Unit 3 □ Computer Based Cataloguing and Character Encoding

Structure

3.0 Objectives

The objectives of the Unit are to :
- Examine functions of computer based cataloguing
- Review standardization of bibliographic and related standards
- Examine data structures of the catalogue database
- Identify subsystem and corresponding functions of each subsystem in integrated computer based library management system
- Study various character encoding systems

3.1 Introduction

The catalogue is the most valuable surrogate to information resources of a library. It can easily be developed as a byproduct of the computerized acquisition system. The primary purposes of the catalogue are to :
- Provide easy and efffective access to the library's collection
- Enable rapid and effective retrieval of information as per request of the user.
- Maintain accuracy, completeness and consistancy in description of library resources
- Enable cooperative collection development programmes.

3.2 Benefits

Cataloguing, classification and subject indexing are cost intensive processes in a library. These are done by highly trained professional. The rationale of computer based cataloguing are :

● Computer based cataloguing can reduce the cost of cataloguing. This is primarily possible through sharing of catalogue records in a consortium and importing cataloguing data from library utilities like OCLC and/or other cooperative networks (INFLIBNET/DELNET—strictly speaking such facilities are not yet available from any network in India)

● Enable integrating Internet resources in a library's catalogue.

● Provide remote access to the library catalogue.

● Reduce unnecessary duplication in intellectual cataloguing efforts.

● Generate spine labels and catalogue cards etc.

However, computer based cataloguing can provide the greatest benefits when the concerned library resolves to adhere to international standards-MARC 21,ISO 2709, Z39.50 and ISO ILL protocols etc.

3.3 Infrastructure

For a computer based cataloguing system, the necessary infrastructures are as follows :

● Access to external cataloguing resource databases

● Access to internal databases and files

3.3.1 Access to External Databases

The kind of external databses that are useful to have in computer based cataloguing system are sources of original catalogues and authority data. These databases may be on the Internet, CD-ROM, available in library utility like OCLC. Access to these databases may be free or may need membership.

Even if a library decide to do the original cataloguing, access to standard authority files of personal names, corporate names, conference names, serials, geographic name, language code etc. will be immensely helpful in maintaining consistancy in the library catalogue. The Library of Congress authory files are freely searchable on the Internet.

Access to machine-readable thesaurus or list of subject headings during the cataloguing process would help cataloguers. However, not all integrated library automation software, which are available in India, provide such tools.

3.3.2 Access to Internal Databases/Files

In absence of access to external authority files, a library must be able to create its own authority files for personal names, corporate names, conference names, subject terms etc. Such authority files should be created and updated as an integral part of the cataloguing process.

Cataloguing module must have access to the OPAC so that the cataloguers may verify important data to ensure consistancy, accuracy and completeness of the catalogue.

3.4 Subsystems and their functions

A computer based cataloguing system can be broken down into several subsystems based on the functions. A typical computer based cataloguing system will have the following functional subsystems :

Subsystems	Description/Functions
Bibliographic Control/ Cataloguing system	• It must allow records for any type of material in any format to be created, migrated, searched and displayed, modified, exported, and deleted. • Changes resulting from MARC format integration must be accommodated without reprogramming. • Cataloguing wizards should streamline the process of adding a brief title record, editing existing titles, duplicating an existing title, removing title, call number/volume or copies, creating and editing call number/volume records, adding or editing copies (includes global edits), offering authority control options, linking order line holdings to titles, capturing and editing records from Z39.50 sources using a Z39.50 copy cataloguing client. • When removing a title, call number, and/or item record, the Bibliographic Control/Cataloguing module must alert staff if a bill or hold is associated with the material. Staff must be able to immediately investigate the open transaction, without closing the removal process.

Subsystems	Description/Functions
	• When using the Bound-with process to link bibliographic descriptions for items bound together, the Bibliographic Control/Cataloguing module must use parent and child call number records. A child call number with copies must not be bound with a parent call number. A child call number must be linked to only one parent, but staff must be able to link a parent call number to an **unlimited** number of child call numbers.
	• In the transferring title, call number/volume, and/or copies process, authorized staff must be able to transfer all copies to an existing call number, transfer only selected copies, transfer all volumes, transfer only selected volumes, remove a title automatically after staff elect to transfer the last copy or volume to a new title, search and display bibliographic information without exiting the transfer process.
	Items on reserve or in transit must not be transferred. Inactive, available title-level holds linked to a call number must also block the transfer. The Bibliographic Control/Cataloguing module must alert staff to such exception conditions when staff attempt to transfer records.
	At Library staff's discretion, records may be hidden from public user display at the title level (to hide bibliographic data and all associated items), the call number level (to hide selected volumes and all associated copies), item level (to hide selected copies), or by assigning the item to a "shadowed" location.
	• It must support MARC format error checking, including error checking (tags, indicators, and subfields) for all formats.
	• A utility should verify uniform resource locator(s), or URLs, catalogued within MARC 856 bibliographic fields.

	• It must support creation, editing and maintenance of Community Information records in the MARC Format for Community Information and in a locally developed format. • It must include a Z39.50 Copy Cataloguing Client that can capture bibliographic records from any Z39.50 bibliographic resource.
Authority Control	• Link all authority-controlled bibliographic headings with corresponding authority records through an ANSI-standard thesaurus. • It should include a machine-proposed authority feature based upon the Standard for Machine-Proposed Authority Records (http://www.loc.gov/catdir/pcc/strawn. html) developed for the Programme for Cooperative Cataloguing, which must generate a new authority record with reasonable values in the fixed field and 001, 005, 040, 1XX, and 670 entries (plus the 022 and 642-646 entries for a series heading). These values must be automatically generated based on the information in the bibliographic record being validated. • It must enable the system administrator to specify whether entry of bibliographic data which does not match an authority record must result in rejection of the input, or in a warning, in which case the system must display a browse list of possible authority headings from which the operator may choose by clicking the desired heading to replace the unauthorized heading in the bibliographic record, automatically flag the unauthorized entry for later display, review, and/or printing. • It must support multiple authority files, including separate authority indexes for LC name and subject headings, NLM subjects, or other locally-defined indexes. • It must enable the Library to define an unlimited number of authority formats or types, and to specify the bibliographic fields and subfields addressed by each authority record type through policy configuration.

	• It must automatically generate appropriate *See* and *See Also* references from authority records for use in the online catalogue. • At the Library's discretion, the Authroity Control module must be configured either to display cross-references but not to verify headings, or to display cross-references and also to verify headings.
Reports	• It must be fully-integrated with all other system modules, and provide a comprehensive suite of library-customizable report templates. • It must enable an authorized operator to schedule production of report output at a specified date and time and on a regular periodic basis, such as daily, weekly, monthly, and at pre-specified times. • It must enable an authorized operator to view completed reports on screen or to e-mail or print the report, at the operator's convenience. • On any database reports involving materials, such as new accessions lists, shelf lists, high/low circulation lists, and bibliographics, the operator can select items for inclusion based on any combination of bibliographic information (using full Boolean word and phrase searching) and on any combination of control information, for example collection, current status, number of circulations, number of holds, classification, and accessions date. • It must enable an authorized operator to specify the starting date-and-time and ending date-and-time that the report is to cover on reports involving historical data. • The fully-integrated Reports module must use the same user interface as other modules. The Reports module must provide lists, counts, and statistical reports for each purchased module, provide row, column, and grand totals in applicable reports, provide reports for all record types within the proposed system, track statistical and management information by counting various staff proceses to measure productivity, identifying items which are likely candidates for weeding, or tracking fund information for budgeting, perform housekeeping tasks by changing the status of groups of users, or

	removing users or items in batches when necessary, allow authorized operator(s) to select, customize, name, save and schedule reports, allow staff to display and/or e-mail finished reports, employ an easy, point-and-click interface with dropdown menus for report criteria selection inventory.
	• The Circulation module must support marking items 'used' in-house for statistical reporting. If an item has a special status (charged, on hold. in transit), a glossary must appear next to the item for displaying more item information.
	• The Circulation module must support the creation of a shelf list from inventory data in report output. Also, staff must be able to list inventory by item number in a report.
	• The Circulation module must support the use of portable inventory scanners, provided the Vendor's software is loaded to the Library's portable scanners.

3.5 Standardization/MARC format

Bibliographic Standards Development is a key element in the Library's cataloguing strategy. Standardization facilitates the exchange of bibliographic records between cataloguing agencies, thereby producing better, faster, cheaper catalogues. The following Table explains the types of standards involved in a computerized library environment.

Cataloguing Standards	Enabling Library to share catalogue records with other libraries, both in the country and overseas. AACR (Anglo-American Cataloguing Rules, 2nd edition) is a major international standard for the cataloguing of all types of materials collected by general libraries. The family of International Standard Bibliographic Descriptions (ISBD) specifies the requirements for description and identification of information resources.
Exchange Formats	Formats are used to transfer data in a structured form. Exchange formats were developed in parallel with the development of computers and other electronic storage devices to facilitate the transfer of bibliographic data between computer systems.

	The MARC 21 formats are standards for the representation and communication of bibliographic and related information in machine-readable form. The Library of Congress in consultation with various user communities maintains the MARC 21 formats.
	The primary purpose of UNIMARC is to facilitate the international exchange of data in machine-readable form between national bibliographic agencies. UNIMARC may also be used as a model for the development of new machine-readable bibliographic formats.
Name Authority Control	Authority Control is the process of establishing and maintaining consistency in headings in a bibliographic file by means of an authority file.
Subject Access	The use of a standardized system of subject headings allows compatible access by subject across different files and catalogues. In libraries document classification enables ordering of material in a helpful way on open shelves, browsing and retrieval of related items in catalogues, meaningful arrangement in subject indexes and bibliographics, analysis of the collections, and increasingly provides structural tools for mapping and organizing Web resources.

3.5.1 Needs for Standardization

In the library context, standards mean a set or code of rules established by national and international organizations for the purpose of bibliographic control, including those providing for the unique identification of bibliographic items, such as the International Standard Book Number and International Standard Serial Number; the uniform description of items, such as the International Standard Bibliographic Description; and the exchange of bibliographic records by means of a bibliographic exchange format, such as MARC (Machine Readable Cataloguing) format.

With the steadily increasing need for information, and the new technical capabilities for handling it, standardization becomes increasingly important-to ensure greater effectiveness of information services in all countries and the use of information and information systems across national, regional and institutional borders. Standardization, therefore, is aimed at achieving uniformity and it serves, in the end, as a quality control tool. Library catalogues which are produced under a certain standard cataloguing code can guarantee a certain level of information retrieval success, because that standard will ensure a certain level of quality so that the record

will be sufficiently complete and will be consistently arranged and hence that the content and the location of the record will be predictable.

A library could devise its own method of organizing the bibliographic information, but it would isolate the library, limit its options, and create much more work for itself. Using the MARC standard prevents duplication of work and allows libraries to better share bibliographic resources. Choosing to use MARC enables libraries to acquire cataloguing data that is predictable and reliable. If a library were to develop a "home-grown" system that did not use MARC records, it would not be taking advantage of an industry-wide standard whose primary purpose is to foster communication of information.

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 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.

3.5.2 Development of Exchange Format

MARC is the acronym for Machine-Readable Cataloguing. It provides the mechanism by which computers exchange, use, and interprets bibilographic information, and its data elements make up the foundation of most library catalogues used today. This general description, however, is rather misleading, as MARC is neither a kind of catalogue nor a method of cataloguing. In fact, MARC is a short and convenient term for assigning labels to each part of a catalogue record so that computers can handle it. While the MARC format was primarily designed to serve the needs of libraries, the wider information community as a convenient way of storing and exchanging bibliographic data has since embraced the concept.

3.5.2.1 Importance of MARC

The information from a catalogue card cannot simply be typed into a computer to produce an automated catalogue. The computer needs a means of interpreting the information found on a cataloguing record. The MARC record contains a guide to its data, or little "signposts," before each piece of bibliographic information.

The place provided for each of these pieces of bibliographic information (author, title, call number, etc.) is called a "field." The records in simpler computer files sometimes have a fixed number of fields, and each field contains a fixed number of characters. However, to allow proper cataloguing of books and other library items, the best file structure allows for records with an unlimited number of fields and unlimited

field lengths. This flexibility is necessary because not all titles are the same length. Some books are part of a series, requiring a field for that information, while others have no series statement. And audiovisual items have much longer physical descriptions than do most books.

The computer cannot expect a certain type of information to begin and end at the same position in every bibliographic record. The statement of responsibility will not always begin with the 45th character of the record and end at the 107th position, for example. Therefore each MARC record contains a "table of contents" to the record, according to a predefined standard.

If a bibliographic record has been marked correctly and saved in a computer data file, computer programmes can then be written to punctuate and format the information correctly for printing a set of catalogue cards, or for displaying the information on a computer screen. Programmes can be written to search for and retrieve certain types of information within specific fields, and also to display lists of items meeting the search criteria.

3.5.2.2 Development of MARC21

The original MARC format was developed at the Library of Congress in 1965-66 leading to a pilot project, known as MARC-I, which had the aim of investigating the feasibility of producing catalogue data in machine-readable form. Similar work was in progress in the United Kingdom where the Council of the British National Bibliography had set up the BNB MARC Project with the objective of examining the use of machine-readable data in producing the printed *British National Bibliography (BNB)*. These parallel developments led to Anglo-American cooperation on the MARC-II project, which was initiated in 1968.

Despite cooperation there emerged several versions, e.g. UKMARC, INTERMARC and USMARC, whose paths diverged owing to different national cataloguing practices and requirements. Since the early 1970s an extended family of more than 20 MARC formats has grown up. Differences in data content means that editing is required before records can be exchanged.

One solution to the problem of incompatibility was to create an international MARC format (UNIMARC), which would accept records created in any MARC format. So records in one MARC format could be converted into UNIMARC and then be converted into another MARC format. The intention was that each national agency would need to write only two programmes-one to convert into UNIMARC and one to convert from UNIMARC-instead of one programme for each other MARC format, e.g. INTERMARC to UKMARC, USMARC to UKMARC etc.

MARC-II established certain principles, which have been followed consistently over the years. MARC-II was to prove 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

MARC-I dealt mostly with books. However, MARC II was planned to cover all types of materials including books and monographs. During 1970-1975, documentation was issued for other materials i.e. in 1972 films; in 1973 serials, maps and French books; in 1975 records for German, Spanish, and Portuguese material were covered.

MARC became USMARC in the 1980s and MARC21 in the late 1990s. MARC21 is not a new format. After having discussions and making minor changes to both formats that accommodated USMARC and CAN/MARC users' specific needs, the USMARC and CAN/MARC (Canadian MARC) formats were "harmonized" into MARC21 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 maintain the MARC21 formats. Input for development is provided by MARC21 users from around the world, including libraries, library networks and utilities, and library system vendors.

The *MARC21 Format for Bibliographic Data* is designed to be a carrier for bibliographic information, such as titles, names, subjects, notes, publication information, and physical descriptions of items. The *MARC21 Format for Bibliographic Data* contains data elements for the following types of material :

- **Books-** textual items, monographic in nature, that are in bound book, electronic, or microform.
- **Continuing resources-** textual items with a recurring pattern of publication, e.g., periodicals, newspapers, and yearbooks.
- **Computer files-** Computer software, numeric data, computer-oriented multimedia, online systems or services. Other types of electronic resources are coded for their most significant aspect, such as textual ("books" or "serials"), cartographic ("maps"), etc.
- **Maps-** all types of cartographic materials, including sheet maps and globes in printed, manuscript, electronic, and microform.
- **Music-** printed and manuscript notated music
- **Sound recordings-** nonmusical sound recordings and musical sound recordings.

◆ **Visual materials-** images and objects, e.g., projected media, motion pictures, two-dimensional graphics, three-dimensional artifacts, naturally occurring objects.

◆ **Mixed materials-** primarily archival and manuscript collections of a mixture of forms of material.

MARC21 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 five MARC21 communications formats are widely used standards for the representation and exchange of bibliographic, authority, holdings, classification, and community information data in machine readable form. The five MARC21 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 MARC21 Format for Bibliographic Data is an integrated format 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 enconding 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, programmes, services, etc. So that this information can be integrated into the same public access catalogues as data in other record types.

The standards present a generalized structure for records, but do not specify the content of the record and do not, in general, assign meaning to tags, indicators, or data element identifiers. Specification of these elements is provided particular implementations of the standards.

3.5.2.3 Organization of MARC21 Record

A MARC record is composed of the following three elements :

- **Record structure :** is an implementation of ISO 2709, *Information and Documentation-Format for Information Exchange.*
- **Content Designation/Tagging :** is the set of tags and codes that identifies and further characterize the data elements within a record and support the manipulation of the data content.
- **Data content of the record.** The **content** of the data elements that comprise a MARC record is usually defined by standards outside the formats, such as cataloguing rules, classification schemes, subject thesauri, code lists, or other conventions used by the organization that creates a record. The content of certain coded data elements (e.g., the Leader, and field 008) is defined in the MARC formats themselves.

3.5.2.3.1 Record Structure

MARC is a specific implementation of ISO 2709, an international standard that specifies the structure of records containing bibliographic data. It specifies that every bibliographic record prepared for exchange conforming to the standard must consist of a :

- ◆ RECORD LABEL : consisting of 24 characters
- ◆ DIRECTORY consisting of a 3-digit tag of each data field, along with its length and its starting character position relative to the first data field,
- ◆ DATA FIELDS of variable length, each separated by a field separator,

The record layout is as follows :

Record Label	Directory	Data Fields	Record Terminator

ISO 2709 further specifies that the data in fields may optionally be preceded by indicators and subdivided into subfields.

Record Label

It is the first field in all MARC records. ISO 2709 prescribes that each record start with a 24-character Record Label. This contains data relating to the structure of the record, which are defined within the standard ISO 2709, and several data elements that are defined for this particular implementation of ISO 2709. These implementation-defined data elements relate to

- The type of record,
- Its bibliographic level and position in a hierarchy of levels,
- The degree of completeness of the record.

55

The data elements in the Record Label are required primarily to process the record and are intended only indirectly for use in identifying the bibliographic item itself. A key data element in the Record Leader is the type of item being described in the record. It identifies the following item types.

Language (textual) material	Nonmusical sound recording
Manuscript (textual) language Material	Musical sound recording
Computer file	Projected medium
Cartographic material	Two-dimensional non projectable graphic
Manuscript cartographic material	Three-dimensional artifact or natural objects
Notated music	Kit
Manuscript music	Mixed material

Directory

A directory entry in MARC21 is made up of a tag, length-of-field, and field starting position. The directory begins in character position 24 of the record and ends with a field terminator. It is of variable length and consists of a series of fixed fields, referred to as "entries." One entry is associated with each variable field (control or data) present in the record. It contains a series of entries that contain the tag, length, and starting location of each variable field within a record. Each entry is 12 character positions in length. The Directory entry for the Record Control Number (001) appears first. Subsequent entries for variable data fields follow, arranged in ascending order according to the first character of the tag. The stored sequence of the variable data fields in a record is not necessarily the same as to the order of the corresponding Directory entries. The Directory ends with a field terminator character.

Each directory entry is 12 characters in length; and it is represented schematically below. The numbers indicate the character positions occupied by the parts of the entry.

Structure of a directory Entry in MARC 21 Records

TAG	LENGTH_OF_FIELD	STARTING_CHARACTER_POSITION
00-02	03-06	07-11

Tag (character positions 00-02), consists of three ASCII numeric characters or ASCII alphabetic characters (uppercase or lowercase, but not both) used to identify or label an associated variable field. The MARC 21 formats have used only numeric tags. The tag is stored only in the directory entry for the field; it does not appear in the variable field itself.

Length of field (character positions 03-06), contains four ASCII numeric characters which give the length, expressed as a decimal number, or the variable field

to which the entry corresponds. This length includes the indicators, subfield codes, data and field terminator associated with the field. A field length number, of fewer than four digits, is right justified and unused positions contain zeroes (zero fill).

Starting character position (Character positions 07-11), contains five ASCII numeric characters which give the starting character position, expressed as a decimal number, or the variable field to which the entry corresponds relative to the base address of data of the record. A starting character position, of fewer than five digits, is right justfied and unused positions contain zeroes (zeor fill).

Order of entries, Directory entries for control fields precede entries for data fields. Entries for control fields are sequenced by tag in increasing numerical order. Entries for data fields are arranged in ascending order according to the first character of the tag, with numeric characters preceding alphabetic characters.

Variable-Fields

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 reocrd is not necessarily the same as the order of the corresponding directory entries.

Control Fields in MARC 21 formats are assigned tags beginning with two zeroes. They are comprised of data and a field terminator; they do not contain indicators or subfield codes. The control number field is assigned tag 001 and contains the control number of the record. Each record contains only one control number field (with tag 001), which is to be located at the base adress of data.

Data fields in MARC21 formats are assigned tags beginning with ASCII numeric characters other than two zeroes. The tag is stored in the directory entry for the field, not in the field itself. The data in a MARC 21 record is organized into fields. Such fields contain indicators and subfield codes, as well as data and a field terminator. There are no restrictions on the number, length, or content of data fields other than those already stated or implied, e.g., those resulting from the limitation of total record length. The variable fields follow the leader and the directory in the record and consist of control fields and data fields. The structure of a data field is shown schematically below.

Structure of a Variable Data Field in MARC 21 Records

INDICATOR_1	INDICATOR_2	DELIMITER	DATA_ELEMENT_IDENTIFIER_1
	DATA_ELEMENT_1 . . .	DELIMITER	DATA_ELEMENT_IDENTIFIER_n
	DATA_ELEMENT_n	FT	