INTEROPERABILITY BETWEEN DUBLIN CORE, UNIMARC, MARC21, WITH AACR2R AS THE STANDARD FRAMEWORKS FOR CATALOGUING IN THE DIGITAL ENVIRONMENT

K S Chudamani ***H C Nagarathna***

Abstract

The technological development of the past 25 years, such as the electronic database, online services, CD-ROMs and the introduction of internet has radically transformed the process and access to information through catalogues. AACR2 rules cover the description of and the provision of access points for all library materials. The rules for description are based on the general frame work for the description of library materials the general international standard bibliographical description (ISBD(G)). The transformation of ISBD(G) to AACR2 is one phase. The second phase leads to creation of machine readable metadata. The most popular are Dublin core, UNIMARC, MARC21. Another local library software LIBSYS which is used in Indian libraries has also been used here for presentation of the mapping between different formats to draw conclusions to make them interoperable. As far as metadata interoperability is concerned, the OPAC display, now almost completely dependant on the AACR2 the world over requires minimum data aggregation, mainly the author, title, series, subject, etc. aggregation which are the mandatory fields prescribed by IFLA.

Keywords : MARC 21; Interoperability; AACR2R; Libsys; Metadata

1. Introduction

Libraries and Information centers are the intermediaries between the information, information sources and users. In order to make information accessible libraries perform several activities. Amongst them, cataloguing is one of the most important and fundamental activity to make information accessible. While making the bibliographic records for an item in hand, a lot of intellectual processing takes place through cataloguing in every organization, large or small. This happens whenever materials of information contents are collected and organized to form surrogates systematically in order to serve the needs of the user. The technological development of the past 25 years, such as the electronic database, online services, CD-ROMs and the introduction of internet has radically transformed the process and access to information through catalogues.

A bibliographic record for any document might be defined in terms of four major components.

1. A Physical description of the item itself (in terms of such elements as author, title, publisher and place of publication)

 © INFLIBNET Centre, Ahmedabad

2. For the complete description of the document Selection of elements from the complete description (such as personal author or corporate author) which become access points. From this the record can be retrieved.
3. A unique identifier for the document
4. selection and rendering of terms representing subject matter dealt with, which might be controlled terms drawn from an authority such as Library of Congress subject heading, or words and complex set of words or phrases drawn from the document itself.

According to the Anglo-American Cataloguing Rules (AACR2R), "a catalogue is a list of materials in a collection. It contains bibliographic descriptions, created in accordance with known and presumably universal rules"

Cataloguing rules are not static; They must respond changing needs. They are interpreted by experts to keep in tune with the changing needs but preserving Cataloguing guidelines. A number of continuing issues affecting implementation of AACR2 have been compounded in recent years by the fast growth in technology and metadata requirements and its impact on publishing patterns as well as access requirements. AACR2 rules cover the description of and the provision of access points for all library materials. The rules for description are based on the general frame work for the description of library materials the general international standard bibliographical description (ISBD(G)).

It is important to bring out all aspects of the item being described including its content, its Physical format, the type of publication, the items bibliographic associations and information such as whether it is published or unpublished. The cataloguer should follow the more specific rules applying to the item being catalogued, whenever they differ from the general rules.

As it is well known, a cataloguer obtains information which can be used in a bibliographic description from a "chief source of information". With a printed book, this is primarily the book's "title page" - both its recto (front) and its verso (back). With AACR2R, the units of the bibliographic description rules and choice of rendering were designed with automation in mind from the outset. The units of the description translate directly into "fields" in a book's Machine Readable Cataloguing (MARC) record (cf. ISO 2709).

In formulating the rules and statements regarding the formulation of recommendations, the authors had continued challenges which had to be understood and adjustments were made to these changing realities. Also one has to foresee what effect anticipated changes need to be made and its impact on the standards which will guide decisions, and visualization of the changes required to improve the recommendations that would work in a web-based catalog. The transformation of ISBD(G) to AACR2 is one phase. The second phase leads to creation of machine readable metadata. The most popular are Dublin core, UNIMARC, MARC21. Another local library software LIBSYS which is used in Indian libraries has also been used here for presentation of the mapping between different formats to draw conclusions to make them interoperable.

2. Standards

2.1 DUBLIN CORE

Dublin Core (DC) was primarily designed to provide a simple resource description format for networked resources [4]. Dublin core standard comprises fifteen element. Each elements is optional and may

be repeated also has a limited set of qualifiers, attributes that may be used to further refine the meaning of the element. Dublin core with a view to its automatically producing a catalogue record in another format? Even this does not seem possible as Dublin core does not have any coding to provide the necessary detail for the specification of a record that could be converted to any machine readable coding like UNIMARC, Marc21, etc.

2.2 UNIMARC

UNIMARC: an introduction is available from the format maintainers: the Universal Bibliographic Control and International MARC Core Programme of the International Federation of Library Associations (IFLA UBCIM)[2] . The UNIMARC mappings are based on the information given in the 2nd edition of the UNIMARC manual (with update 1), published in 1996 [3]

Unimarc records consist of data formulated by highly controlling cataloguing codes : Dublin Core data elements are less highly specified. The data elements reflect this in that they cover broader categories of data. The UNIMARC format[2]," like any other version of MARC, the unimarc has three elements of the bibliographic record namely the record structure, content designation and data content.

Further, certain standard conventions are followed in order to identify the data elements within records. Elements such as author, title and subject access are further characterised wherever necessary. This supports the manipulation of the data for a variety of purposes:

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

2.3 MARC21

The Library of Congress' Network Development and MARC Standards Office has developed a framework for working with MARC data . This framework can be made both flexible and extensible to allow users to work with MARC data in ways specific to individual library needs. "The framework itself includes many components such as schemas, style sheets, and software tools. Library of Congress Network Development and MARC Standards Office has developed a schema for a bibliographic element set that may be used for a variety of purposes, and particularly for library applications"[5]. To recapitulate, library cataloguing systems need MARC records. So, if a MARC record could be extracted from a Web page which contains an electronic document which is worthy of cataloguing, so much the better. For ex. Now, MARC has introduced new field 856. It has been adapted for the location of electronic resources.,

3. LIBSYS

Libsys is a Library automation software. Its Cataloguing System provides for data entry which closely adheres to broad guidelines of AACR2 code. Online catalogues or indexes to the collection by author, title, class no., subject, etc for a database are created in it. It has searching facility for word-based combination searches using Boolean operators. Other facilities of the system include current awareness lists, bibliographies, and SDI. It is possible to import / export data in varying formats.

The system provides following functions for the maintenance of the library Catalogue[1].

- Titles In-process
- Enter new title
- Update Title
- Update Holdings
- Merge titles
- Remove title
- Change accession no.
- Maintain classified subject directory

4. Mapping from Dublin Core to UNIMARC, MARC21 and LIBSYS

The next section provides a mapping of Dublincore, UNImarc, Marc21 and Libsys.

From the Table-1 it can be noted that though MARC21 is related to AACR2, it is codifying each and every example shown in AACR to ease the burden of programming the computer. But in practice, it is the computer software that needs to be developed which can handle multiple occurrence of a field and we are cataloguing a book as a manifestation of a work. The system should at its best give data in a specified order. According to Sengupta[8], a cataloguer need not spend too much time on cataloguing. If the book in hand uses only initials of the author it can be entered as such, if there is no conflict between various entries. Also, some times if the cataloguer finds contradictory information he has to build up the authority file for future information.Some of the sections need not be too specifically coded; ex notes. The conventional notes are to specify certain information that appears at the at the head of the title or to indicate works that are bound with it, etc. It is usually not a searchable field. Hence only visible information may be recorded and coding reduced. Even though the card catalogue is no more considered as an important tool, its ideas and principles are still valid and can be the guiding force in computerized catalogues. When indexes are searched, it is possible to match a query with the existing data only when all its repeating fields are indexed. Libsys follows this type of an aggregation with multiple repetitions. The study of Libsys for its handling of repetitive data is an useful exercise. For example, any number of author, title, series entries can be provided in the system. Later these are merged into one index for the specific broad field like author, title, etc.

Table-1: Mapping from Dublin Core to UNIMARC, MARC21 and LIBSYS

S.N	Dublin Core	UNIMARC	21-Mar	LIBSYS
1.	Title	200 $a Title Proper 200 $e Other Title Information (for subtitle) 517 $a Other Variant Titles (for other titles)	245 title proper statement 210 abbreviated title 222 key title 240 uniform title 242 collective uniform title 246 varying form of title 247 former title	Title, Sub title, Alternate title, Uniform title

S.N	Dublin Core	UNIMARC	21-Mar	LIBSYS
2.	Creator	700 $a Personal Name - Primary Intellectual Responsibility, or if more than one: 701 $a Personal Name - Alternative Intellectual Responsibility 710 $a Corporate Body Name - Primary Intellectual Responsibility, or: 711 $a Corporate Body Name, Alternative Intellectual Responsibility 200 $f First Statement of Responsibility	100 main entry personal name 110 main entry corporate name 111 main entry meeting name 130 main entry uniform title	Main entry Corporated body., etc Personal Author Title
3.	Subject	610 $a uncontrolled Subject terms 606 topical name used as subject (for lcsh and mesh) 675 udc 676 ddc 680 lcc 686 other Classification systems	072 - Subject Category Code (R) 074 - GPO ITEM NUMBER (R) 080 - Universal Decimal Classification Number 082-Dewey Decimal 084-Other Classification Number (R) 086-Government Doc. Classification Number(R) 088-REPORT NUMBER (R) 600 - subject added entry personal name (r) 610 - subject added entry corporate name (r) 611 - subject added entry meeting name (r) 630 - subject added entry uniform title (r) 648 - subject added entry chronological term (r) 650 - subject added entry topical term (r) 651 - subject added entry geographic name (r) 653 - index term uncontrolled (r)	Keywords, subjects

S.N	Dublin Core	UNIMARC	21-Mar	LIBSYS
			654 - subject added entry faceted topical terms (r) 655 - index term genre/form (r) 656 - index term occupation (r) 657 - index term function (r) 658 - index term curriculum objective	
4.	description	330 $a summary or abstract	520 summary note	abstracts
5.	publisher	210 $c name of publisher, distributor, etc.	260 - publication, distribution, etc. (imprint) (r) 263 - projected publication date (nr) 270 - address (r)	publishers, distributers, address
6.	contributors	701 $a personal name - alternative intellectual responsibility 711 $a corporate body name - alternative intellectual responsibility 200 $g subsequent statement of responsibility (if role known)	100 main entry personal name 110 main entry corporate name 111 main entry meeting name 130 main entry uniform title	editor, translator, compiler,
7.	Date	210 $d date of publication, distribution, etc.	263 - projected publication date	date of publications
8.	Type	608 form, genre or physical characteristics heading	characteristics—general information (r) 006—books 006—computer files/ electronic resources 006—maps 006—mixed materials 006—music 006—continuing resources 006—visual materials	physical form
9.	Format	336 $a type of computer file (provisional)	007—physical description fixed field—general information (r) 007—map 007—electronic resource	

S.N	Dublin Core	UNIMARC	21-Mar	LIBSYS
			007—globe 007—tactile material 007—projected graphic 007—microform 007-nonprojected graphic 007—motion picture 007—kit 007—notated music 007-remote-sensing image 007—sound recording 007—text 007—videorecording 007—unspecified	
10.	identifier	001 (mandatory for unimarc) 010 (isbn) 011 (issn) 020 (national bibliography number) 300 $a general note (for url)	010 - library of congress control number (nr) 013 - patent control information (r) 015-national bibliography number (r) 016-national bibliographic) agency control number (r) 017 - copyright or legal deposit number (r) 020-international standard book number (r) 022-international standard serial number (r) 024-other standard identifier(r) 025-overseas acquisition number (r) 026-fingerprint identifier(r) 027-standard technical report number (r) 028-publisher number (r) 030-coden designation (r) 031-musical incipits information (r) 032-postal registration number (r) 033-date/time and place of an event (r) 034-coded cartographic mathematical data (r)	ISBN No., LC No.

S.N	Dublin Core	UNIMARC	21-Mar	LIBSYS
			035-system control number 036-original study number for computer data files(nr) 037-source of acquisition(r) 038-record content licensor(nr) 040-cataloging source (nr) 041-language code (r) 042-authentication code(nr) 043-geographic area code (nr) 045-time period of content (nr) 046-special coded dates (r) 047-form of musical composition code (nr) 048 - number of musical instruments or voices code	
11.	Source	324 original version note		
12.	Language	101 language of the item 300 general note	044 language	language
13.	Relation	300 general note	record links and notes general information 760-main series entry(r) 762-subseries entry(r) 765-original language entry(r) 767-translation entry (r) 770-supplement/special issue entry(r) 772-supplement parent entry(r) 773-host item entry(r) 774-constituent unit entry(r) 775-other edition entry(r) 776-additional physical form entry (r) 777 - issued with entry(r) 780 - preceding entry(r) 785 - succeeding entry(r) 786 - data source entry(r) 787 - nonspecific relationship entry (r)	series

S.N	Dublin Core	UNIMARC	21-Mar	LIBSYS
14.	Coverage	300 general note	500 - general note (r) 501 - with note (r) 502 - dissertation note (r) 504 - bibliography, etc. note (r) 505 - formatted contents note (r) 506-restrictions on access note (r) 507-scale note for graphic material (nr) 508 - creation/production credits note (r) 510 - citation/references note (r) 511 - participant or performer note (r) 513 - type of report and period covered note (r) 514 - data quality note (nr) 515 - numbering peculiarities note (r) 516 - type of computer file or data note (r) 518 - date/time and place of an event note (r) 520 - summary, etc. (r) 521 - target audience note (r) 522 - geographic coverage note (r) 524 - preferred citation of described materials note(r) 525-supplement note (r) 526 - study program information note (r) 530 - additional physical form available note (r) 533-reproduction note (r) 534-original version note(r) 535-location of originals/ duplicates note (r)	note

S.N	Dublin Core	UNIMARC	21-Mar	LIBSYS
15.	Rights	300 general note	536 - funding information note (r) 538-system details note(r) 540-terms governing use and reproduction note (r) 541 - immediate source of acquisition note (r) 544 - location of other archival materials note (r) 545 - biographical or historical data (r) 546 - language note (r) 547 - former title complexity note (r) 550-issuing body note (r) 552 - entity and attribute information note (r) 555 - cumulative index/ finding aids note (r) 556 - information about documentation note (r) 561 - ownership and custodial history (r) 562 - copy and version identification note (r) 563-binding information(r) 565 - case file char acteristics note (r) 567 - methodology note (r) 580 - linking entry complexity note (r) 581 - publications about described materials note(r) 583 - action note (r) 584 - accumulation and frequency of use note (r) 585 - exhibitions note (r) 586 - awards note (r) 59x - local notes	notes

As far as metadata interoperability is concerned, the opac display, now almost completely dependant on the AACR2 the world over requires minimum data aggregation, mainly the author, title, series, subject, etc. being the mandatory fields as prescribed by IFLA[10]. This opinion is gained after a study of the approaches taken by staff towards information retrieval (submitted to CGLA Conference on

digital Libraries 2006) There is not much variation in the data among author, title, series information as pointed out by Zing: but as he further points out the subject aggregates may differ largely. However, their aggregation can rather augment searching than limiting searching.

The aggregation should be of elements based on the displayed record at the 2nd level of description which most libraries follow. A bibliometric analysis of the aggregated keyword/ subject distribution may show some significant patterns.

5. Conclusion

The above analysis points out that though interoperability is possible between various machine readable formats there is a need for further refinement in coding and expressing the bibliographic data to make it more simpler and easy to handle based on AACR2(R). Record aggregation at the web level based on the 2nd level of description if displayed may be the most appropriate, both for data exchange and searching.

References

1. LIBSYS Electronic manual.
2. IFLA Universal Bibliographic Control and International MARC Core Programme, UNIMARC: an introduction. <http://www.ifla.org/VI/3/p1996-1/unimarc.htm>
3. IFLA Universal Bibliographic Control and International MARC Core Programme. UNIMARC manual: bibliographic format. 2nd ed., with update 1. UBCIM publications, n.s., vol. 14. München: K.G. Saur, 1996. See also the UNIMARC concise bibliographic format (May 1998) on the Web: <http://www.ifla.org/VI/3/p1996-1/concise.htm>
4. Dublin Core Metadata. <URL:http://purl.oclc.org/metadata/dublin_core/>
5. http://www.loc.gov/marc/bibliographic/ecbdhome.html
6. AACR2R 2002 release, 2nd ed. ALA, chicago,2002
7. M. Lib. Sc. course material, IGNOU, 1997
8. Sengupta, B. Cataloguing: its theory and practice.
9. http://www.dlib.org/dlib/january99/bearman/01bearman.html
10. Zeng, Marcia Lei. Metadata Interoperability and Standardization – A Study of Methodology Part II: Achieving Interoperability at the Record and Repository Levels. D-Lib Magazine June 2006 Volume 12 Number 6
11. http://www.ifla.org/VII/s13/pubs/ISBD-consolidated-July2006.pdf