INFS 427: AUTOMATED INFORMATION RETRIEVAL

(1st Semester, 2018/2019)

Session 04 – BIBLIOGRAPHIC FORMATS

Lecturer: Mrs. Florence O. Entsua-Mensah, DIS
Contact Information: fentsua-mensah@ug.edu.gh

UNIVERSITY OF GHANA
College of Education
School of Continuing and Distance Education
2014/2015 – 2016/2017

Session Overview

- The exchange and sharing of information across different user communities require the use of standard formats to facilitate creation and exchange of bibliographic data.
- In the case of libraries, to facilitate international library resource sharing, interlibrary loan and global networking by use of ICT, there is the need for the development of several bibliographic formats.
- A designer of a bibliographic database must choose a format suitable to the needs of the target user community.
- This session therefore seeks to deliberate on the various bibliographic formats that are used in modern day IRS.

Session Outline

The key topics to be covered in the session are:

- Bibliographic Records
- Bibliographic Formats

Recommended Reading

Chowdhury, G. G. (2010). *Introduction to modern information retrieval*. London: Facet publishing. – **Chapter 3**.

Introduction

An information retrieval system may include 3 categories of information:

- Factual
- Bibliographical
- Institutional

- Exchange and sharing of these categories of information across different user communities require the use of standard formats to facilitate creation and exchange of bibliographic data.
- In the case of libraries, to facilitate international library resource sharing, interlibrary loan and global networking by use of ICT, hence the development of several bibliographic formats.
- A designer of a bibliographic database must choose a format suitable to the needs of the target user community.

Topic One

BIBLIOGRAPHIC RECORD

What is a bibliographic record?

- All the elements that can be used “to describe, identify, or retrieve any physical item of information content” (Chowdhury, 2010).

OR

- A collection of data elements organized in a logical way to represent a bibliographic item
- Bibliographic item – Any record of human communication treated as an entity, such as a book, a document, part of a document, or group of documents.

Components of a bibliographic format

Efficient exchange of bibliographic data between institutions/agencies can be accomplished if their record conform to the ff components:

1. **Physical structure** – Rules for the arrangement of the data to be exchanged on a computer storage medium, e.g. Floppy disk or CD-ROM
2. **Content designators** – These are codes or tags to identify or define the different data elements in the record, i.e., author, title, date of publication etc.
3. **Content** - contents of the record and the rules that govern the formulation of the data elements of different formats. For example all bibliographic IR systems must follow some form of cataloguing rules to ensure uniform data presentation, display and printed output.

Topic 3

BIBLIOGRAPHIC FORMATS

Types of International exchange formats

The lack of uniformity in national standard formats has led to the development of international standard exchange formats such as:

- **UNIMARC** – This format used by national libraries for exchange purposes and uses the International Standard Bibliographic Description (ISBN) to describe the data elements of each item.
- **UNISIST (**United Nations International Scientific Information System) – It is UNESCO World Scientific Information programme to facilitate scientific information. It includes a number of programmes.

Programmes of UNISIST

- **ISSN (International Standard Serial Number)-** This is an eight digit number for identifying a journal.
  - The number is associated with the tile of the journal and therefore changes when the name of a journal is changed.
  - The purpose of ISSN is to ensure bibliographic control, i.e., organization of recorded information according to established standards to make it readily retrievable.

- **Reference Manual –** A manual of bibliographic description used mainly by abstracting and Indexing services.

Format for bibliographic information exchange- ISO 2709

- It is an international standard that specifies the requirements for a generalised exchange format suitable for all forms of bibliographic descriptions.
- **Characteristics**
  - It is a framework for communication between data processing systems and also for use as a processing format within a system.
  - It does not define the length or the content of individual records
  - Does not assign any meanings to tags, indicators or identifiers. Such specifications are the functions of an implementation format

Basic structure of ISO 2709

- Record label – It has a fixed length field of 24 characters in total and holds the basic information of a record. It includes *record length* (total no. of characters in the records), *record status* (e.g. new record), *implementation tag* (e.g. Record type) and *identifier length*
- **Directory** – a variable length field which provides the entry positions to the fields in the records together with the field tags. It is made up of field tag (3 characters), length of the field (4 characters), starting character position of the field (5 characters), occurrence of the field, and number of segments containing the field. It ends with a terminating symbol.

Types of fields in the ISO 2709 record

- **Record identifier** – a variable-length field given by the organization that creates the record. Its purpose is to identify the record.
- **Reserved fields**- used to hold reference data of a given record that may be required for processing.
- **Bibliographic data fields** – it holds the actual bibliographic data together with its indicators or tags
- **Field separators** – each data field is terminated with a field separator symbol

MARC FORMAT

- **MARC** – means Machine Readable Catalogue or Cataloguing
- **Machine-readable** – means a machine or a computer can read and interpret the data in the cataloging record
- **Cataloging record** – means a bibliographic record or the traditional information on a library catalogue card, i.e. Main and added entries, subject headings, classification or call no., etc.
- **Purpose** – The purpose of the MARC format is to employ a set of conventions to identify and arrange bibliographic data so that it handled by a computer
- ***MARC adhere to ISO 2709 record structure***

Brief history of MARC

MARC was developed by the Library of Congress in 1960.

There were 2 slightly different types of MARC formats during the 1980's and 1990's, the US version, USMARC and the Canadian version CAN/MARC. In 1999, the two versions were blended together into a single version called MARC 21.

It is maintained by the Standards and Support Office at the National Library of Canada, and the Network Development and MARC Standards Office at the Library of Congress.

Advantages of using MARC as a common bibliographic standard

- It avoids duplication of work and allows libraries to share bibliographic data
- It enables libraries to acquire reliable cataloging data

- MARC format is compact, therefore saves space.
- MARC is the standard format used by most library computer programs and systems , and therefore:
  - It enables libraries to use commercially available library automated systems to manage their operations
  - Libraries are able to benefit from the latest advances in computer technology
  - Libraries have the flexibility of replacing one system with another without fear of incompatibility with their data
  - There is easy communication and exchange of information

MARC 21-Guidelines for managing and formatting electronic records of different information resources

- *MARC 21 Format for Bibliographic Data* – contains specifications for encoding data elements required to describe and retrieve all forms of bibliographic data.
- *MARC 21 Formats for Holdings Data* – contains specifications for encoding data elements relevant to locations and holdings information for all formats
- *MARC 21 Format for Authority Data* – contains format specifications for encoding data elements relating to records subject to authority control
  - Authority control is the establishment and maintenance of consistent forms of terms—names, subjects, and titles—to be used as headings in the bibliographic records of the library catalog. Headings must not only be consistent, they must also be unique.
    - For e.g. two authors who happen to have published under the same name can be distinguished from each other by adding middle initials, birth and/or death dates

MARC 21-Guidelines for managing and formatting electronic records of different information resources

- *MARC 21 Format for Classification Data – contains* specifications for encoding data elements relating to classification numbers
- *MARC 21 Data for Community Information* – provides format specification for encoding records relating to information about events, programmes, and services to enable their integration into OPAC
- Others are *MARC Code List for Languages, Countries, Geographic Areas, Organizations, etc.*

Understanding MARC terms and their definitions

- A field – it is a bibliographic record such as author, title etc. It is a term used to describe the various sections of cataloging information. In MARC the fields are by 3-digit tags supplied by the system software
- A tag – It is the first 3-digit number that identifies the field.

Components of a MARC record

Field tags (Stands for numeric values from 00 to 99	Component descriptions
0xx	Control fields
1xx	Main entries
2xx	Tile, edition and imprint information
3xx	Physical description
4xx	Series statements
5xx	Notes
6xx	Subject access entries
7xx	Added entries and linking fields
8xx	Series added entries and holdings information
9xx	Fields for local use

Definitions of components

- Control fields- holds information on bibliographic control numbers (and coded information used for processing MARC records). This number is assigned by the organization creating, using, or distributing the record. For e.g., 001-006 contain control numbers and coded information about date and time of processing and type of material e.g., e-resources or books.
- Main entry fields (1XX) – used for storing information on the main entry heading of a record, example 100 –personal name (NR-non repeated)

Definitions of components

- Title and title-related fields- stores title of item and related information. E.g., 210=abbreviated title, 222=key title, 245=title statement
- Edition, imprint, etc. Field (250-270)- stores information on edition, imprint, address, etc. Example 250=edition statement, 260=publication distribution
- Physical description (3XX)- stores information on physical characteristics, publication frequency, price etc. For e.g., 310=publication frequency

Comparison of same record with textual information with MARC tags

Components	Data	MARC
Main entry, personal name with a single surname: The name:	Arnosky, Jim.	100 1# $a
Title and Statement of responsibility area, pick up title for a title added entry, file under "Ra..." Title proper: Statement of responsibility:	Raccoons and ripe corn / Jim Arnosky.	245 10 $a $c
Edition area: Edition statement:	1st ed.	250 ## $a
Publication, distribution, etc., area: Place of publication: Name of publisher: Date of publication:	New York : Lothrop, Lee & Shepard Books, c1987.	260 ## $a $b $c

Summary

- In this class session we have established the role of standardised bibliographic formats in information search and retrieval.
- We looked at some of the international standards, such as:
  - ISSN
  - UNIMARC
  - UNISIST

Activity 3.1

- Define with examples the remaining components of a MARC record, i.e., series statement, notes, subject access added entry, series added entry fields.

References

Chowdhury, G. G. (2010). *Introduction to modern information retrieval*. London: Facet publishing.

Library of Congress (n.d.). What is MARC record and why is it important? Retrieved from http://www.loc.gov/marc/umb/um01to06.html#part2