

## METADATA: A PATHWAY TO ELECTRONIC RESOURCES

---

**Kanta Kapoor**

GGS Indraprastha University  
University Information Resources Centre  
Kashmere Gate  
Delhi

*Gives an overview of metadata and its importance in the present era of information explosion. The paper also discusses metadata standards and lists various metadata systems that have evolved.*

### INTRODUCTION

As the personal computer and subsequently, the Internet have revolutionized the way in which people look for information, a need has arisen to evaluate, organize and integrate information. With the proliferation of digital technology and electronic information, there are almost as many interfaces as there are resources. It has always been the mission of the library professionals to bring order to information. Now, the domain is paper as well as electronic media.

With the web growing day by day and the escalating difficulty in finding specific information, search engines such as Yahoo, Lycos, Altavista, Google, Search etc. facilitate information search. Thus, not only is there an increase in the number and size of resources but also is an overabundance of system interfaces, leading to information overload or "interfaces chaos" for information professionals. Moreover, the internet has led to transformation of libraries from paperbased, location specific, on demand services to the ones that are available to their clientele at any time and from any location.

The wealth of information and the quick access available to it provides a frustrating dilemma for libraries and information seekers alike. The information is available, but how to find it, how to organize it to be found again? The user is

overwhelmed, when confronted by **disorganized** and often un-indexed information.

This availability of vast sources of information on the net initiated a need to have a tool to **organise** this information, i.e. metadata.

Several researchers are now engaged in finding ways and means of cataloguing and classifying materials available on the Internet and other online networks. Several metadata schemes have been created by library and information specialists like the MARC formats, the AACR II cataloguing formats, subject headings lists (such as LCSH) and classification schemes such as LC, DDÇ, UDC and so on. Each of these schemes has been constructed by experts in the relevant field from an understanding of the specific domain, information resources, needs and the requirements for describing documents. While these schemes have been used for bibliographic access and control for decades, **there** remains the question of how to catalogue and index all the materials available on internet using these schemes.

While metadata has become a buzzword in the information business, the concept is important for both authors and seekers of electronic information. Used effectively, it makes information accessible by labeling its contents **consistently**.

### WHAT IS METADATA?

Metadata is data about data. It **describes** the contents of an original document or work.

Metadata in other words can also be described as "data associated with objects, which relieves their potential users of having to have full advance knowledge of their existence and characteristics. In other words, standard bibliographic information, summaries, indexing terms and abstracts are all surrogates for the original material, hence metadata.

The term is generally applied to electronic sources and refers to "data" in the broadest sense of datasets, textual information, graphics, and anything else that is likely to appear electronically. While the concepts include indexing and cataloguing information, it can go far beyond conventional document representation, such as MARC records. Information about authenticity, availability and accessibility, digital signatures, copyright, reproduction, etc is also metadata.

Examples of metadata include: the catalogue records used by libraries, museums; and an index database.

Metadata describes various aspects of data, including the following:-

### **Identification**

What is the name of the data set? Who developed the data set? What geographic area does it cover? What themes of information does it include? How current are the data? Do restrictions exist for accessing or using the data?

### **Data Quality**

How good is the data? Does the information available allow a user to decide if the data are suitable for his or her purpose? What is the positional and attributal accuracy? Is the data complete? Was the consistency of the data verified?

### **Spatial Reference and Organization Information**

How is the data organised? (data models, topology)

### **Entity and Attribute Information**

What geographic information is included? How is the information encoded? Were codes used? What do the codes mean?

### **Distribution**

From whom can I obtain the data? What formats are available? What media are available? Is the data available online? What is the price of the data.

### **IMPORTANCE OF METADATA**

As increasing numbers and types of objects were being made available digitally, it was recognized that raw data was of little value without information about how it was collected, the purpose for which it was indented, formats, platforms for viewing and manipulation, and restriction on reproduction and uses, aside from more conventional information such as authors, or producer, title, subject and abstract.

Metadata acts as a surrogate for a larger work. It characterizes the original work sufficiently covering its purpose, source and condition of use. Besides this, metadata serves many important purposes including the following:-

- \* Data browsing
- \* Data transfer
- \* Data documentation

Metadata can be organized into several levels ranging from a simple testing of basic information about available data to detailed documentation about an individual data set. At a fundamental level, metadata may support the creation of an inventory in the data. Metadata is also important in the creation of spatial data clearing house, where potential users can search to find the data they need for their intended application.

### **Creating Metadata**

Metadata can be created at the time of creation of an object, either by or under the auspices of its

creator. It can also be added later as part of the traditional cataloguing process. The former mode of creation is expected to predominate, largely because the cataloguing and indexing method simply cannot cope with the massive and the rapid growth of electronic objects in existence.

Many agencies now have started providing web-based forms for entry of metadata, completion of one of these forms by an object creator or owner would result in the creation of metadata records for that object.

Many libraries are also making efforts to "catalogue the internet". e.g OCLC'S Internet Cataloguing Project (InterCat). Generally, these efforts are designed to cover the best sites or those, which are used for a particular purpose.

Information retrieval on the Web using Metadata searching is essentially a process of matching the query terms to the words in documents. If the terms don't match, then the document is not retrieved, no matter how much it is about the subject of the query. Therefore, one of the greatest barriers to finding information is the difficulty of coming up with right terminology.

Metadata helps the retrieval in the following ways:-

1. It makes sure that all the materials about the same subject are found together in one online database.
2. It singles out important concepts from those that are merely incidental to the work.
3. It ensures proper use of indexing vocabularies and field structures both in searching and in cataloguing. It increases the precision and minimizes the change of false drops.
4. Metadata also helps to solve language problems that cause poor precision and minimizes the change of false drops e.g. polysems, synonyms, etc. A good metadata helps the search engine to understand the meaning, not just be able to match the spelling of a word.

5. Metadata can not only improve precision; it can also increase recall of pertinent documents by using the same standardized term for each occurrence of a subject. Thus, a document will be retrieved from properly applied metadata even if it never uses the controlled term in its text.

## METADATA STANDARDS

*The International Organization for Standardization (ISO)* has recently set up a Metadata Working Group to take responsibility for standards for specification and management of metadata. The scope of the Working Group includes metadata elements, classification, and coding schemes and metadata management and exchange.

*National Committee on Information technology Standards* formed by *American National Standards Institute (ANSI)* covers naming, identification, definition, classification and registration of metadata.

### *World Wide Web Consortium (W3C)*

W3C'S metadata activity is concerned with ways to model and encode metadata with ways to model and encode metadata. A particular priority of W3C is to use the web to document the meaning of metadata. W3C has promoted the development of the *Resource Description Framework (RDF)* and its relative *PICS (Platform for Internet Content Selection)*. PICS is now complete and work on RDF continues.

### *PICS: Platform for Internet Content Selection*

PICS consist of a suite of specifications, which enable people to distribute metadata about the content of digital material in the form of labels. These contain information about the content in simple, computer readable form. Information can be given a label, which computers can then process in the background according to setting previously specified by the user, filtering out undesirable material or directing users to sites that may be special interest to them.

**RDF: The Resource Description Framework**

PICS work led to the development of Resource Description Framework (RDF), which provides a more general treatment of metadata. RDF provides a framework in which independent communities can develop vocabularies that suit their specific needs and share vocabularies with other communities. The description of these vocabulary sets is called RDF scheme. Dublin Core is also an example of RDF scheme. W3C has created the RDF specifications as a framework for application specific vocabularies.

**METADATA PROJECTS**

Many metadata systems have already been evolved. One of the major metadata systems is:-

1. The Dublin Core, which was developed, by OCLC and the National Centre for Supercomputing Applications, in 1995. Content creators originally developed the Core for use but interest has become widespread among specialized description groups such as museums and libraries.

The Core contains 15 metadata elements in three groups.

(i) Content:-

- \* Title
- \* Subject
- \* Description
- \* Sources
- \* Language
- \* Relation
- \* Coverage

(ii) Intellectual Property

- \* Creator
- \* Publisher
- \* Contributor
- \* Rights

(iii) Instantiation

- \* Data
- \* Type

- \* Format
- \* Identifier

2. Meta Web Project:- Sponsored by a number of major Australian organizations, it has the goal of developing indexing services, user tools and metadata element sets, to promote use of metadata.
3. The Computer Interchange of Museum Information (CIMI) consortium has initiated a metadata project that will be an extension of the Dublin Core which includes a more detailed level of description, in order to satisfy other requirements, such as research or rights and use management.
4. The Instructional Management Systems (IMS) project is collaboration between EDUCOM and the National Institute for Standards and Technology Information. It consists of a metadata dictionary that incorporates and extends the Dublin Core elements and sets of values defined in the dictionary.
5. Warwick Framework: It was the outcome of the second Dublin Core Workshop in 1996. The framework accommodates more than one type of metadata, for instance, it may incorporate MARC record and a Dublin Core record as well as description of the document. The Warwick Framework allows different types of description to be appended to a document.

However, it does not solve the problem of mapping from one set of metadata to another, keeping consistency across descriptions.

Apart from these, there are many more metadata projects, all of them having different standards.

**CONCLUSION**

Metadata is an essential tool, which should be developed as a standard in this information era. However, the problem in this area is the sheer number of different metadata projects. As no one is in charge of the Internet, any group may start up its own metadata definition effort, and creators

are free to use whatever ways happen to come to their mind. Any group that is developing a metadata set is free to limit its work to its narrow interests; it need not take a broader view unless it voluntarily chooses to do so.

Even if common metadata elements are used, there is no guarantee that the vocabularies, and the content of the elements will be compatible. For instance, the defined list of resource types in the Dublin Core is strongly oriented to the needs of libraries and similar agencies and does not fully meet the needs of other communities, such as software or geospatial data. There is a serious possibility that the situation may grow more chaotic and metadata users will have to learn a different set of conventions for each kind of data. Some of the efforts are being made towards the use of Library of Congress Subject Headings (LCSH) as a resource, even if not as an authority. Use of LCSH is probably determined more by the fact that it is broad in scope and available for use, rather than by its inherent quality and suitability for electronic searching. In fact, its use of broad terms may make it less useful for searching. Yet there are many questions remaining to be answered:-

- \* Who will make a final decision about which fields to use and which not to support among competing proposals?

- \* Who will apply the metadata?
- \* Will we have controlled vocabularies and how can we create that to catch every subject and idea?

## REFERENCES

1. CHOWDHURY(G G). Introduction to modern information retrieval. 1999; Library Association Publishing; UK.
2. ELLIS (D). In search of the unknown user: indexing, hypertext and the World Wide Web, *Journal of Documentation*, 54, 1; 1998; 28-47.
3. DATE (C J). An introduction to database systems; the system programming series, 1999; Addison-Wesley; UK
4. ELMASRI (RAMEY), Fundamentals of database Systems. 2000. Pearson Education International; U K
5. [http://purl.org/metadata/dublin\\_core](http://purl.org/metadata/dublin_core)
6. <http://www.lub.lu.se/egibin/nmde.pl>