Volume: 02, Issue: 02, March –April 2015 *eISSN NO. 2394-2479*
Impact Factor (IIFS) - 0.331

METADATA HARVESTING- AN INTRODUCTION

Gopnarayan S.R
Research Student
DLIS
Dr. Babasaheb Ambedkar Marathwada University,
Aurangabad. Maharashtra. India.

Dr. Adhe G.D
Associate Professor & Head,
DLIS
Dr. Babasaheb Ambedkar College of Arts & Commence,
Aurangabad. Maharashtra. India.

Bawaskar S.S
Research Student
DLIS
Dr. Babasaheb Ambedkar Marathwada University,
Aurangabad. Maharashtra. India.

Abstract

T*oday more and more information is created in digital form. One of the key challenges faced by today's information managers is the need to inter-relate different sources and types of information with different formats, data structures and description standards. Using metadata to record data about information sources allows an initial assessment of compatibility and provides an avenue for merging information or for exchanging information between systems. Another service has gained momentum nowadays is the Metadata Harvester. Metadata Harvester provides indexes or harvests metadata, from different open access archives and open access journals. This paper describes Metadata, Open Access Initiative Protocol for Metadata Harvesting (OAI-PMH) and major Metadata Harvesting Services in India.*

Keywords: Metadata, Metadata Harvesting, Open Access Initiative Protocol for Metadata Harvesting (OAI-PMH, Metadata Harvesting Services in India Search Digital Libraries (SDL), SJPI Cross Journal Search Service SEED (Search Engine for Engineering Digital-repositories) Open J-Gate and Knowledge Harvester@INSA

Introduction

There has been a dramatic change in the world of learning, scholarship, business and governance brought about by Information and Communication Technologies. The ways and means of creating, accessing, distributing and managing information not only in text but in other forms such as audio-video and multimedia materials have undergone major changes. Today more and more information is created and managed in digital form. One of the key challenges facing information managers today is the need to inter-relate different sources and types of information, whether it is in an internet search across a range of resources with different formats, data structures and description standards or an e-commerce system that needs to exchange data between proprietary applications in order to complete a transaction. Using metadata to record data about information sources allows an initial assessment of compatibility and provides an avenue for merging information or for exchanging information between systems. Interoperability is the ability of two or more systems to exchange information and to use the information that has been exchanged.

National Information Standards Organization (NISO) defines interoperability as "the ability of multiple systems, using different hardware and software platforms, data structures and interfaces, to exchange and share data" [1] Providing access to information free of charge in electronic formats is the concept that is gaining momentum. Open access holds the promise to remove price and permission barriers to the scientific communication by using Internet. Open access literatures are available in open access journals, institutional repositories, subject repositories, digital archives and so on. Another service has gained momentum nowadays is the Metadata Harvester. Metadata Harvester 2 provides indexes or harvests metadata, from different open access archives and open access journals.

Metadata:-

What Is Metadata? Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information. The term metadata is used differently in different communities. Some use it to refer to machine understandable information, while others use it only for records that describe electronic resources. In the library environment, metadata is commonly used for any formal scheme of resource description, applying to any type of object, digital or non-digital. Traditional library cataloging is a form of metadata; MARC 21 and the rule sets used with it, such as AACR2, are metadata standards. Other metadata schemes have been developed to describe various

types of textual and non-textual objects including published books, electronic documents, archival finding aids, art objects, educational and training materials, and scientific datasets.

There are three main types of metadata:

• Descriptive metadata describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author, and keywords.

• Structural metadata indicates how compound objects are put together, for example, how pages are ordered to form chapters.

• Administrative metadata provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it. There are several subsets ofadministrative data; two that sometimes are listed as separate metadata types are: −

Rights management metadata, which deals with intellectual property rights, and − Preservation metadata, which contains information needed to archive and preserve a resource. Metadata can describe resources at any level of aggregation. It can describe a collection, a single resource, or a component part of a larger resource (for example, a photograph in an article). Just ascatalogers make decisions about whether a catalog record should be created for a whole set of volumes or for each particular volume in the set, so the metadata creator makes similar decisions. Metadata can also be used for description at any

level of the information model laid out in the IFLA (International Federation of Library Associations and Institutions) Functional Requirements for Bibliographic Records: work, expression, manifestation, or item. For example, a metadata record could describe a report, a particular edition of the report, or a specific copy of that edition of the report. Metadata can be embedded in a digital object or it can be stored separately.

Metadata is often embedded in HTML documents andin the headers of image files. Storing metadata with the object it describes ensures the metadata will not be lost, obviates problems of linking between data and metadata, and helps ensure that the metadata and object will be updated together. However, it is impossible to embed metadata in some types of objects (for example, artifacts). Also, storing metadata separately can simplify the management of the metadata itself and facilitate search and retrieval. Therefore, metadata is commonly stored in a database system and linked to the objects described.

Data

Data are the basic individual items of numeric or other information, garnered through observation; but in themselves, without context, they are devoid of information. Information is that which is conveyed, and possibly amenable to analysis and interpretation, through data and the context in which the data are assembled. Knowledge is the general understanding and awareness garnered from accumulated information, tempered by experience, enabling new contexts to be envisaged. [7] (Quentin L. Burrell)

Volume: 02, Issue: 02, March –April 2015 *eISSN NO. 2394-2479*
Impact Factor (IIFS) - 0.331

Data are (or datum is) an abstraction. I mean, the concept of 'data' or 'datum' suggests that there is something there that is purely given and that can be known as such. The last one hundred years of (late) philosophic discussion and, of course, many hundred years before, have shown that there is nothing like 'the given' or 'naked facts' but that every (human) experience/knowledge is biased. This is the 'theory-laden' theorem that is shared today by such different philosophic schools as Popper's critical rationalism (and his followers and critics such as Kuhn or Feyerabend), analytic philosophy (Quine, for instance), hermeneutics (Gadamer), etc. Modern philosophy (Kant) is very acquainted with this question: experience ("Erfahrung") is a product of 'sensory data' within the framework of perception ("Anschauung") and the categories of reason ("Verstand") ("perception without concepts is blind, concepts without perception are void"). Pure sensory data are as unknowable as "things in themselves"."

Data. It depends on your framework. If you are a Kantian, it is the foundation for the a priori categories of the understanding. If you are a computer programmer it is preprocessed information (data collected according to some algorithm for some purpose) or post-processed information (e.g., tables of such information). In this latter case data cannot be defined apart from information, because it is dependent on it. If you are a biologist, it might be stimuli, but these scientific approaches are built on a faulty understanding of perception (e.g., perception is sensations (i.e., stimuli) glued together—which is false).

Data are dynamic objects of cultural experience having the aspect of being meaning-neutral and a dual nature of description and instruction. Data are observations and measurements you make on objects (artifacts, sites, seeds, bones) and on their contexts. Data are theory-laden. Regarding the theory of knowledge organization we may say that knowledge is not organized by elements called data combined or processed according to some algorithmic procedure. What data are is domain specific and theory-laden. At the most general level what is seen as data is depending of the epistemological view that one subscribes to. Data are raw evidence, unprocessed, eligible to be processed to produce knowledge. Information is the process of becoming informed; it is dependent on knowledge, which is processed data. Knowledge perceived, becomes information. Knowledge is what is known, more than data, but not yet information. Recorded knowledge may be accessed in formal ways. Unrecorded knowledge is accessible in only chaotic ways. [38] (Richard Smiraglia).

Metadata Harvesting:-

Major Metadata Harvesting Services in India A metadata harvesting service harvests or indexes metadata from OAI-compliant archives or repositories through harvesting software that supports a protocol known as OAI-PMH (Open Access Initiative Protocol for Metadata Harvesting). Some Indian institutions have been experimenting with metadata

harvesting services and installed metadata harvesters. Major metadata harvesting services in India are [4]

• Search Digital Libraries (SDL)

• SJPI (Scientific Journal Publishing in India)Cross Journal Search Service

• SEED (Search Engine for Engineering Digital-repositories)

• Open J-Gate

• Knowledge Harvester@INSA 4. Details about Metadata Harvesting Services in India Table No. 1 covers general information of Metadata Harvesting Services in India like URL, host and software used.

Table No. 01: Metadata Harvesting Services in India

Sr.No.	**Name**	**URL**	**Host**	**Software used**
1	Search Digital Libraries (SDL)	http://drtc.isibang.ac.in/sdl	DRTC, Bangalore	PKP System
2	SJPI Cross Journal Search Service	http://144.16.72.144/harvester/	NCSI, IISc	PKP System
3	SEED	http://eprint.iitd.ac.in/seed/	IIT, Delhi	PKP System
4	Open J-Gate	http://www.openj-gate.com/	Informatics India Ltd.	-----
5	Knowledge Harvester@INSA	http://61.16.154.195/harvester/	INSA	PKP System

Metadata Analysis:-

1. Definition of meta-analysis (from Glass, 1976): The statistical analysis of a large collection of analysis Results for the purpose of integrating the Findings.
2. The basic purpose of meta-analysis is to provide the same methodological rigor to a literature review that we require from experimental research.
3. We refer to the direct investigation of human or animal data as primary research." Providing a report Of primary research using statistical methodology and analysis is called \quantitative synthesis" or \meta-analysis." A report of primary research using traditional, literary methods is called a \narrative Review."
4. Meta-analyses are generally centered on the relationship between one explanatory and one response Variable. This relationship, \the e®ect of X on Y," de¯nes the analysis.
5. Meta-analysis provides an opportunity for shared subjectivity in reviews, rather than true objectivity. Authors of meta-analyses must sometimes make decisions based on their own judgment, such as when de¯ning the boundaries of the analysis or deciding exactly how to code moderator variables. However,
6. Meta-analysis requires that these decisions are made public so they are open to criticism from other Scholars.
7. Meta-analyses are most easily performed with the assistance of computer databases (Microsoft Access,Paradox) and statistical software (DSTAT, SAS).

Conclusions

Metadata is a key part of the information infrastructure necessary to help create order in the chaos of the Web, infusing description, classification, and organization to help create more useful stores of information. Sources of metadata, like the sources of the resources themselves, will be of different quality and organized around different purposes to reflect the different objectives and business models of information providers. The Open Archives Metadata Harvesting Protocol opens many new possibilities which are yet to be explored. OAI metadata harvesting offers a new bridge to bring new innovation in networked information services and applications, out of the research community more rapidly. Researchers who want to explore new ways of organizing, presenting or using the large data resources now have a standardized way of extracting content without much disruption or cost to existing operational systems. Metadata harvesting services are powerful mechanism for enabling development of new applications and services that have never been possible.

References:

- Chaim Zins (2007). Knowledge Mapping Research, 26 Hahaganah Street, Jerusalem 97852, Israel (479-485).
- NISO Press National Information Standards Organization 4733 Bethesda Avenue, Suite 300

- Mangala Hirwade (2006). Metadata Harvesting Services in India(1-7) hirwade2004@indiatimes.com
- National Information Standards Organization (NISO). "Understanding Metadata" http://www.niso.org/standards/resources/UnderstandingMetadata.pdf (Last accessed on 18.08.2006)
- What is Metadata? www.web2.gov.mb.ca/mli/glossary/m_z.html (Last accessed on 22.08.2006) 3. Rajshekhar, T.B. "Open Access initiatives in India" http://www.istl.org/05-spring/article3.html#8 (Last accessed on 24.08.06)
- Ghosh, S.B. and Das, Anup Kumar "Open access and institutional repositories – a developing country perspective: a case study of India" http://www.ifla.org/IV/ifla72/papers/157-Ghosh_Das-en.pdf (Last accessed on 22.08.2006)