

all educational institutions (including colleges), a pre-requisite for providing access to e-resources.

### Current Status

The joint project proposal submitted by the UGC-Infonet Digital Library Consortium and the INDEST-AICTE Consortium to the MHRD under the National Mission on Education through ICT has been sanctioned and the Ministry has released Rs.15.00 crores as first installment. Dr. Jagdish Arora, Director, INFLIBNET Centre and

Prof. Surendra Prasad, Director, IIT Delhi are the Principal Investigators of the Project.

The INFLIBNET Centre is currently evaluating the availability of ICT infrastructure in Govt. / Govt.-aided colleges. For further information on this initiative, visit the project Web site at <http://www.inflibnet.ac.in/n-list/>. Colleges eligible for the scheme (12 B / 2 F) are invited to register online by fill-in the Registration Form.

## Visitors at INFLIBNET Centre

This quarter, following professionals visited to the INFLIBNET Centre

1) Dr. Pawan Kumar Gupta, Deputy Librarian, Rajasthan University, Jaipur

2) Prof. M R Rawtani, Head, Department of Library and Information Science, Rajasthan University, Jaipur

3) Dr. Usha Munshi, Librarian, Indian Institute of Public Administration, New Delhi

4) Dr. R K Chadha, Joint-Secretary, Parliament Library, New Delhi

## Topics on Cutting-Edge Technology in LIS

The article on "Metadata Crosswalks" is a part of the series of Topics on Cutting-Edge Technology in LIS. The authors of this article Mrs. Vaishali Shah, STO-I and Dr. Jagdish Arora, Director, INFLIBNET Centre, Ahmedabad explain about the concept of the metadata and its crosswalks. The article also enumerates about the different metadata standards. Mrs. Vaishali Shah and Dr. Jagdish Arora can be contacted at [vaishali@inflibnet.ac.in](mailto:vaishali@inflibnet.ac.in) and [jarora@inflibnet.ac.in](mailto:jarora@inflibnet.ac.in), respectively.

Metadata, by definition, is simply "data about data". Metadata can be information about objects available in a library, museum, personal collection or a digital library. These objects may be publications in print or electronic formats in a library or physical artefacts like paintings, statues, sculptures, etc. in a museum. The primary function of metadata is to facilitate information access, search and retrieval. To achieve this goal, the metadata provides information about the document, such as its title, creator (author), publisher, and date of publication, etc. It usually includes information about the intellectual content of the document (i.e.

subject keywords or descriptors). In the case of digital objects, metadata is usually structured textual information that, in addition to describing intellectual content on object, also provides digital representation data (file formats), security or rights management information, software used for creation of digital objects, context of an image (e.g. date made, subject matter, location of digital file, etc).

In the traditional library world, bibliographic records in a library catalogue is an example of metadata, as it contains information about items available in a library, i.e. books, journals, microforms, CDs, DVDs, microfilms, audio and video tapes, etc. Metadata, in a library environment, is commonly used for describing a resource using a formal scheme of resource description applicable to different types of digital or non-digital objects. Metadata records accessible through a Library Management Software (LMS) in a traditional library fulfil several functions for users such as searching for an item, its location in the library, its status (issued, lost, damaged, reserved, on shelf, etc.). Besides, metadata in an LMS

allows librarians to administer the collection more effectively and efficiently. The same principles apply to objects available in a digital library. There are a number of standards that are available for handling bibliographic metadata. Other metadata schemes have also been developed to describe various types of textual and non-textual objects including published books, electronic documents, learning objects, archival finding aids, art objects, educational and training materials, and scientific datasets.

A standard metadata scheme should be used to ensure interoperability, inter-change and broadcast of metadata to a large community of information seekers. With availability of multiple numbers of metadata schemes developed by different communities of users, generating metadata for different standard becomes more and more repetitious, time-consuming, tedious and complex task. There is, therefore, a need to create and maintain metadata in one standard format that can made be accessible using related content metadata standards with an aim to minimize the time required for creating and maintaining metadata compatible to different standards and to maximize its usefulness to the widest community of users.

Several metadata standards have been developed by different communities of users. Each of these metadata standards has a unique focus; for example, CHIN's Humanities Data Dictionary is designed for describing and managing object collections; MARC and its variants are designed for resource discovery in a library, Dublin Core is designed for resource discovery of digital objects in Web environment; METS is a metadata standard for encoding metadata about objects in digital libraries; ETDMS (Electronic Theses and Dissertation Metadata Standard) is developed based on Dublin Core by NDLTD especially to handle metadata for theses and dissertations; Object ID is designed to identify museum objects; Federal Geographic Data Committee (FGDC) is designed for GIS products and services; Consortium for the Computer Interchange of Museum Information (CIMI) is a standard metadata schema for museum objects; CanCore and Learning Object Metadata (LOM) are designed to describe learning objects. However, many of these standard formats have commonalities. For example, Dublin Core is a subset of MARC and as such it is possible to convert metadata in MARC format into Dublin Core or vice versa. Similarly, Common Communication Format (CCF) consists of essentially same data fields / sub-fields that are available in a MARC record. As such, it is possible to convert bibliographic metadata from one standard format to another. Metadata crosswalks are essentially software tools developed to convert metadata from one standard format to another. A fully specified crosswalk provides the ability to create and maintain one set of

metadata, and to map that metadata to any number of related metadata standards. Fully automated crosswalks will enable search engines to function with any given family of content metadata standards.

### **What is Crosswalks?**

Crosswalks (sometimes called "tag mapping" or "metadata translation") are used for "translating" between metadata formats. It is a mapping of the elements, semantics and syntax, from one metadata scheme to those of another. The elements (or fields) in one metadata set are correlated with the elements of another metadata set that have the same or similar meanings. The process is also sometimes called "semantic mapping" since it essentially provides a mapping of metadata elements from one metadata standard to another. The prerequisite to a meaningful mapping requires a clear and precise definition of the elements in each standard. A crosswalk allows metadata created by one community to be used by another group that employs a different metadata standard. The degree to which these crosswalks are successful at the individual record level depends on the similarity of the two schemes, the granularity of the elements in the target scheme compared to that of the source, and the compatibility of the content rules used to fill the elements of each scheme.

A crosswalk is a set of transformations applied to the content of elements in a source metadata standard that result in the storage of appropriately modified content in the analogous elements of a target metadata standard. (Piere and Laplant, 1998). A complete or fully specified crosswalk takes care of semantic mapping as well as metadata conversion. The metadata conversion specifications contain instructions for transferring content of a field / sub-field defined for source metadata standard into corresponding fields / subfields of the target metadata standard.

A metadata crosswalk is a software tool that incorporates specifications for mapping one metadata standard to another. It enables transfer of content of fields / sub-fields defined in one metadata standard into corresponding fields / sub-fields of another metadata standard. Development of crosswalk utility, therefore, requires in-depth knowledge and expertise in the associated metadata standards. Given the fact that metadata standards themselves are often developed independently and are targeted to a specified community of users, developing expertise in different metadata standards and developing crosswalks for them is a challenging task. Moreover, maintaining crosswalks as the metadata standards change over a period of time becomes even more problematic due to the need to sustain a historical perspective and ongoing expertise in the associated standards.

Crosswalks are important for virtual collections, like union catalogues, and search engines, where resources are drawn from a variety of sources and are expected to act as a whole, perhaps with a single search engine applied. Development of crosswalks between metadata standards involve in-depth study of different types of metadata and metadata standards that exist.

### Interoperability and Exchange of Metadata

Interoperability is a critical requirement in the network environment given the fact that retro-conversion of bibliographic records, digitization and digital conversion activities are distributed amongst libraries that hold traditional print-based resources. Moreover, the products and services of retro-conversion and digitization is required to be made accessible universally through union catalogues and digital repositories. Collaboration amongst participants is, therefore, necessary in order to adopt common standards and protocols including metadata standards.

There are a number of standards that are available for handling bibliographic metadata. It is important to remember that different schemes are designed to serve distinct and diverse needs and audiences. Complementary schemes can be used to describe the same resource for multiple purposes and to serve a number of user groups. For example, a technical report could have a record in MARC21 format in the library's online catalogue and an embedded set of Dublin Core elements in its web version. The Resource Description Framework (RDF), developed by the World Wide Web Consortium (W3C), is a data model for the description of resources on the Web that provides a mechanism for integrating multiple metadata schemes. In RDF a name-space is defined by a URL pointing to a Web resource that describes the metadata scheme that is used in the description. Multiple namespaces can be defined, allowing elements from different schemes to be combined in a single resource description. Multiple descriptions, created at different times for different purposes, can also be linked to each other. RDF is generally expressed in XML. The interoperability and exchange of metadata is further facilitated by metadata crosswalks.

### Needs for Crosswalks

Crosswalks are tools that facilitate migration of metadata from one standard format to another. Crosswalks are required for reasons mentioned below:

- ▶ Different organizations are using different standards for creating machine-readable bibliographic records. Metadata crosswalks are required to convert bibliographic records in different formats into a common format with an aim to achieve universal availability of resources scattered amongst different organizations through union catalogues.

- ▶ Bibliographic Utility Networks (like OCLC / INFLIBNET / DELNET) require catalogue information from various participating libraries in a standardized format. While different libraries may use different metadata schema, crosswalks are necessary for Bibliographic Utility Networks to convert bibliographic records from different libraries into a common format before porting it to their union catalogues.
- ▶ Skeletal bibliographic information can be used to enhance bibliographic records with additional information.
- ▶ Libraries are large systems with valuable information available in various physical and electronic formats. The documents in libraries are gathered over a long period of time and their bibliographic records are created using different formats. With availability of greater number of Web-based electronic resources over the past few years, new syntaxes have also been considered for providing metadata. It is important for systems to be able to search metadata in different syntaxes and databases and have commonality in the definition and use of elements.

### Metadata Standards

There are a number of metadata standards (also called metadata schema or scheme). Some of these schemes are applicable to documents received in a library, others have broader scope. These metadata standards attempt to describe the author, the work, and the context in which the work was produced in a way that will be useful to the researcher as well as the librarians and / or technical staff maintaining the work in its electronic form.

#### 1. CCF

The Common Communication Format (CCF) was developed by the Adhoc Group set-up for the Establishment of Common Communication Format in order to facilitate exchange of bibliographic data between organizations. The first edition of the format was published in 1984, the second in 1988 and the third in 1992. The format has been developed as an ISO-2709 exchange format and adapted the second revised edition of the Anglo-American Cataloguing Rules (AACR) as a standard for rendering of information. The CCF enables an information provider to have a common format into which all data could be converted, and recipients of information would need to develop only one conversion program for incorporating incoming data from whatever source into their information system. In addition, if two or more organizations wish to exchange records with one another, it will be necessary for each of these organizations to agree upon a common standard format for exchange purposes. Each must be able to convert to an exchange-format record from an internal-format record, and vice versa. If in any network of organizations, whether national or international, there is a single standard exchange format, information interchange within that network will be greatly facilitated, both technically and economically. But if each network

has a different standard format then information interchange between different networks and among various bibliographic agencies will still be so complex as to be uneconomical, because of the number of computer programs that must be written to accommodate the translation of records from one format to another. For more information on CCF visit:

<http://unesdoc.unesco.org/images/0009/000924/092449eb.pdf>

## 2. MARC

MARC is a standard for recording bibliographic data at the logical level. A MARC record provides for elements for content, physical and process description. MARC is the established standard for the creation of machine readable cataloguing records, and underlies virtually all online library catalogues. It consequently has extensive features for describing bibliographic and copy-specific information, but has very limited structural facilities and administrative metadata which is heavily biased towards the needs of library operations. It is of limited use for incunabula or manuscripts, and other objects which may be included in a digital collection. Mappings to MARC are incorporated into most metadata systems, so that MARC records can be readily generated to allow linking from these to library catalogues. MARC is not a single standard, but rather a framework within which each country has developed an individual standard. The MARC21 is becoming a "de facto" standard as it is being adopted as a common format by various National libraries. For more information on MARC visit:

<http://www.loc.gov/marc/>.

## 3. MARCXML

All MARC-formats use the same technical framework format, ISO 2709, an extremely flexible format for the wrapping of bibliographic data. The conceptual model is not outdated, however, physical structure reflects the age of punched cards and magnetic tapes. But now information can be exchanged in a more modern XML-wrapping – MarcXchange – that builds on the same conceptual model as ISO 2709. This initiative also comes from USA where the XML-schema MARCXML was developed in 2003. It is closely associated with MARC21 and cannot be used for the other MARC-variants. The International Standardisation Committee for Information and Documentation, ISO TC46, therefore, decided in May 2003 to prepare a general XML-schema, following the same principles as MARCXML, but generalised it so as to be able to contain all MARC-formats. For more information on MARCXML visit:

<http://www.loc.gov/standards/marcxml/>.

## 4. Dublin Core (DC)

The Dublin Core refers to a set of metadata element that may be assigned to web pages so as to facilitate discovery of electronic resources. Originally conceived for author-generated description of

web resources at the OCLC/NCSA Metadata Workshop held at Dublin, Ohio in 1995, it has attracted the attention of formal resource description communities such as museums, libraries, government agencies, and commercial organizations. The Dublin Core Workshop Series has gathered experts from the library world, the networking and digital library research communities, and a variety of content specialists in a series of invitational workshops. The building of an interdisciplinary, international consensus around a core element set is the central feature of the Dublin Core. A set of 15 core elements in Dublin Core include: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights.

Dublin Core is being expanded with "qualifiers" for each core elements. For example, core element "creator" can further be qualified as "creator.author" or "creator.compiler" or "creator.editor" to specify that creator is an author, or a compiler or an editor. For more information on Dublin Core visit:

<http://dublincore.org/>

## 5. Metadata Encoding & Transmission Standard (METS)

A newly devised standard, which refines and extends the earlier Making of America II (MOA), system, METS is designed specifically to encode descriptive, administrative, and structural metadata for objects within a digital library. One of the few systems designed specifically for digital libraries, it can fulfil all basic requirements of electronic collections, albeit, in a rather verbose and clumsy manner. METS has already been used by a number of projects, including Harvard's Harvard / Radcliffe Online Historical Reference Shelf, and will undoubtedly become a standard for many projects. METS is written in XML Schema, a new way of describing XML systems, and so requires software that can handle this new format. METS depends on a complicated system of cross references within documents, and is, therefore, better generated automatically, instead of being manually edited. For more information on METS visit:

<http://www.loc.gov/standards/mets/>.

## 6. Metadata Object Description Schema (MODS)

The Metadata Object Description Schema (MODS) is a descriptive metadata schema that is a derivative of MARC 21 and intended to either carry selected data from existing MARC 21 records or enable the creation of original resource description records. It includes a subset of MARC fields and uses language-based tags rather than the numeric ones used in MARC 21 records. In some cases, it regroups elements from the MARC 21 bibliographic format. Like METS, MODS is expressed using the XML schema language. Although the MODS standard can stand on its own, it may also complement other metadata formats. Because of its flexibility and use of XML, MODS may potentially be used as a Z39.50 Next Generation specified format, an extension schema to METS, a

metadata set for harvesting, and for creating original resource metadata records in an XML syntax. Rich description of electronic resources is a particular focus of MODS, which provides some advantages over other metadata. For more information on MODS visit:

<http://www.loc.gov/standards/mods/>.

## Types of Metadata

Metadata supports efficient and effective organization, access and retrieval of information contents in a digital library. Besides providing access to intellectual contents of a document, a function analogous to bibliographic records, digital objects also require metadata about applications and formats used for creating a digital object. Such metadata is required to provide long-term access to a digital resource. The following four types of metadata are associated with the digital objects:

- a. **Descriptive Metadata:** Descriptive metadata is used to describe textual / non-textual contents of a digital object. It includes content or bibliographic description consisting of keywords and subject descriptors that may be assigned using controlled vocabulary or thesaurus like MESH, INSPEC, Library of Congress Subject Headings (LCSH).
- b. **Administrative or Technical Metadata:** Consists of information necessary to allow a repository to manage digital objects contained in it. Administrative metadata incorporates details on original source, date of creation / scanning, version of digital object, file format used, compression technology used, object relationship, etc. Administrative metadata also includes copyright and licensing information and information that is necessary for the long-term preservation of the digital objects. Administrative data may reside within or outside the digital object and is required for long-term collection management to ensure longevity of digital collection.
- c. **Structural Metadata:** Elements within digital objects that facilitate navigation, e.g. table of contents, index at issue level or volume level, page turning in an electronic book, etc.
- d. **Identification Metadata:** Used for tracking different versions and editions of same digital work, i.e. pdf, HTML, PostScript, MS Word, etc. and TIFF, JPG, BMP, etc. in case of images.

Generally, only descriptive metadata is visible to the users since descriptive metadata facilitates searching and browsing operations, and indicates the value of items in the collection. Administrative or technical metadata is used for long-term maintenance of digital collection.

## Methodology

Following steps are involved in the process of evolving of metadata crosswalks:

- i) **Identification of Metadata Format:** This process involved identification of standards metadata format used. In case of standard metadata, the tags are already assigned to each bibliographic fields / subfields. However, in case of non-standard formats, fields have to be tagged, before tag mapping. Sequential numeric tagging can be assigned in cases on non-standard formats.
- ii) **Tag Mapping:** Tags of the fields / sub-fields of source file are mapped with tags of fields / sub-fields of targeted file. For example, if the source data is in non-standard format (xls), the sequential numeric tags of fields / subfields in source file have to be mapped with the tags of fields / subfields of standard format (CCF / MARC) in target file. Getty Standards and Digital Resource Management programme provides tag mapping for all important bibliographic formats.
- iii) **Embedding Delimiters or Subfields:** In case of metadata in non-standards format, delimiters or subfield have to add before migration of data from one metadata format to another.
- iv) **Migration / Transfer the Data in Specific Format:** The last step in the process of developing a cross-walk is migration of data from its native format (source file) to target format.

## Crosswalk, Data Conversion Tools and Utilities

A number of crosswalk and data conversion tools are readily available. Some of these data conversion tools are mentioned below:

- ▶ **Excel to dBase:** Excel facilitate conversion the data from .xls format to .dbf using 'save as' command.
- ▶ **DB3ISO:** This programme converts the data from dbf to ISO. DB3ISO does not offer any tool to do so, this must be done with an external editor. The FST allows renumbering the fields and indicating the occurrence delimiter of a repeatable field. Using this programme one can convert the entire data in ISO format.
- ▶ **ISODB3:** Keeping in view the wide-spread use of CDS/ISIS database management software for creating bibliographic database of the library resources, for transferring the bibliographic data from the CDS/ISIS to the dBase based library management software this program was developed. It is still being used as utility for converting CDS/ISIS data into dBase format.

- ▶ **DB3ISO to CCF:** Use CDS/ISIS Software to import dBase ISO file and create FST (Field Selection Table) for tag mapping based on Common Communication Format.
- ▶ **ISIS-ASCII:** The utility facilitate conversion of data from ASCII files (symbol-separated files or pre-fixed fields) to ISIS. This utility is available at [http://portal.unesco.org/ci/en/ev.phpURL\\_ID=11063&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/ci/en/ev.phpURL_ID=11063&URL_DO=DO_TOPIC&URL_SECTION=201.html).
- ▶ **CCF to MARC:** INFLIBNET Centre has recently developed Data Conversion Utility (CCF to MARC). It is available on <http://www.inflibnet.ac.in/downloads/>. Import your data in CCF and Export in MARC format. This facility is in-built in SOUL 2.0.

## Conclusion

Metadata crosswalks, as software tools, are necessary to convert bibliographic records from one format to another or from different formats to one common format. The libraries need metadata crosswalks when they move from one LMS package to another. Metadata crosswalks are more relevant now because bibliographic standards like MARC21 and Dublin Core are being adopted as default standards by most of the contemporary LMS and digital library software such as SOUL 2.0, KOHA, Dspace, NewGenLib, etc.

## References

- Chandrakar Rajesh (2005); An approach to making CCF to Dublin Core; The Electronic Library, Vol 23, No.5, pp.577-590
- Pierre, Margaret St. and LaPlant, William P. Jr. Issues in crosswalking content metadata standards.  
([http://www.niso.org/publications/white\\_papers/crosswalk/](http://www.niso.org/publications/white_papers/crosswalk/))
- Godby, Carol Jean, Young, Jeffrey A. and Childress, Eric. A repository of metadata crosswalks. D-Lib Magazine, 10 (12), December 2004.  
(<http://www.dlib.org/dlib/december04/godby/12godby.html>)

## Web sites (last visited on 27<sup>th</sup> May, 2009)

CCF/B: Common Communication Format for bibliographic Information.

(<http://unesdoc.unesco.org/images/0009/000924/092449eb.pdf>)

Crosswalk between NOAA FGDC, MARC21, and Dublin Core Metadata Standards

(<http://coris.noaa.gov/backmatter/examples/MetadataCrosswalk.pdf>)

Dublin Core Metadata Initiative: making it easier to find information.

(<http://dublincore.org/>)

Learn more about the metadata crosswalk repository.

(<http://www.oclc.org/research/researchworks/schematrans/default.htm>)

MARC Standards

(<http://www.loc.gov/marc/>)

MARC: MARC 21 XML Schema

(<http://www.loc.gov/standards/marcxml/>)

Metadata Reference Guide

(<http://libraries.mit.edu/guides/subjects/metadata/mappings.html>)

Metadata Standards cross walk

([http://www.getty.edu/research/conducting\\_research/standards/intrometadata/crosswalks.html](http://www.getty.edu/research/conducting_research/standards/intrometadata/crosswalks.html))

METS: Metadata Encoding and Transmission Standards

(<http://www.loc.gov/standards/mets/>)

MODS: Metadata Objects Description Schema

(<http://www.loc.gov/standards/mods/>)

# INFLIBNET Celebrated Republic Day

The INFLIBNET Centre celebrated 61<sup>st</sup> Republic Day on Monday, 26<sup>th</sup> January 2009. The National Flag was hoisted by Dr. Jagdish Arora, Director of the Centre and thereafter he addressed the gathering regarding the activities of the Centre. Staff members with their family attended the function. A number of games were organized for staff and their children during the event.

