Chapter – 4

AUTOMATING DOCUMENT DESCRIPTION PROCESS

4.1. INTRODUCTION

Automation implies the use of automatic equipments and machines to do routine and repetitive nature of works or operations with little or no intervention by human beings. In an automated library, computer is used to manage one or several key functions of the library such as acquisition, circulation, cataloguing, serial control and Online Public Access Catalogue (OPAC). Today, the scope of library automation has been extended which includes all modern technologies those are used by the library and information centres for collecting, processing, storing, retrieving, disseminating and transmitting all types of information at local, regional, national, and international levels.

The computers were first applied for automating document description vis-à-vis cataloguing procedures in the early part of 1960’s. A number of libraries in the United States of America (USA) were started producing computer generated printed card catalogue and Computer Output Microform (COM) catalogue in the later part of 1960’s. This has increased the efficiency and effectiveness of use of library catalogue. The introduction of offline, computer produced book and COM catalogue in the 1960’s was a remarkable success in the evolution of library catalogue in terms of its production and maintenance. Gradually computer system has been employed for the production of online public access catalogues (OPACs) in 1980’s. The
computer produced catalogues have solved many problems those are inherited in the production and maintenance of card catalogue. The use of computerized catalogue had eliminated many labour intensive task associated with the production of card catalogues. In computerized machine-readable catalogue, only one unit entry record has to be created for each item and then computer will expand this record into as many individual catalogue entries as necessary.

Introduction of OPACs in the libraries in 1980's made a profound impact on the user interface with the library catalogue. It extends the access point effectively by eliminating the physical constraints associated with the card catalogue. The online catalogue indexes a variety of additional data field of a record thereby extends the scope of searching as compare to the traditional card catalogue, which has fixed searching capability. Online technology has also had a significant impact on the way catalogue data is displayed and thereby replacing the conventional 'unit record' display of the card catalogue. These new capability of online catalogue has substantially changed the underlying structure of the library catalogue. The structure of the card catalogue is effectively predetermined by the format of 'unit record', which has a fixed structure for rendering heading and reference entries and filing of individual entries are made within a established sequence. With the introduction of online catalogue, the entire scenario has been changed. The online technology has provided the opportunity for extending access to the data stored in the catalogue and also it facilitates a wide range of display options. The web-based OPACs began to appear in the libraries in the late 1990's. The web OPACs serve as a gateway to the resources not only held by the particular library but also to the holdings of other
linked libraries and further to regional, national and international resources (Ramesh Babu and O'Brien: 2000: 316).

4.2 BACKGROUND OF AUTOMATED DOCUMENT DESCRIPTION

The process of automated document description was started in 1965 - 66, when the Library of Congress (LC) developed the MARC format. The MARC format brought a revolutionary change in automated document description, which enabled libraries to produce catalogues in machine-readable form. It is a framework standard to which cataloguing contents are to be added so that catalogue records can be handled by computers. The format is primarily designed as a means of sharing and exchanging bibliographic record between the libraries. It sets a standard for identifying, storing and communicating cataloguing data. The development of MARC promoted the concept of co-operative cataloguing and it allows libraries to create centralized library database for sharing cataloguing information.

MARC forms the foundation of moving libraries in to the computer age by providing a common syntax for recording and transferring bibliographic data between the computer systems of different libraries. It is the instrument in defining the concept of communication format. The development of MARC format by LC has influenced the other national and international organizations viz. International Organisation for Standardisation (ISO), International Federation of Library Associations and Institutions (IFLA), United Nation, Educational, Scientific and Cultural Organisation (UNESCO) etc. to developed other standard formats for construction of bibliographic record in order to meet the basic requirements of
library catalogues. These organizations have made independent as well as joint efforts in formulating appropriate standards to facilitate international co-operation and exchange of bibliographic records among different libraries. A brief discussion of the formats of automated document description developed by these organizations has been made in the next section.

4.3 BIBLIOGRAPHIC RECORD FORMAT

A bibliographic record is a group of related data elements pertain to a bibliographic item structured in a specific manner and created as a unit. The data elements are pieces of data, which describe or help to identify an item. The bibliographic record can be created for any independent bibliographic item such as books, journals, newspapers, audio cassettes, video cassettes, article from the journals, or any documents or other records of human communications. The bibliographic record pertain to a bibliographic item is consist of several data elements like author, title, publisher etc. which helps to identify the item. It is a collection of data elements, organized in a logical way, which describe and identify a bibliographic item. The bibliographic record contains data elements of distinctive features, which helps to identify the bibliographic item and thereby facilitate access to the document. The selection of data elements for the bibliographic record should be drawn up as per the general principles and guidelines given by the international bodies for standardization of bibliographic record.

The term ‘bibliographic’ record is used in the automated environment which has a similar meaning with the term ‘entry’. The distinction between the bibliographic record and entry is that the master bibliographic record is held in the machine in automated system from where different entries are generated by one or
more computer programmes. The entries might appear in printed bibliographies or they may be intended for screen display in Online Public Access Catalogues (OPAC).

A machine-readable bibliographic record is arranged according to a standard set of rules and represented in a particular format. The bibliographic record format is usually provided a formalized framework used to describe the arrangement and structure of computer readable records of various bibliographic items such as books, periodicals, journal articles etc. The data elements of a traditional bibliographic record have to be described, arranged or structured in a particular format for identification of data or data elements of the record for computer handling so that the bibliographic data or records can be transferred in machine-readable form and can be exchanged between different computer systems. The record structures of bibliographic record may be fixed or variable. The creation of a bibliographic record is an intellectual process, which requires following certain standard set of rules presented in a particular format that enables the computers to recognize data. The record format enumerates all components of bibliographic record necessary for describing and identification of all possible library materials. Standard for creating bibliographic records in the form of various cataloguing codes have designed by librarian and library scientists to ensure consistency in practice and to have maximum retrieval efficiency within a library or information service. ‘Bibliographic record not only describe and identify bibliographic item but also indicate or specify the subject of the intellectual work contained in those items’ (Kashyap:1993:295).

The objectives of bibliographic record format are:

- To ensure the international exchange of bibliographic records
To promote the conversion of the bibliographic record into machine-readable format

To provide standards for creation of databases so as to reduce incompatibility among different computer systems for sharing bibliographic record.

To promote library automation and also to access online library catalogues

Bibliographic record format is necessary to facilitate sharing and exchanging data in efficient and effective way. Creation of a standardized bibliographic database for sharing and exchanging of data contained in the files of a database through networks has become a common feature among the libraries and bibliographic agencies worldwide. The standardization of information processing and exchange procedure is very essential in order to facilitate international exchange of information. In the automated environment, standardization of content and structure of descriptions of bibliographic records is very important for merging and exchanging information originating from different computer systems.

Standardization in the format for the creation of bibliographic database is essential to facilitate exchange of data in efficient and effective way between the libraries. The adoption of different standards for creation of database creates incompatibility in sharing cataloguing data, which act as a major barrier in the use of bibliographic related information. Format compatibilities are necessary for computerized cataloguing data and these are being standardized by the international
organizations such as ISO, LC, IFLA, UNESCO etc. The formats developed by these organizations are being discussed in brief below.

### 4.3.1 The ISO 2709 Format

The International Organization for Standardization (ISO), founded in 1947, is an official international agency engaged in promoting the development of standardization and related activities in the world with a view to facilitate the international exchange of goods and services, and to develop co-operation in the spheres of intellectual, scientific, technological and economic activity. The ISO comprises of national standard bodies from more than 130 countries, one member from each country, and constitutes various Technical Committees (TC) to carry out the job of formulating standards in specific fields. The Technical Committee 46 (ISO TC/46) has been constituted for the library and information field. The scope of ISO TC/46 is to standardize the practices relating to libraries, documentation and information centers, indexing and abstracting services, archives, information science and publishing.

The international standard exchange format ISO 2709 has been developed by the ISO TC/46 in 1981 for bibliographic information interchange on magnetic tape. The format describes a generalized structure for the exchange of bibliographic records contained in databases of all kinds of document. This international standard only specifies the requirements of a generalized communication format, which will hold records describing all forms of bibliographical materials as well as related records. The format does not define the length or content of individual records and does not assign any content designators (i.e. tags, indicators or identifiers). These
specifications are to be assigned or developed by an internal format of bibliographic record, which may be developed by an information organization according to their local needs.

The ISO 2709 was developed from the record structure of LC MARC. The LC MARC, on the other hand, developed on the basis of an American standard ANSI Z39.2. The features added to the later version of ISO 2709 published in 1981 are that it permits alphanumeric tags and each entry in the record directory can be extended up to nine characters. Although the format has been designed for magnetic tapes, its structure may be used for other data carriers such as floppy diskettes, CD-ROMs etc.

i) Record Structure of ISO 2709

The ISO 2709 specifies that a record should be composed of the following four main elements and its general structure is presented in Fig. 4.1 below:

<table>
<thead>
<tr>
<th>Record Label</th>
<th>Directory</th>
<th>Variable Data Fields</th>
<th>Record Separator</th>
</tr>
</thead>
</table>

Fig. 4.1: ISO 2709 Record Structure

The record label includes data, which identify the type of record (i.e. monographs, serials etc.) and also contain information necessary for the processing of the record such as the total number of characters in the record and the length of various elements in the record.
The directory consists of a content designator for each data field followed by an indication of the position in the record where the data relating to that field start and the length of the field. If a field is repeated, two entries for that field are found in the directory, one for each appearance.

The record label and directory are both control segments, which are used to process data contained in the third segment, the variable data field segment. A record may be of any desired length. It ends with record terminator or separator.

### 4.3.2 The MARC Format

The LC initiated the MARC pilot project in 1965-66 with an aim of investigating the feasibility of producing cataloguing data in machine-readable form. A similar work was also undertaken by the Council of the British National Bibliography (BNB) in the United Kingdom, known as BNB MARC project, on the use of machine-readable data for producing the printed BNB. These parallel developments led to the Anglo-American co-operation on the MARC II project, which was initiated in 1968. The MARC II was the instrumental in defining the MARC as a communication format. It is the archetype of all subsequent MARC formats. The American National Standard Institute (ANSI), the British Standard Institute (BSI) and the International Organization for Standardization (ISO) later adopted the record structure of MARC II. MARC format adhere to the ISO 2709 record structure. The MARC format, both by its structure and the content designators, has brought a revolution in the creation of bibliographic database worldwide. Many national formats were later developed on the principles based on the MARC II format. MARC-II established certain principles, which have been
followed consistently over the years. MARC-II was instrumental in defining the concept of MARC as a communication format. In general terms, the MARC communication format is intended to be:

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

The success of MARC format was brought a keen interest to the other countries to develop their own national MARC format based on MARC. As a result in early 1970’s as many as 20 more MARC formats were developed in different countries whose data contents were found different owing to different national cataloguing practice and requirements. Few examples of different MARC formats developed in different countries are – CANMARC in Canada; AUSMARC in Australia; FINMARC in Finland etc.

i) MARC 21 Format

MARC became US MARC in the 1980s and MARC 21 in the late 1990s. MARC 21 is not a new format and it was evolved from the original LC MARC. After discussion and making minor changes to both USMARC and CAN/MARC to accommodate users’ specific need in both the format, the USMARC and CAN/MARC were harmonized into MARC 21 in 1997. The Network Development and MARC Standards Office at the Library of Congress and the Standards and the Support Office at the National Library of Canada are maintaining the MARC 21 formats. Input for development in the format can be provided by the users of MARC
21 from around the world, including libraries, library networks and utilities, and library system vendors. The LC maintains the MARC forum, an electronic discussion list for the formats that provides a conduct for broad, open discussion of proposed changes and other issues for all interested users around the world. Along with individual MARC user input, the Library of Congress and the National Library of Canada hold open meetings for incorporating changes to the MARC 21 formats.

MARC 21 is an implementation of the American national standard, information interchange format (ANSI Z39.2) and its international counterpart, Format for Information Exchange (ISO 2709). These standards specify the requirements for a generalized interchange format that will accommodate data describing all forms of materials susceptible to bibliographic description, as well as related information.

The MARC 21 format is defined for five types of data, which are widely used standards for the representation and exchange of information in machine-readable form. The five MARC 21 communication formats are:

- **Bibliographic Data**: It contains format specifications for encoding data elements needed to describe, retrieve, and control various forms of bibliographic material. The MARC 21 Format for Bibliographic Data is an integrated format defined for the identification and description of different forms of bibliographic material. MARC 21 specifications are defined for books, serials, computer files, maps, music, visual materials, and mixed material. With the full integration of the previously discrete
bibliographic formats, consistent definition and usage are maintained for different forms of material.

- **Authority Data**: It contains format specifications for encoding data elements that identify or control the content and content designation of those portions of a bibliographic record that may be subject to authority control.

- **Holdings Data**: It contains format specifications for encoding data elements pertinent to holdings and location data for all forms of material.

- **Classification Data**: It contains format specifications for encoding data elements related to classification numbers and the captions associated with them. Classification records are used for the maintenance and development of classification schemes.

- **Community Information**: It provides format specifications for records containing information about events, programs, services, etc. so that this information can be integrated into the same public access catalogs as data in other record types.

MARC 21 format is a set of codes and content designators defined for encoding a particular type of machine-readable record. Content designator is an inclusive term used to refer to tags, indicators and sub-field codes. The three types of content designators - tags, indicators and sub-field codes are the keys to the MARC 21 notational system. The MARC21 formats as a group serves as a vehicle for authority, bibliographic, classification, community information, and holdings data of all types. These formats are intended to be communication formats and are
primarily designed to provide specifications for the exchange of information between systems.

a) Record Structure of MARC 21

A MARC record consists of three main sections: the Leader, the Record Directory and the Variable Fields.

**Leader**: Each MARC record contains a leader consists of data elements to process the record. The data elements in the leader define parameters for processing the records. A leader composed of fixed field consisting of a total of 24 characters used to indicate length, status, type and bibliographic level of the ensuing records. Length indicates that total number of characters used in the record, status indicates whether it is new, changed or deleted record, type indicates whether it is printed or not while bibliographic level states whether the work is a monograph, part of a series, serial publication, a collection of manuscripts, pamphlets or other items catalogued as a single unit.

**Record Directory**: The directory contains the tag, starting location and length of each field within the record. It is an index to the location of various fields, both central and variable length, within the record. The number of fields in the directory will be equal to the number of fields in the record. As the number of variable length fields can vary, the length of record directory can also vary. It is at the same time a series of fixed fields, which contain the field tag, length of the field and starting character position of each of the major types of data in the variable fields. A field terminator is used to mark the end. The record directory facilitates the retrieval of selected fields from within a MARC record.
Variable Field: The variable data fields within each MARC record consists of indicators, sub-field codes, data elements and field terminators. Each variable field is represented by a 'tag' number and that tag is stored in the directory. For example — tag 100 for personal name, tag 110 for corporate name and 130 for uniform title heading etc. used preparing for main entry in a MARC record. Indicator is a code supplying additional information about the field and it is located at the beginning of the field. The variable fields follow the leader and the directory in the record and consist of control fields and data fields. Control fields precede data fields in the record and are arranged in the same sequence as the corresponding entries in the directory. The sequence in which data fields are stored in the record is not necessarily the same as the order of the corresponding directory entries.

b) Salient Features of MARC 21

If a library developed its own method of organizing the bibliographic information that does not use MARC format, it would be isolating its library, limiting its options, and creating much more work for the library. Using the MARC standard prevents duplication of work and allows libraries to better share bibliographic resources. The format enables libraries to acquire cataloguing data that is predictable and reliable. The MARC formats are communication formats, primarily designed to provide specifications for the exchange of bibliographic and related information between systems. Salient features of MARC are:

- MARC is a standard, which is used by most of the library automation systems. If a library use the MARC format, it will be able to move its
database from one library system to another without converting every
record into a new format whenever change to a new system.

- MARC allows sharing of records between the systems. There are more
  than tens of millions of MARC records available for copying. If a library
  use MARC format, it won’t have to make original cataloguing for every
  record in the collection. The MARC enables libraries to acquire
  cataloguing data that is predictable and reliable.

- MARC allows sharing of resources

- MARC allows development of union catalogues or linked in a virtual
  union catalogue, which make relatively easy to do Inter Library Loan
  (ILL).

- MARC allows data entry in UNICODE.

- It allows record linking.

- It allows migration from one system to another.

Using the MARC standard also enables libraries to make use of
commercially available library automation systems to manage library operations.
Many systems are available for libraries of all sizes and are designed to work with
the MARC format. Systems are maintained and improved by the vendor so that
libraries can take benefit from the latest advances in computer technology. The
MARC standard also allows libraries to replace one system with another with the
assurance that their data will still be compatible.
c) User of MARC 21 in India

Developing Library Network (DELNET) began using CCF in 1998. The CCF incorporates only necessary and sufficient fields and can’t be used internationally as a detailed bibliographic format. As a result for compatibility with international and national bibliographic database DELNET started using MARC 21 in 1999. As on December 31, 2003, DELNET has created 44,132 records for its Union Catalogue of books based on MARC 21 formats.

The INFLIBNET has decided to adopt MARC 21 formats instead of CCF because the MARC 21 format is comprehensive, constantly updates and revised and used majority of the countries in the world for exchanging data with other countries.

The National Library, India has decided to move from the UNIMARC format to the MARC 21 bibliographic formats to catalogue records in 2003. The National Library is using a US based integrated library software ‘VIRTUA’ which is fully supported the MARC 21 formats. The National Library has prepared conversion tables of UNIMARC records created for the Indian National Bibliography to MARC 21 formats in order to support the consistent and accurate mapping of UNIMARC data to MARC 21.

The MARC has been the most popular communication format which aims to provide universally accepted bibliographic records in machine-readable form. The MARC format adheres to ISO 2709 standard through which bibliographic records can easily be transferred from one system to another. The MARC 21 format is an implementation of the Information Interchange Formats ANSI Z39.2. The format also incorporates other relevant ANSI standards. The information retrieval protocol
ANSI/NISO Z39.50 supports the MARC record, which enables search and retrieval of bibliographic information over the Internet. The protocol can be used for searching MARC records from a Z39.50 client to a Z39.50 server. The MARC 21 format also fully supports the record transfer on Floppy Diskette which standardized the process of exchanging MARC records on diskettes rather than on magnetic tapes which is incompatible in the present generation micro computers.

The MARC standard enables libraries to make use of commercially available library automation systems to manage different library operations. The MARC 21 formats are comprehensive, constantly reviewing and revising by the MARBI (Machine-Readable Bibliographic Information) and the MARC Advisory Committee. Because of its constant updating and comprehensiveness, majority of the countries in the world are using MARC 21 for exchanging information with other countries.

4.3.3 UNISIST Reference Manual

The UNESCO in collaboration with the International Council of Scientific Union’s Abstracting Board (ICSU – AB) developed an international format ‘Reference Manual for Machine-Readable Bibliographic Description’ for description and international exchange of scientific information. The first edition of the manual was published in 1974, the second edition came out in 1981 and the third revised edition in 1986. The Reference Manual (RM) provides the essential elements for an adequate description and identification of bibliographic items. The objective of the RM is to provide a complete working manual for the international standardization of the form and content of computer readable bibliographic descriptions prepared by
abstracting and indexing services and others. The RM is intended for use by producers of computerized databases with a hope that its adoption will improve both intellectual and physical access to information.

The Working Group on Bibliographic Description of UNISIST (World Science Information System) and ICSU-AB worked together to develop a standard set of data elements for the exchange of bibliographic data for secondary information services in machine-readable form. The outcome of their joint efforts led to the design and publication of ‘UNISIST Reference Manual for Machine-Readable Bibliographic Descriptions, in 1974 by UNESCO.

The RM basically contains essential bibliographic data elements and their description and not providing any cataloguing rules for rendering data in specific form. It contains tables, which support essential and optional data elements for each relevant combination of bibliographic level and type of document. One of the significant features of RM is its explicit use of bibliographic levels and formal distinction of several levels of bibliographic description in the format make the identification and description of bibliographic item very easy. Though, the RM was conceived in practical as an exchange format but in actual practice it has also been used as a source for local bibliographic description manual and as an internal format.

The UNISIST International Centre for Bibliographic Descriptions (UNIBID) is maintaining the RM in support of UNESCO. An Advisory Committee is set up to monitor the work being done by UNIBID in maintaining the RM. The RM defines a minimum set of data element agreed by the abstracting and indexing services for
exchange of records. Many secondary organization utilized the RM additionally as cataloguing rules and as a source of bibliographic description.

This manual belongs to the category of the guides for making bibliographic references, which is usually developed by the editors of academic and professional journals working with abstracting and indexing services. It was designed for the Abstracting and Indexing Services which neither follow any standard application software package nor an established cataloguing rules. It embodies its own set of cataloguing rules, which differ from the international standard developed around ISBD and AACR2. The use of RM is not widespread.

4.3.4 Common Communication Format (CCF)

In 1978 the UNESCO under its PGI Programme sponsored an International Symposium on Bibliographic Exchange Format, which was held in Taosmina, Sicily, organized by the UNISIST International Centre for Bibliographic Descriptions (UNIBID) in co-operation with the ICSU-AB, IFLA and ISO. The purpose of the symposium was to develop a more flexible communication format that permits maximum compatibility among the existing bibliographic exchange format.

After the Symposium, the UNESCO/PGI formed an Ad-hoc Group for the development of a Common Communication Format that would be useful for universal exchange of bibliographic record. The Group decided to follow the following principles while designing the format:
• The structure of the format would conform to the international standard ISO 2709

• The bibliographic record would consist of a number of mandatory and optional data elements essential for bibliographic description for the purpose of exchange between two or more computer based systems

• A standard technique would be devised for accommodating levels, relationships and links between bibliographic entities.

The Ad-hoc Group compares the data elements and the content designator system of the format with the data elements of other standard formats. The format was published in 1984 under the editorship of Peter Simmons and Alan Hopkinsons and the second edition was published in 1988, with certain modifications. There are two versions of CCF: CCF/B for database of bibliographic records and CCF/F for database of factual records.

i) Purpose of CCF

The CCF was designed in order to meet the needs of all types of library and information centers for three major purposes:

• To permit the exchange of bibliographic record between groups of libraries and abstracting and indexing services.

• To permit a bibliographic agency to use a single set of computer programmes to manipulate bibliographic records received from both libraries and abstracting and indexing services.
• To serve as the basis for a format for an agency's own bibliographic database, by providing a list of useful data elements.

The chief purpose of CCF is to provide detailed and structured method of recording a number of mandatory and optional data elements in a machine-readable record for exchange purpose. It is designed to meet the requirements of all types of libraries and information organizations allowing them to design their database as per their local requirements.

A bibliographic agency may choose to use for local implementation of any format or formats, which are convertible by computer programme to the CCF. Alternatively, any agency may wish to use the CCF directly, to contain the bibliographic records it creates. This is one of the main purpose of CCF.

4.3.5 UNIMARC Format

Despite co-operation, several versions of MARC formats were emerged in 1970's whose data contents were found different owing to different national cataloguing practice and requirements. In early 1970's an extended family of more than 20 MARC formats, derived from the same MARC family, were came out. Difference in data content in different MARC format made it essential to edit the records before they can be exchanged with others. The proliferation of national MARC formats has been a major concern for exchange of data between the bibliographic agencies for many years. One solution to the problem of incompatibility resulted from the proliferation of MARC format was to create an Universal MARC format (UNIMARC) which would accept records created in any MARC format. The aim was that the records in one MARC format could be
converted into another MARC format with the intention that each national agency would need to write only two programmes, one to convert to UNIMARC and one to convert from UNIMARC, instead of one programme for each other MARC format. To overcome the incompatibility among the different national MARC formats, IFLA developed the UNIMARC format which was published in 1977. The primary purpose of UNIMARC is to facilitate the international exchange of bibliographic data in machine-readable form between national bibliographic agencies. A second edition of UNIMARC format was published in 1980, which was restricted to monographs and serials only. The necessity of expanding UNIMARC to cover all different sorts of documents was envisaged in the mid 1980s and in 1987 a new description of the format was published as a UNIMARC manual. It was also explicitly mentioned in the manual that ‘UNIMARC’s objectives would not only be the carrier, or universal MARC for conversion purposes, but also the model for the development of new machine-readable bibliographic format’ (Campos: 1995: 449). The second edition of the UNIMARC Manual was published in 1994.

UNIMARC conforms to the ISO 2709 and ISBD standard, which can be used as a model for the development of new machine-readable bibliographic database. The creation of an international MARC format (UNIMARC) would accept, in principle, record created in any MARC format and act as a common format in terms of conversion. The important feature of UNIMARC is that it should have one set of tags or sub-field identifiers covering all materials. Since 1977 several national libraries have undertaken projects to convert from an existing national format to UNIMARC or have simply adapted UNIMARC as their national format. It covers monographs, serials and cartographic materials, music, sound
recordings, graphics, projected and video material, with provisional fields for computer files.

i) Elements of UNIMARC Format

The UNIMARC format, like any other format consists of three elements:

- Record Structure
- Content Designation
- Data Content

• **Record Structure:** Record structure refers to the process in which the data elements in a record are identified. It is designed to control the representation of data elements in a record by storing them in the form of strings of characters known as fields, which are identified by three numeric characters called tags. The record structure established by UNIMARC is an implementation of international standard format for bibliographic information interchange on magnetic tape (ISO 2709 : 1981).

• **Content Designation:** Content designation refers to the codes and conventions established explicitly to identify and further characterize the data elements within a record and to support manipulation of that data. It is the means of identifying data elements or providing additional information about a data element. It is consists of tags, indicators, and sub-field identifier which support the manipulation of data for a variety of purposes such as:

  - To provide multiple access points for searching
• To allow the typography and layout to be varied

• To permit certain elements of the record to be omitted where this is required

**Data Content:** The content of the data that comprise a UNIMARC record is the data, which is stored in the fields within the record. The data found in the record are comprises of coded data and bibliographic data. In UNIMARC record the functional block 0xx - 1xx holds the coded data while 2xx - 8xx contains the bibliographic data. Coded data is used to represent the items such as control numbers, publication type, main language of the text etc. Bibliographic data is usually defined by standards outside the formats, such as ISBD, Subject Heading List, AACR or other cataloguing rules, subject thesauri and classification schedules used by the organization to create record.

### 4.4 METADATA FORMAT

Literally metadata means structured data about other data, information that qualifies other information. It is structured data describing the attributes and characteristics of an entity, usually digital web resources, to facilitate effective indexing and retrieval of information available on the World Wide Web (WWW). It is used in the Internet context, to refer to data describing attributes of a digital resource. It helps to organize and discover information resources by supporting the precision of identifying, locating, evaluating, selecting, accessing information about the resources available on the local intranet or global Internet.
In Internet, metadata is a structured description of attributes of a document like object (DLO). Metadata structure is defined in formats, which again are derived from cataloguing rules. Metadata may support a number of functions, such as location, discovery, documentation, evaluation, and selection of DLO’s.

The libraries have been dealing with metadata since long time, even before the emergence of Internet. The library catalogues could be an example of metadata as it contains structured data, which describe a document properly in order to retrieve it. Any form of catalogue, contents list, inventory, review, abstract or index is a collection of metadata, as long as it may be used to describe other sources or documents. Gradmann (1999:88) in his article Cataloguing vs. Metadata cited a quotation from R. Heery paper on Metadata Formats published in 1996 that ‘the familiar library catalogue record could be described as metadata in that the catalogue record is data about data’. However, metadata is machine understandable information about web resources. There are many sharp differences between metadata and catalogues. The differences get more clearer if we observe the Dublin Core workshop on metadata initiative which objective was that the authors and publishers could supply their own metadata records while creating their works. It means that the originator or producer of meta information are not basically the library cataloguer. They are either author or publisher or bookseller. The context of usage of metadata are different from library catalogues, they are typically not created by professional cataloguer, they are intended to be produced more efficiently than catalogue records and they cover a specific kind of materials i.e. electronic resources.
The actual Dublin Core (DC) metadata records may have some sorts of similarity with cataloguing records, which can be mapped to a MARC format but the context of production and usage of DC record is substantially different with that of traditional catalogue records. 'Considering the process of metadata creation to be some kind of simplified cataloguing thus probably would be a serious misunderstanding' (Gradmann : 1999 : 89). Heery (1996) pointed out differences of metadata with catalogue as metadata also differs from traditional catalogue data in that the location information is held within the record in such a way to allow direct document delivery from appropriate application software, in other words the record may well contain detailed access information and the network address.

i) Dublin Core Metadata Initiative (DCMI)

The schema that has been most widely used in recent years to describe Internet resources is the Dublin Core Metadata Element Set (DCMES). The Dublin Core (DC) has been developed by an informal group of librarians, networking people, and content specialists. The DC is a core list of metadata elements agreed in a metadata workshop in March 1995 held at Dublin, Ohio, convened by Online Computer Library Centre (OCLC) and National Centre for Super Computer Application (NCSA) in order to solve the problem of indexing web resources and to set up a standard. The DC was proposed as the minimum number of metadata elements required to facilitate the discovery of document like objects (DLO) like HTML (Hyper Text Markup Language) pages, PDF (Portable Document File) files, graphic images etc. available in Internet.

The objective of DC was to define a simple set of data elements so that authors and publishers of Internet documents could create their own metadata records without
the need of extensive training. The standard provides a minimum set that may be augmented with specialist data by publishers or metadata producers. A syntax for expressing DC metadata in HTML (Hyper Text Mark-up Language) has also been developed later. The DC consists of 15 core elements divided into three classes shown in Fig. 4.2 below.

<table>
<thead>
<tr>
<th>Content</th>
<th>Intellectual Property</th>
<th>Instantiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Creator</td>
<td>Date</td>
</tr>
<tr>
<td>Subject</td>
<td>Publisher</td>
<td>Type</td>
</tr>
<tr>
<td>Description</td>
<td>Contributor</td>
<td>Format</td>
</tr>
<tr>
<td>Source</td>
<td>Rights</td>
<td>Identifier</td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coverage</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4.2: Dublin Core Element Set.

These elements are relatively easy to define and can be understood easily. All elements are optional and repeatable. The DC has been adopted by a number of communities such as museums, libraries, government agencies, and commercial organizations. It has achieved international recognition for interdisciplinary resource description.

The DC has been developed so far informally through a series of workshops and five workshops have already been conducted at several places to improve the syntax and semantic aspects of DC. The Workshops were held at Dublin, Ohio; Warwick, England; Dublin, Ohio again; Canberra, Australia; Helsinki, Finland. The sponsors of
these meetings were OCLC, NCSA, the National Science Foundation and the Coalition for Networked Information in the United States, United Kingdom Office for Library and Information Networking (UKOLN); the Australian National Library; the National Library of Finland, and many others. Stuart Weibel of OCLC has been the leader of the effort since beginning.

It is to be mentioned that the World Wide Web Consortium (W3C) has been working on a format, which they call as Resource Description Framework (RDF), a set of standards for supporting the exchange of metadata on the web. The W3C has published the public specification of a general model for metadata with its underlying encoding syntax in the eXtensible Mark-up Language (XML). The XML is expected to become a primary format for document encoding on the web. XML has been introduced as the substitute of HTML as a way to distribute document on the web. Different information suppliers are using this format as a way to integrate their data and documents with those owned by libraries and documentation centre. The RDF defines the general mechanisms for attaching metadata of all kinds of web pages composed using the new XML defined by the W3C, including DC metadata. This work is likely to be central to facilitate the large scales use of DC metadata within the web.

4.5 AUTOMATED DOCUMENT DESCRIPTION: THE RECENT TREND

We have noticed a very fast technological development in the field of computer and communication technology since the development of MARC in 1965-66. When the MARC was developed, the Internet was just mere an idea, a very few people had access to web and digital resources were very limited in numbers.
MARC is a forty years old format developed as a record structure based on ISO 2709 for exchanging records between different computer systems. This format is exclusively used for creation of bibliographic data as the record structure of the format is based on ISO 2709, which is designed for bibliographic data and not for indexing full text bibliographic documents available on the Web. The MARC formats are rely on AACR for the description of their data elements. AACR has had a very strong influence on MARC and MARC in turn has had a very strong influence on the development of system that supports AACR so that their databases are compatible with each other. There are now huge numbers of bibliographic records around the world in the MARC format. These records can easily be transferred from one system to another as they all use ISO 2709 format and AACR for description of data elements. Strict adherence of these standards in MARC formats have ensured bibliographic data transfer between systems and it would not have been successful without the standard like MARC. This standard has been used for a long time and strict adherence to these standards have led to some fossilization of bibliographic data processing which has provided a stable environment. But if do not have a standard that is universally accepted, we could have probably greater variation of practices than we have now. It is very encouraging to know that the Joint Steering Committee (JSC) is working towards a new edition of AACR in order to accommodate new rules for describing newly emerging multi-media resources. The new edition of AACR will be published in 2008 under a new title 'Resource Description and Access' (RDA).

The rigidity in standards of the MARC format makes them an unsuitable structure for the World Wide Web for accessing full text bibliographic record. The
massive growth of Internet in 1990's had brought tremendous increase of digital resources on the web and these digital web resources become major sources of information for users, which can be accessed through the Internet. The amount of digital web resources such as electronic text, images, audio and video files etc. are increasing on the web constantly, which play a crucial role in fulfilling the information requirements of the users. As the trend is moving towards accessing full text digital Web resources and the MARC format's unsuitability in the Web environment, the organization and cataloguing of these newly emerging digital web resources require new tools and techniques. The failure of the traditional communication format like MARC for describing Web resources has led to the development of new format by the Internet Community which is commonly known as metadata format. These recent formats are generally called DTDs (Document Type Definitions) encoded usually in Mark-up language like SGML (Structured Mark-up Language) format. Significance of metadata has enhanced in the context of enormous growth in the volume of information on Internet.

Many organizations have shown keen interest on metadata and developed their specific metadata schema in order to serve their specific purposes only. Some references may be made in this regard such as –Government Information Locator Service (GILS) developed by US Federal Government in order to provide the general public and its own employees with a device for locating information generated by many government agencies; Text Encoding Initiative (TEI) for humanities related data; Encoded Archival Description (EAD) for describing archival materials etc. Keeping in mind the necessity of a standard general metadata schema, which can be used by different communities such as libraries, museums,
archives, OCLC in association with National Centre for Supercomputer Application (NCSA) on March 1–3, 1995 convened an invitational metadata workshop in Dublin, Ohio to develop a general metadata schema which is commonly known as Dublin Core (DC) Metadata Element Set. The DC contains 15 data element sets, which can be used by different communities viz. libraries, archives, museums and publishers to describe a broad class of information object. It is now emerged as a standard metadata schema widely used by the Internet Community.

4.6 AUTOMATING UNIVERSITY LIBRARIES IN INDIA: THE INFLIBNET INITIATIVE

The process of development of in-house library databases in the university libraries in India was started during 1990's when UGC established INFLIBNET centre in 1991. The centre became an independent autonomous inter-university centre in 1996 with its headquarter located at Gujrat University campus, Ahmedabad. The purpose of establishment of INFLIBNET centre was to promote automation and creation of in-house databases in the university libraries in India so as to create union databases of holdings of university libraries in India for the purpose of sharing resources by establishing a single library network linking the library and information centres in universities, deemed to be universities, colleges, UGC information centres, institutions of national importance, R&D institutions etc. The centre plays a very important role for automating university libraries in India and also to create an IT culture in the universities. With the support of UGC, the Centre has spent several crores of rupees by giving initial grants and also subsequent recurring grants for five years to the universities in India to procure hardware and software for library automation activities. It funded 142 universities so far of rupees
6.5 lakhs each for automation and networking. The beginning of the university library automation has been made by the INFLIBNET centre by providing necessary funds, technical guidance, developing Software for University Libraries (SOUL), organizing training programmes for the professionals etc. INFLIBNET is set out to be a major player in promoting scholarly communication among academicians and researchers in India (Murthy and Cholin : 2003).

4.6.1 Union Catalogue

INFLIBNET centre has taken the initiative of building union catalogues of academic libraries in India. The union catalogue contains bibliographic description of location and holding information for books, serials and thesis in all subject areas held in university libraries in India.

The union catalogue of INFLIBNET integrates a group of individual catalogue into one single place in order to facilitate identification and location of records. The union database developed by the INFLIBNET serves as a key to resource sharing among university libraries, which represent the holdings of different type of materials owned by the participating university libraries in India. It provides vital linking to individual library collections and functions as a central pool of data, which can be used for collection development and Inter Library Loan (ILL) by participating libraries.

INFLIBNET centre has been developing bibliographic union databases on books, serials and thesis and also non-bibliographic databases on research projects, experts etc.
4.6.2 Software Development

To promote automation in the university and college libraries, INFLIBNET centre has developed a software SOUL (Software for University Library). It has been designed on Window based Client-Server Architecture using MS-SQL Server, which imparts extra strength to storage capacity, multiple access to the databases, various levels of security, back-up and re-storage facilities etc. It also provides Web access. SOUL supports barcode technology, international standards such as MARC 21, CCF, AACR 2, ISO 2709 etc. The software is very economical and it has more than 450 installations so far.

4.6.3 Support in Manpower Training

The centre has been organizing training for the professionals working in the university and college libraries in India to promote use of IT in these libraries. The centre under INFLIBNET Regional Training Programme on Library Automation (IRTPLA) has been organized more than 35 regional training programme in collaboration with the universities across the country to train college librarians and more than 800 library professional have been trained so far under these training programme. The centre conducting an annual two days convention “PLANNER” (Promotion of Library Automation and Networking in North Eastern Region) to provide special attention to the problems and issues of university and college library automation in North Eastern states. The first and second PLANNER was held in 2003 and 2004 at North Eastern Hill University, Shillong and Manipur University, Imphal respectively. The PLANNER 2005 was held at Assam University, Silchar in November 2005.
4.6.4 Bibliographic Format and Standard

To maintain consistency and quality in databases, a Task Group comprising experts was constituted by UGC with a responsibility to design the standard to be followed by the participating libraries. The group designed the standard under the document titled "INFLIBNET Standards and Guidelines for Data Capturing" which is made available to all participating libraries for creating standard records in the database. The standards designed by the INFLIBNET centre are to be followed by the participating member libraries in order to create quality records in the databases.

i) Cataloguing Standard

In order to make the database the best possible database, which can be accessed easily, it is very essential that the participating libraries should input only accurate and standard cataloguing data to the union databases of INFLIBNET. For this, each member library has to accept the responsibility of abiding by the standards adopted by the INFLIBNET centre for the member libraries. For the content of the information, the standard recommended are:

- Anglo-American Cataloguing Rules, 2nd ed., Current Revision for bibliographic description
- Library of Congress Subject Headings (LCSH) to assign subject headings.

In spite of the common standards for all member libraries, it is found that catalogue records in the database do not conform to one specific cataloguing standard. These records exhibit a variety of cataloguing standards because of local
variations of the participating libraries. Lack of consistency, typographic errors, uniformity, records contain in various format – all these factors are major concern for INFLIBNET, which leads to much complication and editing of records in the union databases.

ii) Communication Formats

For converting printed manual catalogue into machine-readable bibliographic records, it is essential to follow certain international standard communication format by the libraries. CCF was the initial choice for INFLIBNET databases and all the participating member libraries have been advised to use this format for conversion of printed catalogue into machine-readable form. Simplified guidelines regarding the use of this format have been distributed to all participating libraries by the centre. Later on, the centre decided to adopt MARC 21 because of its comprehensiveness and wide use around the world. For this, the INFLIBNET developed the new interface for transfer of data from CCF based records to MARC 21 and vice versa.

4.6.5 Retrospective Conversion

The INFLIBNET has initiated a major project on “Retrospective Conversion” in 2000 and selected five major libraries under this project with an expectation to avoid duplication of work and also to achieve the expected level of computerization. The university libraries included under this project are –

- University of Bombay, Mumbai
- University of Madras, Chennai
There are approximately 10 millions of documents which include text books, reference books, PhD thesis, conference proceedings, serials, technical reports etc., available in the 142 universities in India and out of these 50 lakhs are already converted in to machine-readable form. Out of these 50 lakh records, it is estimated that 45% records are duplicated (Prem Chand... et al: 2003). Under this project these five libraries will create good quality records of their entire holdings as per the standards recommended by the INFLIBNET and contribute these records to union databases created at INFLIBNET. The main objective of this project is to use these high quality records created by the above libraries for Retro-conversion and shared cataloguing of other member libraries, thereby minimizing the laborious efforts and cost involved in retro-conversion (Murthy and Cholin: 2003).

4.7 AUTOMATING DOCUMENT DESCRIPTION PRACTICES IN UNIVERSITY LIBRARIES IN NORTH EAST INDIA

The exponential growth of literature in various physical forms and users varied demands for information have forced the information professionals to look for effective and efficient methods for processing, storing, and retrieving of information. The enormous workload, the ever-increasing number of documents and their nature, the astronomical growth of information and their characteristics – all these factors have posed several problems for efficient management, storage, and retrieval of information in the university libraries in North East. It has become
imperative and inevitable to use computers in the university libraries in North East for processing, storing, and retrieving of information efficiently. The advances in Science & Technology (S&T) and the impact of Information Technology (IT), together, led to the tremendous growth and diversification of knowledge and information in every field. The growth of literature is so high that it becomes impossible for any library, whether big or small, to collect and catalogue all materials that are useful to its users. Automating document description process could provide solution to that problem. Libraries are now realizing that the only way through which they could satisfy the users need is to automate and to participate in resource sharing.

The progress in automated document description in some of the university libraries in North East is found to be very slow. In order to have a bird eye view on the progress of automation in the university libraries in North East, an investigation is carried out in eleven university libraries and the only IIT library of the region by distributing questionnaires and paying personal visit to some of the libraries. The results of the investigation have been discussed in Chapter - 5 (Section 5.3.2)

4.7.1 Library Application Software

The university libraries of North East have been used three different types of library application software viz. SOUL, Libsys and CDS/ISIS for the creation of databases (See Chapter – 5, Section – 5.3.2, Table: 5.14). The basic features of these software have been discussed in brief.
i) SOUL (Software for University Libraries)

The advancement in computer and communication technologies have brought revolutionary changes in the libraries for information acquisition, processing, storage, retrieval and dissemination. Keeping in view the latest development in Information Technology (IT), the INFLIBNET centre has developed a state-of-the-art library automation software, the SOUL, to facilitate automation function of the university, college and other participating libraries of the INFLIBNET programmes. SOUL has been designed on Window based Client-Server Architecture using MS-SQL server as back end tool, which imparts extra strength to storage capacity, multiple access to the databases, various levels of security, back-up and re-storage facilities. The software provides a total solution to the university library automation as it has been designed after a comprehensive study of different library related functions, practiced in university libraries. It provides web access and also supports barcode technology to generate labels. The software designed on the platform, which supports international standards such as MARC 21, CCF, AACR 2, ISO 2709 etc. Soul is very user friendly and economical software as it has been developed and distributed by non-profit making organization INFLIBNET. Strong features of the software are:

- It is window based user friendly software, very easy to work, not required a comprehensive training.
- It is integrated software having six modules viz. Acquisition, Cataloguing, Circulation, Serial Control, OPAC and Administration.
- The screens are well designed with logically arranged functions and extensive help messages.
- Designed on window based client-server architecture.
- It has bar coding and web access facilities.
- Multi-user and multi language facilities
- Supports internationally known standards such as MARC 21, CCF, AACR 2, and ISO 2709.
- OPAC is very user friendly and accessible over the web using GUI (Graphic User Interface) browsers.
- Cost of the software is comparatively low in comparison to other commercially available software in the market.
- Manual is available in different regional languages.
- It includes Boolean logic for searching information in the database.
- It has the facility to display and print record as per AACR II format
- Easy and quick searching facility and search result can be saved and print.
- User can access his/her current status by entering borrowing number

To strengthen the SOUL software after sales services, the INFLIBNET centre has established service centres at Mumbai, Patiala, Ahmedabad, Hyderabad, Bangalore, and many more service centres are likely to be established in different places of the country. The six university libraries of the region are using the software and Gauhati University library is planning to install the SOUL.
ii) Libsys

Libsys is an integrated library management software package designed and developed by Libsys Corporation, New Delhi in the year 1992 - 1993. It is a fully integrated multi-user library system designed to run on wide spectrum of hardware/software platforms in Client-Server environment. The software is very easy to operate and the library staff need not to go for extensive training to handle the software. Some of the strong features of the software are that it does not requires maximum data entry; it integrates various functions of the library and it has a very powerful search and query facilities. Libsys is built around its own centralized bibliographic database based on ANSI Z39.50 format which supports almost all activities relating to acquisition, cataloguing, circulation, serial control, article indexing, and it has also a powerful and user-friendly OPAC interface.

The acquisition system deals with ordering of library materials, receive monitoring, invoice processing and accessioning. It also maintains expenditure and budget analysis by respective account heads.

The cataloguing system has a powerful data entry facility which provides option to accept data in different standard machine-readable format such as CCF (ISO 2709); MARC (ANSI Z39) etc. The system makes available the catalogue of the holdings of the library online for ready references. It provides facilities to generate bibliographies, Current Awareness Services (CAS), Selective Dissemination of Information (SDI) and export/import of bibliographic data in standard exchange formats, meeting specific requirements of a library.
The circulation system maintains the library membership records, status of the memberships and library collection meant for circulation. Circulation related all functions like checking of status of a book; overdue; membership status; bindery record management; display of recent addition in the library etc. can be performed in this system.

The serial control provides control of periodical subscription and schedule checking of arrival of individual issues of different titles. The system provides complete budgetary control of serials by maintaining records of budget sanctioned for serials under different categories, amount allotted and expenditure made on purchasing journals for different fields. It also handles the serials, which may receive on grants or exchange. The article indexing and abstracting provides facility to create separate article database from which different services like SDI, documentation, bibliographic etc. can be provided. In this database, the article records can be added, deleted and modified as per requirement of the library.

The OPAC provides online searching facilities to users to search information in a variety of ways such as using Boolean search operators, free text searching etc. The OPAC also provides the periodic list of recent additions to library collection and allow users to find material issued out to them as well to put materials on reserves etc.

In India, the Libsys software was being used in the libraries which started automation during 1994 –1999 viz. Jawaharlal Nehru University (JNU), IITs, Indira Gandhi National Open School (IGNOU) etc. as during that period there was very few integrated library software package available in Indian market. The initial cost
of the software was Rs.1 lakh to 1.5 lakh, which is now crossed Rs. 2 lakh for a standard version having customization facility @ Rs.20,000/- per module. The software had tried a lot to capture the Indian market and simultaneously invested in its research and development but failed – a) in controlling the cost of the software; b) in meeting the local needs (customization); c) lack of multi-lingual facility (incorporated in later version) and d) after sales service.

The Tezpur University Library; North Eastern Hill University (NEHU) library; the Arunachal University Library and the Indian Institute of Technology (IIT), Guwahati have been using this software.

iii) CDS/ISIS (Computerised Documentation System/ Integrated Set of Information System)

The micro CDS/ISIS is a menu driven generalized information storage and retrieval system, which was designed specially for the computerized management of structured non-numerical databases containing mainly textual data. It is developed by UNESCO and distributed in India almost free of cost (Rupees 1,500.00 only charged for the manual) by National Information System for Science and Technology (NISSAT), Department of Scientific and Industrial Research, Govt. of India, New Delhi.

The lack of suitable software for information storage and retrieval was a major hurdle for computerization of library and information services in the developing countries like India. The introduction of CDS/ISIS by UNESCO in 1985 solved many difficulties in library automation faced by the libraries specially for choosing suitable software. The libraries were shown keen interest in choosing this
software as it was being distributed free of cost to the libraries and the software has very strong features, which allowed libraries to develop the library databases as per the users requirements. The database developed in CDS/ISIS, using standard exchange format like CCF, can be exchanged or merged with other database. In India, the CDS/ISIS was used widely for automation by the libraries in 1990’s when they first started the library automation process. The INFLIBNET centre was also initially used this software for data capturing and storage before the development of SOUL.

This software consists of a set of programmes written in Pascal language, which is designed specially for computerized management of structured non-numerical databases whose major constituent is text. The major advantage of this software is that the same set of computer programme is able to manipulate an unlimited number of databases each of which may consist of completely different data element. CDS/ISIS can handle variable data field and record length, which fulfills one of the major requirements of bibliographic databases. It is a menu driven user-friendly system. It can support up to 999 fields/records. ‘CDS/ISIS has many additional capabilities like -- powerful search language; worksheet design for data input and data updating, formulation of output/display formats, defining and searching on sub fields, powerful indexing capabilities, handling of repeating groups and multi lingual menu/messages for English, French, Italian and Spanish’ (NAIR 1992:179).

Some of the important features of CDS/ISIS software are given bellow:

- Database can be designed as per users selected data elements.
• Creation of new records and modification, correction, or deletion of existing records can be done in the database.

• Automatic creation and maintenance of fast access files such as the inverted file using different techniques to each database.

• Retrieval of records using sophisticated search expressions including Boolean logic viz. 'AND', 'OR' & 'NOT'.

• Sorting of records in a desired sequence.

• Run over LAN (Local Area Network).

• Display records according to users defined formats.

• Printing out partial or full listing of the record of any given database.

• Exchanging or merging of records of two or more databases whose records are compatible in ISO 2709 format.

• A wide range of services such as bibliographic, SDI, CAS, document indexing etc. can be provided.

• It supports ISO 2709 format structure for data representation, which enables healthy transmission of data among the MARC family of databases.

• Capacity of the software can be enhanced through advance programming written in CDS/ISIS Pascal which is an unique feature of the software.

The software CDS/ISIS was been used by seven university libraries of this reason viz. K.K.Handique Library, G.U.; Dibrugarh University Library; Assam agricultural University Library; Manipur University Library; Tezpur University library; Assam University Library; and Nagaland University Library. Out of these seven university libraries, Dibrugarh university library and Manipur university library have already converted the databases created in CDS/ISIS in to SOUL successfully. The Assam University Library and Tezpur University Library both have ceased entering new records in the CDS/ISIS database as they purchased SOUL and Libsys software later for carrying out various house keeping operations.
of the library and using the CDS/ISIS database in parallel to the new database created in SOUL and Libsys. The K.K.Handique library, G.U is at present using CDS/ISIS software and the library is planning to purchase SOUL very soon. Preparation of data entry sheet is going on.

It is found that databases created by the university libraries in this region using CDS/ISIS were later either converted in to SOUL or stop addition of new records in the existing CDS/ISIS database. In spite of having strong features, the CDS/ISIS is not preferred by the university library because the software does not support integrated housekeeping operations of the library, multi-user facility is not available and also handling of this software requires technical knowledge and experience of computers.

4.8 CONCLUSION

The machine-readable bibliographic record formats are meant for exchange of bibliographic data between different systems. The bibliographic record format, therefore, should be conformed to a common international standard and it should be acceptable to exchanging agencies. The format should be hospitable to all kinds of materials an agency wishes to exchange and it should be flexible enough to cope up with the needs of many different software systems. It should facilitate the exchange of bibliographic data, which are to be used in wide range different bibliographic applications - from the production of traditional catalogue cards to production of records in databases to be used for online access. But lack of an international agreement on standards for exchanging bibliographic data seems to be a major problem to the system analysts who write programmes to convert records from an exchange
format in to a systems internal format. Although IFLA developed ISBD, it is not accepted by many organizations outside the sphere of libraries. The IFLA through its several programmes contributed a lot to achieve the uniformity in bibliographic description at national and international level in order to bring the concept of Universal Bibliographic Control (UBC) more closer. Despite the achievements made by IFLA, cataloguing rules in different countries lacked a common basis for the assignment and form of access points and adhered to different descriptive practices. Fundamental change in cataloguing rules would of course have an impact on exchange formats.

Effective exchange of bibliographic data between bibliographic agencies can be accomplished only if the exchanging bibliographic record conforms to in respect of three elements – the record structure, the content designation and the data elements. The universal acceptance of the ISO 2709 record structure as a basis for exchange format has enormously benefited the information community. The ISO 2709 format owes its position as a universally recognized standard for exchange of bibliographic records. The format was designed for the exchange of bibliographic data on magnetic tape, though conversion of data to other physical media such as floppy disk, CD-ROM can be done successfully. The record structure of the format has now been criticized for many imperfections and it is expected that one day new technology will make the record structure of ISO 2709 inadequate and the situation may lead to the formulation of a new standard format.

The library community has been using the MARC format, which is based on ISO 2709 as a bibliographic data exchange structure for long 40 years. The MARC format is, by any standards, is said to be an historic achievement in the field of
machine-readable cataloguing. It has been the main force in international standardization, which brought a revolutionary advancement in modern librarianship. It forms the foundation for moving libraries into the computer age by providing a common syntax for recording and transferring bibliographic data between computers and between national bibliographic agencies. In association with the AACR, MARC allowed libraries to share cataloguing on a massive scale, and thus greatly increase the efficiency of cataloguing task as well as set the stage for the creation of centralized library databases. It is, literally, the engine that brought the concept of UBC more closer. There are more than tens of millions of MARC records are available in the world, which are used throughout the world. MARC is the basis for almost all automated bibliographic systems including commercially produced systems. It is the foundation of cost-saving copy cataloguing and it enables libraries to automate in an integrated manner. The usefulness of the MARC format has been built over the years as systems, tools, training, and globalization made the MARC standard into a keystone for automation and development.

With the advancement of technology, the information sources have gone through many changes since the development of MARC format in last 40 years. In the early MARC days the challenges was to achieve consistent bibliographic control of textual materials only, then the non-textual materials such as maps, music, graphics, moving images etc. were developed. By the late 1970’s the computer file became an important library resource and various forms of expression such as text, graphic, cartographic, and sound began to appear in electronic and digital forms. But an explosion took place in the last 10 – 15 years with the development of a communication vehicle for electronic data on the Web, the Internet, and today’s
greatest challenge of the information community is to provide appropriate tools for finding and retrieving the burgeoning web resources. The MARC format is trying to accommodate the newly emerged networked web resources and in 1993 a MARC field 856 has been introduced which identify the path of a electronic resource. The field 856 was added to describe the location and access of the electronic resources.

The bibliographic description within the MARC fields strictly followed the rules set out in AACR2 and ISBD (G). However, strict adherence to AACR2 rules in MARC record reduces the access points which is becoming increasingly problematic as multiple access points is more important than description. The increasing computing power and storage capacity of the present generation computer makes it possible to have more access points in each record with sufficient descriptive content. Strictly controlled description and display of records become meaningless, as the technology has now been completely changed. In addition, a record describing network resources might require details of several locations, versions and editions. AACR2 is not fit for this purpose. It is a code of the age of the catalogue card and printed bibliography. AACR2 is completely failed to describe Internet resources. Strict adherence to AACR2 by MARC format limits its scope and purposes for describing the newly emerging digital networked resources.

The advent of the WWW marks profound changes in how we use sources of information such as databases, indexes, and archives and how we use representation of knowledge such as maps, pictures, sounds, books, and journal articles. In the today’s changing technical environment, the bibliographic records are being used in a variety of computer systems within libraries. The technical environment has completely
changed from the first days of MARC but the cataloguing rules and formats have not essentially adapted with the technology. It is often debated in the recent literature that the record structure of MARC format is not fit for describing burgeoning web resources. The advent of full document markup system in last 5 to 10 years such as SGML, XML and more recently a framework RDF developed by W3C would replace MARC one day. The Internet provides unprecedented access to globally distributed information through WWW. Metadata formats can organise the web resources and provide access to such information more effectively than the traditional communication format like MARC.

Though the MARC format gradually become unsuitable to describe the web resources, it is assumed that the format will continue to play its role to transfer bibliographic data between different bibliographic agencies. Experiments are going on to extent the MARC format to make it suitable in the web environment. The MARC format can provide a vehicle for the description of web and network resources and it has to be kept up-to-date with developments in the technology. Millions of MARC records are available worldwide and huge amounts have invested in developing the MARC format as an international standard exchange format. The AACR and other relevant standards those are strictly followed by MARC, should be revised in accordance with the changing technology and should support the MARC format to keep alive in the current changing environment.

The mission of the university library is to acquire, organize and make readily accessible the recorded knowledge that fulfills the information needs of the university community. In fulfilling this mission, the university libraries of the
present age are facing many challenges due to very rapid development in Information Communication Technology (ICT). Modern technologies have a major impact on every spheres of human life, including libraries. The modern technologies are fostering and providing new ways for communicating and exchanging information and knowledge. The libraries of the present age are experienced with enormous explosion of information in every discipline in both printed and digital forms. This is because of wide application of IT for producing and communicating information. The university libraries have to be well equipped with the modern computer and communication technologies in order to provide better services to the university community and also to survive in this technological age. The information explosion, users varied demand for information, and the development of modern technology – all these factors have compelled the university libraries of the present age to apply computer and communication technology for organization and dissemination of information pin-pointedly, exhaustively and expeditiously.

In an integrated automated environment, a wide variety of library functions can be carried out, including the most generally accepted house keeping operations such as acquisition, cataloguing, serial control, authority control, and circulation. As an automated system supports a full range of library operations from acquisition and serial control through circulation and online public access catalogue, the cost of the system is spread over the whole library managerial operations making per unit cost of the library correspondingly low. Integration of the system provides much more improvement in reader services particularly retrieval of information from the system, which could not be possible in conventional manual system. Automated document descriptions have phenomenally improved the effective retrieval of information as
per users’ need. The automated document description system allows libraries to retrieve information effectively from the system as per the user’s requirement, which could not be possible previously in the conventional card cataloguing system. The OPAC when it becomes a part of the integrated system is able to furnish status information about whether a material is on order, in processing or circulation. Another significant advantage of automated document description system is that the system has the facility to place terminal at distant locations where a duplicate card catalogue could not be provided. Library automation systematized and speedy the process of collection, processing, storage, and access to information on a massive scale, which could not even dream a decade ago. It increases productivity and quality in both work as well as services and economy in expenditure. It made a revolutionary change in the functioning of libraries and supporting the staff in performing their jobs and duties effectively. Database created in the automated system can be made accessible through network to the users of inside as well as outside throughout the world. Services like – bibliographic service, indexing service, CAS, SDI etc. can be provided in an automated system and it also enable libraries to share resources through networks and also to access network resources at cheaper cost which is very essential in the environment of constant escalating in the prices of printed document and decreasing library budget day by day.