Standards in Library Automation and Networking

L J Haravu
Trustee, Kesavan Institute of Information & Knowledge Management

1 Introduction

We live in an increasingly globalized and interconnected world. Barriers in trade, commerce, banking, industry, science, technology, education, etc., are breaking down and will continue to become more and more borderless with time. Globalization also means increased dependency on one other in various domains. Globalization also has increased competition; consumers of all kinds now, more than ever before, have a number of choices. Globalization is in a sense synonymous with networking. Computers, telecommunications infrastructure and the Internet are not only essential but indispensable in the world we live in today.

The library and information world has traditionally been borderless long before others. However, the Internet and the worldwide web have pushed libraries to be even more interdependent than before, because of the bewildering array of sources and providers of information from which they draw their information. Equally, libraries are now facing competition from other players. The place of libraries as centres for the acquisition, description, organization, preservation and access to information has been challenged by new players such as Google, Yahoo!, etc., and subject portals, digital libraries and open access repositories. These services use new methods to acquire, analyze, display and organize information and are undoubtedly a threat to the traditional role of libraries and librarians.

How libraries can get an edge over competing services and continue to be relevant, different, wanted and useful has engaged the attention of many library professionals. Even before the advent of the Internet and the web, libraries sought to automate and network with each other so that their clients could have easier and wider access to resources. The proliferation of resources such as other library catalogues, online databases, full text publications, electronic serials, learning resources, discussion forum postings, etc., on the web has posed a considerable challenge to librarians. How to cope with these new resources and newer forms of older resources is a question that requires to be addressed by all libraries.

Library automation and networking, which until the 1990's was considered to be adequate, today do not provide all the answers. Library automation and networking is all about connecting to other libraries and information centres, search engines, peer groups, databases and experts. Thus the sceanrio today is one of many dispersed digital collections. The need is to move to inter-operable

digital libraries. This requires that we focus on standards of many kinds. Of these, metadata standards have assumed the highest importance.

To put it in simple words, libraries and networks of all kinds use different hardware and software platforms. To enable interconnectivity to such a diversity of systems and to enable mutual sharing of resources and exchange of data between them requires that all of them follow internationally agreed upon standards.

Networks operate at different levels in respect of:

- the variety of resources they handle,
- the computer hardware they use,
- operating system platforms,
- The language used to describe subject content,
- the support to protocols and standards they provide, and
- the back-end software in use.

The one element in all library automation and networking today that has assumed highest priority is the use of (wherever possible) international standards for the following.

- Metadata standards which includes standards for data elements as well as identification, description and representation standards
- Information exchange standards
- Communication standards
- Content representation standards
- Interoperability standards.

The above categorization is only used for convenience and it must be emphasized that there is considerable overlap in the purpose and function of these standards. It is important that these should be understood and used in library systems if there is to be effective automation and networking.

This tutorial attempts to:

- describe briefly the various metadata and interoperability standards, and
- emphasize the importance of these standards when planning library automation and more importantly networking of libraries.

What is metadata

Metadata may be described as having the following characteristics.

- Structured Data about data used to aid identification, description, management and location of resources

- Also may include physical resources, e.g., metadata embedded in web pages

The earliest form of metadata records are 5” x 3” card catalogue records

Metadata standards

Metadata standards comprise the following types:

- standards for data elements needed to describe different types of information objects -- some well known, some new and some newer forms of existing objects,
- library cataloguing standards such as the AACR2R and ISBD comprising rules on how to describe the data elements constituting information objects and rules for the determination and establishment of headings (or access points) (and sequence with separating symbols as in ISBD) under which the descriptive information is to be presented to catalogue users, and with the making of references to those headings.
- standards (or languages) on how to represent the metadata in computers,
- standards to represent subject content

The MARC-21 bibliographic metadata format

Records structured in the MARC-21 format are the most widely used ones in bibliographic databases. MARC-21 is a data element standard as well a standard that specifies how different data elements are recognized by a computer via numeric tags, numeric indicators and alphanumeric subfield codes. This format is ideally suited for variable length records, and to accommodate all the idiosyncrasies of bibliographic records such as the presence of a variable number of repeatable fields, presence of fields in some records and not in others. The MARC record was designed as a standard format for the exchange of bibliographic data between organizations, e.g., two national libraries, a national library and other academic libraries.

MARC formats exist for:

- bibliographic data,
- authority data and
- holdings data.

The MARC format has undoubtedly been one of the most successful and durable metadata standards in the library environment. Most western library automation software support the MARC format for import of records into a local catalogue database (also called copy cataloguing) into their local databases. More than a billion metadata records are available in the MARC format with large national libraries such as the Library of Congress (LOC) and cooperative networks such as OCLC and RLN. Many sources of MARC-21 catalogue records are available

free for use and download by other libraries worldwide. It makes good sense therefore for Indian libraries to use these resources within their library's. There are several advantages in doing this:

- The efficiency of cataloguing is greatly improed and therefore the cost of cataloguing can be considerably reduced
- Metadata imported from an authoritative MARC source ensures more completeness and accuracy in library catalogues.
- Use of authority files with MARC sources enhances the consistency of library catalogues.
- Exchange of a library's catalogue records is possible with other libraries via a common standard format.
- Moving from one integrated library management software (ILMS) to a more sophisticated one is more easily accomplished than if a library has to move from a non-standard structure to a standard one such as MARC-21.

Irrespective of how an integrated library management software (ILMS) represents its catalogue database records internally, e.g., using a relational database management system (RDBMS) or as a flat file structure, it is important that the software should allow import of records from one or other MARC data source, e.g., the national library or a cooperative network such as OCLC. In other words when considering library automation, it is important to examine the extent to which the The MARC format is supported by the ILMS. The MARC format has been an important element in the development of many union catalogues and for ensuring the exchange of data between and across library systems through out the world. In fact, the MARC format is the bedrock on which global cooperative networks such as OCLC and the RLN are founded.

Most Indian libraries by default do not use the MARC format in their catalogue databases. Similarly, most Indian library automation software do not provide the capability to seamlessly import and/or export records from and to the MARC format. This has led to many shortcomings. Some of these are given below.

- It is not possible to use copy cataloguing by downloading MARC-21 cataloguing data from authoritative sources such as the Library of Congress and other freely available web sources to build online catalogues.
- There is considerable duplicated effort across libraries apart from inconsistent entries in catalogues leading to difficulties in building union databases.
- Large libraries in India such as University libraries have not been able to act as providers of authoritative cataloguing data to smaller libraries, e.g., colleges. Common use of the MARC format across libraries would enable such sharing.

- Indian libraries cannot participate in global cooperative networks such as OCLC and RLN. Indian intellectual input is thus not adequately represented in global databases.

Closely linked with the MARC-21 format are standards (actually rules) such as AACR2R and ISBD. ISBD is a set of rules produced by the International Federation of Library Associations (IFLA) to describe a wide range of library materials, within the context of a catalogue. ISBD defines a number of areas, e.g., title and statement of responsibility, edition, material-dependent information (for example, the scale of a map or the duration of a sound recording). These rules organize the bibliographic description of an item. Librarians use the AACR2R and ISBD in conjunction with MARC.

Authority and holdings records

It is important to emphasize that MARC-21 is a format not only for bibliographic records. MARC-21 formats exist for authority records and holdings records. Good ILMS software provide support to both these formats. There are several advantages of adhering to these formats in addition to the bibliographic formats.

- Local authority file records can be created via imports from an authoritative source such as the LoC.
- An ILMS could ensure that new entries of data are validated against the local authority data to ensure better consistency in the use of personal names, subject terms, etc.

It is high time that Indian libraries embrace the MARC format and look for software solutions that support this format if they wish to effectively contribute and benefit from networked resources within India and outside it. Equally importantly, today's ILMS should also provide crosswalks to and from other metadata standards.

Dublin Core Metadata Element Scheme (DCMES)

The emergence of a variety of new digital information objects and digital formats of traditional materials (e.g., images, text, sound) required adaptations of traditional descriptive methods and this triggered the development of a metadata description scheme which would be simple and also cover a wide range of materials. Work on defining such a metadata element scheme which later became known as the Dublin Core Metadata Element Scheme (DCMES) began in 1995 in an invitational workshop convened by the Online Computer Library Center (OCLC) and the National Center for Supercomputing Applications (NCSA), at Dublin, Ohio. The workshop was attended by professionals from librarianship, computer science, text encoding, and related areas to advance the

state of the art in the development of resource description of records for networked electronic information objects. The inititaive, now known as the Dublin Core Metadata Intiative (DCMI) has had the benefit of consultations from specialists from many countries over the years. DCMES is now an international standard (ISO 15836:2003) and is also used as a national standard in countries such as New Zealand, Australia, USA, and the UK among many others.

Some of the other reasons for a new metadata schema for today's needs cited are:

- New types of information objects, (e.g., web pages, computer programs, discussion forum postings) require new methods of description to accurately describe their unique content.
- Increasing technological capabilities allow delivery of richer and more complex 'objects' (e.g., an interactive learning object may have text, video and sound objects embedded into it). Such objects will need more complex descriptive methods.
- Search methods have changed. Users rely more and more on keywords than on field-directed searches, e.g., personal name or title searches. Advanced searching, browsing, text-mining algorithms and relevance ranking are being employed in addition to the use of hyperlinking between two or more objects to ensure better search results and resource discovery.
- Users no longer feel limited to the resources of their institution or institutional network. They expect virtual access to resources wherever they may be available.
- The preservation needs of digital objects are very different from those of physical objects. Digital objects require to be monitored actively and detailed documentation about these is useful in helping to preserve these objects as compared to paper-based objects.
- Many new systems for direct use by authors and other professionals are now becoming available, e.g., document management systems (DMS) and digital archives applications. These require that authors of various types of information objects should themselves create the metadata for the objects they author. A simpler scheme is called for in such situations and DCMES is an answer to such a need.

DCMES is a simple yet effective element set for describing a wide range of networked resources. The Dublin Core (DC) standard defines two levels: Basic and Qualified. *Basic Dublin Core* comprises fifteen elements; *Qualified Dublin Core* includes an additional element, Audience, as well as a group of element refinements (also called qualifiers) that refine the semantics of the elements in ways that may be useful in resource discovery. The semantics and grammar of Dublin Core have been established by an international, cross-disciplinary group of professionals from librarianship, computer science, text encoding, the museum community, and other related fields of scholarship and practice. Most elements

also have a limited set of qualifiers or refinements, attributes that may be used to further refine (not extend) the meaning of the element. Each of the elements is optional and may be repeated as many times as needed to describe an instance of a digital resource, e.g., the full text of a journal article or an annual report; a digital image of an insect pest. Standard ways to refine elements have been defined by the Dublin Core Metadata Initiatitive (DCMI) and they encourage the use of encoding and controlled vocabulary schemes.

A metadata scheme is expected to serve the following purposes:

- Description of the intellectual content (descriptive metadata) of an information object. The purpose of this metadata is to facilitate resource discovery.
- Management of the digital file and the maintenance of a digital object. This is especially important for successful preservation – that is to insure that it will be accessible even after the original hardware and software become obsolete.
- Definition of the internal structure of an information object (structural metadata). This metadata provides the connection between the intellectual organization of the information object and the digital representation of that organization.
- Provide information related to rights and use, i.e., the social and legal restrictions on how the information object can be used, including copyright information, intellectual property and breeders rights, and usage guidelines (administrative metadata). Information on rights of use is important since many copyrighted digital objects have become accessible via the web.

Although Dublin Core (DC) was originally developed to describe document-like objects, DC metadata can be applied to other resources as well. Examples of the use of DC in many projects, some of which are quite unique in that they integrate text, images, videos, sound, and multimedia resources using Dublin Core as the basis for description and resource discovery may be seen at http://dublincore.org/projects/.

Application Profiles

The DCMES is more like a framework and allows domain-specific extensions. In other words, an information agency that needs to describe resources in a specific domain need not re-invent metadata elements. They can use the basic elements, the element refinements of the DCMES and add domain specific extensions together with content rules, encoding schemes and controlled vocabularies. These domain-specific schemes also have come to be known as schemas or **Application Profiles.** The use of DCMES as the underlying basis of different metadata schemas or Application Profiles (AP) enables the easy exchange of data between two or more metadata databases. For instance, if closely related

information centres dealing with rural development, agricultural research, agricultural statistics, fisheries, agricultural universities, agricultural extension centres, etc., all use a common AP such as the AGRIS-AP, they will all be able to exchange information (at least at the level of common elements) with each other much more easily than if each of them follow their own metadata schemes.

An Application profile is defined as a type of metadata schema comprising data elements drawn from one or more namespaces, combined together by implementors, and optimised for a particular local application.
This model allows different communities to use the DC elements for core descriptive information with extensions which make sense within a more limited area or discipline. The idea is to ensure many metadata databases to be able to exchange data with each other because all of them have the same core elements.

In view of the fact that there are more than one internationally accepted metadata scheme in use, e.g., MARC-21, DCMES, there are computer programs which allow the transformation of data in one scheme to another. This allows the exchange of data at the base level. Inevitably, however, when data from one metadata schema is converted into another, there is bound to be loss of data.

In today's globalized, networked world, the use of DC is widespread in many digital repositories and especially those subscribing to concepts of and the software compliant with the protocols of the Open Archives Initiative (OAI). Effective information networking today requires that libraries utilize software that supports newer metadata standards such as Dublin Core and become capable of contributing to and participate in such initiatives.

As a result library automation solutions now need to provide support to inter-operable standards such as DCMES Libraries in India seeking automation solutions need to take into account the need for software to support

Markup languages

How should metadata be actually represented within a computer is another area which has received considerable attention. For instance in MARC-21, the metadata is represented using tags, indicators and subfield codes. The idea is that it should be possible for a computer program to easily identify the title of a resource, its creator, kind of resource, etc. The tagged structure was invented in the 1950's when on-line storage was expensive. In fact this structure was well suited for variable length records such as bibliographic records on magnetic tape. However, the emergence of the web and formats such as HTML which are Markup langauges and the fact that many computer applications, especially Internet browsers, can easily interpret data in such formats. The fact that on-line storage has become cheap thanks to hard disk technology, has made markup languages as the preferred way of representing data in computers.

More recently the development of XML and its advantages over HTML have made XML as the *lingua franca* of data on the web. XML representation of data in a computer has several advantages.

- There is clear demarcation between content and display unlike with HTML
- Each application area (e.g., banking, e-commerce, libraries, chemistry, geophysics, etc.,) can define Markup tags specific to their area. The XML tags and their definitions in each application area is agreed upon internationally (or as W3C standards) and these definitions are all provided as what is known as a Document Type Definition (DTD). The DTD is accessible at a web site and all XML documents in the area are validated against such DTDs to ensure consistency of data transmitted across the web.
- XML data is textual data and is therefore not subject to the quirks of proprietary software such as word processing. Thus XML data is readable by computer programs irrespective of the operating system platform being used by a computer that connects to the data over the web. In a sense therefore XML data is future-proof.
- Due to the wide spread use of XML in web-based systems, there are standard open source computer programs (freely downloadable from the web) that can handle (search, index) such data.
- Many library-related metadata standards such as MARC-21 and Dublin Core have already developed DTDs for the storage of data in XML form and for inter-conversion of data, e.g., there is a DTD which allows MARC-21 data to be converted into MARC-XML form using a DTD that has been developed by the LoC and vice-versa.
- Of special interest to libraries: XML's fixed character set is Unicode (and its extensions), which allows diacritics, special characters, and non-Roman data to be handled like ordinary text. HTML was heavily oriented to English and did inhibit its use in other languages.
- Due to Unicode and platform neutrality, XML offers the greatest promise of data longevity (or future proofing), as hardware, software, and network protocols continue to change.
- XML also provides for the unambiguous identification of complex data structures that can be treated as objects, well suited for bibliographic data.

Many of the new metadata standards such as Dublin Core and Application Profiles, expose their data as XML files due to the advantages of XML mentioned above. The LoC also has made available a DTD to convert MARC records to XML. New ILMS are expected to provide the import of data in XML formats to local structures.

As of today, the thinking is that it is not a question of choice between XML and MARC, or between XML and relational databases. The attempt will be to see how these technologies can be used synergistically to enable greater and

seamless integration of library resources with the many other web resources. There is little doubt that effective library and information networking requires that libraries pay more and more attention to software that will allow them to use the new markup languages in their database and networking transactions.

Interoperability standards

A simple definition of the term interoperability is the ability of software and hardware on different machines (and on different networks) from different vendors to share data. Interoperability may also be said to be an approach for dealing with the diversity and heterogeneity of the marketplace. Interoperability includes the exchange of data, records, and messages between computer systems across different hardware, operating systems, and networks. The greater the ease and seamlessness of the exchanges, the greater is the interoperability. Interoperability is sought to be achieved by establishing standards that different vendors of software and hardware can adopt so that they can share data and information.

> When two or more organizations exchange records on a regular basis, such as the transfer between shared cataloguing services and local systems, these arrangements are manageable. Similarly, when libraries migrate from one library automation system to another, they can plan and implement a one-time data transfer. Today's requirements are much more complex, however. Ideally, libraries would like to treat multiple bibliographic databases as comprising an online, real-time, virtual digital library. Such a structure would enable libraries to exchange records on a casual, ad hoc basis, and would enable information seekers to search one or more external digital libraries as though all were one single resource. Only recently has it become possible to support generic client-server relationships between information systems. The most advanced is the z39.50 standard that is being implemented for online catalogues and other bibliographic retrieval systems. The standard and its implementation are described below.

The z39.50 standard

ANSI/NISO z39.50 is the American National Standard Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection. Version 3 of the standard with new facilities and services was released in 1995. ANSI/NISO z39.50 defines a standard procedure for two computers to communicate for the purpose of information retrieval. z39.50 makes it easier to use large information databases by standardizing the procedures and features for searching and retrieving information. Specifically, z39.50 supports information retrieval in a distributed, client and server environment, where a computer operating as a client submits a search request (i.e., a query) to another computer acting as an information server. Software on the server performs a search on one or more databases and creates a result set of records that meet the criteria of the search request. The server returns records from the result set to the client for processing. The power of z39.50 lies in its ability to separate the

user interface on the client side from the information servers, search engines, and databases. Z39.50 provides a consistent view of information from a wide variety of sources, and it offers client implementors the capability to integrate information from a range of databases and servers. ANSI/NISO z39.50 can be implemented on any platform. In other words z39.50 enables different computer systems—with different operating systems hardware, search engines, database management systems—to interoperate and work together *seamlessly*. A z39.50 implementation enables one interface to access multiple systems, providing end users with nearly transparent access to other systems. Users access multiple systems with the familiar commands and displays of their own local systems, without having to learn new commands and search techniques. Such searches are called Federated Searches. The results of the searches are presented on the local system - again, in the formats and styles that users are accustomed to.

Originally developed for use in computer communications networks using Open System Interconnection (OSI) protocols, z39.50 is now successfully operational in the environment of the Internet. Interoperability testing of various z39.50 implementations has made it clear that the transport services (i.e., the TCP/IP) of the Internet could be used successfully for the protocol. Using z39.50 over the Internet is a major step forward.

Z39.50 implementations are available both as standalone applications as well as integral parts of an integrated library automation system (ILMS). In today's networked world, a library automation system that provides federated searching capabilities for its users will obviously be at a great advantage than others.

Web services protocol-based applications for library networking

Although the z39.50 protocol has been around since before the Internet era, there are several problems with this implementation of federated searching. Some of these are:

- Z39.50 is a complex and high-level protocol
- The implementation barrier is high for present web-based systems
- Federated search results from different servers are retrieved at various times depending on the availability of the server and network resources. In other words, this protocol is very information resource and bandwidth-sensitive.

As a result of these limitations, a new generation of services have been developed using web services protocols for federated searching. The two new services are SRW (Search and Retrieve Web Service) and SRU (Search and Retrieve URL Service).

These two new protocols have several advantages:

- All the advantages of Z39.50
- Implementation barrier is very low for web based systems, since this is based on high level web-based protocols
- SRW works with all high level protocols, viz., http, ftp and smtp

SRW and SRU are web Services-based protocols for querying Internet-accessible databases and returning search results. SRW and SRU requests and results are similar, their difference lies in the ways the queries and results are encapsulated and transmitted between client and server applications. SRW uses existing, well tested and easily available technologies such as Simple Object Access Protocol (SOAP) and XPath in order to perform what has been done in the past using proprietary solutions. The design has had the benefit of 20 years of experience with the Z39.50 information retrieval protocol, and is both robust and easy to understand while still retaining the important aspects of its predecessor.

A SRW query is represented in CQL, the "Common Query Language", designed for human readable, human writeable, intuitive queries. It supports very simple queries -- for example an unqualified single term (e.g. "cat") -- but maintains the expressiveness of more complex languages, to represent arbitrarily complex queries. Many libraries in the west using specialized software and/or a new generation ILMS have begun implementing SRU/SRW protocols. Similarly, many libraries have also made their databases accessible via SRW/SRU servers. Indian libraries must become aware of these developments as they will undoubtedly increase a library's capability to provide easier access to specialized databases in single or federated mode. Provision of such access is a clear way of differentiating libraries from their competitors.

OpenURL

More and more full text resources are now becoming accessible on web sites of publishers, libraries, professional bodies. These resources are identified by a URL. However, when the web site decides to move the resource to some other location within the site or even remove it from their site, the old URL captured in a database will no longer point to the resource. This leads to frustration for the user. This is particularly so when users search journal article databases or when they access e-journals. Access to the full text of articles is sought and this is where the concept of a persistent URL becomes necessary. However, users, working as they do in an organizational context, have defined access and view privileges for e-serials and databases. For instance, the organization they work for may not have access to an e-journal's full text but only to the abstract of articles.

The OpenURL standard is a protocol for interoperability between an information resource, e.g., an on-line or CD-ROM database and a service component, referred to as a link server, which offers localized services. The underlying concept of the OpenURL standard is that links should lead a user to appropriate

resources. A link server, such as the SFX server from the ILMS, ExLibris, defines the context of the user. When the link server accepts an OpenURL as input, it acts on it to provide users with services that comply with their institution's collections and policies.

The OpenURL standard enables a user who has retrieved an article citation, for example, to obtain immediate access to the "most appropriate" copy of that object through the implementation of extended linking services (also called context-sensitive linking). The selection of the best copy is based on user and organizational preferences regarding the location of the copy, its cost, and agreements with information suppliers, and similar considerations. This selection occurs without the knowledge of the user; it is made possible by the transport of metadata with the OpenURL link from the source citation to a "resolver" (the link server), which stores the preference information and the links to the appropriate material.

A *link resolver* interprets incoming OpenURLs, takes the local holdings and access privileges of that institution (usually a library) into account, and displays links to appropriate resources. A link resolver allows the library to provide a range of library-configured links and services, including links to the full-text, a local catalogue to check print holdings, document delivery or ILL services, databases, search engines, etc. Many of today's information providers can generate and output OpenURLs. Libraries that have access to an OpenURL-compliant link server such as SFX, must furnish an OpenURL for their full text resources, e.g., their own monographs, reports, etc., in digital form.

The OpenURL system works in conjunction with a body called CrossRef, an independent membership association, founded and directed by publishers. CrossRef's mandate is to connect users to primary research content, by enabling publishers to do collectively what they cannot do individually. CrossRef establishes Digital Object Identifiers (DOI), an identifier using an open standard, to scholarly and professional publications. CrossRef operates a cross-publisher citation linking system that allows a researcher to click on a reference citation on one publisher's platform and link directly to the cited content on another publisher's platform, subject to the target publisher's access control practices. The citation-linking network today covers millions of articles and other content items from several hundred scholarly and professional publishers.

CrossRef DOIs link to publisher response pages, which include the full bibliographic citation and abstract, as well as full-text access (for authenticated users or at no charge, as determined by the publisher). The publisher response page often includes other linking options, such as pay-per-view access, journal table of contents and homepage, and associated resources. CrossRef is a collaborative membership network, and not a product for purchase.
DOI, an open standard, is an alphanumeric identifier for digital content, such as a book or journal article. The DOI is paired with the object's electronic address, or

URL, in an updateable central directory, and is published in place of the URL in order to avoid broken links while allowing the content to move as needed. DOIs are distributed by publishers and by CrossRef, and there is no end-user charge associated with their use. As an identifier, the DOI can be incorporated into many different systems and databases. The DOI system provides a framework for persistent identification, managing intellectual content, managing metadata, and linking customers with content suppliers.

The key benefits of the CrossRef DOI for libraries are:

- Truly persistent links
- Increased usage of acquired resources
- Expanded access to content not owned
- Enhanced and customizable localized linking
- Metadata retrieval privileges at no cost to most libraries

Although these new fangled standards seem complex at first glance, it is important that Indian libraries become familiar with these and look for solutions which will support these new standards. The move towards provision of access to full text, the fact that more and more digital resources are on the web, the emergence of the open archives initiatives (OAI) and digital repositories has made it absolutely vital that libraries embrace these new standards and technologies if they wish to differentiate themselves from their competitors.

Open Access Initiative – Protocol for Metadata Harvesting

The OAI-Protocol for Metadata Harvesting (OAI-PMH) defines a mechanism for harvesting records containing metadata from digital repositories. The OAI-PMH gives a simple technical option for data providers to make their metadata available to service providers, based on the open standards HTTP and XML. The metadata that is harvested may be in any format that is agreed by a community, although unqualified Dublin Core is specified to provide a basic level of interoperability. Thus, metadata from many sources can be gathered together in one database, and services can be provided based on this centrally harvested, or "aggregated" data. The link between this metadata and the related content is not defined by the OAI protocol. OAI-PMH does not provide a search across this data, it simply makes it possible to bring the data together in one place. In order to provide services, the harvesting approach must be combined with other service mechanisms, e.g., a search, alerting services.

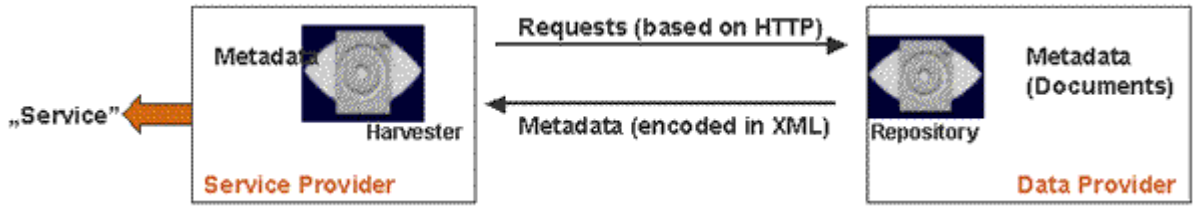
Two players in the Open Access movement are Data Providers and Service Providers. A Data Provider maintains one or more repositories (web servers) that support the OAI-PMH as a means of exposing metadata. A Service Provider, on the other hand, issues OAI-PMH requests to data providers and uses the metadata as a basis for building value-added services. A Service Provider in this manner is "harvesting" the metadata exposed by Data Providers. They use the harvested metadata for the purpose of providing one or more services across all the data. The types of services that may be offered include a search interface,

peer-review system, etc. One organization can play both roles, offering both data for harvesting and end-user services.

Libraries now have the possibility to selectively harvest useful metadata from one or more OAI-PMH compliant archives and into their own database and provide search and access not only to metadata but also to digital content for the benefit of their users. .

Similarly, library's can also become data providers and allow other libraries and information centres to harvest metadata for their users.
OAI-PMH is a simple protocol based on HTTP and XML, it allows for rapid deployment. The figure below is a graphical representation of the way OAI-PMH works.

Basic functioning of OAI-PMH



Conclusion
Infrastructure may be likened to the foundation of a building and standards to the bricks. Only when these two are cemented by appropriate software that useful applications can result.

Libraries and information centres in India have always been faced with difficulties when it comes to choosing software to automate their libraries. This problem is even more exacerbated today when they are facing competition from other players as they will need to justify the choice of a software even more convincingly to their managements than before. In such a scenario, the best approach is to concentrate not so much on criteria that will help the library to become more efficient but on how software will make it possible for the library to utilize networked resources for the benefit of their users. In doing so, libraries would do well to ensure that the software they choose will support networking and exploitation of web resources apart from automating their internal operations. In other words, the software that they choose must support international data and open interoperability standards that were discussed in the previous section of this paper, and should be web-based because only such software will be flexible, have the possibility of the integrating the access to the library's resources along with other web resources that users have become accustomed to. Some of the criteria in choosing appropriate software are given below.

- Does the software support international data standards such as MARC-21 and the DCMES?
- Does the software support copy cataloguing using authoritative MARC-21 cataloguing sources via the web?

- Will the software allow the library to create its own network of libraries cost-effectively? Will it be possible to use the web to allow networked libraries to utilize a single server/software infrastructure? This will be cost-effective and will expand access to resources for the users.
- Will the software allow federated searching of one or more library and specialized subject databases using an interoperability standard such as z39.50 or newer ones such as SRU/SRW? This will allow access to useful resources beyond that of the library.
- Will the software allow the building of an OAI-PMH compliant institutional repository of self-archived materials and will it allow the library to harvest metadata from other OAI-PMH compliant repositories? This will allow specialized resources (journal articles, reports) and full text to become accessible via the library's interface.
- Will the software allow the entry, storage and searching of data in non-Roman scripts? This is particularly important for Indian academic libraries which deal with materials in one or more Indian languages in their collections.
- Will the software allow other libraries to download metadata into their databases?
- Will the software allow access to digital content such as full text, images, audio, video, web pages, and databases and full text on CD-ROMs (that may be mounted on an institution's CD Server).
- Does the software allow the description of a wide range of materials including analytics, web resources, learning objects, audio, video, etc., and not only conventional monograph and serial items?
- Does the software's OPAC allow flexible searching of library holdings and does it empower users by offering access to other services, e.g., interlibrary loans, document delivery services, alerting services and access to other web resources, e-serials, search engines?