PROGRESS IN WEB-CATALOGUING USING STANDARD METADATA SCHEMAS

Jiban K Pal ***Falguni Pal***

Abstract

Web cataloguing' or 'metadata creation' is an effective strategy to enhance the resource-discovery on digital collection that can be accomplished by standard metadata schema. Technological capabilities now allow multiple metadata schemas for discovering web resources. This paper aims to present the progress of web-cataloguing using 'metadata' and its impact on precision of search results, in discovering web resources. It also proposes various means of creating metadata in order to pursue with the large databases and/or dynamic web pages. Several metadata initiatives provide glimpses of recent activities on 'metadata'. Here, an attempt has been made to discuss about different metadata schemas of which Dublin Core (DC) & MAachine Readable Cataloguing (MARC) seem to be widely accepted and used in cataloguing web resources. It also seems imperative to make a few modifications of the existing crosswalk between the DC & MARC-21 elements to be used in conversion of metadata from one into another. This discussion brings out the idea of OAI-PMH to make the metadata available to harvesters and attempts to explore the mechanism of DSpace for exposing metadata in a digital repository.

Keywords: Web-Cataloguing; Resource discovery; Metadata Standards; Metadata harvesting; Metadata research.

1. Introduction

With the rapid advancement of computer and communication technology, bibliographic items have been published in increasingly diverse forms. Disks in different formats, interactive multimedia, and Web pages are some of the examples. Some unique genres (document types) are only applicable to digital resources like the Web, but not to traditional print resources (and vice versa). For instance, boundaries in document types and editions are blurred and more difficult to establish in electronic resources than print. So, clear and standard guidelines are needed for helping the users as well as cataloguers[1]. These standard guidelines are essential to recognize both the resources. Otherwise, standards ensure compatibility and facilitate interchange ability of information sources across the global network system. It also improves the quality of information services and reduces economic and technical barriers in information flow. Adoption of good set of standards by the constituents of a network is a pre-requisite to aim at a certain level of quality consistency. It contributes to optimization in utilization of resources and facilities the total network[2].

In order to cope with these new developments, recent past, cataloguing rules have undergone more frequent and drastic changes than ever[3]. Nancy Olson summarizes what has been going on in the cataloguing field to accommodate electronic resources in new formats. In her article, Olson describes both the new rules and the unchanged rules that are applicable to the cataloguing of remote access electronic resources[4,4] In 2000, the Library of Congress hosted a conference on bibliographic control for the new millennium. This conference highlighted several issues related to the cataloging of electronic resources and library standards for bibliographic control such as Z39.50 interoperability, DDC for organizing web resources, effectiveness of different metadata standards, online-serial cataloguing, etc. In 2001, Gorman raised the fundamental issue – is it worth cataloguing Internet resources? He argues that librarians should select only those worthy of cataloguing and catalog fully the selected resources. He also asserts that cataloguing electronic resources should be done following standard rules in order to facilitate an efficient information retrieval5. Gradually many new ideas, standards, and proposals have been launched in an attempt to make the organization of electronic resources easier and more efficient. Although some exiting ideas and projects are in progress, some concerns about the future directions in cataloguing of Web resources remain.

Web resources are growing at an exponential rate. Semantic based search engine and meta-search engine stimulates resource discovery on digital collections. Practically, there are enormous sources that are relevant to any user's query, uncountable stacks of web resources (so called Cyber jungle – Ding & Marchionini, 1996) drastically creates a hurdle for accessing the information effectively and efficiently. In fact, a considerable noise exists in retrieval of information. This is basically due to the uncountable number of heterogeneous resources available in large cyberspace. Moreover, search engine scalability and retrieval effectiveness is likely to decline, which stimulates to consider an alternative or additional resource discovery mechanism. A 'metadata' schema is the way to improve resource discovery on web resources and can be accomplished by embedding structured metadata in web-resource headers; through installing a metadata search engine, for searching on individual/ multiple metadata elements (e.g. HotMeta – developed by Distributed Systems Technology Centre). Technological capabilities now allow multiple metadata schemas for discovering Web resources.

This paper aims to present the progress on web cataloguing using 'metadata' and intended to discuss about the impact on precision of search results, in discovering web resources. It proposes various means of creating metadata in order to pursue with the large databases and/or dynamic web pages. Here an attempt has been made to discuss about the growth of different metadata initiatives. Among various metadata schemas 'Dublin Core' (DC) and 'MAachine Readable Cataloguing' (MARC) seems to be most widely accepted and used in cataloguing web resources. It also seems imperative to make a few modifications of the existing crosswalk between DC and MARC-21 elements to be used in conversion of metadata from one into another. This discussion brings out the idea of OAI-PMH to make the metadata available to the harvester in a digital achieve. It also attempts to explore the mechanism for exposing the metadata in Dspace enabled digital repositories. In practice, it is really difficult to organize a single metadata creation procedure over the distributed network, but it has been realized by the information community that the metadata could have some ability to enhance the efficiency of the search engines.

2. Metadata – The Concept

Until last decade, only the philosophers are used the term 'metadata', but today it has become a buzzword in the information society. Really, it is difficult to get any literature on electronic-cataloguing that discards the concept of metadata. This concept is equally important for librarians, authors, digital archivists, database developers, and seekers of electronic information. It implies metadata is inevitable for searching – i.e. a process of matching query terms with the terms embedded in the source contents. However, "metadata is expected to improve matching by standardizing the structure and content of indexing or cataloguing information". This phenomenon receives an increasingly greater importance in order to enable the mechanism for efficient retrieval of information in a complex and large distributed environment.

The term 'metadata' has an ambiguity and difficult to define –"it is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource[6]". Dr. Warwick Cathro defines metadata as –

> "An element of metadata describes an information resource, or helps provide access to an information resource… For example, a library catalogue record is a collection of metadata elements, linked to the book or other item in the library collection through a call number. Information stored in the "META" field of an HTML Web page is metadata, associated with the information resource embedded within it. The indexing data held by Web crawlers is also metadata (though not very good metadata), hyper-linked to the information resource through the URL[7] ".

The classic definition of metadata is "data about data". It describes the attributes and contents of an original document or object. Say for example, if a Web page has an author, a title, date of creation, and unique address – these elements constitute metadata about the page. Metadata is basically considered to denote the information about digital object only. "It is an Internet-age term for information discovery that librarians traditionally have put into catalogs, and it most commonly refers to descriptive information about Web resources"[8].

Many researchers agree that the term metadata is useful to describe the electronic information resources. Milstead & Feldman emphasized – "the term is generally applied to electronic resources (though it doesn't have to be) and refers to 'data' in the broadest sense; datasets, textual information, graphics, music, and anything else that is likely to appear electronically. While the concept includes indexing and cataloguing information (information for 'resource discovery' in web-speak), it can go far beyond conventional document representations, such as MARC records"[9]

3. Metadata Creation for Web Resources

Growing amount of Web resources, available in diverse electronic format, demands for creating 'metadata' with adequate quality. In the absence of good quality metadata it would be very difficult to discover 'relevant documents' timely. Virtually the question comes, how we can create metadata for dynamic resources? Who can create a better quality metadata? – Metadata for web resources (mostly dynamic) can be generated through automatic or traditional means. In this regard traditional techniques (using human efforts) are highly labor-intensive and limiting when large databases or dynamic pages are involved. The problems of traditional techniques have insisted to develop the interest in techniques for generating metadata by automatic means, which pose a challenge to traditional one. In practice, search engine spiders, Web crawlers, HTML & XML editors, etc. produce various types of metadata via automatic means. Such devices can generate fairly accurate metadata for a few specific elements, (say for date, language, etc.) but these tools failed to produce metadata appropriately when it is more intellectually demanded for certain elements like creator/author or title of the object or subject, etc. However, in automatic means there are no consistent filtering practices to ensure the quality/credibility of metadata and certain structural factors in generating software's/ search engine spiders hamper the production of better quality metadata. Therefore, often many systems prefer traditional processing exploiting human-intellectual efforts to generate schema-specific metadata.

Further question rose, who will generate metadata? – "Metadata professionals and Resource authors represent two main classes of metadata creators". Metadata professionals (read as cataloguers and indexers) have an intellectual ability, which they achieved through training and experience, and obviously gained their proficiency in the use of content-value and descriptive standards. "Although few researchers have noted problems with inter indexer consistency"10. Ideally professional metadata creators could ensure the efficiency but they are limited in their availability and they never satisfy the law of parsimony. Certainly these professionals can produce high quality metadata11. Notionally 'resource author/s' make them solely responsible to create the intellectual content of an object. These resource-authors might also be involved for creating acceptable quality of metadata. "Yet there is a perception that author-generated metadata will be of poor quality and may actually hamper rather than aid to resource discovery"12. But, a counter logic imposed through a study of Greenberg et al and reported that resource authors have an ability to create adequate quality of metadata as –

> "…creators are intimate with their work, they want their work to be discovered and consulted, and they know their audience and can thus describe their resources appropriately. These factors support the hypothesis that resource authors can create acceptable metadata when working with the Dublin Core, a schema initially designed for resource authors… and in some cases they may be able to create metadata that is of better quality than what a metadata professional can produce"[13].

So, the creators like scholars, painters, artists and even class-teachers regularly creating metadata for their technical or artistic works or objects in the form of abstract, keyword, and many other means like web-forms, web-templates, posting their objects to repositories to make their object more accessible on the web. In real situation most of the digital repositories or open archives (viz. NDLTD, NEEDS, etc) prefer author generated metadata. Obviously this practice makes sense to produce a consistent and quality metadata in consideration with the phenomenal increase of the web-resources and in terms of the economics of hiring professional metadata creators. In such an orientation resource-author normally creates metadata (either by him or under his supervision) at the time of object creation. Several agencies (e.g. FGDC, EPA, etc.) have taken a dominating role in developing web-based metadata entry forms to generate metadata for the particular object. Sometimes the agencies provide a guideline to web-developers on use of "Meta tagging for search engines"[14]. Moreover, a good number of initiatives (often voluntarily by libraries or by specialists) have been taken so far to catalogue the web resources. In this juncture OCLC's "InterCat" project may be considered as landmark15. Basically such initiatives have taken to determine the most useful and relevant sites to satisfy a specific requirement.

4. Functions of Metadata

Multiple types of metadata has been discussed in various literatures and each type has their own functions – say 'descriptive metadata' enumerates the object to discover or identify the information resources, whereas, 'administrative metadata' depicts information to administrate and manage the resources that includes legal rights to access (IPR), when & how created, version control, etc. Similarly, 'structural metadata' describes the way of bringing similar resources or compound objects together; 'technical metadata' indicates the system functions and technical behavior of metadata (viz. formats, compression ratio, data authentication, encryption keys, etc); and 'preservation metadata' provides information required for preservation management like archiving the resources, physical features, survival challenges, etc.; and many other types of metadata varies in their functions[16].

So, various contexts could emphasize different functions, however, the prime functions of metadata are to help in data management and to ensure an efficient retrieval from a large digital collection. In 1994, Strebel et al describes three main functions of metadata – data access, data management, and data analysis. In fact, metadata can make it possible for users to determine the availability of Information (whether the information objects exist - how many and where are the objects - are the objects same) and usefulness of same information (whether it is useful/ good/ authentic or not). Metadata functions can also be described in two different levels – one is system level, where the metadata provides facility for interoperability and integrity of resource discovery tools. Another is end-user level, where the metadata ensures the capacity to determine – type of data available, how to acquire it, whether meets the requirement, and how to capture at user-end, etc.[17]

Although, function of metadata varies with its category and use, still a significant number of writings focused on the function to support 'resource discovery'. This function ensures to support in retrieval or discovery or access to digital information. Objectively it provides an effective mechanism to select or locate the data/ information that is highly relevant to the user's query, and make it possible to users to determine. In principle metadata acts as surrogate for a larger whole. However, a metadata schema also establishes the standard structure and terminology for the resources available in the large distributed network; where the abstracts, keywords, indexes, and other bibliographic data make the original-resources available to the users – hence the metadata is functionally justified. Moreover, metadata not only support to resource-discovery but also promises — rights management, links to e-resources, interoperability using standard schemas and protocols (e.g. cross search system through Z39.50 protocol / metadata harvesting using OAI protocol), digital object identification (DOI), and archiving to make the metadata accessible into the future.

5. Metadata Initiatives and Trends

In the orderly development of metadata research, perhaps too many formal initiatives are underway. While there are disparate sets of needs to formalize and standardize metadata, several attempts have been made by libraries (as LC), federal agencies (as FGDC), voluntary organizations (as W3C), and other groups to satisfy the perceived interests of those communities. This widespread interest among different metadata standard groups results the growth of conflicting standards and projects for associating diverse types of metadata with the diverse electronic resources. A few of them emerged from the library and information research community and others have come from the peoples those realized the need for certain standard. Therefore, "metadata takes a variety of forms, both specialized and general — new metadata sets will develop as the networked information infrastructure matures — different communities will propose, design, and be responsible for different types of metadata"[18].

This situation has stimulated metadata communities to meet and talk earnestly all over the world through various workshops (Dublin Core workshops), conferences (IEEE conferences on metadata), seminars (OCLC seminars on metadata, offered regularly) and meetings. Several projects have been initiated, such as, DESIRE project (gives a typology of metadata formats that includes three categories19), OCLC's CORE project, Alexandria project, MetaWeb, Nordic Metadata, Learning Object Metadata/ SCORM, CIMI, EAD, CSDGM, and so many20. Various standard setting bodies, working groups, task forces have been emerged in different organizations like ISO, NISO, DCMI, ANSI, NCITS, FGDC, LC, OCLC, UKLON, IFLA, NCSU, etc. In this regard, WWW-Consortium has taken a strong initiative in metadata and its standards are very simple (meta-meta level) with an attention to make it highly compatible to a variety of designs. For the purpose, W3 Consortium has developed Resource Description Framework (RDF) and PICS specifications to be used to encode and transmit the metadata produced from DC & Warwick Framework[21].

Therefore, various metadata formats became popular in their use. Such a format that became in practice since 1960s in libraries is the MARC. In view of its' comprehensiveness, interoperability, and maturity - MARC is highly specific and holds semantically enriched metadata. In 1990's, remarkable growths of digital repositories on the web have been noticed. Practically it is very difficult to develop a single digital repository (though not impossible) to navigate with unified interfaces and through common search algorithms – it means organizational repositories not yet be seamlessly accessed from a single site and terabytes of web-information cannot be explored using MARC alone. To cope with this prevalent situation different professional communities have made their debut to introduce new standards, guidelines, and architectures for managing those digital resources. For instance – Content Standard for Digital Geospatial Metadata (CSDGM) was initiated by FGDC in 1992; NCSU Libraries introduced Encoded Archival Description (EAD) in 1993; Dublin Core Metadata Initiative began in 1995 with an invitational workshop at Dublin, Ohio. A recent survey entails the current projects and initiatives on metadata research for organizing web resources[22].

Similar occurrences found in mid 1990s – Computer Interchange of Museum Information (CIMI) consortium has initiated a metadata test-bed project as an extension of DC elements; CDWA guideline provide detailed guidelines for scholarly description of art objects and their visual surrogates; Text Encoding Initiative (TEI) was emerged by humanities & linguistic research communities; Government Information Locator Service (GILS) has developed a complex metadata format with the intention to identify the US Government information resources; ROADS project has undertaken to design and implement the user-oriented resource discovery system by UKLON; SCORM (Sharable Content Object Reference Model) uses the IEEE Learning Object Metadata element set for descriptive metadata. SCORM draws on a variety of standards to create reference model specifically for learning objects; and so many. It is necessary to mention that, in the way of natural evolution of metadata schemas three patterns have evolved, as identified below – a) Metadata schemas that evolved in different professional traditions (e.g. MARC, CSDGM, CDWA, etc.), b) Metadata schemas that evolved in flexibility & scalability in metadata structures (e.g. Dublin Core), and c) Metadata schemas that evolved in adoption of a common formal language to support different applications (e.g. SGML, XML in application with EAD & TEI).

6. Metadata Standards

Above discussion entails that – the idea "one size fits all" does not work with metadata; it means a single standard cannot suit for all. Hence, numerous metadata standards have been developed and multiple metadata types can be traced in a hierarchy of complexity. Jan Smits described thoroughly about the need for various levels of metadata in one his articles entitled The Creation and Integration of Metadata for Spatial Map Collections. It can be summarized as – "those who are trying to describe complex GIS data sets will probably need to work with the FGDC/ISO metadata… MARC records can be used with less complex data sets, and Dublin Core, as well as MARC, is suitable for raster images and simple vector data sets that do not require a lot of description". Moreover the demand for uniformity and linkage persists within metadata standards. Suppose the map librarians generally like to create a link between FGDC and MARC or FGDC and DC, minimizing the data entry efforts for OPAC. The inherent cause to keep the records in different formats is basically to enable the interchange of information. Frequently, the librarians are needed for switching the metadata available on their hands into their required standard/s like MARC, DC or FGDC or etc. Virtually a number of mappings or crosswalks have been noticed – such crosswalks among various metadata standards are available from UKOLN23.

So, more than a dozen of standards available for each conceivable digital objects, like ETD, E-Learning, E-Governance, Geo-Spatial Data, Museum Items, Architectural Drawings, etc. These metadata standards include Dublin Core, Meta tags, RDF, TEI, CIMI, GLIS, METS, MODS, MARC, VRA Core, SCROM, LOM, GEM, EAD, PB Core, IMRC, CDWA, CSDGM / FGDC, MIDAS, VERS, DDI, PREMIS, CIDOC, ETDMS, AGLS, ONIX, and so many. These metadata standards have been developed gradually in order to facilitate the organization of web resources and used in different digital library projects as well as search services to describe their resources. Among various metadata standards, Dublin Core and Meta tags are the most common and widely implemented schemes for describing the content of web resources. Although DC is widely accepted and used in general, while MARC is popular in the research sector24. Dempsey and Heery in 2000 divided these metadata standards into three bands – First band includes full-text indexes (eg. search engines as Google), Second band includes several formats emerged to support search and directory services (eg. Spires, Whois++, and even DC too), Band three has more complex metadata structures like TEI, MARC, GILS, EAD, etc. Each and every standard has its certain specialty. To make an understanding a few of them have taken in discussion.

6.1 FGDC Metadata

It denotes the information content of metadata for digital geospatial objects and becomes a most established metadata standard in US. Basically it is an outcome of Federal Geographic Data Committee (FGDC), which is a nineteen member inter-agency committee, intended to promote the coordinated sharing, development, and dissemination of geospatial data. This nationwide data publishing effort gained its popularity through the National Spatial Data Infrastructure (NSDI). The federal government and many other agencies use this standard. In fact, this standard is mandatory for federal agencies to generate FGDC based metadata for their digital objects. The CSDGM (Content Standard for Digital Geospatial Metadata) is an extension of FGDC and has emerged with the objectives to provide a common set of terminology and definitions for the documentation of digital geospatial data. Its scope includes – "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure", that was signed on 11th April 1994, by the President William Clinton. I am clearly citing the site at http://fgdc.er.usgs.gov/ for further information on any of this metadata standard viz. CSDGM, FGDC/ISO, FGDC framework, etc It is worth to mention here that several initiatives are underway to harmonize the FGDC standard with European standards, therefore, a new international standard with the help of ISO may supercede the current FGDC standard immediately.

Anyway, FGDC metadata has received its widespread acceptance; although it is not sophisticated enough to suit everyone's needs and seldom criticized, as it is too cumbersome and more difficult to apply. Even sometimes it is overkill the simple digital files. To overcome this problem, a good number of agencies/ states in US is trying to adopt a simple version of FGDC metadata (called as 'metadata lite'). This new version concentrates on the essential elements that can provide better flexibility and facility for data translation. It is worth to mention that the FGDC metadata is featured by its incompatibility with MARC. In reverse, FGDC records are more extensively indexed and more information rich than MARC, and can be searched more effectively on the Web. No doubt, this standard is inevitable for the agencies that are intended to develop an efficient graphical interface, like Alexandria Project or Harvard's "Liboratory".

6.2 Dublin Core Metadata

The original Dublin Core emerged as a small set of descriptors that quickly drew global interest from a wide variety of information providers in the arts, science, education, business, and government sectors. It was initiated by the OCLC in 1995. In 2000, DC got the formal recognition by the Centre for European Normalization (CEN), the European standardization body. In 2001, it was ratified under the auspices of the National Information Standard organization (NISO) & Dublin Core Metadata Initiative (DCMI) as ANSI standard (Z39.85-2001)25. Actually, DCMI is an organization dedicated to promote the widespread adoption of interoperable metadata standards and developing specialized metadata vocabularies for describing resources. DCMI began in 1995 with an invitational workshop in Dublin (Ohio), to enable more intelligent information discovery systems[26].

DC is applied to organize electronic information even in the government sector. The Minnesota state government uses DC to provide effective tools for citizens to discover the environmental and natural resource information they need and to integrate access to diverse information resource types across multiple domains27. Web pages are one of the most common types of resources to utilize the Dublin Core's descriptions; usually within HTML's metadata tags, which are first becoming de facto standard (Weibel, 1996) as they are easy and quick to include at the beginning of WWW HTML files. An example encoding of DC using Meta tags is shown in Appendix – I. Increasingly there are many digital archives of physical objects that are starting to make use of the DC. Dublin Core metadata is often stored as "name-value" pairs within the Meta tags, which are placed within the HEAD elements of an HTML document. However, it can also be located in an external document or loaded into a database enabling it to be indexed and manipulated from within a propriety application. Guinchard reports the results of her e-mail survey on who uses DC and why and how it is used[28].

The Dublin Core was primarily developed to be simple and concise, and to describe Web-based documents. The current standard (finalized in 1996) defines fifteen metadata elements for resource description in a cross-disciplinary information environment. These elements are: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, and Rights29. All the elements discussed above are unqualified Dublin Core (having 15 elements) but the Qualified Dublin Core has more or less 65 elements. A detailed description of the elements can be found in Appendix – II. Basically, these Core elements can be categorized into three groups on the basis of the type/ scope of information stored in them30. These are as follows; i) Content: Elements related mainly to the content of the resource – Title, Subject, Description, Source, Language, Relation (to another resource), and Coverage; ii) Intellectual Property: Elements related mainly to the resource when viewed as intellectual property – Contributor, Creator, Publisher, and Rights; iii) Instantiation: Elements related mainly to the instantiation of the resource – Date, Type, Format, and identifier.

Although the metadata concept has been largely ignored in the probabilistic search world, but certainly in digital collection management it has great merit. Notionally, if all resource objects under the search engines carry unified fields and expose the same controlled vocabularies then the users can initiate an improved searching. Unexpectedly, most commercial search engines avoid attaching any significance to DC and other standard metadata elements. A few of the search engines allow for inclusion of limited metadata at the HEADER part, but this metadata could be useful when it follows the recommended syntax for that particular search engine.

6.3 MARC – A Signatory of Metadata

MARC is an acronym stands for Machine Readable Cataloguing and was primarily designed to serve the needs of libraries as a convenient way of storing and exchanging bibliographic information. The Library of Congress designed original MARC format in 1965-66 leading to a pilot project, known as MARC-1. Gradually several versions of MARC format have emerged, such as USMARC, UKMARC, CANMARC, INTERMARC and as many as twenty formats has come up. These formats differ from each other in various ways31. The problem of inconsistency among different MARC formats stimulated to develop the UNIMARC, which would accept records created in any MARC format. So records from one MARC format could be converted into UNIMARC and then UNIMARC to another.

MARC-21 is the first harmonized version of USMARC & CANMARC. It establishes a common taxonomy for defining all types of content, including print and electronic. Johnson remarks that despite the great potential of XML, MARC is still an important and broadly accepted encoding system32. Now it has great potential to describe the computer-readable bibliographic records in libraries. MARC is useful in library automation software as the basis of manipulating library records for display, indexing and easy retrieval. Majority of library automation systems allow for input and retrieval in MARC format, even if the records are stored internally in another format. The Z39.50 protocol can be used to execute searches of MARC fields from a Z39.50 client to a Z39.50 server fronting a database of MARC records, and retrieved records can be returned in MARC format[33].

MARC-21 Formats are standards for the representation and communication of bibliographic and related information in machine-readable form. A MARC-21 format is a set of codes and content designators defined for encoding machine-readable records. These formats are defined for five types of data – bibliographic, holdings, authority, classification, and community information. A MARC record involves three elements: the record-structure, the content- designation, and the data-content of the record[34]

a) Record structure: MARC record structure consists of three main components: the Leader, the Directory, and the Variable Fields. The structure of MARC records is an implementation of national and international standards, e.g., Information Interchange Format - ANSI Z39.2 and ISO 2709.

b) Content designation: The codes and conventions established to identify explicitly and characterize further the data elements within a record and to support the manipulation of those data, is defined in the MARC-21 formats. The goal of content designation is to identify and characterize the data elements that comprise a MARC-21 record with sufficient precision to support manipulation of the data for a variety of functions like, display and information retrieval, etc

c) Data Content: The content of most data elements is defined by standards outside the formats, e.g., AACR, LCSH, NLM Classification. The content of other data elements (e.g., coded data) is defined in the MARC-21 formats.

7. Metadata Crosswalk - DC vs. MARC 21

A crosswalk is a semantic and/or technical mapping (sometimes both) of one metadata framework to another metadata framework. Crosswalks (sometimes called "mappings") are used to "translate" between different metadata element sets. The elements (or fields) in one metadata set are correlated with the elements of another metadata set that have the same or similar meanings. This is also sometimes called "semantic mapping"[35][35]

Too many metadata standards have been developed in different communities. Each of these metadata standards has a unique focus; say, Dublin Core is designed for resource discovery; MARC is designed for the representation and communication of bibliographic and related information in machine-readable form. However, many of these standards also have commonalities. As many different standards are used in different information systems, it seems imperative to make crosswalks among the databases possible so that users can find information across different databases without worrying about differences in the systems. To respond to such demand in the LIS communities, an effort has been made to make a crosswalk between the Dublin Core element set and MARC-21 bibliographic data elements.

This crosswalk may be used in conversion of metadata from DC syntax into MARC and making it possible for different systems to communicate with each other. In fact, this crosswalk might interact with MARC records in various ways, as the following; a) "Enhancement of simple resource description record – A cataloguing agency may wish to extract the metadata provided in Dublin Core style (presumably in HTML or XML) and convert the data elements to MARC fields, resulting in a skeletal record. That record might then be enhanced as needed to add additional information generally provided in the particular catalogue. Some projects convert data and use as basic record for reporting to national bibliography", b) "Searching across syntaxes and databases – Libraries have large systems with valuable information in metadata records in MARC format. Over the past few years with the expansion of electronic resource over the Internet, other syntaxes have also been considered for providing metadata. It will be important for systems to be able to search metadata in different syntaxes and databases and have commonality in the definition and use of elements" 36. Appendix-III is showing the crosswalk from DC into MARC-21 in a simplest manner for setting an example of the discussion. For conversion MARC-21 into Dublin Core, many fields may be mapped into a single DC element and best described in the original source[37].

8. Metadata Harvesting [Creation & Capture]:

Metadata harvesting may be exhaustive or selective and a harvester is operated by a service provider as a means of collecting metadata from repositories or archives. It is worthy to mention, a harvester is a client application that issues OAI-PMH requests. The Open Archives Initiative - Protocol for Metadata Harvesting (referred as OAI-PMH) is basically provides an application-independent interoperability framework based on metadata harvesting. Perhaps two classes of clients are available to support the OAI-PMH framework – viz. 'Data Providers' (for exposing metadata) and 'Service Providers' (for building value-added services). Recently OAI-PMH is becoming more popular with the popularity of Open Access movement. So, a good number of OAI metadata harvesting tools have been noticed, of which, PKP-Harvester from John Willinski at University of British Columbia proved an excellent open source metadata harvesting and presentation tool38. This multi-platform web-based tool can effectively extracts data and have an intuitive user interface to organize data. Another such interesting tool is Virginia Tech Perl Harvester that can promise to add as a module in a metadata retrieval and browsing program. A few others are OAICat, UIUC Java/ VB Harvester, DLESE, myOAI, etc. and sometimes less tested. Importantly, this OAI-PMH data provider software can handle various metadata formats. Noteworthy is the fact that any repository's metadata can be exposed using either by Z39.50 protocol or OAI-PMH. Now a day, almost all digital repositories and open archives are introducing OAI-PMH to make their metadata available to search engines and harvesters. Even many digital repositories have some inbuilt mechanism to expose metadata using OAI metadata harvesting protocol.

Metadata vs. D-Space: DSpace is well established as an institutional repository tool (digital library software) in which scholarly publications in different forms (video, images, etc) are archived and preserved. It uses OAI-PMH through OAICat (an open-source product of OCLC) tool for harvesting metadata, which can be easily extendable to other metadata schemas by developing java programs. More over Dspace by default uses Qualified Dublin core set (has about 65 elements) for furnishing metadata, and exposes metadata using Unqualified Dublin core (has 15 elements) format for the purpose of OAI-PMH. The recent versions (1.2.2 beta onwards) allow clienteles to define their own metadata formats by using XML input-forms, i.e, these versions allow users to extend to Non-DC formats. Expectedly future versions of DSpace may permit a more integrated use of specialized metadata. Keeping such an intention MIT's SIMILE project is investigating Semantic Web technologies. Perhaps the support for multiple metadata formats (like VRA core, IMS/ IEEE-LOM, etc.) may greatly enhance the use of DSpace for archiving the digital objects. Prasad in DSpace User Meet at Cambridge has made a detailed discussion.[39] Furthermore, DSpace deals with three types of metadata about archived content[1] – namely Administrative (for preservation, authorization policy data, etc.), Descriptive (for description), and Structural (for presentation i.e, implementation of METS).

Metadata vs. Search Engines: Despite the different methods of discovery of information on the web, the 'search engine' and 'meta-search engine' has a great importance – to search the web by exploring traditional and advanced information retrieval techniques. Search-engine allows the users to search and access the resources from particular databases, whereas 'meta-search-engine allows users to access many search engines at a given time and obviously retrieve ranked result following global merging techniques (Mohamed, 2004). Development of multiple search engines (as Yahoo, Google, Infoseek, etc.) has expanded the access to digital information from a distributed environment. Users of search engines can easily retrieve the information in any digital collection that has been catalogued using standard metadata. An experiment by Turner and Brackbill (1998) found that the addition of a keyword meta tag enormously improved the retrieval ability of web pages on Infoseek, Alta Vista, etc. In reverse the meta tags does not improve the retrieval ability of web pages on HotbBot. Alternatively, Yahoo, Exite, etc. do not look at the meta tags at all in their search techniques, as Yahoo follows on human indexing and Exite uses full-text indexing. It is worthy to mention that common metadata standards may offer many promises for describing the content of web resources, of which Meta-tags and Dublin core has great implementation value. Metadata should make search engines more efficient though in practice it is really difficult to organize efficient metadata creation procedures for distributed contributors. In a paper, Alan and Val discussed about the impact of Google and a summary of recent trend by commercial publishers as well as other information publishing societies to make ensure their online content more visible to search engines i.e. though indexed by Google. The fact is that the AIP, IoP, IEEE, OCLC-WorldCat, etc have taken major initiatives (though by different means) to ensure their digital content indexed by Google41. Commercial publishers are taking mostly similar steps through cross-publisher citation linking system, began a pilot collaboration with Google in 2004 to allow indexing of full-text content from more than 29 academic publishers. Such activities indicate the willingness to work with search engines rather to criticise them. Interestingly, a few pages may appear near the top of the search results of most search engines, without any deliberate effort to achieve this. The reasons for the phenomenon are well understood – In a study Mohamed has examined the impact of using metadata in discovering the web resources. The study claimed that – using metadata elements influence the page rank order. Even the rank order of the pages that contain meta tags is higher than those that include Dublin Core and those that do not contain any metadata, though the difference is not significant[42]

9. Conclusion

Internet' has wrought a dramatic change in generation, communication & dissemination of information. Though it is treated as self-publishing media and not a library of evaluated publications, but it must be catalogued for its appropriateness for use. Phenomenal increase of Internet resources invites an efficient method for web-cataloguing and content organization of the web publishing's. If all the web resources carry the same cataloguing fields, and also use the same controlled vocabularies, then we should be able to improve the resource discovery on the Internet. Therefore, different metadata schemas have dramatic effect on how the web is indexed and will improve the discovery of resources on the Internet. These metadata sets differ in potentialities to meet users needs and it is very difficulty to establish an exact relationship between two metadata standards. The preceding study between DC and MARC-21 indicates the potential difficulties in identifying appropriate web resources. Again, this mapping will allow to flow the elements seamlessly between these two metadata standards and to enhance user services through building a more sophisticated gateway to any collection. It is necessary to mention that a metadata scheme must be sufficiently flexible to capture useful information about a wide variety of resources for a range of purposes. A standard schema should concern what set of information is to be captured by the metadata, production of metadata, and how metadata is accessed and used. Metadata actually tags the information so that the data can be recognized whenever it is asked for. Using meta-tags we can increase the retrieval performance and the rating/ relevance of the page can be easily calculated. Finally, metadata enormously increase the precision of a search and reduces recall. No doubt, these metadata schemas are critical mechanism for representing the information of digital collections.Here, the most basic question lies that – how

Internet documents will come with any metadata standards? Who will take the responsibility to create the metadata of web documents? Though the documents are hosted from several hosts in several ways and forms, the responsibility of describing the documents perhaps lies with the creator of the documents. In many cases, authors are not aware about it. We suggest that the consortium, the publishers or the librarians should take the responsibility of creating high-quality metadata for cataloguing web resources. In fact, joint effort of both the cataloguer and metadata coordinator on the effectiveness of metadata representation would be a great frontier in future information science research. Integrating different metadata sets together and building metadata production tools will be another challenge to make these standards potentially useful in finding information on the Internet.

APPENDIX-I

(DC Metadata Description Embedded in HTML file)

```
<html>
<head><title>Progress in Web-Cataloguing</title></head>
<link rel = “schema.DC”
  href = “http://purl.org/DC/elements/1.0/”>
<meta name = “DC.Title”  content = “Standard Metadata Schemas”>
<meta name = “DC.Creator”  content = “Pal, Jiban Krishna”>
<meta name = “AC.Email”   content = “jiban@isical.ac.in”>
<meta name = “DC.Subject”  content = “Resource discovery, Metadata”>
<meta name = “DC.Type”  content = “review article”>
<meta name = “DC.Date”  content = “2006”>
<meta name = “DC.Format”  content = “text/html”>
<meta name = “DC.Language”  content = “en”>
<meta name = “DC.Identifier”  content = “www.isical.ac.in/~jiban/webcat.pdf”>
<body>
Body of the document…
</body>
</html>
```

APPENDIX - II

(Dublin Core Metadata Elements with examples)

TITLE [A name given to the resource]
Comment : Typically, a Title will be a name by which the resource is formally known.
Example: <meta name = “DC.Title” Content = “Progress in WebCataloguing”>

SUBJECT & KEYWORD [A topic of the content of the resource]

Comment: Typically, a subject will be expressed as keywords, key phrases or classification codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.

Example: <meta name = "DC.Subject" Content = "Web-Cataloguing">
<meta name = "DC Subject"
Scheme = "DDC"
Content = "025.3">

DESCRIPTION [An account of the content of the resource]
Comment: Description may include but is not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content.

Example: <meta name = "DC.Description"
Content = "Metadata enhance the resource discovery in digital collection. . .">

SOURCE [A reference to a resource from which the present resource is derived]
Comment: The present resource may be derived from the Source in whole or in part. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.

Example: <meta name = "DC.Source"
Content = "http://www.loc.gov/marc/">

LANGUAGE [A language of the intellectual content of the resource]
Comment: Recommended best practice for the values of the Language element is defined by RFC 1766 [RFC1766] that includes a two-letter Language Code (taken from the ISO 639 standard [ISO639]), followed optionally, by a two-letter Country Code (taken from the ISO 3166 standard [ISO3166]). For example, 'en' for English, 'fr' for French, or 'en-uk' for English used in the United Kingdom.

Example: <meta name = "DC.Language" Content = "en">

RELATION [A reference to a related resource]
Comment: Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.

Example: <meta name = "DC.Relation"
Content = "http://www.dublincore.org/">
<meta name = "DC.Relation.Version"
Content = "http://www.dublincore.org/standard/2003/">

COVERAGE [The extent or scope of the content of the resource]
Comment: Coverage will typically include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity).

Example: <meta name = "DC.Coverage" Content = "Metadata harvesting, MARC, DC, etc.">
<meta name = "DC.Coverage"
Content = "Best practice metadata standards - libraries">

CREATOR [An entity primarily responsible for making the content of the resource]
Comment: Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.
Example: <meta name = "DC.Creator" Content = "Pal, Jiban Krishna">

PUBLISHER [An entity responsible for making the resource available]
Comment: Examples of a Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.
Example: <meta name = "DC.Publisher"
Content = "NACLIN - DELNET">

CONTRIBUTOR [An entity responsible for making contributions to the content of the resource]
Comment: Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.
Example: <meta name = "DC.Contributor.Co-author" Content = "Pal, Falguni">
<meta name = "DC.Contributor.Editor" Content = "Kaul, H. K">

RIGHTS [Information about rights held in and over the resource]
Comment: Typically, a Rights element will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights.
Example: <meta name = "DC.Rights"
Content = "© 2006, DELNET all right reserved">

DATE [A date associated with an event in the life cycle of the resource]
Comment:Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [W3CDTF] and follows the YYYY-MM-DD format.
Example: <meta name = "DC.Date" Content = "2006">
<meta name = "DC.Date.Created"Content = "2006-07-31">

TYPE [The nature or genre of the content of the resource]
Comment: Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary (for example, the working draft list of Dublin Core Types [DCT1]). To describe the physical or digital manifestation of the resource, use the FORMAT element.
Example: <meta name = "DC.Type" Content = "Article">
<meta name = "DC.Type" Content = "Digital Indexing - Metadata">

FORMAT [The physical or digital manifestation of the resource]
Comment: Typically, Format may include the media-type or dimensions of the resource. Format may be used to determine the software, hardware or other equipment needed to display or operate the resource. Examples of dimensions include size and duration. Recommended best practice is to select a value from a controlled vocabulary (for example, the list of Internet Media Types [MIME] defining computer media formats).
Example: <meta name = "DC.Format" Content = "Text/Html">

<meta name = “DC.Format” Content = “Image/MPEG”>

IDENTIFIER [An unambiguous reference to the resource within a given context]
Comment: Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Example formal identification systems include the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL)), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).

Example: <meta name = “DC.Identifier”
Scheme = “ISSN or ISBN or DOI”
Content = “0018-8441”>
<meta name = “DC.Identifier”
Content = “http://www.isical.ac.in/~jiban/web-cat.pdf”>

APPENDIX - III

(Describe the Crosswalk from DC into MARC-21 for example setting)

Dublin Core	MARC-21
Title: A name given to the resource.	**245 00 $a** (Title statement/ Title proper) **246 33 $a** (Varying form of title/ Title proper) ***e.g.*** *245 00 $aWeb database applications with PHP & MySQL/$cHugh E. Williams & David Lane.*
Subject: The topic of the content of the resource.	***653 ## $a*** *(Index term- uncontrolled)* ***e.g.*** *653 ## $aData Mining.*
Description: An account of the content of the resource.	**520 ## $a** (summary, etc., note) **505 0# $a** (Formatted contents note) ***e.g.*** *520 ## $aThe author presents a tutorial introduction to PHP programming examples with extensive examples on regular expressions*.
Source: A Reference to a resource from which the present resource is derived.	**786 0# $n** (Data source entry/ note)
Language: A language of the intellectual content of the resource.	**546 ## $a** (Language note) ***e.g.*** *546 ##$aEnglish, Bengali, Hindi.*
Relation: A reference to a related resource.	**787 1# $n** (Nonspecific relationship entry/note) **773 0# $n** (Host item entry/ Note) **775 0# $n** (Other edition entry/ Note) **780 0# $t** (Preceding entry) **785 0# $t** (Succeeding entry) **786 0# $n** (Data source entry/ Note)

Coverage: The extent or scope of the content of the resource.	**500 ## $a** (General note) **504 ## $a** (Bibliographic, etc) ***e.g.*** *500 ## $aIncludes index.* **e.g.** 504 ## $a*Includes bibliography: p.231-238*
Contributor: An entity responsible for making contributions to the content of the resource.	**100 1# $a** (Main entry- personal name) **110 1# $a** (Main entry- corporate name) **111 1# $a** (Main entry- meeting name) **700 1# $a** (Added entry- personal name) **710 2# $a** (Added entry- corporate name) **711 2# $a** (Added entry- conference name) **720 ## $a** (Added entry-uncontrolled name/name) ***e.g. 700 1#*** *$aLane, David*
Publisher: An entity responsible for making the resource available.	**260 ## $b** (Publication, description, etc. (Inprint)/ Name of publisher, distribution, etc.) ***e.g.*** *260 ## $aCambridge:$bO'Reilly,$c2002.*
Creator: An entity primarily responsible for making the content of the resource.	**100 1# $a** (Main entry- personal name) ** **110 1# $a** (Main entry- corporate name) ** **111 1# $a** (Main entry- meeting name) ** **700 1# $a** (Added entry- personal name) ** **710 2# $a** (Added entry- corporate name) ** **711 2# $a** (Added entry- conference name) ** **720 ## $a** (Added entry-uncontrolled name/name) ***e.g.*** *100 1# $aWilliams, Hugh E.*
Rights: Information about rights held in and over the resource.	**540 ## $a** (Terms governing use & reproduction note) **856 42 $u with $3** (Electronic Location & Access/ URL) with $3=Rights ***e.g.*** *856 42 $uhttp://www.sciencedirect.com $3©Elsevier 2002*
Date: A date associated with an event in the life cycle of the resource.	**260 ## $c** (Date of publication, distribution etc.) ***e.g.*** *260 ## $c2002*
Type: The nature or genre of the content of the resource.	**655 #7 $a** (Index term- genre/form)
Format: The physical or digital manifestation of the resource.	**856 ## $q** (Electronic location & access/ Electronic format type)

Identifier: An unambiguous reference to the resource within a given context.	**856 4# $u** (Electronic location and access/URL) ***e.g.*** *856 4# $uhttp://isical.ac.in/~jiban* **020 ## $a** (ISBN) ** **022 ## $a** (ISSN) **
** *Not approved by DCMI but can be added*	

References

1. Campbell, D. Grant (2001): Straining the standards – how cataloging websites for curriculum support poses fresh problems for Anglo-American Cataloging Rules, in Journal of Internet Cataloging, V.3 (1), p.79-92.
2. Persico, John Jr. (1991): Leadership and empowerment in a total quality organization, in Total Quality Management, V.2 (1), p.57-67.
3. Sandberg-Fox, Ann M (2001): The Microcomputer revolution, in Cataloguing & Classification Quarterly, V.31 (2), p.85-99.
4. Olson, Nancy (2001): Cataloguing remote electronic resources, in Cataloguing & Classification Quarterly, V.31 (1), p.101-137.
5. Gorman, Michael (2001): Bibliographic control or chaos – an agenda for national bibliographic services in the 21st century, in IFLA Journal, V.27 (5/6), p.307-313.
6. NISO: Dublin Core Metadata Element Set Approved, Visited on 2nd April 2004, Source: http://www.niso.org/news/releases/PRDubCr.html
7. Cathro, Warwick (1997): Metadata – an overview, visited on 2nd May 2004, Source: http://www.nla.gov.au/nla/staffpaper/cathro3.html
8. DCMI – Introduction: What is Metadata? Visited on 14th July 2006, Source: http://dublincore.org/documents/usageguide/
9. Milstead, J & Feldman, S (1999): Metadata – Cataloging by any other name… in Online, (© Information Today Inc.) January issue, visited on 30th April 2004, Source: http://www.infotoday.com/online/OL1999/milstead1.html
10. Chan, L. M. (1989): Inter-indexer consistency in subject cataloging, in Information Technology & Libraries, V.8 (4), p.349-358.
11. Weinheimer, J. (2000): How to keep the practice of Librarianship relevant in the age of the Internet. Vince (special issue on Metadata, part 1), V.116, p.14-27.
12. Thomas, C & Griffin, L. (1999): Who will creat the Metadata for the Internet?, in First Monday – a peer reviewed journal of the Internet, visited on 30th April 2004, Source: www.131.193.153.231/issues/issue3_12/Thomas/index.html
13. Greenberg, J. & et. al (2001): Author-generated Dublin Core Metadata for web Resources – a baseline study in an organization, in Proceedings of the International Conference on Dublin Core and Metadata Application, held at NII, Tokyo, Japan on October 24-26, 2001, p.38-46
14. Richmond, Alan: META Tagging for Search Engines, in Web Developer's Virtual Library, visited on 2nd May 2004, Source: http://www.wdvl.com/Search/Meta/Tag.html
15. OCLC: InterCAT Project, visited on 3rd May 2004, Source: http://www.oclc.org/research/projects/archive/intercat.htm

16. NISO (2004): Understanding Metadata. Bethesda, MD: NISO Press, p.1-12

17. Bor Ng, Kwong and Burnett Kathleen: Control or Management: A Comparison of the Two Approaches for Establishing Metadata Schemes in the Digital Environment visited on 12th May 2005 at http://www.scils.rutgers.edu/~sypark/asis.html

18. Lagoze, Carl (1996): "The Warwick Framework – a container architecture for diverse sets of metadata, in D-Lib Magazine, July/August issue, visited on 4th May 2004, Source: http://www.dlib.org/dlib/july96/lagoze/07lagoze.html

19. DESIRE – Development of a European Service for Information on Research and Education (1998): Specification for resource description methods (Part 1) - a review of metadata: a survey of current resource description formats (in para, some characteristics of investigated metadata formats), Work Package 3 of Telematics for Research project DESIRE (RE 1004), UKOLN Metadata Group, visited on 4th May 2004, Source: http://www.ukoln.ac.uk/metadata/desire/overview/rev_01.htm

20. Milstead, J & Feldman, S (1999): Metadata Projects and standards in Online, January issue (©Information Today Inc.), visited on 30th April 2004, Source: http://www.infotoday.com/online/OL1999/milstead1.html

21. World Wide Web Consortium (W3C): Source: http://www.w3.org/Metadata/ visited on 4th may 2004.

22. Hunter, Jane L (2003): A survey of metadata research for organizing the web, in Library Trends, V.52 (2), Fall issue (Organizing the Internet), p.318-344.

23. Michael Day (2002): Metadata - Mapping between metadata formats, May 2002, UKOLN: The UK Office for Library and Information Networking, University of Bath, UK. Visited on 12th May 2004, Source: http://ukoln.ac.uk/metadata/interoperability/

24. Polydoratou, P & Nicholas, David (2001): familiarity with and use of metadata formats and metadata registries amongst those working in diverse professional communities within the information sector, in Aslib Proceedings, V.53 (8), p.309-324.

25. Dekkers, Makx & Weibel, S. L. (2002): Dublin Core metadata initiative progress report and workplan for 2002, in D-Lib Magazine V.8 (2), p.1-9. Source: www.dlib.org/dlib/february02/weibel/02weibel.html

26. DCMI: Dublin Core Metadata Initiative, visited on 5th May 2004, Source: http://dublincore.org/

27. Quam, Eileen (2001): Informing and evaluating a metadata initiative – usability and metadata studies in Minnesota's Foundations Project, Government Information Quarterly, V.18, p.181-194

28. Guinchard, Carolyn (2002): Dublin Core use in libraries – a survey, in OCLC Systems & Services, V.18 (1), p.40-50.

29. DCMI: Dublin Core Metadata Element Set, Version 1.1: Reference Description, visited on 5th May 2004, Source: http://dublincore.org/documents/dces/

30. Burnett, K & Ng, K. B. & Park, S (1999): A comparison of the two traditions of metadata development, in Journal of the American Society for Information Science, V.50 (13), p.1209-1216.

32. ELAG – European Library Automation Group (1995): Organizing the electronic library, 19th Library Systems Seminar, Trondheim/ Norway, 14-16 June 1995, (Torbjørg Breivik ed., TULN Pub, p.291, ISBN: 82-471-0298-6)

33. Johnson, Bruce C. (2001): XML & MARC – which is right? in Cataloguing and classification Quarterly, V.32 (2), p.107-126.

34. Library of Congress: MARC Standards, Visited on 7th May 2004, Source: http://www.loc.gov/marc/

35. Library of Congress (MARBI): The MARC 21 Formats: Background and Principles, Revised November 1996, Visited on 7th May 2004, Source: http://www.loc.gov/marc/96principl.html

36. Metadata Standards Crosswalks: visited on 5th may 2004, Source: http://www.chin.gc.ca/English/Standards/metadata_crosswalks.html

37. Library of Congress - Network Development and MARC Standards Office (2001): Dublin Core/ MARC/ GILS Crosswalk, visited on 4th May 2004, Source: http://www.loc.gov/marc/dccross.html

38. Library of Congress - Network Development and MARC Standards Office (2001): MARC to Dublin Core Crosswalk, visited on 4th May 2004, Source: http://www.loc.gov/marc/marc2dc.html

39. Kellogg, David (2004): Open Source OAI Metadata Harvesting Tools, visited on 12th July 2006, Source: http://www.diglib.org/aquifer/oct2504/harvesting.pdf

40. Prasad, A R D (2005): Using Multiple Metadata formats in DSpace, presented at an User Meet on 6-8th July 2005, University of Cambridge, UK

41. DSpace Federation at MIT: DSpace System Documentation - Metadata. Visited on 12th July 2006, at http://www.dspace.org/technology/system-docs/functional.html

42. Dawson, Alan and Hamilton, Val (2006): Optimizing metadata to make high-value content more accessable to Google users. In Journal of Documentation, v.62 (3), p.307-327

43. Mohamed, Khaled A. F. (2006): The impact of metadata in web resource discovering. In Online Information Review V.30 (2), p.155-167