergence of FRBR is proof of the need to apply a common conceptual framework to cataloguing processes. As Elaine Svenonius observes in a recent book, the emergence of global cataloguing makes an ontology necessary. FRBR now in itself is a factor driving the development. It has contributed to the theoretical understanding of the cataloguing activity among cataloguers around the world, and it has become a framework, or an inevitable point of reference, for catalogue revision projects undertaken since its publication.

On remit from the Joint Steering Committee of AACR, a working group has experimented with Expression based cataloguing, and their experiences are available on the home page of AACR. Their findings, so far, do not support the Expression oriented approach. Instead, they are turning back to the Manifestation, as the solid ground for a record, and envisage different means to derive the Expression and Work information as a distinct layer, when needed, by other methods.

FRBR might ‘provide the uniformity of perception needed to automate the operations involved in organizing information.’
Table 1: Dublin Core and FRBR Comparison

<table>
<thead>
<tr>
<th>Dublin Core</th>
<th>FRBR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ScholarlyWork</td>
<td>Work</td>
</tr>
<tr>
<td>Copy</td>
<td>Item</td>
</tr>
<tr>
<td>Agent</td>
<td>Corporate Body</td>
</tr>
<tr>
<td>isExpressedAs relation</td>
<td>'is realized through'</td>
</tr>
<tr>
<td>isManifestedAs relation</td>
<td>'is embodied in'</td>
</tr>
<tr>
<td>isAvailableAs relation</td>
<td>'is exemplified by'</td>
</tr>
<tr>
<td>isCreatedBy relation</td>
<td>'is created by'</td>
</tr>
<tr>
<td>isPublishedBy relation</td>
<td>'publisher' attribute of</td>
</tr>
</tbody>
</table>

References


8. Ibid, p. 34.

About Author

Dr. Kanchan Kamila, Deputy Librarian, The University of Burdwan, Burdwan
E- mail: Kanchan_lis@yahoo.com