KARNATAKA STATE OPEN UNIVERSITY

Mukthagangothri, Mysuru – 570 006

MASTER OF LIBRARY AND INFORMATION SCIENCE

(SECOND SEMESTER)

MLIDSC-2.2: DIGITAL LIBRARIES

BLOCK-2

M.Lib.I.Sc.

MLIDSC-2.2: DIGITAL LIBRARIES

BLOCK – 2: DIGITAL LIBRARY: STANDARDS

Programme Name: M.Lib.I.Sc. **Year/Semester: II[nd]** Semester **Block No :** 2

Course Name: MLIDSC-2.2: Digital Libraries **Credit:** 4 **Unit No :** 5-8

COURSE DESIGN AND EDITORIAL COMMITTEE

Editorial Committee

COURSE WRITER	UNIT	COURSE EDITOR
Dr Parthasarathi Mukhopadhyay Professor, Dept of Library and information Science, University of Kalyani, West Bengal	5-8	**Dr. Shilpa Rani N. R** Assistant Professor, DoS&R in LISc, KSOU, Mukthagangotri, Mysuru

Copyright

Developed by the Department of Studies and Research in Library and Information Science, under the guidance of Dean (Academic), KSOU, Mysuru.
Karnataka State Open University, November-2022.

M.Lib.I.Sc.

MLIDSC-2.2: Digital Libraries

Introduction

In a library set up standards bring order in chaos. The three most important conditions for effective information services in this digital era are organization, collaboration, and interoperability. Libraries across the world have a long history of using voluntarily agreed upon standards, for example, the adoption of standardized cataloguing codes (AACR2, RDA, etc.), standard bibliographic data elements (ISBDs), subject access systems (LCSH, SLSH), and standardized subject classification schemes (DDC, CC) by libraries all over the world are extremely important events in the history of library standards. Karen Coyle (2006a) reported that the very first instance of the library technology standards was the decision at the first annual ALA meeting in September of 1877 to standardize the catalogue card at 7.5 X 12.5 cm. The purpose of this dimensional standard was to enable large-scale production and distribution of cards. Libraries across the world followed this standard, and in 1898, the Library of Congress (LoC) began its printed cards service.

The design and development of digital libraries centres on a few major issues such as technical architecture, collection building processes and methods (including digitization), metadata (document description), resource identifiers (for global identification of resources), preservation, and perpetual access. All of these activities centre around one objective: the sophisticated retrieval of digital objects.

In this block following units are discussed in detail-

Unit-5: Digital Library: Standards

Unit-6: Creating Web documents

Unit-7: Digital library architecture

Unit-8: Institutional repositories

Dr. Shilpa Rani N.R.

UNIT- 5: DIGITAL LIBRARY STANDARDS

Structure

5.0 OBJECTIVES

After going through this Unit you will be able to:

- Evaluate importance of standards in automated and digital library systems;
- Explore nature, types of digital library standards;
- Trace the path of development and future directions of digital library standards;
- Identify a set of common minimum standards of the domain; and
- Relate the symbiosis between a digital library software and domain standards.

5.1 INTRODUCTION

In a library set up standards bring order in chaos. The three most important conditions for effective information services in this digital era are organization, collaboration, and interoperability. Libraries across the world have a long history of using voluntarily agreed upon standards, for example, the adoption of standardized cataloguing codes (AACR2, RDA, etc.), standard bibliographic data elements (ISBDs), subject access systems (LCSH, SLSH), and standardized subject classification schemes (DDC, CC) by libraries all over the world are extremely important events in the history of library standards. Karen Coyle (2006a) reported that the very first instance of the library technology standards was the decision at the first annual ALA meeting in September of 1877 to standardize the catalogue card at 7.5 X 12.5 cm. The purpose of this dimensional standard was to enable large-scale production and distribution of cards. Libraries across the world followed this standard, and in 1898, the Library of Congress (LoC) began its printed cards service. Coyle (2007) also opined that the card-size standard was the key to interoperability of this card catalogue service, and the LoC cards service may be considered as the precursor of bibliographic data services like the MARC-formatted record service or bibliographic data exchange standards like ISO-2709. And we all know the revolutionary role of standard bibliographic formats like ISO-297 and MARC family of standards in library automation. Apart from these functions, library standards are also necessary to achieve interoperability, not only for the exchange of bibliographic data but also to help library users navigate different library systems and services without learning new skills to perform core bibliographic functions – to find, to identify, to select, to obtain and to navigate. But a few questions arise at the same time, such

as what do we mean by the term standard? Why are standards important for service institutes like libraries and information centres? Why are standards addressed by unique identifiers? Why are some library standards prefixed with "ISO" or "Z"? Why are the same library standards identified by different names and identifiers? Who is developing library standards? How and to what extent can libraries adopt standards? How to know the standards compatibility of a library's software or information products? Is there a common minimum set of standards that can be followed in libraries irrespective of size or type? This unit is an attempt to provide answers to such questions that you need to know about standards. This unit is an attempt to locate major players in the domain of library standards; to identify layers of library technical standards through logical grouping; to understand the process of developing standards; and to suggest a set of minimum technical standards that are required to be followed by libraries of any type or size, keeping in view the need for consistency.

5.2 LIBRARY STANDARDS - GENESIS

The word *standard* has been used for many years in library literature. An extensive literature search shows that New York State University first used the term library standard in a statement issued in 1894 entitled "*Minimum Requirements for Proper Library Standard*". In 1916, the ALA council suggested that library legislation needed to ensure and safeguard standard library services (Doren, 1917). In 1917, the editorial section of an issue of *Library Journal* (Vol. 42, Issue 81) emphasized the importance of library standards. This section deals with brief history of library standards along with the process for development of these entities.

Standards are essential for manufacturing sectors and industries. In recent years, service institutes like libraries and information centres have also increasingly depended on standards for developing new services, improving existing services, and achieving uniformity, order, and interoperability. The process of developing and attaining standards for libraries has been taken seriously by library professionals since the Second Great War. The movement started in America and has gained strength and support from countries across the globe. Standards can be *de jure* or *de facto*, proprietary or open. In the library world, some standards are merely best-practice guidelines, and some standards are developed through a

formal process by professional associations or national institutions in the domain of library activities. Library professionals are generally interested in their own national standards (e.g. Bureau of Indian Standards (BIS), British Standard Institutes (BSI), etc.), in ISO standards (International Organization for Standardization), and in NISO standards. For the purpose of developing standards in library and information services, the National Institute of Standards Organization (NISO) is the institute that ANSI (American National Standards Institute) accredits. Hopkinson (2006) grouped standards in the domain of library services by their applications: describing and identifying information resources; information exchange; managing collections; and delivering services. However, in software-centric library systems standards, in areas like descriptive metadata, metadata harvesting, crosswalk and interoperability, RFID technology and circulation data exchange are emerging rapidly across the globe.

5.2.1 Definitional scopes

Standards are essential for different walks of human activities. As reported in Kent's encyclopedia (*Encyclopedia of Library and Information Science*, edited by Allen Kent, et al., vol. 30, p.176–190), standards are necessary to ensure reproducibility of research and reliability of research results in research and development (R & D). ISO/IEC Guide 2:2004 defines "*a standard as a document, established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines, or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context.*" The South African Library Association (1968) opined that library standards may be viewed as a crucial aid for library professionals and authorities in order to: i) act as the best practice guideline; ii) act as a model to follow; iii) show path for evaluation; iv) help future development; and v) support decision making process.

5.2.2 Terminological scopes

Library stakeholders often use the terms standards, guidelines and specifications synonymously. But these terms differ significantly in the definitional plane. A "guideline" is a set of procedures for a domain of activity by an expert (e.g. IFLA's Standards for Public Libraries, 1973), whereas a "standard" is an acknowledged measure of comparison for

quantitative or qualitative value, criterion, norm, or level of requirement, excellence or attainment. Standards and specifications indoctrinate or advocate minimum levels of performance, quality of products and services, operational procedures for production, evaluation, distribution and utilization of materials, products and services (Encyclopedia of Library and Information Science, V. 30, p. 178). Both standards and specification stipulate acceptable features and attributes of materials, products, and services (Doren, 1917). But the scope of applicability of a standard is usually much wider than a specification. A standard is formed by global-scale cooperation involving many stakeholders. A specification, on the other hand, is a sort of guideline. It does not have to cover subjects of wide use or even existing objects (Hirsh, 1975). A standard is a specification agreed upon by the stakeholders. Standards undergo changes with the technical and socio-technical changes (Withers, 1970). In short, it may be noted that all standards are specifications (of some type) but all specifications are not standards.

5.2.3 Types of standards

As per their intrinsic attributes, standards are broadly divided into two groups – (a) standards for uniformity and (b) standards for quality. Standards belonging to the first group ensure recommended features such as speed, size, and other attributes. The standards for uniformity facilitate interoperability of products, components, data and computer programs. Quality standards stipulate the minimum levels of product performance and are helpful in assessing quality of products. However, these two broad groups of standards can again be arranged in different ways by following different parameters (as prescribed by LINFO http://www.linfo.org/index.html) such as – i) Process of formation (*de jure* or *de facto*); ii) Domain of standard (unique or competitive); iii) Geographical scope (universal, national, local, etc); iv) Degree of compulsion (voluntary or mandatory); v) Nature of standard (proprietary i.e. commercial or open); vi) Duration of standard (time validity and review process); and vii) Areas of application (product, activity, industry, service etc).

5.3 STANDARDS DEVELOPMENT PROCESS

Although standards development organizations are mainly responsible for producing

standards in different areas of human activity, some voluntary standards are offered for use by professional associations, groups of people, regulatory authorities, or industry leaders. Some of these standards achieve global recognition and acceptance, become the *de facto* standard in a given area of application, and are sometimes adopted by national or international standards organizations. In the domain of automated and digital library systems, there is always a balance between de jure and de facto standards. Most of these standards are developed by professional groups and later adopted by national and international standards development agencies. Similarly, a few standards are first developed by these agencies and subsequently adopted by libraries.

5.3.1 De jure vs. De facto standards

A considerable number of library standards are voluntary standards developed by library associations (IFLA in particular) and national libraries (Library of Congress in particular). At the same time, it is necessary to report here some globally adopted library standards that are not standards by law but standards by practice (Mukhopadhyay, 2012, 2014). Let's consider the case of AACR. It has never been made into a formal standard through the standardization process but is accepted by library professionals in many countries as a standard cataloguing code. The case of ISBD (International Standard Bibliographical Descriptions) is slightly different from AACR. Hopkinson (2006) reported an interesting story in *Digitalia* in this regard. ISBDs are intended to be incorporated into national cataloguing codes. For some countries, incorporation of ISBDs into national cataloguing codes required that the ISBDs be enacted as international standards. In the late 1970s, there was an initiative by IFLA to make the ISBDs into formal standards. During the process of making ISBDs into ISO standards, the respective ISO committee wanted to incorporate a few changes that were not acceptable by IFLA. As a result, IFLA withdrew ISBDs from the international standardization process. Again, there are some instances where cooperative initiatives are adopted on an as-it-is basis as national or international standards. The classic example is DCMES (Dublin Core Metadata Elements Set). This descriptive metadata schema has been acting as the de facto global standard since 1995. It was then accepted as a NISO standard in 2001 *(*ANSI/NISO Z39.85-2001) (revised in 2007 as ANSI/NISO Z39.85-2007) before being accepted as an ISO standard in 2003 (ISO 15836:2003). In the area of library classification, DDC is acting as the de facto standard, whereas English-language versions of UDC (complete/medium/abridged and print/online) are published by BSI, and these are

British standards.

5.3.2 Standards Development Organizations (SDOs)

As per the Wikipedia definition (http://en.wikipedia.org/wiki/Standard), "a standards development organization or SDO address the interests of a wide base of users outside the standards development organization." Although standards development organizations are mainly responsible for producing standards in different areas of human activity, some voluntary standards are offered for use by professional associations, groups of people, regulatory authorities, or industry leaders. Some of these standards achieve global recognition and acceptance, become the de facto standard in a given area of application, and are sometimes adopted by national or international standards organizations. The classic case is DCMES (Dublin Core Metadata Elements Set). However, standards development organizations may be studied under: administration level (e.g. national, regional, or international) and authority level (which agencies are involved – government entitities or others):

- **International Standards Development Organizations:** Many international standards organizations exist today but the first name comes in mind is the International Organization for Standardization (ISO), founded in 1947 at Geneva. It's a non-governmental federation of national standards bodies from 150 countries. ISO is not an acronym. The word *iso* is derived from Greek and means equal (e.g. isotope, isomer, isometric etc.). The standard development process of ISO is properly distributed and carried out through a hierarchy of technical committees, subcommittees, and working groups.
- **Regional Standards Development Organizations:** Regional SDOs are cooperative initiatives of governmental and/or non-governmental organizations for a group of countries in a geographical region e.g. CEN (European Committee for Standardization) in Europe; PASC (Pacific Area Standards Congress) in Asia-Pacific; COPANT (Pan American Standards Commission); and ARSO (African Regional Organization for Standardization) in Africa.
- **National Standards Development Organizations:** Almost all countries have their own national standards bodies. In India, we have the Bureau of Indian Standards (BIS). Generally, a national SDO is the only recognized institution in a country. They may be either public or private sector initiatives, or sometimes public-private

partnership-based initiatives. For example, the Bureau of Indian Standards is a governmental organization, whereas the American National Standards Institute (ANSI) is a non-profit organization with members from both the private and public sectors.

5.3.3 Players and paths of development

Hopkinson (2006) opined that LIS professionals consider their own national standards organizations and the library standards developed by ISO, NISO, and BSI. The library standards developed by NISO are American national standards, but in many cases, these standards are used by libraries and related organizations across the globe (e.g., Z 39.50). Now the question arises, how do standards come into being? Library professionals in India are generally concerned with the Indian National Standards, American Standards, British Standards, and international standards related directly or indirectly to the domain of library and information services. Let's discuss the activities of the division/section/committee related to library standards and working under BIS (India), NISO (US), BSI (UK) and ISO (International) respectively.

MSD 5 committee under BIS

MSD stands for Management and Standards Department, working under the Bureau of Indian Standards (BIS). The activities of MSD are coordinated through the Management and Systems Division Council (MSDC). Presently, there are eight Sectional Committees under MSDC. Out of the eight committees, four deal with the service sector, namely, documentation, education, social responsibility, banking, and financial services. MSD 5 is the Sectional Committee for Documentation and Information. MSD 5 is also entrusted to coordinate related committees/sections working under ISO.

TC 46 of the International Organization of Standards (ISO)

The committee named TC 46 under ISO is entrusted with development of library related standards. AFNOR (Association Française de Normalization) of France is the head office of TC 46 committe. There are 3 working groups (WG), 4 sub committees (SC), and 1 coordinating group (CG). It maintains a close relationship with national and global standards organizations and other technical committees of ISO. To date, there are 127 published (and 25 under development) ISO standards related to TC 46 and its SCs.

Z39 under NISO

National Standards Institute (ANSI) entrusted NISO (National Information Standards Organization) to identify, develop, maintain, and publish technical standards for managing information and documentation activities (http://www.niso.org/). NISO is also given the responsibility to represent the US in ISO (TC 46). The 3 entities under NISO develop standards - X3 (Information processing systems); Z85 (Library equipment); and Z39 (LIS discipline as a whole). The Z 39 group is the most active and the most productive unit for LIS standards.

IDT/2 of the British Standards Institute (BSI)

IDT/2 is the BSI British Standards is the National Standards Body (NSB) of UK. BSI is an independent and non-profit-making standards organization. It serves both the private and public sectors. BSI has close liaison with many standards bodies to facilitate the production of British, European and international standards. BSI standards, like those of other global standards bodies, are developed according to strict rules and processes.

5.4 STANDARDS IN AUTOMATED LIBRARY SYSTEMS

Integrated library systems or ILSs are standards-centric entities. LIS professionals need to understand that what standards must be and should be supported by an ILS before its selection or implementation. Standards perform an important role in an ILS for:

- developing workflows and in designing the modules;
- standardizing the procedures related to a workflow;
- supporting interoperability for cross-system interaction;
- evaluating a library system;
- encouraging future development; and
- helping in decision-making process.

LIS professionals in India follow standards as developed or prescribed by the BIS, India, ISO, ANIS/NISO, US and BSI, UK. The library standards developed by NISO are American national standards, but in many cases, these standards are used by libraries and related organizations across the globe (e.g., Z 39.50). You already know how SDOs at different levels develop standards related to libraries, however, it is better to know that a typical ILS should support the following standards:

- A standard for exchange of bibliographic data across the library systems (ISO-2709

developed from Z 39.2);

- Bibliographic and authority data formats (MARC 21 family of standards, UNIMARC, CCF/Bibiographic);
- A standard for fetching catalogue records (MARC/CCF/UNIMARC records) – Z 39.50;
- A standard for holdings (like serials holdings) – Z 39.71;
- EDIFACT for electronic transaction of financial activities;
- Exchange of circulation data from system to system – ANSI/NISO Z 39.83-1, Z 39.83-2
- Standards for RFID based circulation and collection management - ISO/CD 28560 (available in three parts);
- Multilingual document management (Unicode).

The other global-scale de facto standards (mostly developed by Library of Congress and IFLA) that almost all library follow are:

- MARCXML – replacing ISO-2709 in ILSs;
- MODS, MADS and METS for managing metadata object schema, authority data and metadata encoding respectively (developed by LoC);
- SRU/SRW – the next generation application of Z 39.50 protocols (developed by LoC); and
- OAI/PMH for metadata harvesting (equally important for a digital library);
- ILS/DI – A standard for seamless integration of ILS with Discovery Interface (DI). For example, libraries which are using a discovery service (like VuFind, ED, etc.) fetch real-time item-level status from the backend ILS through the ILS/DI standard.
- REST/API – REST is a generic protocol for API-based content negotiation. REST/API is a more advanced protocol in comparison with ILS/DI for real-time interaction between an ILS and a DI.

5.5 STANDARDS IN DIGITAL LIBRARY SYSTEMS

Many of the standards in the domain of automated library systems are of equal importance in digital library systems like OAI/PMH and REST/API. For example, the latest

version of the DSpace digital library (version 7.x) has changed a lot to achieve REST/API-based architectural design. Now you know the areas where DL systems should attempt to achieve standardization. Application of global standards and protocols in designing and developing DL systems supports interoperability in two major areas – i) Syntactic interoperability; and ii) Semantic interoperability. However, the specific standards related to the interoperability of digital library systems and services, apart from the basic metadata standards, may be discussed under the following heads (Mukhopadhyay, 2015):

- Metadata level interoperability for transferring/sharing metadata (OAI/PMH and Z 39.50);
- Content level interoperability for supporting multiple deposits and sharing of digital objects (SWORD (Simple Web-service Offering Repository Deposit) for multiple deposit and OA-RJ (Open Access Repository Junction);
- Identifier level interoperability for unique identification of resources and contributors (identification of contributors - ORCID and AuthorClaim; identification of digital objects - DOI, Handle system, PersID; identification of datasets - DataCite);
- Usage data level interoperability for sharing and aggregating usage statistics (Counting Online Usage of Networked Electronic Resources or COUNTER and Standardized Usage Statistics Harvesting Initiative or SUSHI);
- Network level interoperability for cross-system data transfer (OpenAIRE - Open Access Infrastructure Research for Europe and DRIVER - Digital Repository Infrastructure Vision for European Research);
- Object level interoperability for transferring compound/multimedia digital objects (OAI-ORE - Open Archive Initiative / Object Reuse and Exchange); and
- Semantic level interoperability (RDF and RDF serialization formats like TURTLE).

5.6 EMERGENCE OF SEMANTIC INTEROPERABILITY

Open standards like OAI/PMH, OAI/ORE, SWORD, REST/API, JSON, are presently considered important in view of distributed library services, increasing use of open access resources, the exorbitant scope of object integration and information mashup in library services, the need for interoperability at different levels, and the availability of technologies that support participation, interaction, and collaboration (Lagoze, 2005; COAR, 2012). These

are mainly instances of syntactical interoperability. Two new generations of interoperability are emerging rapidly are object-level interoperability (that allows transfer of full-text object along metadata from system to system), and the other one is semantic interoperability that allows RDF-based transfer of objects and data across the systems. Semantic-level interoperability is dominated by RDF serialization formats and an array of open source software like GraphDB (RDF triple store), Fuseki (RDF data store), VocBench (Semantic vocabulary) are using different semantic interoperability standards.

5.7 CHECK YOUR PROGRESS

1) The American Library Association established the physical dimension of a catalogue card (7.5 x 12.5 cm) in 1877, making it the first instance of a library standard in history.

2) IFLA once attempted to include ISBDs as ISO standards but later withdrew the ISBDs from the process of standardization for some technical reasons. ISBDs are adopted by libraries all over the world voluntarily and are considered as de facto standards.

3) The Dublin Core Metadata Elements Set (DCMES) is a metadata schema that was created collaboratively by a group of librarians and computer scientists in 1995 and quickly adopted as the de facto global standard by libraries.It was then accepted as a NISO standard in 2001 *(*ANSI/NISO Z39.85-2001) (revised in 2007 as ANSI/NISO Z39.85-2007) before being accepted as an ISO standard in 2003 (ISO 15836:2003).

4) The Bureau of Indian Standards (BIS), Government of India, has charged the MSD 5 committee with overseeing the standardisation process in the domains of documentation, information systems, and services.It is acting as the guiding post for Indian libraries in adopting standards in different activities of libraries and information centres.

5) Z 39.50 is an ANSI/NISO standard for distributed cataloguing.Almost all ILSs

nowadays include Z39.50 clients to get MARC formatted bibliographic data from Z39.50 servers. A global directory of Z 39.50 servers is available from irspy.indexdata.com/.

6) REST/API is a standard for retrieving data from REST-enabled systems in real time.It generally retrieves data in a light-weight exchange format called JSON. This leads to the possibility of cross-system interaction to develop many new generation information services in the LIS domain.

7) ORCID (Open Researcher and Contributor Identifier) is a digital identifier that authors, researchers, and others can use.It has been available since 2012 and is a very popular author ID. Anyone can freely register to get an ORCID ID for unique identification of his or her authorship.

8) OAI/PMH is a metadata interoperability standard that has six verbs and can fetch metadata of a bibliographic record easily, whereas OAI/ORE is a standard for obtaining compound digital objects like the metadata of an item along with the item itself.

9) MARC 21 standards are available to anyone, not chargeable and allows participation for suggestions, comments etc. - therefore, MARC 21 is an open standard.

10) An open standard has three major characteristics: it is free to access and use, it is well documented, and it is participatory in nature.

5.8 KEYWORDS

- **AuthorClaim:** an author identification system like ORCID.
- **CNRI Handle:** a digital object identifier developed by Corporation for National research Initiatives (CNRI).
- **COUNTER (Counting Online Usage of Networked Electronic Resources):** a major standard for digital resources usage statistics storing and transfer.
- **DRIVER (Digital Repository Infrastructure Vision for European Research):** a network level interoperability standard for digital libraries.
- **NEEO (Network of European Economists Online):** a COUNTER-based standard for

managing digital objects usage statistics.

- **OA-RJ (Open Access Repository Junction):** a protocol for multiple deposits across the repositories.
- **OAI-ORE (Open Archives Initiative – Object Reuse and Exchange):** a standard for transfering digital objects along with metadata.
- **ORCID (Open Researcher & Contributor ID):** a standard for author identifier.
- **PersID:** allows persistent identification of knowledge objects.
- **RDF (Resource Description Framework)**: a greater metadata architecture for semantic interoperability.
- **SWORD (Simple Web-service Offering Repository Deposit):** a standard to support cross-repository data transfer in real-time.
- **VIAF (Virtual Internet Authority File)**: a digital archive of name authority datasets from national libraries of the world – an initiative by OCLC.

5.9. QUESTIONS FOR SELF STUDY

1) What is a standard? Why an ILS should support global standards? List the standards required for a globally competitive ILS.
2) "All standards are specifications but not all specifications are standards." - Elucidate the statement.
3) Discuss in detail the roles of SDOs in developing library standards.
4) Make a comparison of Z 39.50 and OAI/PMH as library standards.
5) Comment on the minimum essential standards for a digital library system.
6) What is considered as the first instance of a library standard?
7) What is the status of an ISBD standard – de jure or de facto?
8) Why is DCMES considered as a unique example in the domain of ibrary standards?
9) Why is MSD 5 committee important for Indian libraries?
10) What is Z 39.50?
11) "REST/API is the future standard in LIS domain". - Elucidate the statement.
12) What is ORCID?
13) What is the major difference between OAI/PMH and OAI/ORE?
14) Do you think MARC 21 family of standards are open standards?
15) What are major features of an open standard?

5.10 REFERENCES

Avram, H.D. Maccallum S.H. and Price M.S. (1982). Organizations contributing to development of Library Standards. *Library Trends*. 31, 2, 197-224.

Bureau of Indian Standards. (2008). *Programme of work: management and system department.* New Delhi: Bureau of Indian Standards.

Campbell, D. (2004). How the use of standards is transforming Australian digital libraries. *Ariadne*, 41. Retrieved August 25, 2005, from http://www.ariadne.ac.uk/issue41/campbell/intro.html

COAR. (2012). *The current state of open access repository interoperability*. Retrieved November 11, 2013 from http://coar-repositories.org

Coyle, K. (2004) 'The 'Rights', in Digital Rights Management', *D-Lib Magazine*, 10 (9) Retrieved July 12, 2008 from http://www.dlib.org/dlib/september04/coyle/09coyle.html

Coyle, K. (2006a). Libraries and standards. *Journal of Academic Librarianship*, 31, 3, 280-283

Coyle, K. (2006b). Standards in a time of constant change. *Journal of Academic Librarianship*, 31,4, 373-376

Coyle, K. (2007). *Open Source, Open Standards.* Retrieved August 17, 2008, from ITAL: *Information Technology and Libraries*, V.21, N.1: http://www.ala.org/ala/lita/litapublications/ital/volume21no1.cfm

Doren, E. (1917). Standardization of library service. *ALA Bulletin , 11*, 19-24.

Hirsch, F.E. (1975). Library Standards" In *Encyclopedia of Library and Information Science*, edited by Allen Kent, et al., vol. 16 p. 43-62.

Hopkinson, A. (2006). Introduction to library standards and the players in the field. Digitalia. Retrieve September 25, 2008 from http://digitalia.sbn.it/upload/documenti/digitalia20062_HOPKINSON.pdf

ISO. (2008). *The TC 46 group*. Retrieved September 12, 2008 from http://www.iso.org/iso/standards_development/technical_committees/list_of_iso_technical_committees/iso_technical_committee.htm?commid=48750

Lagoze, C., Payette, S., Shin, E., & Wilper, C. (2005). An architecture for complex objects and their relationships. *International Journal on Digital Libraries*, 6(2), 124-238. Retrieved October 11, 2013 from doi:10.1007/s00799-005-0130-3.

Martin Fenner, (2011). Author identifier overview. *LIBREAS. Library Ideas*, 18. Retrieved from http://libreas.eu/ausgabe18/texte/03fenner.htm

Mukhopadhyay, P. (2015). *Interoperability and retrieval*. - UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000232199

Mukhopadhyay, P. (2014). Library standards: Players, layers and followers. *Charaibeti: Golden Jubilee Commemorative Volume*, 130–150.

Mukhopadhyay, P.. (2012). Marching with mashup: application of information mashup for developing open library system. *Challenges in Library Management System (CLMS 2012)*: Proceedings of the national seminar held on 24-25 Feb. 2012 (pp. 39–47).

Mukhopadhyay, Parthasarathi. (2011). *Unit 2: Standards.* In Ph. D. coursework in library and information science, IGNOU, Course 2: ICT in LIS Research; Block 3 Software and Standards (edited by Uma Kanjilal), New Delhi: IGNOU.

NISO. (2008). Retrieved September 12, 2008 from http://www.niso.org/standards/

Olsen, Sara. (2003). Social return on investment: standard guidelines. *Center for Responsible Business, Working Paper Series*. Paper 8 (2003). Retrieve September 20, 2008 from http://repositories.cdlib.org/crb/wps/8

Park, Margaret L. (1977). Bibliographic and Information Processing Standards. In *Annual Review* of *Information Science and Technology*, vol. 12,edited by Martha E. Williams, White Plains, N.Y.: Knowledge Industry Publications, pp. 59-80.

Qureshi, N. (1980). Standards for libraries In *Encyclopedia of Library and Information Science,* edited by Allen Kent, et al., vol.28 p.470-499.

Rusbridge, Chris, and William J. Nixon (2010). *Setting up an institutional ePrints archive—what is involved*? Unpublished paper, UKOLN Meeting. Retrieved July 12, 2010 from http://www.lib.gla.ac.uk/eprintsglasgow.html.

South African Library Association. (1968). *Standards for South African public libraries.* Potchefstroom: SALA.

Standards and Specifications In *Encyclopedia of Library and Information Science*, edited by Allen Kent, et al., vol. 30 p.176-190.

UNIT 6: CONTENT DEVELOPMENT AND METADATA ENCODING

Structure

6.0 OBJECTIVES

After going through this Unit you will be able to:

- Asses the importance metadata in a typical digital library system;
- Justify use of generic and domain-specific metadata schemas in describing digital resources ;
- Trace the differences in metadata encoding for html, xhtml and xml documents;
- Identify the role of Dublin core as a defacto global standard; and
- Manage metadata in various digital library software.

6.1. INTRODUCTION

You already have an idea about the digital library in Unit 5 of this module. The design and development of digital libraries centres on a few major issues such as technical architecture, collection building processes and methods (including digitization), metadata (document description), resource identifiers (for global identification of resources), preservation, and perpetual access. All of these activities centre around one objective: the sophisticated retrieval of digital objects. We are going to cover in this unit two important activities, as mentioned above: content development and metadata encoding. The content of a digital library is the main focus, but without proper descriptions of the content (metadata encoding), retrieval won't be possible. In simple words, metadata is data about data. Every object under the sun has its own set of metadata. For example, the metadata of a student are name, registration number, session, class, age, address, etc. The term "metadata" began to appear in the context of describing information objects on the network. Library professionals realized immediately that they had been creating data about data in the form of cataloguing since the time of Panizzi.

6.2 CONTENT IN DIGITAL LIBRARY SYSTEM

A web resource is anything that can be accessed via a URI (Uniform Resource Identifier), and metadata is just structured data about other material. Tim Berners-Lee defined metadata as information about web resources or other things that can be understood by machines. A dynamic environment, the World Wide Web constantly adds new resources. Without adhering to any set structure, interested parties produce and maintain these materials. As a result, it can be challenging to obtain pertinent information online, and despite all of their drawbacks, search engines are frequently the only method to navigate the Internet and discover resources. In such cases, metadata is obviously necessary, but information must be in a form that is understandable by search engines and by humans.

6.2.1 Features

Content in a digital library system is a combination of bitstream (documents in various formats like HTML, XHTML, and XML, PDF,DOC, PPT, along with image and multimedia files), metadata associated with the bitstream, a thumbnail image, and an attached license (Creative Commons or another). Library professionals are generally responsible for managing the metadata associated with the content in a digital library system.

6.2.2 Mark-up languages

HTML is the lingua franca of the Web, but the current activities of the W3C are focused on the development and standardization of two important projects: XML and RDF. The Extensible Markup Language (XML) is a data format for structured document interchange on the web. XML permits web authors to add tags as necessary. It is intended to make easy and straightforward use of SGML on the Web. The extensible feature of XML will make the encoding of metadata easier and more flexible. But this strength of XML leads to a serious problem with standardization. Anyone can create a set of tags for describing resources. It reduces the scope of harmonization among various metadata schemas. As a result, in addition to XML, the web requires a unifying architecture to accommodate various metadata schemas from various communities. The Resource Description Framework (RDF) is a W3C initiative in this direction. DC metadata (IETF RFC 2413) and RDF are two distinct specifications, but both communities have a number of members in common and have

evolved side-by-side. In fact, RDF is based on the Warwick Framework, a major recommendation of the Second DC Workshop at Warwick in 1996. The co-evolution of DCMES and RDF forms a natural complement within the web's greater metadata architecture. The DC has provided a semantic focus for RDF, and in turn, RDF has clarified the importance of a formal underlying data model for DC metadata. RDF is a meta-language for representing information and serves as a key piece of the technical framework underlying Semantic Web activities. RDF defines its statements in "triples": the subject is what is being described, the predicate is an indication of what property of the subject is being described by the statement, and the object is the value of the property. A simple RDF model has three parts called RDF triples. It says that a fact represented has three parts: a subject, a predicate (i.e., verb), and an object. The subject is what's at the start of the edge, the predicate is the type of edge (its label), and the object is what's at the end of the edge. The subjects, predicates, and objects in RDF are always things: concrete things or abstract concepts. The things that names denote are called resources, nodes, or entities. Predicates indicate relationships between two things. RDF also specifies that names for subjects, predicates, and objects must be expressed in uniform resource identifiers (URIs).

6.2.3 Tools and editors

Digital library software generally include a metadata editor, where a library professional can add/edit/modify metadata of a digital resource on the basis of an inbuilt metadata schema (e.g. Dublin Core). A few digital library software also include a facility to build domain-specific metadata schema, for example, Greenstone Editor for Metadata Schema (GEMS) – an additional tool in Greenstone software suite allows creation of a specific metadata set and integration of the newly created metadata schema with the metadata editor panel of Greenstone. There is a very sophisticated stand-alone metadata editor (open source) that can be used online (https://nsteffel.github.io/dublin_core_generator/generator_nq.html) or can be downloaded for local use (https://github.com/nsteffel/dublin-core-generator).

6.3 METADATA

With the rise of the Internet and the Web as global publishing media, the concept of metadata began to flourish. However, there is inconsistent use of the term "metadata" even within the library community. Some are using it to refer to the description of both digital and

non-digital resources, and others are restricting the term to the description of electronic resources. For example, the definitions given by IFLA (the International Federation of Library Associations) and W3C (the World Wide Consortium) are restrictive in nature. IFLA defines (IFLA, 2002) metadata as "any data used to aid the identification, description, and location of networked electronic resources." According to W3C (W3C, 2003), "Metadata is the machine-understandable information for the Web." In contrast, the definitions given by the Getty Research Institute (GRI) and UKOLN (U.K. Office for Library and Information Networking) are fairly liberal. GRI says (Murtha, 2002) that metadata is "data associated with either an information system or an information object for purposes of description, administration, legal requirements, technical functionality, use and usage, and preservation." Similarly, UKOLN (UKOLN, 2002) says, "Metadata is normally understood to mean structured data about digital (and non-digital) resources that can be used to help and support a wide range of operations." These might include, for example, resource description and discovery, the management of information resources (including rights management), and their long-term preservation.

6.3.1 Types and importance

This unit takes a liberal stance in terms of the definition and scope of the term "metadata." Metadata is here used to mean structured information about an information resource of any media type or format. Metadata, by definition, is a description of an entity or object. Recently, different uses of metadata have led to the erection of a very broad typology of metadata as being descriptive, administrative, and structural (Hadge, 2001).

a) Descriptive metadata

It is meant to serve the following purposes -

- Discovery (how one can find a resource);
- Identification (how a resource can be distinguished from other similar resources);
- Selection (how to determine that a resource fills a particular need);
- Collocation (bringing together all versions of a work);
- Obtain (obtaining a copy of resource, or access to one); and
- Other related functions (evaluation, linkage and usability).

b) **Administrative metadata** is information intended to facilitate the management of resources such as date of creation, rights and restrictions of access and archiving, control or processing activities etc.

c) **Structural metadata** is concerned with recording of relationships that holds compound digital objects together.

Metadata has many different applications in diverse areas of human activities. The major applications are as follows –

- Metadata is used for resource description;
- Resource description leads to speed up and enrich searching for resources;
- Metadata provide additional information to users of the dataset it describes;
- Metadata helps to bridge the semantic gap in information retrieval;
- Certain metadata is designed to optimize data transfer and compression;
- Some metadata enable variable content presentation;
- Descriptive metadata can be used to automate workflows;
- Metadata is important for electronic resource discovery; and
- Metadata is essential for document retrieval on the World Wide Web because of the need to find useful information from huge mass of information available.

6.3.2 Metadata vs cataloguing

There are strong similarities between traditional library cataloguing and the description of Web resources by using a set of metadata. Modern cataloguing theory and practice evolved over the last 150 years (recall Panizzi's efforts). It evolved as a tool for organizing information resources for retrieval in libraries. Library catalogues typically consist of a collection of bibliographic records that describe library resources such as printed books, cartographic materials, music scores, manuscripts, etc. Gradually, the scope of cataloguing codes and resource description standards has expanded to include a range of newer publishing media such as sound recordings, microfilms, video recordings, films, and computer files. The increasing use of the Internet as a publishing medium also influenced these standards and cataloguing codes. As a direct result of this impact, MARC21 (the current

version of MARC) has identified a new tag 856 for representing the URI of the resources to be described. The intellectual content of Web resources is primarily text. Therefore, the metadata required for describing digital information resources will bear a strong resemblance to the metadata that describes traditional printed texts. But the question arises as to whether the relatively heavyweight codes developed for traditional library cataloguing (e.g., MARC 21 bibliographic format) are really suitable for the mobile and ephemeral resources on the Web. As a result, practical solutions to the problem of web retrieval should avoid the use of traditional cataloguing codes or formats. The ideal solution may be the use of a resource discovery system that makes use of a relatively simple metadata format. In view of this discussion, the major similarities and differences may be earmarked in Table 6.1.

Table 6.1: Metadata encoding vs Cataloguing

	Metadata encoding	Cataloguing
S i m i l a r i t i e s	Description of information resources of any types or formats	Description of knowledge objects available mainly in print formats
	Objective is to support findability of resources	Objectives are to find, to select, to locate and to obtain
	Encoding is based on rules and specifications	Cataloguing is based on catalogue codes, description standards (ISBDs) and Framework standards (MARC, CCF etc.)
	There are common elements of description (generic metadata) and elements for specific document types (domain-specific elements)	Some elements are generic (e..g. ISBD-G) and some elements are for specific document types (ISBD-M, ISBD-CM, ISBD-ER and so on)
D	Generally deployed for electronic resources	Generally deployed for print resources
	Encoding rules are simple in nature	Encoding rules are complex and demand skills
	Target users are contributors and creators	Target users are skilledlibrary professionals

i	of documents	
f	Every element is optional and repeatable	There are specific rules for determining which elements of description are mandatory, optional or repeatable
e r e n c e s	Includes only basic elements of description and not quite suitable for comprehensive description of resources	Includes an array of elements (tags, subfields etc) for simple and complex levels of description and suitable for comprehensive treatment of a resource

6.3.3 Metadata schema

Metadata schemas are set of metadata elements and rules for their use that have been defined for a particular purpose. A metadata schema specifies three independent but related aspects of metadata – semantics, content rules and syntax.

Semantics refers to the metadata elements that are included in the scheme by giving each of them a name and definition. A metadata schema also specifies whether each element is mandatory, optional or conditionally required and whether the element may or may not be repeated.

Content rules indicate how values for metadata elements are selected and represented. For example, semantics of a metadata schema may define the element “author” but the content rules would specify which agents qualify as author (selection) and how an author’s name should be recorded (representation).

Syntax of a metadata schema is concerned with the encoding of metadata elements in machine-readable form. Syntax also specifies the way of transmission, transport and communication of metadata between different systems.

Metadata registries are used whenever data must be used consistently within an organization or group of organizations. A metadata registry typically has the following characteristics:

It is a protected area where only approved individuals may make changes

It stores data elements that include both semantics and representations

The semantic areas of a metadata registry contain the meaning of a data element with precise definitions

The representational areas of a metadata registry define how the data is represented in a specific format such as within a database or a structured file format such as XML

Library and information professionals are generally associated with descriptive metadata. As per their applications, metadata schemas are of two types: generic and domain-specific. Generic metadata schemas are intended to be generally applicable to all types of resources (e.g., Dublin Core Metadata Elements Set), whereas domain-specific metadata schemas are primarily designed to describe items related to a particular category (e.g., VRA [Visual Resource Association] Core for visual resource collection, FGDC (Federal Geographic Data Committee) metadata schema for geospatial data, etc.). Both of these types of schemas have different types of metadata elements, i.e., descriptive data elements, administrative data elements, and structural data elements.

6.3.4 Generic metadata schema

The Dublin Core Metadata Element Set (DCMES) or Dublin-core is a small set of resource description categories that is notably different from many of the other metadata schemas due to its ease of use and interoperability. The Dublin Core Metadata Initiative (DCMI), an international community, has led the development of metadata components that enhance cross-disciplinary resource discovery. The mission of DCMI is to develop an easy and seamless mechanism for searching and indexing web resources through: (i) developing metadata standards for cross-domain resource discovery; (ii) defining frameworks for the inter-operation of metadata sets; and (iii) facilitating the development of discipline-specific metadata sets that work within the frameworks of cross-domain resource discovery and metadata interoperability. The DC element set is today's de facto standard for metadata on the web. The DC metadata set has 15 major elements (total 22 elements for description – see Figure 6.1), and these metadata elements fall into three groups: (i) elements related mainly to the content of the resource; (ii) elements related mainly to the resource when viewed as intellectual property; and (iii) elements related mainly to the instantiation. DC elements are flexible enough for the description of a variety of resources in different subject areas.

Moreover, the meanings of the elements will be understood by most users. DC metadata achieved this level of quality by adhering to six principles:

Intrinsically: DC metadata is based on intrinsic data. These data refer to the property that could be identified from the intellectual content and physical form of the resource.

Extensibility: It allows the inclusion of extra descriptive materials for specialised requirements;

Syntax Independence: It is applicable to a wide range of disciplines and application programmes;

Optionality: All the DC elements are optional.

Repeatability: All the DC elements are repeatable. For example, a resource with multiple authors may use the "Creator" element repeatedly to accommodate all the authors; and

Modifiability: Each element in the Dublin Core has a self-explanatory definition. Each element can be modified by an optional qualifier, and in such cases, the definition of the element is modified by the value of the qualifier.

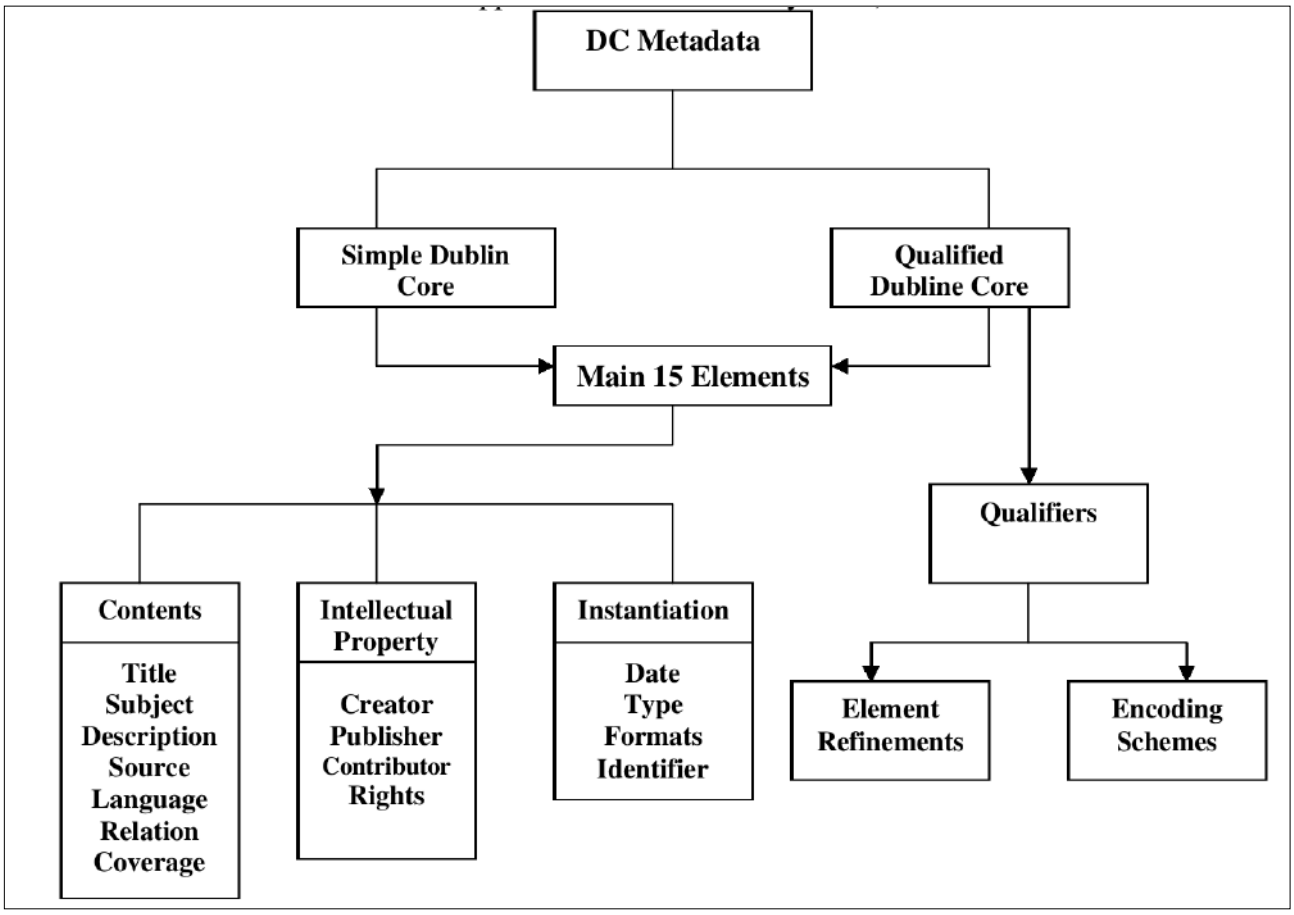

Figure 6.1: Categories of Dublin Core metadata elements

In simple terms, the DC metadata schema follows the dumb-down principle (i.e., any element may be ignored or simplified in describing an object) and the one-to-one principle (i.e., a single metadata record describes a single object). As a bibliographic data model, the DC metadata schema has the following properties: 1) machine processability through a coherent data model; 2) provides explicit definitions of resources; 3) relates DC principles and practises to the developments outside the Dublin Core Metadata Initiative (DCMI); and 4) makes clear the relationship of DC "packages" of information to other metadata "packages." A list of major metadata elements of DCMI alongside scopes, element refinement and encoding schemes are given in Table 6.2.

Table 6.2: Major DC metadata elements

Sl. No	DC Elements	Element Refinement	Element Encoding Scheme(s)

	(Scope)	**(Comments, if any)**	**(Comments, if any)**
1	Title (Name given to the resource)	Alternative (Substitute to the formal title)	-
2	Creator (Entity that created the content)	-	-
3	Subject (Topic or Keywords)	-	LCSH, MESH, DDC. LCC. UDC
4	Description (Account, summary or abstract of the content)	Table of Contents (A list of sub-units of the content of the resource) Abstract (A summary of the contents of the resource)	-
5	Publisher (Entity that made the resource available)	-	-
6	Contributor (Other entity that made a contribution)	-	-
7	Date (Date of an event in the life of the resource)	Created (Date of creation of the resource) Valid (Date/period of validity of a resource) Available (Date/period that the resource will become or did become available) Issued (Date of formal publication/issuance of the resource)	DCMI Period (A specification of the limits of a time interval. Available at: http://dublincore.org/documents/dcmi-period/) W3C-DTF (W3C encoding rules for dated and times based on ISO 8601. Available at: http://www.w3.org/TR/NOTE-datetime)

		Modified (Date on which the resource was changed)	
8	Type (Nature, genre or category of the resource)	-	DCMI Type Vocabulary (A list of types used to categorise the nature or genre of the content of the resource. Available at: http://dublincore.org/documents/dcmi-type-vocabulary/)
9	Format (Physical or digital manifestation of the resource)	Extent (The size or duration of the resource) Medium (The material or physical carrier of the resource)	- IMT (The Internet media type of the resource. Available at: http://www.isi.edu/in-notes/iana/assignments/media-types/media-types)
10	Identifier (An unambiguous reference to the resource within a given context)	-	URI (A uniform resource identifier. Available at: http://www.ietf.org/rfc/rfc2396.txt)
11	Source (Reference to the resource's origin)	-	URI (A uniform resource identifier. Available at: http://www.ietf.org/rfc/rfc2396.txt)
12	Language (Language of the content of the resource)	-	ISO 639-2 (Codes for the representation of names of languages. Available at: http://www.lcweb.loc.gov/standards/iso639-2/langhome.html)
13	Relation (Reference to a related resource)	Is Version Of (The described resource is a version, edition, or adaptation of the referenced resource) Has Version (The described resource has	

		a version, edition or adaptation, namely the referenced resource) Is Replaced By (The described resource is supplanted, displaced or superseeded by the referenced resource) Replaces (The described resource supplants, displaces or superseds the referenced resource) Is Required By (The described resource is required by the referenced resource either physically or logically) Requires (The described resource is a physical or logical part of the referenced resource) Is Part Of (The described resource is a physical or logical part of the referenced resource) Has Part (The described resource includes the referenced resource either physically or logically) Is Referenced By (The described resource is referenced, cited or otherwise pointed to by the referenced resource) References (The described resource references, cites, or otherwise points to the	URI (A uniform resource identifier. Available at: http://www.ietf.org/rfc/rfc2396.txt)

		referenced resource) Is Format Of (The described resource is the same intellectual content of the referenced resource, but presented in another format) Has Format (The described resource pre-existed the referenced resource, which is essentially the same intellectual content presented in another format)	
14	Coverage (Extent or scope of the content of the resource)	Spatial (Spatial characteristics of the intellectual content of the resource e.g. place name or geographic coordinates) Temporal (Temporal characteristics of the intellectual contents of the resources e.g. a period label or date range)	DCMI Point (Identifies a point in space using its geographic coordinates. Available at: http://www.dublincore.org/documents/dcmi-point/) ISO 3166 (Codes for the representation of names of countries. Available at: http://www.din/de/germien/nas/nabd/is03166ma/codlstp/index.html) DCMI Box (A specification of the limits of a time interval. Available at: http://dublincore.org/documents/dcmi-box/) TGN (The Getty thesaurus of geographic names. Available at: http://shiva.pub.getty.edu/tgn_browser) DCMI Period (A specification of the limits of a time interval. Available at: http://dublinecore.org/documents/dcmi-period/)

			W3C-DTF (Rules for encoding dates and times, based on ISO 8601. Available at: http://www.w3.org/TR/NOTE-datetime)
15	Rights (Information about rights held in and over the resource)	-	-

Apart from these 15 basic elements, there are DC elements like Audience, Provenance and Rights Holder. But please take a note that these three elements are not part of the Simple Dublin Core. You may use Audience, Provenance and Rights Holder only when using Qualified Dublin Core. Another four newly added DC elements are – Instructional Method (represents ways of presenting instructional materials or conducting instructional activities etc.), Accrual Method (represents the method by which items are added to a collection), Accrual Periodicity (represents the frequency with which items are added to a collection), Accrual Policy (represents the policy governing the addition of items to a collection).

Note: See author's previous publication for more details - Mukhopadhyay, P. (2015). Interoperability and retrieval. - UNESCO.

https://unesdoc.unesco.org/ark:/48223/pf0000232199

6.3.5 Domain-specific metadata schemas

The core function of a digital library setup is to deliver the right contents to users at the right time. Metadata plays a critical role to fulfil this core function. The generic and domain-specific metadata standards that are in use in different digital library initiatives are listed herewith alphabetically for your ready reference:

- **ABCD - Access to Biological Collection Data**
  (http://www.dcc.ac.uk/resources/metadata-standards/abcd-access-biological-collection-data): An evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data) sponsored

by Biodiversity Information Standards TDWG - the Taxonomic Databases Working Group.

- **AGLS** (Australian Government Locator Service, http://www.naa.gov.au/records-management/create-capture-describe/describe/AGLS/index.aspx): AGLS is Australian government metadata standard intended for the description of government resources on the Web. It uses DCMI Terms properties with a few additional metadata elements such as function and mandate.
- **AgMES - Agricultural Metadata Element Set** (http://www.dcc.ac.uk/resources/metadata-standards/agmes-agricultural-metadata-element-set): AgMES, developed by the Food and Agriculture Organization (FAO) of the United Nations enables description, resource discovery, interoperability and data exchange of different types of information resources in all areas relevant to food production, nutrition and rural development.
- **CanCore** (http://cancore.athabascau.ca/en/index.html): CanCore is a set of guidelines for the implementation of the IEEE LOM metadata standard for describing learning resources. It is originated in Canada for managing learning objects in Canadian universities.
- **CSMD-CCLRC** (Core Scientific Metadata Model, http://www.dcc.ac.uk/resources/metadata-standards/csmd-cclrc-core-scientific-metadata-model): It is designed by Science and Technologies Facilities Council to support data collected within a large-scale facility's scientific workflow but the model is also designed to be generic across scientific disciplines.
- **Cataloguing Cultural Objects** (CCO, http://cco.vrafoundation.org/): A schema for cultural objects, developed by the US-based Visual Resources Association with significant input from the Getty Research Institute.
- **Categories for the Description of Works of Art** (CDWA, http://www.getty.edu/research/conducting_research/standards/cdwa/): An extensive metadata schema for cataloguing objects held by art museums developed in the US in the 1990s by the Getty Research Institute.
- **Darwin Core** (http://www.dcc.ac.uk/resources/metadata-standards/darwin-core): A metadata schema developed Biodiversity Information Standards (TWDG) by to cover elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity.

- **DataCite** Metadata Schema (http://www.dcc.ac.uk/resources/metadata-standards/datacite-metadata-schema): A set of mandatory metadata elements prescribed by DataCite consortium that to support persistent approach to access, identification, sharing, and re-use of digital research datasets.
- **DDI - Data Documentation Initiative** (http://www.dcc.ac.uk/resources/metadata-standards/ddi-data-documentation-initiative): A globally recognized standard for describing data from the social, behavioral, and economics and statistics. The XML based DDI metadata specification supports the entire research data life cycle.
- **DIF - Directory Interchange Format** (http://www.dcc.ac.uk/resources/metadata-standards/dif-directory-interchange-format): A domain-specific schema for Earth sciences community, intended for the description of scientific data sets. It inlcudes elements focusing on instruments that capture data, temporal and spatial characteristics of the data.

- **e-GMS** (http://www.govtalk.gov.uk/): A schema dedicated to e-governance developed in UK for describing information resources to ensure maximum consistency of metadata across public sector organizations in the UK.

- **Encoded Archival Description** (EAD, http://www.loc.gov/ead/): A well known schema that provides an encoding for archival descriptions. It adopts a multi-level approach to description, providing information about a collection as a whole and then breaking it down into groups, series and (if significant) individual items. grew out of work done at UC Berkeley in the mid 1990s and was influenced by TEI and ISAD(G).

- **ETD-MS** (https://sites.google.com/a/ndltd.org/ndltd/standards/metadata): NDLTD is the developer of ETD-MS. The initial goal of NDLTD was to develop a standard XML DTD for encoding metadata elements for ETDs. ETDMS is based on the Dublin Core Element Set, but includes an additional element specific to metadata regarding theses and dissertations.
- **EXIF** (Exchangeable Image File Format, http://www.exif.org): A technical metadata standard that can be written to and read from a still image file itself (and formats). It was developed by JEITA (Japan Electronics and Information Technology Industries Association).

- **FGDC/CSDGM** - Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata (http://www.dcc.ac.uk/resources/metadata-standards/fgdccsdgm-federal-geographic-data-committee-content-standard-digital-ge): A widely-used, schema for digital geospatial data required by the US Federal Government. It is sponsored by the US Federal Geographic Data Committee.
- **FOAF** (Friend of a Friend, http://www.foaf-project.org/): FOAF is a RDF-enabled schema for describing people and intended to be used on the Semantic Web. It includes features for encoding names, email addresses, personal interests, home pages, and various online identities. In future traditional library authority files may be translated into FOAF but it needs to settle two very important issues – i) each individual has only one FOAF identity; and ii) FOAF focuses on online presence for current living persons.
- **Genome Metadata** (http://www.dcc.ac.uk/resources/metadata-standards/genome-metadata): A schema dedicated to the field of Genomics. It consists of 61 different metadata fields covering broad categories: Organism Info, Isolate Info, Host Info, Sequence Info, Phenotype Info, Project Info, and Others.
- **GEM** (Gateway to Educational Materials, http://www.thegateway.org/about/ documentation2/ schemas/index_html/): GEM is an RDF-enabled metadata vocabulary designed for the description of educational resources. The GEM model includes all the properties available in DCMI Terms, with a few additional education-specific elements such as educational standards and pedagogical methods.
- **GILS** (http://www.gils.net/): Global Information Locator Service or GILS is a schema governments, companies, or other organizations to support citizen/customer facing information services. GILS was an early metadata standard for the encoding of descriptive information for government records.
- **IEEE-LOM** The IEEE Learning Object Metadata (http://ltsc.ieee.org/wg12/index.html): It aims to develop technical standards, recommended practices, and guides for learning technology. The LOM standard mainly builds on the Dublin Core and is based on the recommendations of IMS and ARIADNE project. It is a multi-part standard contains a description of semantics, vocabulary, and extensions. LOM has a wide set of globally agreed metadata elements

which are grouped into nine descriptive categories: General, Life cycle, Meta metadata, Technical, Educational, Rights, Relation, Annotation, and Classification.

- **IMS The IMS Global Learning Consortium** (http://www.imsglobal.org/): It develops and promotes the adoption of open technical specifications for interoperable learning technology. IMS is based on LOM and Dublin Core metadata.
- **International Virtual Observatory Alliance Technical Specifications** (http://www.dcc.ac.uk/resources/metadata-standards/international-virtual-observatory-alliance-technical-specifications): A schema for astronomical objects developed by the IVOA (International Virtual Observatory Alliance) to enable interoperability between and the integration of astronomical archives across the world into an international virtual observatory (last modified in 2009).
- **ISO 19115** (http://www.dcc.ac.uk/resources/metadata-standards/iso-19115): An internationally-adopted schema for describing GIS (geographic information and services). It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data (last modified in 2009).
- **MIDAS** (http://www.english-heritage.org.uk/server/show/nav.8331): It is a UK standard for describing cultural heritage assets that form the historic environment (buildings, archaeological sites, shipwrecks, areas of interest, artefacts and ecofacts).
- **MIX** (http://www.loc.gov/standards/mix/): It is an XML based schema for encoding the Technical Metadata for Digital Still Images standard developed by NISO group on Metadata for Images in XML ((last modified in 2009).
- **NewsM**L (News Markup Language, http://www.newsml.org/): The NewsML aims to design a complex schema for describing textual news, articles, photos, graphics, audio, and video — the components that make up or express news items.
- **ONIX** (http://www.editeur.org/onix.html): A schema developed by book industry to support Online Information Exchange - international standard for representing and communicating book industry product information in electronic form.
- **PBCore** (http://www.pbcore.org/): Public Broadcasting Metadata Dictionary or PBCore is intended for use by television, radio and web broadcasters and hopes to describe and retrieve broadcast contents efficiently (last modified in 2011).
- **PREMIS** (http://www.loc.gov/standards/premis/): A technical metadata schema that

provides a "dictionary" of core metadata elements that can be used to support the digital preservation of a resource. A key feature of the PREMIS model is the definition of Objects as made up of Representations, Files, and Bitstreams. It was particularly influenced by a conceptual model called the Open Archival Information System. The Library of Congress is the official PREMIS maintenance agency (last modified in 2006).

- **SPECTRUM** (http://www.collectionslink.org.uk/manage_information/spectrum): A key UK standard for museum documentation (last modified in 2005).
- **SDMX** - Statistical Data and Metadata Exchange (http://www.dcc.ac.uk/resources/metadata-standards/sdmx-statistical-data-and-metadata-exchange): A set of common technical and statistical standards and guidelines to be used for the efficient exchange and sharing of statistical data and metadata (last modified in 2012).
- **SWAP** (Scholarly Works Application Profile, http://www.ukoln.ac.uk/repositories/digirep/index/Scholarly_Works_Application_Profile): SWAP is a DCMI-compliant application profile for the description of scholarly works, developed by UKOLN. It aims to support quality metadata encoding of knowledge objects in Green OA. SWAP is based on the FRBR conceptual model, and therefore differentiates between Works and their Manifestations.
- **Text Encoding Initiative (TEI) Header** (http://www.tei-c.org): It is a scheme for marking up electronic text. It also specifies a header portion to accommodate metadata about the object to be described. TEI headers can be used to record bibliographic information of both electronic and non-electronic sources. The TEI header can be mapped to and from MARC.
- **VRACore** (Visual Resources Association Core Categories, http://www.vraweb.org/organization/committees/datastandards/index.html): A widely used metadata schema for describing art or cultural images, providing 17 core categories (last modified in 2007).
- **XrML** (eXtensible Rights Markup Language, http://www.xrml.org/): XrML is an XML language for the encoding of rights information. It is focused on the action of "granting" authorizations between Principals, Rights, Resources, and Conditions.

This is only an illustrative list of metadata standards available in different domains. Metadata elements must be encoded in a standard manner for the purpose of using, processing, and retrieving objects in DL systems. Encoding standards act as a kind of container that holds the precious metadata content; it can be thought of as a way of carrying the metadata from one place to another. The encoding of all these metadata elements takes place in three formats – HTML, XHTML and RDF/XML.

Note: See author's previous publication for more details - Mukhopadhyay, P. (2015). *Interoperability and retrieval*. - UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000232199

6.4 DUBLIN-CORE METADATA: ELEMENTS AND ENCODING RULES

The DC element set is today a defacto standard for metadata on the web. The DC metadata is a set of 15 main elements and presently consists of 22 elements of description. These metadata elements fall into three groups which indicate class or scope of information stored in them (see Figure 6.1). These element may broadly be divided into three groups - 1. elements related mainly to the Content of the resource; 2. elements related mainly to the resource when viewed as Intellectual Property; and 3. elements related mainly to the Instantiation. Some of these elements may be refined using qualifiers. "Simple Dublin Core" is DC metadata that uses no qualifiers. It only applies the main 15 elements, with no qualifiers. On the other hand, "qualified Dublin Core" uses additional qualifiers to increase the specificity or precision of the metadata. Qualified Dublin Core includes three additional elements (audience, provenance, and rights holder), as well as a group of element refinements (also called qualifiers). The objective is to refine the semantics of the elements for effective resource discovery. For example, a "Date" is a DC element that may be specified or refined to identify a particular kind of date (date of last modification, date of publication, etc.) and may be standardized by using a date representation scheme like the ISO date scheme. The DCMI presently admits two broad classes of qualifiers: (i) element refinement (these qualifiers make the meaning of an element specific); and (ii) encoding schemes (these qualifiers identify schemes that aid in the interpretation of an element value; these schemes include controlled vocabularies and formal notations, e.g., a term from a set of subject headings or a standard expression of a date like "2022-12-25"). The rule sets for encoding metadata in digital documents are based on Request for Comments (RFCs) – the standards of World Wide Web Consortium (W3C). This section illustrated encoding of DC metadata elements in HTML/XHTML and RDF/XML.

6.4.1 Metadata encoding: HTML and XHTML

In the metadata community, namespaces are used to identify "newly defined" elements and their qualifiers. A DCMI namespace is a collection of DCMI terms. Each DCMI namespace is identified by a URI. DCMI uses XML-namespace mechanism for the identification of all DCMI terms. The encoding syntax of DC metadata may be grouped into three categories – HTML encoding, XHTML encoding and RDF/XML encoding. In HTML (Standard 4.0) the syntax standard uses <META> tag to place the description within the page's <HEAD>…. </HEAD> area. Although not advised by DCMI, a recognised good practice is evolving, whereby the DC element name is given in sentence case, preceded by an identifier in upper case to denote the element is from Dublin Core (e.g. DC.Creator). The Network Working Group of Internet Society issued a memo in December 1999 (RFC: 2731) for encoding DC metadata in HTML. As per this memo the general syntax is

<meta name = "PREFIX.ELEMENT_NAME" content = "ELEMENT_VALUE">

It may be illustrated as (see - https://www.dublincore.org/specifications/dublin-core/dc-html/2008-08-04/) -

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/html4/strict.dtd">
<html>
<head profile="http://dublincore.org/documents/2008/08/04/dc-html/">
<title>metadata</title>
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/">
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/">
<meta name="DC.Title" content="Digital resource management">
<meta name="DC.Creator" content="Karnataka State Open University">
<meta name="DC.Subject" scheme="DCTERMS.LCSH" content="Digital library">
<meta name="DC.Subject" scheme="DCTERMS.LCSH" content="Metadata">
<meta name="DC.Subject" scheme="DCTERMS.LCSH" content="Digital asset management">
<meta name="DC.Description" content="This unit describes steps and activities related to digital resource management.">
<meta name="DC.Publisher" content="Karnataka State Open University">
<meta name="DC.Contributor" content="Parthasarathi Mukhopadhyay">
<meta name="DC.Date" scheme="DCTERMS.W3CTDF" content="2022-11-30">
<meta name="DC.Type" scheme="DCTERMS.DCMIType" content="Text">
<meta name="DC.Format" scheme="DCTERMS.IMT" content="text/html">
```

```
<meta name="DC.Language" scheme="DCTERMS.ISO639-2" content="en">
</head>
<body>
</body>
</html>
```

Similarly the XHTML based encoding of DC elements may be illustrated as below:

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head profile="http://dublincore.org/documents/2008/08/04/dc-html/">
<title>metadata</title>
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
<meta name="DC.Title" content="Digital resource management" />
<meta name="DC.Creator" content="Karnataka State Open University" />
<meta name="DC.Subject" scheme="DCTERMS.LCSH" content="Digital library" />
<meta name="DC.Subject" scheme="DCTERMS.LCSH" content="Metadata" />
<meta name="DC.Subject" scheme="DCTERMS.LCSH" content="Digital asset management" />
<meta name="DC.Description" content="This unit describes steps and activities related to digital resource management." />
<meta name="DC.Publisher" content="Karnataka State Open University" />
<meta name="DC.Contributor" content="Parthasarathi Mukhopadhyay" />
<meta name="DC.Date" scheme="DCTERMS.W3CTDF" content="2022-11-30" />
<meta name="DC.Type" scheme="DCTERMS.DCMIType" content="Text" />
<meta name="DC.Format" scheme="DCTERMS.IMT" content="text/html" />
<meta name="DC.Language" scheme="DCTERMS.ISO639-2" content="en" />
</head>
<body>
</body>
</html>
```

6.4.2 Metadata encoding: RDF/XML

As discussed earlier, the W3C is focusing on important projects namely XML and RDF. The following is an example of RDF/Xml encoding.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:dcterms="http://purl.org/dc/terms/">
 <rdf:Description rdf:about="http://example.org/123">
  <dcterms:title xml:lang="en">Learning metadata</dcterms:title>
 </rdf:Description>
</rdf:RDF>
```

We already know that an expression in RDF is a "triple," consisting of a subject (the object being described e.g., the sky), a predicate (an element or field describing the object e.g., colour), and an object (the value that the predicate takes on e.g., blue). A set of RDF triples is called an RDF graph. Let's see an example showing the representation of the Web site of the University of Burdwan by using DCMES as schema and RDF as framework.

Subject (Resource)	Predicate (Attribute/property)	Object (Value of attribute)
The University of Kalyani http://www.klyuniv.ac.in	dc:title	The University of Kalyani site
	dc:creator	Sarkar, B.
	dc:subject	Academic Institute
	dc:descriptipon	The University established in the year 1960 under UGC Act……..
	dc:publisher	The University of Kalyani
	dc:contributor role=content writer	Central Library, KU
	dc:date	20220101
	dc:format	text/html
	dc:identifier	http://www.klyuniv.ac.in
	dc:coverage	Education and Research
	dc:rights	The University of Kalyani

Table 6: RDF modeling of DCMES

The encoding of DCMES in RDF structure on the basis of above data model may be entered as:

```
<?xml version="1.0"?>

<rdf: RDF xmlns: rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"

xmlns: dc = "http://purl.org/dc/elements/1.1/">

<rdf:Description about = "http://www.buruniv.ac.in">

<dcterms:title xml:lang="en">Welcome to the Home page of the University of Kalyani </dcterms:title>

<dcterms:creator xml:lang="en"> Sarkar, B.</dcterms:creator>

<dcterms:subject xml:lang="en">Academic Institute</dcterms:subject>

<dcterms:description xml:lang="en">The University of Kalyani , a university under UGC…
</dcterms:description>

<dcterms:publisher xml:lang="en">The University of Kalyani</dcterms:publisher>

<dcterms:contributor role="content writer" xml:lang="en"> Central Library, Kalyani University
</dcterms:contributor>

<dcterms:date xml:lang="en">20220101</dcterms:date>

<dcterms:format xml:lang="en">text/html</dcterms:format>

<dcterms:identifier xml:lang="en">"http://www.klyuniv.ac.in"</dcterms:identifier>

<dcterms:language xml:lang="en">en</dcterms:language>

<dcterms:coverage xml:lang="en">Education and Research</dcterms:coverage>

<dcterms:rights xml:lang="en"> The University of Kalyani </dcterms:rights>

</rdf:Description>

</rdf:RDF>
```

Almost all the advanced level repository management software support RDF based encoding of DCMES. For example, a deposited record in EPrint archive software stores DC metadata elements in RDF format (the metadata of digital resource submitted to EPrint software can also be exported in RDF format).

6.5 METADATA MANAGEMENT

Most of the DL software follows an abstract model for Dublin Core metadata that specifies the components and constructs used in Dublin Core metadata and also prescribes how those components are combined to create information structures. The DCMI abstract model is

equally applicable to various domain-specific metadata schemas. This model is independent of any particular encoding syntax. The DCMI abstract model is a combination of a resource model and a description model. The **resource model** prescribes the following rules:

- Each described resource is described using one or more property-value pairs.
- Each property-value pair is made up of one property and one value.
- Each value is a resource—the physical, digital, or conceptual entity or literal that is associated with a property when a property-value pair is used to describe a resource. Therefore, each value is either a literal value or a non-literal value.
- A literal value is a value that is a literal.
- A non-literal value is a value that is a physical, digital, or conceptual entity.
- A literal is an entity that uses a Unicode string as a lexical form, together with an optional language tag or datatype, to denote a resource.

On the other hand, the **description model** advocates the following rule base:

- A "description set" is a set of one or more descriptions, each of which describes a single resource.
- A description is made up of one or more statements (about one and only one resource) and zero or one described resource URI (a URI that identifies the described resource).
- Each statement instantiates a property-value pair and is made up of a property URI (a URI that identifies a property) and a value surrogate.
- A value surrogate is either a literal value surrogate or a non-literal value surrogate.
- A literal value surrogate is a value surrogate for a literal value and is made up of exactly one value string. The value string is a literal, which encodes the literal value.
- A non-literal value surrogate is a value surrogate for a non-literal value and is made up of zero or one value URI (a URI that identifies the non-literal value associated with the property), zero or one vocabulary encoding scheme URI (a URI that identifies the vocabulary encoding scheme of which the non-literal value is a member), and zero or more value strings. Each value string is a literal, which represents the non-literal value.
- A value string is either a plain value string or a typed value string.
- A plain value string may have an associated value string language that is an ISO language tag (for example, en-GB). Plain-value strings are intended to be human-readable.

- A typed value string has an associated syntax encoding scheme. URI that identifies a syntax encoding scheme

On the basis of this abstract model DCMI recommends encoding of DC metadata elements in three encoding standards:

- Expressing Dublin Core metadata using the DC-Text format (see https://www.dublincore.org/specifications/dublin-core/dc-text/)
- Expressing Dublin Core metadata using HTML/XHTML meta and link elements (see https://www.dublincore.org/specifications/dublin-core/dc-html/)
- Expressing Dublin Core metadata using the Resource Description Framework (RDF) (see https://www.dublincore.org/specifications/dublin-core/dc-rdf/)

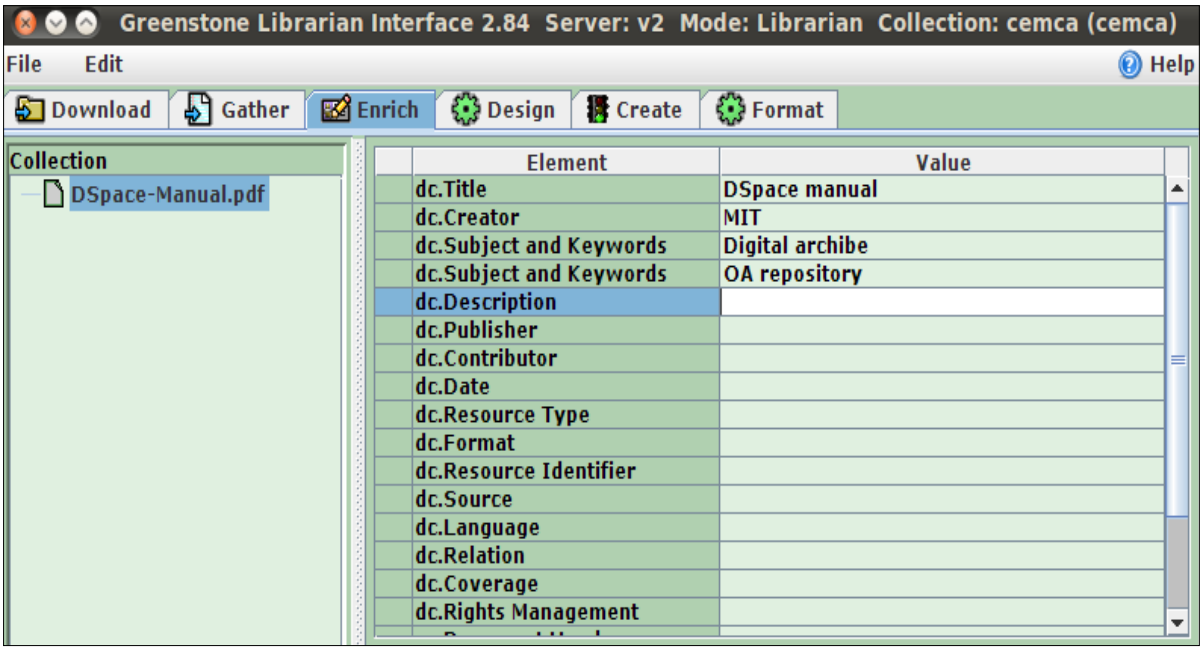

Figure 6.2: Metadata entry interface in Greenstone

DC elements encoding is therefore based on four types of mark-up languages – HTML, XHTML, XML and RDF/XML (see section 6.4 for real-life examples). Most of the digital repository management software (e.g. Greenstone, Eprint, Dspace) include simple DCMES and qualified DCMES by default. The metadata entry interface of Greenstone is given in Figure 6.2. The metadata entered by submitters and/or librarians are stored in repository management software generally in XML format. The metadata as entered in Figure 6.2 are stored inside the Greenstone in the following format as metadata.xml file.

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<!DOCTYPE DirectoryMetadata SYSTEM
"http://greenstone.org/dtd/DirectoryMetadata/1.0/DirectoryMetadata.dtd">

<DirectoryMetadata>

   <FileSet>

      <FileName>DSpace-Manual\.pdf</FileName>

      <Description>

         <Metadata mode="accumulate" name="dc.Title">DSpace manual</Metadata>

         <Metadata mode="accumulate" name="dc.Creator">MIT</Metadata>

         <Metadata mode="accumulate" name="dc.Subject">Digital archibe</Metadata>

         <Metadata mode="accumulate" name="dc.Subject">OA repository</Metadata>

      </Description>

   </FileSet>

</DirectoryMetadata>
```

6.6 SUMMARY

Digital content are the focal point of a typical digital library system. The discovery of digital content depends on efficient description of their attributes. Metadata is a structured scheme to describe properties/attributes of digital resources. There are two categories of metadata standards – generic (applicable to all sorts of resources) and domain-specific (applicable to a certain group of resources). Dublin Core is the global defacto standard for general-purpose metadata, and there is an array of domain-specific metadata schemas for different document types like educational resources, cartographic materials and so on. Cataloguing and metadata encoding share lots of similarities as the objective of both group of activities are same – discovery of resources. There are various rule sets for encoding of DC metadata in HTML, XHTML, XML and RDF/XML environment. These are based on RFCs as proposed by W3C. Almost all digital library systems and software support both simple and qualified Dublin Core metadata on the basis of DCMI Abstract Model. A few digital library software are also

have support for domain-specific metadata schemas apart from Dublin Core e.g. Greenstone extends inbuilt support for AGLS and GILS in addition to simple and qualified Dublin Core.

6.7. CHECK YOUR PROGRESS

Q. 1: DC Date element is available for	
A	Simple Dublin Core only
B	Qualified Dublin Core only
C	Both
D	None of the above

Q. 2: DC Audience element is available for	
A	Simple Dublin Core only
B	Qualified Dublin Core only
C	Both
D	None of the above

Q. 3: Which of the following is true?	
A	Qualified DC uses additional elements
B	Qualified DC uses element refinement
C	Only A is true
D	Both A and B are true

Q. 4: Match the followings:			
a	DC.Contributor	i	entity that made a contribution
b	Element Refinement	ii	aid in the interpretation of an element value
c	Encoding Schemes	iii	metadata that is maintained and stored within the object it describes
d	Embedded metadata	iv	make the meaning of an element specific
Code:			
A	a – i; b – ii; c – iv; d - iii		
B	a – i; b – ii; c – iii; d - iv		
C	a – iii; b – ii; c – iv; d - i		
D	a – i; b – iv; c – ii; d - iii		

Q. 5: Match the followings:			
a	ISO 3166	i	Rules for encoding dates and times
b	W3C-DTF	ii	Codes for the representation of names of languages.
c	ISO 639-2	iii	Standard for encoding DC elements in HTML
d	RFC: 2731	iv	Codes for the representation of names of countries
Code:			
A	a – iv; b – i; c – ii; d - iii		
B	a – i; b – ii; c – iv; d - iii		
C	a – iv; b – iii; c – i; d - ii		

D	a – iv; b – iii; c – ii; d - i

Answer keys: Q 1: C ; Q. 2: B ; Q. 3: D ; Q. 4: D ; Q. 5: A

6.8 KEYWORDS

Authority record: record that registers the preferred form of a personal or corporate name, geographic region or subject term.

Crosswalk: maps the relationships and equivalences between two or more metadata schemes. Crosswalks or metadata mapping support searching across heterogeneous databases.

DCMI term: a DCMI element, a DCMI qualifier or term from a DCMI-maintained controlled vocabulary.

Descriptive metadata: metadata that supports the discovery of an object.

Document Type Definition (DTD): a formal description of the components of a specific document or class of documents for machine processing (parsing) of documents expressed in SGML or XML.

Dublin Core Metadata Initiative: the body responsible for the ongoing maintenance of Dublin Core, hosted by the OCLC Online Computer Library Center, Inc.

Electronic information resource: information resource that is maintained in electronic, or computerized format, and may be accessed, searched and retrieved via electronic networks.

Embedded metadata: metadata that is maintained and stored within the object it describes; the opposite of stand-alone metadata.

Interoperability: ability of different types of computers, networks, operating systems, and applications to work together effectively, without prior communication, in order to exchange information in a useful and meaningful manner.

Namespace: DCMI namespace is a collection of DCMI terms. Each DCMI namespace is identified by a URI.

RDF (Resource Description Framework): is a standard model for web-based data interchange that supports a robust flexible architecture for processing metadata on the Internet.

Record: an instantiation of a description set, created according to one of the DCMI encoding guidelines (for example, XHTML meta tags, XML and RDF/XML).

Request for Comment (RFC): the process of establishing a standard on the Internet by W3C.

Resource: anything that might be identified. Familiar examples include an electronic document, an image, a service (for example, "today's weather report for Los Angeles"), and a collection of other resources. Not all resources are network "retrievable"; for example, human beings, corporations, concepts and bound books in a library can also be considered resources.

Statement: an instantiation of a property-value pair made up of a property URI (a URI that identifies a property) and a value surrogate.

Syntax encoding scheme: a set of strings and an associated set of rules that describe a mapping between that set of strings and a set of resources. The mapping rules may define how the string is structured (for example DCMI Box) or they may simply enumerate all the strings and the corresponding resources (for example ISO 3166).

Term: a property (element), class, vocabulary encoding scheme, or syntax encoding scheme.

6.9 QUESTIONS FOR SELF STUDY

1) Point out the similarities and differences between cataloguing and metadata encoding.
2) Discuss the need of metadata in a digital information retrieval system.
3) What is qualified Dublin Core? How does it differ from that of simplified Dublin Core?
4) Show the RDF/XML encoding of your university website.
5) Write a short note on DCMI Abstract model.

6.10 REFERENCES

American Library Association. (1999). Task Force on Metadata Summary Report, Retrieved March 31, 2003, from http://www.libraries.psu.edu/tas/jca/ccda/tf-meta3.htmlJune 1999

Berners-Lee, T. (1993). Metadata architecture. Retrieved March 13, 2008, from http://www.w3.org/

DCMI. (1996). Dublin core qualifier. Retrieved May 23, 2006, from http://dublincore.org/documents/

DCMI. (2022). Dublin core projects. Retrieved November 13, 2022, from https://www.dublincore.org/specifications/dublin-core/

DCMI. (2022). Namespace policy for the DCMI. Retrieved November 13, 2022, from https://www.dublincore.org/specifications/dublin-core/dcmi-namespace/

Hadge, G. (2001). Metadata made Simpler. Bethesda: NISO Press. Retrieved May 05, 2003, from http://www.niso.org/news/Metadata-simpler.pdf

Heery (R). Review of metadata formats. Program, 30(4);1996, 345-373

IFLA. (2002). Digital libraries and metadata resources. Retrieved March 28, 2003, from http://www.ifla.org/II/metadata.html

Kunze J. (1999). Encoding Dublin core metadata in HTML. Network Working Group, Internet Society, RFC-2731; December 1999. Retrieved March 13, 2008, from http://www.w3.org/TR/

Levan R. (1998). Dublin core and Z39.50. Retrieved March 13, 2008, from http://dublincore.org/documents/

Madalli D P. (2001). A DC approach to display retrieved web resources. Paper AA, DRTC Workshop on Multimedia and Internet Technologies, February 26-27, 2001

Miller P. (2001). Metadata for the masses. Retrieved March 13, 2008, from http://www.ukoln.ac.uk/

Mukhopadhyay, P.S. and Sarkhel J.K. (2002). Towards a semantic web: a metadata approach. Proceedings of the National Seminar on Information Management in Electronic Libraries (ImeL), Kharagpur, 2002. Indian Institute of Technology, Kharagpur. 2002. p. 123-136

Mukhopadhyay, P. (2015). Interoperability and retrieval. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000232199

Murtha, B. (ed.). (2002). Introduction to metadata: Pathways to digital information version 2.0. Retrieved April 04, 2003, from http://www.getty.edu/research/institute/standars/intrometadata.html

UKOLN. (2002). Metadata page. Retrieved March 13, 2003, from http://www.ukoln.ac.uk/metadata/intro.html

Weibel S and Hakala J. (1998). DC-5: the Helsinki metadata workshop. D-LIB Magazine, 4(2);1998. Retrieved March 13, 2008, from http://dlib.org/

Weibel S. (1997). The 4th Dublin core metadata workshop report. D-LIB Magazine, 3(6);1997. Retrieved March 13, 2008, from http://dlib.org/

Weible S and Koch T. (2000). The Dublin core metadata initiative: mission, current activities and future directions. D-LIB Magazine, 6(12);2000. Retrieved March 13, 2008, from http://www.dlib.org/

World Wide Web Consortium. (2003). Metadata and resource description. Retrieved March 31, 2003, from http://www.w3.org/Metadata

UNIT-7: DIGITAL LIBRARY: ELEMENTS & ARCHITECTURE

Structure

7.0 OBJECTIVES

After going through this Unit you will be able to:

- ❖ Asses the importance of the building blocks of a typical digital library system;
- ❖ Explore nature of digital information resources and issues related to ipr;
- ❖ Trace the path of development architectural models of digital library systems;
- ❖ Identify the role of metadata in a digital library system; and
- ❖ Retrieve metadata from large-scale metadata repositories like crossref, unpaywall and like.

7.1 INTRODUCTION

The Internet in general and the Web in particular provide us with a comprehensive multimedia-driven platform for knowledge communication. Digital libraries are major application entities of the Internet and Web technologies. These are considered next-generation library services. In simple words, digital libraries are managed collections of digital objects. These entities enable the creation, organization, maintenance, management, access to, sharing, and preservation of digital knowledge-bearing objects or document collections. Digital libraries are powered by any[6]. It means that these entities are accessible from anywhere, by anyone, at anytime, in any format, and in any language. Digital libraries are being created today by many institutes and agencies for different target groups and in diverse fields like agriculture, cultural heritage, education, health, governance, science, social sciences, social development, etc. The availability of Free/Libre Open Source Software (FLOSS) based digital library software packages, the application of open standards and the sharing of domain knowledge through wikis, blogs, etc., help in designing digital libraries even in developing parts of the world. As a result, the designing and development of digital libraries in India has now become an attractive and feasible proposition for library and information professionals across the country. Now the question comes: what are the advantages of digital libraries? There are some obvious benefits to digital libraries over automated library systems. Some of the key benefits of digital libraries are:

- Traditional libraries are associated with the organization and provision of access to

physical materials like print-on-paper publications;

- Automated library systems are providing improved access to their collections, but online access facilities are limited to the computerized library catalogue (OPAC); and
- Digital libraries differ significantly from such libraries because these entities facilitate online access to and work with digital versions of full-text resources in a multimedia-driven environment.

Many digital libraries also provide access to other multimedia content, like audio and video.

7.2 DIGITAL LIBRARY SYSTEM: AN OVERVIEW

The world is going digital. Libraries are no exception. The availability of distributed information systems like the Web and rapid improvement in database-oriented storage, retrieval, and dissemination of information resources have initiated fundamental changes in almost every aspect of library services across the globe. One of the most significant contributions of Web technology has been the creation of next-generation library automation – digital libraries. Digital libraries, as a new brand of information entity, allow users to access digital information resources in full-text from anywhere at any time. In simple words, a digital library is a managed collection of digital objects.

7.2.1 Definitional scopes

Borgman (2000a) identified two broad groups of definitions on the basis of his analysis (1999) of several definitions of the concept of digital library.

Group I definitions

These definitions are given by digital library researchers. These definitions identify and focus attention on the digital library community and research problems. Wellman et al. (1996) see a digital library of the future in which software agents use principles of artificial intelligence (AI) to perform "monitoring, management, and allocation of services and resources." Nurnberg et al. (1995) identified different aspects of digital libraries for different domains of applications. For example:

- From an information retrieval point of view, it is a large database;
- For people who work on hypertext technology, it is one particular application of hypertext methods;

- For those working in wide-area information delivery, it is an application of the Web; and
- For library science, it is another step in the continuing automation of libraries that began over 25 years ago.

Group-II definitions

These definitions are given by library professionals focusing on real-life challenges involved in transforming information organization and services. Donald J. Waters reported (http://www.clir.org/pubs/) that the definition given by the Digital Library Federation (DLF) is one of the most inclusive definitions in this group. DLF crafted the following definition:

"*Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities."*

Another interesting definition was provided by the Department of Library and Information Science at the University of California, Los Angeles. It says a digital library is "a set of resources and associated technical capabilities ..." and "virtual communities in which individuals and groups interact with data, information, and knowledge resources and systems."

Professor Stephen P. Harter, School of Library and Information Science, Indiana University classified propertied of digital libraries into three groups on the basis of definitions given by stalwarts (Harter, 1996). The grouping is given in Table 7.1.

Table 7.1: Analysis of definitions of digital libraries (source: Harter, 1996)

NARROW VIEW **(based on traditional library)**	**BROADER VIEW** **(concept of Hybrid library)**	**BROADEST VIEW** **(based on digital library)**

Objects are information resources	Most of the objects are information resources	Objects can be anything
Objects are selected on the basis of quality	Some of the objects are selected on the basis of quality	No quality control; no entry barriers
Objects are located in a physical place	Objects are located in a logical place (may be distributed)	Objects are not located in a physical or logical place
objects are organized	hybrid mode of organization	no organization
Objects are subjected to authority control	Some aspects of authority control are present	No authority control
Objects are fixed (do not change)	Objects change in a standardized way	Objects are fluid (can change and mutate at any time)
Objects are permanent (do not disappear)	Disappearance of objects is controlled	Objects are transient (can disappear at any time)
Authorship is an important concept	Concept of author is weakened	No concept of author
Access to objects is limited to specific classes of users	Access to some objects is limited to specific users	Access to everything by everyone
Services such as personalized reference assistance are offered	Both human assisted and software driven services	Services are performed by computer software (ai)
Human specialists (called librarians, etc.) Can be found	Librarian and software agent	There are no librarians
There exist well-defined user groups	Some classes of objects have associated user groups	There are no defined user groups (or alternatively, infinitely many of them)

7.2.2 Digital information resources

A document gives information or facts. Documents are records of human knowledge, observations, and thoughts, available in many forms and formats. A document in any form can be a source of information. Information sources have two components: conduit (the physical facilities used for gathering, storing, processing, and distributing information) and content (the information sources and elements). Information sources become information resources when they are organized and institutionalized in some way, and can thus be reused. Digital information sources are no exception. These need to be identified, evaluated,

organized, archived, and disseminated systematically. In his seminal book on digital libraries, Arms (2000) identified some of the potential benefits of digital information sources. These are as follows:

- **The digital information infrastructure brings the library to the user.**
  Digital information sources, when available over a distributed computer network, can be accessed from anywhere. It overcomes one of the fundamental barriers of information communication i.e. physical space.
- **The digital information infrastructure supports sophisticated searching and browsing.**
  Digital information sources can be searched through various sophisticated search operators like Boolean, relational, and positional operators. Therefore, the primary functions of information retrieval (such as finding, identifying, navigating, and obtaining) are ensured and enhanced in the digital information infrastructure.
- **Digital information can be shared.**
  A single digital information source can be accessed by many users from different places at any given point of time.
- **Digital information sources are easier to keep current.**
  Currency is an important factor for information sources. Digital formats help in keeping information current. Addition, deletion, and modification are less problematic when the definitive version is in digital format and stored on a central server connected to a distributed network.
- **Digital information sources are always available.**
  When organized and disseminated over a distributed network, digital information resources are powered by (any)^6, i.e., any user can access any digital information source from anywhere at any time in any format and in any language.
- **New forms of information have become possible.**
  Digital information infrastructure allows the generation of interactive and multimedia-driven information sources to help in creating a participative global platform for knowledge generation and utilization.

In view of the above characteristics of digital information sources, we may now go for a comparative study of traditional information sources (i.e., printed information sources) and digital information sources (Table 7.2).

Table 7.2: Traditional vs. digital information resources

Traditional information source	Digital information source
Here forms and contents are inseparable	Here contents can be detached from the form
One information source can be accessed by only one user at any given point of time	Digital information source available over a distributed network can be accessed by many users simultaneously
Supports limited search facilities	Supports sophisticated search operators
Access is limited by time and space	Access is independent of time and space
Only metadata (related to source) is searchable	Both metadata and full-text contents are searchable

7.2.3 Characteristic features

You already have an idea of the scopes and advantages of digital libraries in comparison with traditional and automated library systems. However, a quick summary of the essential features of digital libraries may help you understand the services of a typical digital library system. The major benefits are as follows:

- **Digital libraries are available in 24X7 modes.**

  Digital libraries are web-based entities and, therefore, can be accessed by anyone from anywhere at anytime. Digital libraries make information services free from two fundamental barriers of communication, i.e., time and space. In the case of poor network connectivity, digital collections can also be delivered in offline modes (CD-ROM and DVD-ROM) to users.

- **Digital libraries provide improved access.**

  In a digital library setup, users can access document description data sets and source documents. Digital libraries support sophisticated search operators for searching metadata and full-text resources, including features like relevance ranking, hierarchical document browsing, and weighted term searching.

- **Digital libraries support wider access.**

  A digital library system supports access to the same digital resource by many users at

a given point of time. These information entities thereby meet the requirements of a wider user base.

- **Digital libraries facilitate new forms of access.**
  Digital libraries can provide access to digital resources available in different forms and formats (animation, graphical, audio and video formats) and can support post-processing of information resources (e.g. conversion of a pdf file format to html format). New accessibility technologies also help physically disadvantaged users access and use digital collections.
- **Digital libraries allow improved information sharing.**
  Digital libraries, right from the early days of development, support shared understandings, i.e., standards. Almost all the digital libraries are based on internationally agreed upon standards for document description, interoperability and data encoding, e.g. DCMES, OAI/PMH, METS, MODS etc.
- **Digital libraries support improved preservation.**
  Digital libraries can help us in the preservation of special and rare documents and artefacts by providing access to digital versions of these entities.

Keeping in view all the above mentioned benefits of a typical digital library system, the services may be grouped as follows:

- **Access to diverse digital information resources is possible.**
  Users can access remotely located digital collections through browsing and searching.
- **Integration of related documents is an added advantage.**
  A digital library system can provide access to the source documents as well as cited resources can be hyperlinked to help users navigate related documents.
- **Downloading of resources in different formats is allowed.**
  Digital libraries may provide facilities for post-processing of retrieved documents in different forms and formats.
- **Reference services are enhanced.**
  Reference services (virtual reference services) can be integrated with digital library systems seamlessly to provide users an interface for real-time reference services (see Internet Public Library http://www.ipl.org/) .
- **Personal Information Environment (PIE) is possible.**

Digital libraries may provide facilities to each user in developing their own place to store references, digital resources, lists of favourite items, saving of search query options, listing of RSS feeds, rating of retrieved resources, etc.

7.2.4 Comparison of library systems

Library automation activities address two major issues – library housekeeping operations and access to library resources. An automated library system has cataloguing data in digital format but source documents are mostly available in print formats. In a digital library setup both metadata (document description data) and documents are available in digital format. The other major differences are noted in Table 7.3.

Table 7.3: Traditional vs. digital information resources

Automated library system	**Digital library system**
Only metadata(cataloguing data) is finely searchable	Both metadata set and full-text resources are finely searchable
Provides document description data set, not documents.	Provides document description data set and source documents
Based on Z 39.50 standard for cross-system catalogue search/retrieve	Based on OAI/PMH protocol for metadata harvesting
Supports standard bibliographic formats (MARC 21, CCF) for document description	Supports generic and domain-specific metadata schemas (e.g. Dublin Core, LOM, GILS etc) for resource description
Processes global resources for local users	Processes global and local resources for local and global users
Generally follows centralized processing – – distributed access architecture	Generally follows distributed processing – distributed access architecture

7.3 ELEMENTS OF DIGITAL LIBRARIES

Digital libraries enable the creation, organization, maintenance, management, access to, sharing and preservation of digital document collections. Each digital library system depends on a few functional components to provide services as discussed in section 7.2.3.

7.3.1 Functional components

The common functional components of a typical digital library system are: 1) document selection, acquisition, and resource optimization; 2) organization, persistent identification and uploading; 3) indexing and storage; 4) repository and archiving; 5) search and retrieval; 6) Web-enabled access to digital library system and services; and 7) network connectivity.

- **Selection, acquisition, and resource optimization**

  This component supports activities like selection of documents to be included, digitization and/or conversion of selected documents to appropriate digital form, and optimization of born-digital resources available from different locations (servers/URLs).

- **Resource organization**

  This component includes three major processes: Choosing a metadata schema for describing digital resources; determining a metadata encoding standard; and assigning metadata to each document added to the collection.4) persistent identifiers for resources (for example, DOI) and contributors (for example, ORCID)

- **Indexing and storage**

  This component is responsible for 1) the selection of an indexing process; 2) the selection of an indexing tool (many open source text retrieval engines are in use by digital library software like MGPP, Apache-Lucene, Apache-Solr etc.); and 3) the storage of documents and related metadata, for efficient search and retrieval.

- **Repository**

  This core component of a digital library archives document objects, metadata, and indexes for perpetual access.

- **Search and retrieval**

  This component of the digital library acts as a front-end for end-users to browse, search, retrieve, and view the digital resources.

- **The Digital library website**

  This component manages URLs, DNS (Domain Name System), external IP addresses, Web servers, mail servers, and name servers that are required for secured public-domain access of the digital library system.

- **Network connectivity**

  This component offers backbone connectivity for the digital library system, i.e., a dedicated connection to the intranet and/or Internet.

- **IPR issues**

  Digital libraries may also have rights management and e-commerce components to handle security and payment aspects.

Naturally, the issues and challenges in creating digital libraries centre on the successful implementation and seamless integration of the above-stated functional components. The activities related to the application and integration of these functional components may be divided into seven logical groups (Table 7.4).

Table 7.4: Traditional vs. digital information resources

Sl	Group	Focal Area	Major components
1	Group I	Technical architecture	Operating system, Programming environment, Backend RDBMS, Web server software
2	Group II	Copyright/Rights Management and Preservation	Copyright policy, License selection, Digital Rights Management (DRM), Open Access 9Creative Commons)
3	Group III	Digitization	Equipment's, Process, Curtion, Format, Long-term storage, Archiving policies
4	Group IV	Metadata	Generic metadata schema (Dublin Core), Domain-specific metadata schemas (LOM, VRA-Core, FGDC etc.)
5	Group V	Naming, identifiers, and persistence	Digital Object Identifier (DOI), ORCID, AuthorClaim etc for contributors, URI for knowledge organization systems like LCSH, DDC
6	Group VI	Building digital collections	Steps and process associated with a digital library software (like Dspace, GSDL, EPrints)
7	Group VII	User interface and information retrieval	Intuitive user interface design/customization for effective retrieval

7.3.2 5S framework

The important features of digital libraries, as reflected in the above mentioned functional

elements, are as follows:

i. Digital libraries can contain anything from simple text to streaming video;
ii. Digital libraries operate in cyberspace and thereby reduce the requirements of physical space;
iii. Digital libraries are powered by (any)[6] - these entities are accessible from anywhere at any time, and their contents are available in any format and language;
iv. Distributed processing, storage, and access are all supported by digital libraries;
v. Digital libraries provide simultaneous access to same information resources by any user; and
vi. Digital libraries support sophisticated search mechanisms (e.g. fuzzy AND) and search operators (e.g. Boolean, relational, and positional operators).

Fox and Marchionini (1999) identified four dimensions of digital libraries. These are as follows:

Community	socio-legal, political, and cultural issues related to the development of digital libraries (including target-audience selection);
Contents	management of all possible information resource types, forms, and formats (born digital and others);
Technology	networking, information storage and retrieval, search mechanisms, user interface, digital archiving, multimedia support, and multilingual content management issues; and
Services	access and download, reference services, alerting services, feedback mechanisms; personal information environment, download statistics, online help; etc.

Borgman (2000) identified four essential components of a typical digital library system: (i) a collection of services; (ii) an architecture; (iii) a collection of digital information resources (including multimedia objects); and (iv) a collection of tools for locating, accessing, and obtaining available resources. Gonçalves, M. A., Fox, E. A., Watson, L. T., and Kipp, N. A in 2003 proposed the famous **5S framework** for digital library systems (http://doi.acm.org/10.1145/984321.984325) as a formal model for digital library systems. The 5S stands for streams, structures, spaces, scenarios, and societies (5S). The Streams refer

to the primary bit-streams like text, audio, video, and multimedia content that build digital libraries, or in simple words, the content of a digital library system. The Structures refer to the organizational processes (like metadata, catalogues, indexes etc.) and navigational tools (like hypertext, hypermedia, hyperlinking etc.) or in simple words, organizational aspects of a digital library system. The Spaces refer to the front-end of a digital library system, such as the user interface, backend index, retrieval engine in use, retrieval models, and so on, or, to put it another way, the presentation layer of a digital library system. The Scenarios has reference to the service components of a digital library system including the integration and interoperability capabilities of such a system. Finally, the Societies define the roles of system administrator, collection administrator, reviewers,submitters, metadata editors along with the relationship building measures with the users of a digital library system.

7.3.3 Socio-legal issues

Libraries are social organizations. Digital libraries are no exception. Naturally, the design and development of digital libraries requires addressing different socio-cultural, socio-economic, and socio-legal issues.

a) Social issues

On this front, the greatest challenge in implementing digital libraries is the "digital divide." The "digital divide" is a disparity caused by differences in access to ICT resources. Web technologies have made life easier for many and also, at the same time, have increased the gap between the information rich and the information poor. The Digital Divide network database (see http://www.digitaldividenetwork.org) shows that only 6% of the global population has access to online resources. Although the digital divide is a global observable fact, the disparities in online access in the developing world are amazing. The BBC reported in 1999 that more than 80% of the world population had never heard a dial tone, and as per the UNESCO report, the Internet is accessed by only 5% of the world population. Apart from connectivity, Indian libraries face major constraints such as shrinking library budgets, a lack of capital grants, poor bandwidth, a lack of trained manpower, a low rate of information literacy, and the limitation of library automation in a few resourceful institutions. Under such circumstances, one may feel that a digital library system is a distant dream in India. But we should remember that every threat can be converted into an opportunity with proper planning and hard work. Digital library initiatives in India can be a success story if we are able to

address the following issues:

- large-scale information literacy campaigns and programmes
- UNESCO and other philanthropic organizations have provided financial assistance.
- The design and development of Unicode-compliant Indic script-based resources, interfaces, and search mechanisms;
- In developing digital libraries, free and open source software (FLOSS) and open standards are used.
- The design and development of institutional digital repositories (open access repositories) for every Indian university and research institute;
- Development of a country-wide ETD movement;
- organization and extensive use of free online reference sources;
- design of appropriate information services like subject gateways, academic subject directories, and virtual reference services.

b) Economic issues

Application of ICT tools can help in achieving improvements in publishing productivity, better participation and interaction, and cost reduction in publishing. Digital libraries may be instrumental in developing a new model for digital publishing systems. However, the design and development of a digital library system is an expensive venture. The stack-holders of digital libraries should remember that it is not a one-time capital expenditure. There must be provision for recurring expenditure for digitization, hardware maintenance, software maintenance and updation, manpower development, collection development, metadata encoding, etc. It may be easy to obtain funding for discrete projects, but recurrent funds for ongoing activities are harder to come by, and the commitment to maintain digital content for the long run needs to be planned and calculated alongside the initial costs of conversion. Information technology infrastructure and personnel include two critical resources required for digital library projects. Hardware requirements include a server computer for hosting the collection, desktop computers, digitization equipment, network connectivity, and other equipment. Digital library software is another critical technology component. Options include: open source free digital library software, commercial digital library software, and in-house software development. Each of these has advantages and disadvantages. Personnel comprise the most important resource for the digital library apart from technology. The actual number of personnel required for a digital library project depends on the type and volume of tasks to be carried out.

c) Legal issues

Digital libraries are global information entities. They support cross-border access to information resources. Here lies the root of the legal problems facing digital libraries. Each country has its own legal (copyright in particular) system. Some issues concerned with the development, management, and use of digital libraries are legal in one country and may be illegal in another country (or even in another state of the same country). Digital library initiatives should give attention to the following legal matters:

i) IPR

Intellectual Property Rights (IPR) are an important area in digital library design and development. In his paper published in DLib magazine, Petersen (1999) raised a few IPR issues related to digital libraries. These are—(i) to explore the availability of digital resources in the public domain; (ii) to seek permission/license associated with a resource; (iii) to determine ownership/authorship of digital resources; and (iv) to exercise "fair use" and other copyright exceptions judiciously.

ii) License

Digital resources should be attached with a valid license in an accessible format. Policy makers of digital library systems can use their own license or can adopt existing licenses like Creative Commons (see http://www.creativecommons.org). The Creative Commons initiative was created to make it simple for creators who want their works to be freely accessible to anyone. An author can declare that his work is provided under the aegis of one of the various Creative Commons licenses (different Creative Commons licenses have different provisions attached). Authors can decide whether to require ascription or to permit modifications of their work, for example. Each variety of creative commons license has an associated icon, which can be attached to the work and serves as a shorthand emblem to advise users what may and may not be done with the work.

iii) Authenticity

Ensuring authenticity of digital information resources is a challenging area of digital library development. Administrators of a digital library system should consider the following factors (Bearman & Trant, 1998) in terms of authenticity, (i) digital information resources must be unaltered from the original; (ii) the purpose and target audience of digital information resources must be clear; and (iii) digital information resource representations/surrogates must be standard and transparent. The authenticity of digital information resources may be judged by public methods (certificate, registration, checksum method, metadata encoding), secret

methods (stegonography, digital watermarking, digital signature) and functionally dependent methods (encryption, SSL-protocol, cryptolopes). A series of papers on authenticity (Gladney, 1997; 1998, Gladney et al, 1997) are published in DLib magazine.

iv) Privacy

Digital technology can track users, online behavior, search history, and usage patterns easily. Naturally, digital libraries may use these technologies to collect user-related data for improving services. This may sometimes lead to encroachment on the privacy of users. As a result, administrators of digital libraries must be judicious in applying such technologies to the system. A set of recommendations on privacy in digital library services is available from Resource – NDLI (https://ndl.iitkgp.ac.in/privacy-policy).

7.4 BUILDING BLOCKS OF DIGITAL LIBRARY SYSTEMS

Long term archiving and preservation of recorded human knowledge objects is an important area of consideration from the perspective of library professionals. Digital library system offers an achievable solution in this regard. The major problems related with long term digital preservation are of three types –

- Need of content reorganization for digital preservation;
- Problem related with high degree of technology obsolescence; and
- Problem related with high degree of content format obsolescence.

Day in his seminal paper (2001) reported an array of factors that need to be taken care of archiving knowledge objects through digital library system –

Preservation strategy:	A good strategy is required for digital preservation.
Communication:	A good communication system must be there for exchange of views among data creators and system managers.
IPR issue:	Digital rights management needs to be addressed.
Collection management:	Preservation must be supported by sound collection development and management policy.
Metadata:	Standards metadata format must be supported.

Archiving:	Hardware, software, and data formats must be selected judiciously.
Collaboration:	Must follow the best-practice guidelines.
Staff:	Training and participation of staff is necessary.

7.5 ARCHITECTURE OF DIGITAL LIBRARY SYSTEMS

According to OAIS (Open Archival Information System) reference model (oais.info), a typical digital library system is a combination of seven interlinked functions. The scopes of these essential architectural components are given in Table 7.5.

Table 7.5: Architectural components

Functional component	Scope
Ingest	Receives SIP (Submission Information Package) and prepares Archival Information Package (AIP).
Archival Storage	Receives Archival Information Package (AIP) and stores in the system permanently.
Data Management	Receives query and provides responses, along side reports responses as given.
Administration	Manages content, people, content alongside software, hardware, backend DBNS, content negotiating tc.
Preservation	Deals with content format, content migration, content templating and other aspects of content management.
Access	Deals with intuitive user interface, content retrieval, access control and embergo (for protected resources), and handles user-related information services.

The Administration functional component is playing a pivotal role in the architectural diagram as proposed by OAIS (see Fig. 7.1).

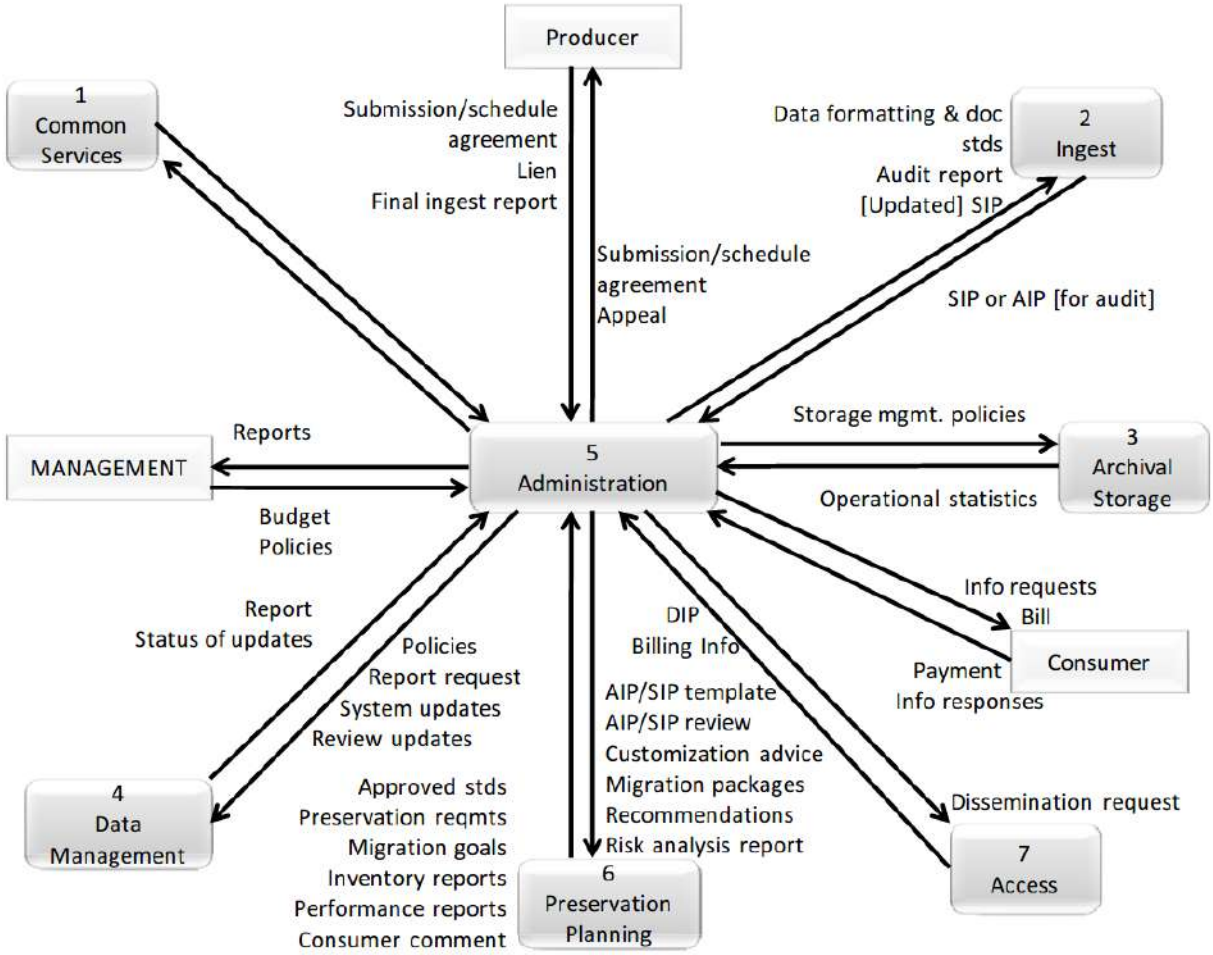

Fig 7.1:Administration of digital content management system (Source: OAIS Reference Model, June 2012)

7.6 IDENTIFIERS IN DIGITAL LIBRARY SYSTEMS

Resource identification on a global scale is required to satisfy five fundamental functions of a distributed digital information retrieval system – i) to find a resource ii) to identify a resource iii) to locate a resource iv) to navigate and v) to obtain a resource. Standard development Organization or SDOs (e.g. ISO, ANSI/NISO, BSI, BIS) including professional associations/ institutions (e.g. IFLA, ALA, British Library, Library of Congress etc.) from different disciplines already proposed and formalized various naming and addressing schemes to identify and locate digital information resources available in the distributed information system. World Wide Web or WWW is based on object modeling. Digital libraries are using WWW as networking platform to deliver services. Therefore, digital library systems like WWW need to adopt a method to identify digital objects.

Resource identifiers are used mainly for the following purposes (as identified by William Arms, 2000) –

- To refer/link objects in databases (e.g. Catalogue database / collection identifier);
- To store and access digital information resources (object identifier); jects
- To link contributors of knowledge objects (person identifiers);
- To provide access management for digital library system (unique identification); and
- To archive digital information resources for the long term (perpetual access).

Internet allocates a numeric identifier for host computers or servers, called IP address. Domain name system is textual representation of IP address. These two schemes uniquely identify servers (or any connected device) in the network. Apart from these interoperability standards for machines, standards like URL, URN, CNRI handle, DOI etc. are in use to achieve interoperability in resource access and exchange. Likewise ORCID, AuthorClaim,RePEc ID etc identify authors and researchers uniquely.

7.7 METADATA REPOSITORIES

Metadata plays an important role in many aspects of digital libraries but is especially important for interoperability. Metadata is often divided into three categories: descriptive metadata is used for bibliographic purposes and for searching and retrieval; structural metadata relates different objects and parts of objects to each other; and administrative metadata is used to manage collections, including access controls. For interoperability, some of this metadata must be exchanged between computers. This requires agreement on the names given to the metadata fields, the format used to encode them, and at least some agreement on semantics. As a trivial example of the importance of semantics, there is little value in having a metadata field called "date" if one collection uses the field for the date when an object was created and another uses it for the date when it was added to the collection. Fortunately, there is an array of metadata repositories from which API-based access to metadata sets is freely available under the Open Data Commons Open Database License (ODbL), like CrossRef (metadata), Unpaywall (metadata and open access status), Open Citation Corpus (OCC - citation profile of a knolwedge object), and so on. The following example shows how metadata elements against a unique resource identifier like a DOI may be obtained:

https://api.crossref.org/works/10.1177/0165551506078404 (10.1177/0165551506078404 is DOI)

(you may check the result by pasting this URL in the address bar of a web browser like Firefoc or Google Chrome).

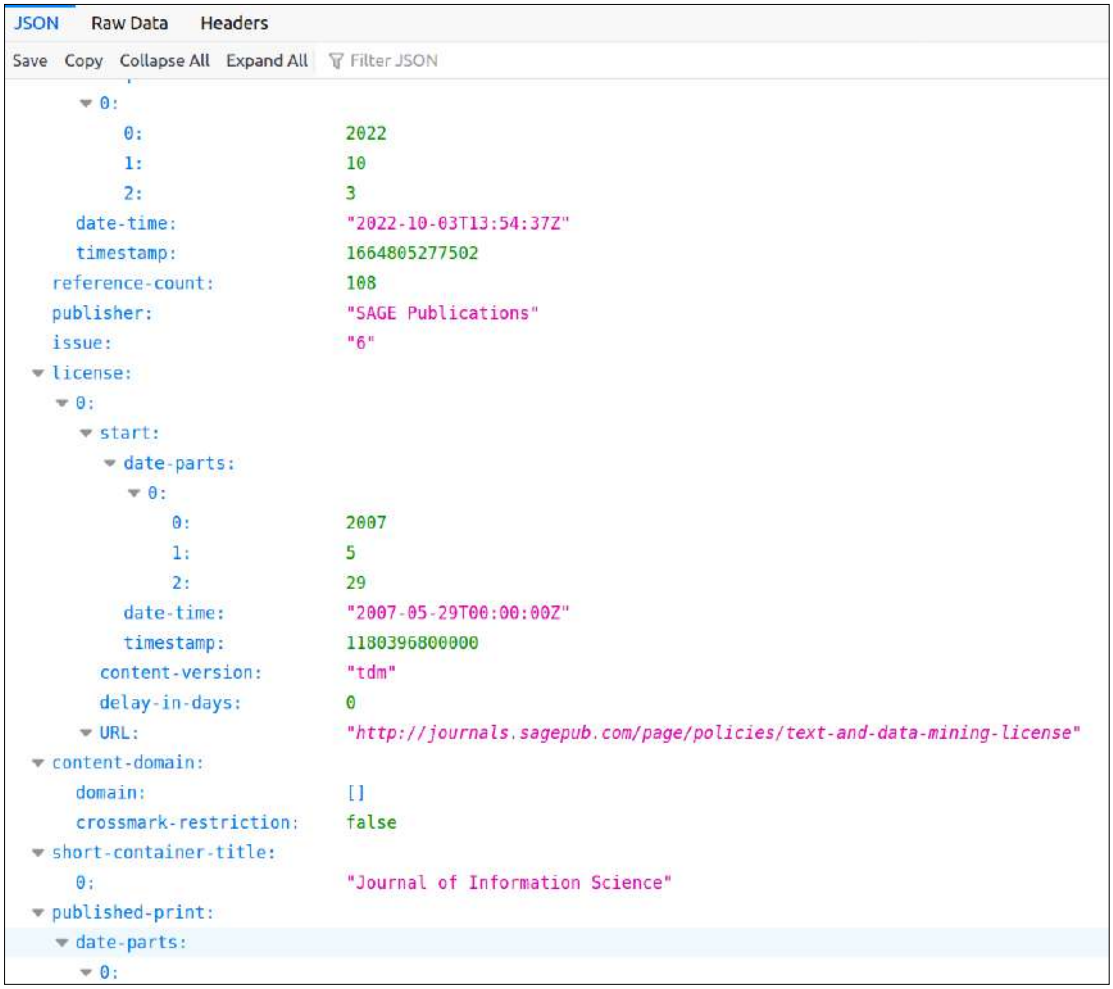

Fig.7.2: API-based access to CrossRef metadata repository

7.8 SUMMARY

A digital library system is a product of a multidisciplinary domain. Library and information science is providing the theoretical backbone for this emerging area of study, and ICT is providing the infrastructural support. Digital libraries help library systems go beyond the four walls of libraries. Users can access full-text resources virtually from anywhere. Apart from technological issues like hardware, software, data formats, content formats, metadata, archival format, etc., digital libraries need to be viewed from different sociological perspectives. In this unit, we studied the definition, scope, and elements of a digital library system, how a digital library differs from that of an automated library system, what is the relationship between a digital library, metadata, and resource identifiers,

how a digital library system is helping long-term preservation of knowledge objects, what are the typical services of a digital library system, and also noted the sociological issues related to the designing and development of digital library systems keeping in mind the needs of a developing country like India.

7.9 CHECK YOUR PROGRESS

Q. 1: What is DRM?	
A	Digital Resource Management
B	Digital Rights Management
C	Digital Regulation Management
D	Digital Restoration Management

Q. 2: Which of the following is related to contributor identification?	
A	ORCHID
B	ISO-2709
C	SWORD
D	CNRI Handle

Q. 3: Who is associated with the famos 5S framework of digital library system?	
A	Ian Witten et al.
B	G G Chowdhury et al.
C	Edward Fox et al.
D	None of the above

Q. 4: Match the followings:			
a	INGEST	i	Metadata-level interoperability
b	CreativeCommons	ii	Open Access Licensing
c	ODbL	iii	Open Database Licensing
d	OAI/PMH	iv	Submission management in DL system
Code:			

A	a – i; b – ii; c – iii; d - iv
B	a – i; b – ii; c – iv; d - iii
C	a – iii; b – ii; c – iv; d - i
D	a – iv; b – ii; c – iii; d - i

Q. 5: Match the followings:			
a	FLOSS	i	Open source software
b	MADS	ii	Authority data interoperability
c	Unpaywall	iii	Open Access status of an object
d	OCC	iv	Citation profile of an object
Code:			
A	a – i; b – ii; c – iii; d - iv		
B	a – i; b – ii; c – iv; d - iii		
C	a – iv; b – iii; c – i; d - ii		
D	a – iv; b – iii; c – ii; d - i		

Answer keys: Q 1: B ; Q. 2: A ; Q. 3: C ; Q. 4: D ; Q. 5: A

7.10 KEYWORDS

- **AIP – Archival Information Package:** an information package for storing content.
- **Copyright:** a component of IPR that protects original authorship(s).
- **Creative Commons:** an open access licensing system.
- **CrossRef**: an open access metadata repository for an array of publishers.
- **DSpace:** an open source digital library software.
- **OCC:** Open Citation Corpus – provides citation profile on the basis of DOI or ISBN against ODbL license.
- **OAI-ORE (Open Archives Initiative – Object Reuse and Exchange):** a standard for transfering digital objects along with metadata.
- **ORCID (Open Researcher & Contributor ID):** a standard for author identifier.
- **RDF (Resource Description Framework)**: a greater metadata architecture for semantic interoperability.

- **SIP – Submission Information Package:** an Information Package in a digital library system for ingesting.
- **Unpaywall**: a digital metadata repository that provides access to open access status of a digital object against API call on the basis of ODbL.

7.11 QUESTIONS FOR SELF STUDY

1. Define "digital library." Compare digital and automated library systems.
2. Digital libraries are helping in the democratization of knowledge – elucidate.
3. What are the essential services of a digital library system?
4. Discuss the IPR issues related to the development of digital libraries.
5. What is a metadata schema? How is it related to the digital library system?

7.12 REFERENCES

Arms, Willaim Y. (2000). Digital libraries. The MIT Press.

BBC. (1999). Information rich information poor, BBC News, Retrieved December 8, 2021, from http://news.bbc.co.uk/2/hi/special_report/1999/10/99/information_rich_information_poor/466651.stm

Bearman, D. & Trant, J. (1998). Authenticity of digital resources: towards a statement of requirements in the research process, D-Lib Magazine, June 1998, Retrieved December 22, 2004, from http://www.dlib.org/dlib/june98/06bearman.html

Borgman, C. (1999). What are the digital libraries? Competing visions, Information Processing and Management, 35(3), 227-43.

Borgman, C. (2000a). From Gutenberg to the Global Information Infrastructure. The MIT Press.

Borgman, C. (2000b). Digital libraries and the continuum of scholarly communication, Journal of Documentation, 56(4), 412-30.

Borgman, C. (2000c). *From Gutenberg to the global information infrastructure: access to information in the networked world*, New York, ACM Press.

Bush, V. (1945). As we may think. *Atlantic Monthly*, July 1945, pp. 101-108.

Chowdhury, G.G. & Chowdhury, S. (2003). *Introduction to digital libraries*, London, Facet publishing.

Crow, Raym. (2002). The Case for Institutional Repositories: A SPARC Position Paper. (Washington, DC: Scholarly Publishing & Academic Resources Coalition). Retrieved January 11, 2021 from https://ils.unc.edu/courses/2014_fall/inls690_109/Readings/Crow2002-CaseforInstitutionalRepositoriesSPARCPaper.pdf

Day, M. (2001). Metadata: Preservation 2000. Ariadne, 26, Retrieved January 12, 2021, from http://www.ariadne.ac.uk/issue/26/metadata/

Greenstein, D. and Suzanne T. (2002). *The Digital Library: A Biography*. Retrieved May 23, 2022, https://www.clir.org/wp-content/uploads/sites/6/pub109.pdf

Guthrie, K. and Wendy L. (1997). The JSTOR solution: accessing and preserving the past. *Library Journal* 122:2 (1997), 42-44.

Harter, Stephen P. "What is a digital library? Definitions, content, and issues." *KOLISS DL'96 (proceedings of the international conference on digital libraries and information services for the 21st century)*. 1996.

Jatowt, A., Morishima, A., Ishita, E., Pang, N., & Zhou, L. (2022, March 16). *Special Issue on Selected Papers from ICADL 2020 - International Journal on Digital Libraries*. SpringerLink. Retrieved November 7, 2022, from https://link.springer.com/article/10.1007/s00799-022-00323-4

Lesk, M. (1996). Going digital. *Scientific American*. March, 1996, 58-60. Also available at: URL: http://www.sciam.com/0397issue/0397lesk.html

Library Trends (2000). Special issue: Assessing digital library services (edited by T.A. Peters), 49(2).

Marchionini, G. & Fox, E. A. (1999). Editorial: Progress toward digital libraries – augmentation through integration, Information Processing and Management, 35(3), 219-25.

Nurnberg, P.J., Furuta, R., Leggett, J.J., Marshall, C., and Shipman III, F.M. (1995). Digital libraries: issues and architectures. In Proceedings of the Second Annual Conference on the Theory and Practice of Digital Libraries. Austin, Texas, June 11-13, 1995, pp. 147-153.

Peter Noerr. Digital Library Toolkit. Sun Microsystems, January 2003 http://www.sun.com/products-nsolutions/edu/whitepapers/digitaltoolkit.html)

Petersen, R.J. (1999). Copyright ownership issues and higher education policies, D-Lib Magazine, June 1999, Retrieved December 22, 2004, from http://www.dlib.org/dlib/june99/06clips.html

Stefik, M. (1997). Trusted systems. *Scientific American*, March, 1997, 78-81. Also available at: URL: http://www.sciam.com/0397issue/0397stefik.html

Stephen, P. (2001). How Do Physicists Use an E-Print Archive? *D-Lib Magazine* Volume 7, Issue 12, Retrieved July 27, 2007, from http://www.dlib.org/dlib/ december01/pinfield/12pinfield.html.

Stephen, P., Gardner, M. and MacColl, J. (2002). Setting up an institutional eprint archive, *Ariadne* 31. Retrieved June 23, 2007 from http://www.ariadne.ac.uk /issue31/eprintarchives/intro.html.

Warner, Beth, and David Barber. (1994). Building the digital library: the university of Michigan's UMLib text project," *Information Technology and Libraries*, 13:20-24, March 1994.

Waters, D.J. (1998). What are digital libraries? *CLIR Issues*, July/August. URL: http://www.clir.org/pubs/issues/issues04.HTML

Wellman, Michael P., Edmund H. Durfee, and William P. Birmingham (1996). The digital library as community of information agents. A position statement, to appear in IEEE Expert, June, 1996. Retrieved July 27, 2022, from http://ai.eecs.umich.edu/people/wellman/ pubs/expert96.html.

Wilkin, John. (1999). *Moving the digital library from "Project" to "Production."* Retrieved July 27, 2022, from http://jpw.umdl.umich.edu/pubs/japan-1999.html.

Witten, I. H., Bainbridge, D., & Nichols, D. M. (2009). *How to build a digital library*. Morgan Kaufmann.

UNIT 8: INSTITUTIONAL REPOSITORIES

Structure

8.0 OBJECTIVES

After going through this Unit you will be able to:

- Asses the importance of institutional digital repositories (IDRS) in open access movement
- Explore nature of IDRS and reasons of their emergence as information entities across the world;
- Trace the path of development IDR movement in India;
- Identify the issues to be considered for developing an IDR; and
- Apprehend the important role a librarian play in sustainable development of an IDR.

8.1 INTRODUCTION

Digital scholarly communication, and more specifically, scholarly publication, is a significant mode of knowledge creation and diffusion that allows for immediate online access to local research outputs around the world. It also captures and preserves knowledge assets more efficiently and transparently than in the past. The major issues with the present scholarly communication process are as follows:

- The exorbitant price rise of journals is making access to knowledge objects difficult even for researchers in institutes of developed-country;
- The current journal publication system is a perfect example of oligopoly, i.e., the entire industry is dominated by five or six major publishers and driven by the profit of those publishers; and
- The phenomenon in which authors write research articles, their senior colleagues review those articles, but they do not have access to knowledge due to a pay wall.

Institutional repositories play an important role in making publicly-funded research reports available in the public domain. These entities gather, organize, index, and allow retrieval of knowledge objects produced by a given institute or a group of related institutes from anywhere at any time by anyone. Institutional repositories form a major part of the global Open Access (OA) movement and are often called the "green path" of Open Access. In this unit, we are going to study aspects and prospects of this important information entity.

8.2 INSTITUTIONAL REPOSITORY: AN OVERVIEW

Institutional Repositories, also known as Institutional Digital Repositories (IDRs), are digital collections that catalogue, index, archive, and make available a particular institution's assemblage of knowledge resources through a single-window retrieval interface. The value of an institution's intellectual property is dispersed over thousands of scholarly publications under the existing system of scholarly communication system. A university's (or an institution's) academics' intellectual output is concentrated in an institutional repository, making it simpler to show its economic, social, and scientific significance.

8.2.1 Definitional scopes

As information entities, institutional digital repositories:

- Offer required infrastructure and facilities to help members of a given institute to share their works across the world;
- May supplement and complement the traditional channels of scholarly communication process;
- Ensures greater visibility of an institution and its intellectual output and thereby increases chances of getting higher citations by its researchers ;
- Supports long-term and secured preservation of knowledge objects produced by a given institute over the years; and
- Collect all of the institutes' research endeavors in a central location and thereby ensures efficient discovery of resources and reduces

In a 2002 SPARC position paper, Crow (2002a) defined an institutional repository (IR) as

"a digital archive of the intellectual product created by the faculty, research staff, and students of an institution and made accessible to end users both within and outside of the institution, with few if any barriers to access".

8.2.2 Elements of Institutional Repository

The traditional model of scholarly publication for journals and books is characterized by the following features -

- Research is publicly funded
    - Personal academic efforts
    - Supported by institutions

- Authors sign away rights with publishers in order to publish
- Given away freely to publishers
- Publishers make huge profits out of the intellectual properties of institutions
- Author gets no tangible reward
- And loses rights to copy material for colleagues, teaching etc…
- Institution potentially loses out on its investment

The exorbitant price rise of journals leads to the situation known as the "serials crisis," in which institutions create intellectual property in the form of journals, papers, book chapters, etc. and give it away to commercial publishers free of charge but need to pay a hefty sum of money to access those resources produced by them and their colleagues in other institutes. According to a study conducted by the Association of Research Libraries (ARL), journal prices have increased by 521 percent, while the consumer price index has increased by only 118 percent (Figure 8.1). Under such circumstances, IDR may act as an alternative platform for distributing intellectual output of a given academic institute. A typical IDR has six basic and essential elements. These are -

- **Institutionally Defined:** Accumulates in a single place all intellectual products of the members of a knowledge producing institute like universities, research organization etc;
- **Global-scale cooperation:** Directories and path finder services allow researchers to identify open access repositories in their domain of research and thereby leads to global cooperation in research;
- **Academic content:** IDRs allow researchers to download and use archived knowledge objects distributed in different institutional repositories (preprints and post prints);
- **Perpetual Access:** IDRs ensure long-term preservation of knowledge objects and supports perpetual access to the archived resources;
- **Single-window Search:** IDRs are distributed across the world and as a result it is a difficult task for a research to search and retrieve content from many such repositories. However, a set of services are emerging that provide one window access to content that are distributed in many repositories. These services are based on metadata harvesting based discovery services like BASE (https://www.base-search.net/), CORE (https://core.ac.uk/) etc; and
- **Legal Framework:** Resources in institutional repositories are archived along with

license (mostly CreativeCommons licenses) and there are support tools like Sherpa/RoMEO (https://v2.sherpa.ac.uk/romeo/) and Sherpa/Juliet (https://v2.sherpa.ac.uk/juliet/) that can guide researchers in selecting licenses.

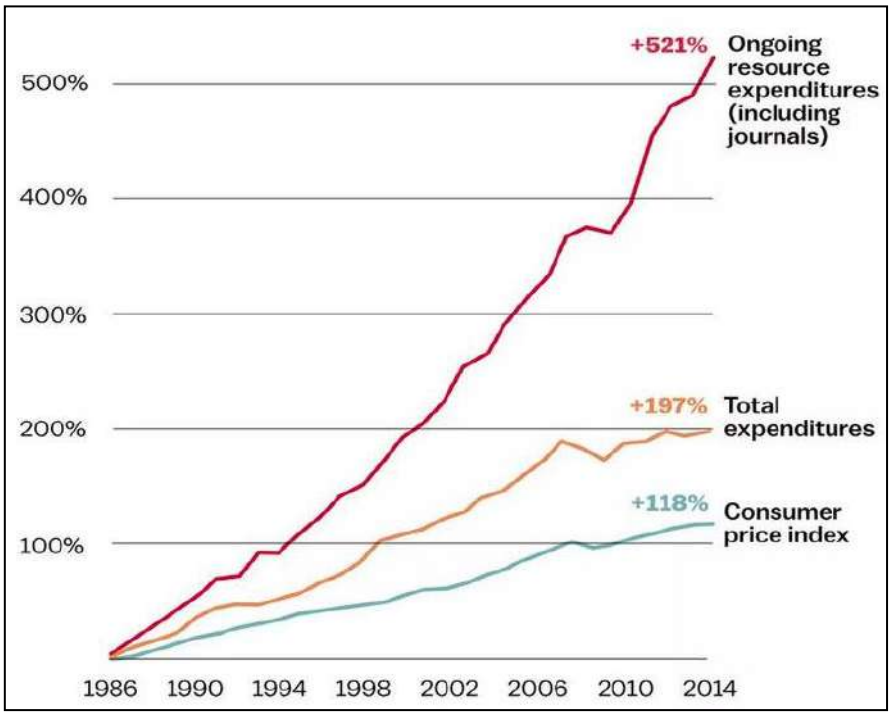

Figure 8.1: Price rise of journals during the period from 1986 to2014

8.2.3 Role in Open Access

Open Access (OA) is evolving as a new model of scholarly communication across the world. OA has two major routes – Gold path and Green path. These two paths are associated with OA journals (Gold path) and the self-archiving. The later one (self-archiving) is known as Green path of OA and is dominated by OA repositories. Institutional digital repositories are part of the Green path of OA. The management of the digital intellectual output that academic institutions produce, such as journal articles, conference papers, reports, theses and dissertations, instructional materials, artwork, research notes, and research data, has been a source of contention. Institutional repositories have been suggested as a tool to assist academia in maintaining and disseminating their digital materials. Universities are discovering new ways to capture, manage, and distribute these intellectual electronic resources. In order to build collections of digital resources and educational materials in the

form of open access repositories that will allow faculty and researchers to upload and download scholarly literature and use them to share resources with each other either within the institution or across the region, numerous universities and research institutions around the world are researching, testing, and developing these systems known as institutional digital repositories.

8.2.4 Institutional Repository vs Digital Library

We often use the concept of IDR and digital library synonymously. Of course, IDR may be considered as a very important component of a digital library system but these two concepts vary in terms of scope, coverage and architectural issues. The major differences are given in Table 8.1.

Table 8.1: Digital library vs IDR

Digital library	**Institutional digital repository**
Digital library may include IDR as a component	IDR may act as a part of digital library system
Scope of digital library system is higher in comparison with IDR	Scope of IDR is specific in comparison with digital library system
Digital library system generally includes resources from different sources and in different formats	IDR generally includes scholarly output of a single institution
Digital library may cover a single discipline or a group of related disciplines	IDR always provides multidisciplinary coverage
Discipline specific	Institution specific

8.3 INSTITUTIONAL REPOSITORY: BASIC COMPONENTS

The framework of a typical institutional digital repository (IDR) centers around six basic questions:

1. *What essential functions does the repository program provide?*
2. *Who is permitted to add items to the repository?*
3. *What type of content is proper?*
4. *How can we coordinate migration and preservation?*
5. *What guidelines for rights management should be implemented?*
6. *What should the criteria be for cataloging (creating metadata)?*

8.3.1 Content management

IDR offers the ability to store and provide access to a much wider variety of material. Researchers produce articles and reports, but also "original art, grant proposals, maps, radio/TV interviews, motion pictures, music scores, photographs, consulting (technical) reports, technical drawings, and poster session displays". All of these, once converted to digital format, might be deposited in the IDR. A global-scale analysis on the basis of OpenDOAR datasets (see Figure 8.2) shows that journal articles and theses form the core content for IDRs.

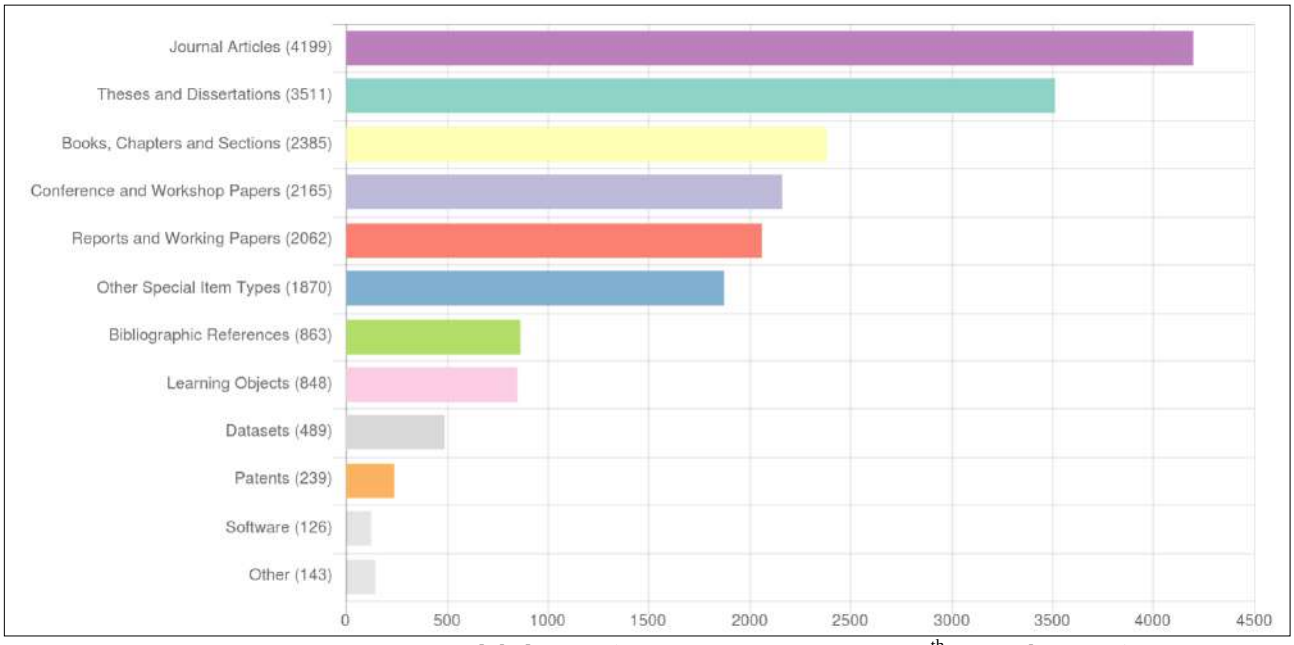

Figure 8.2: Content types in global IDRs (source: OpenDOAR as on 30th November 2022)

Moreover, supplementary material such as supporting evidence and data, interim reports, and draft versions of papers may also be stored. A repository must have a content management policy. In the content management policy, it should be stated what types of materials will be accepted and stored in the repository. There is no hard and fast rule regarding the type of contents because the structure of the archive depends on the software, technical support, vision, and resources of the IDR. An IDR may host varieties of materials depending on the institution's preferences. The actual content type of an IDR will vary according to its perceived functionality. But a repository must be populated with contents and managed in an effective and sustainable way. It is important that policies are put in place that will enable this. The types of contents can range from dissertations and articles to raw research data and data sets, post-prints (peer-reviewed research articles), book chapters, working papers, theses,

etc. So it is important for an IDR system to have a stated contents policy, keeping in mind the interests of the organization.

8.3.2 Preservation and migration

The preservation policy indicates how long contents will be retained in the IDR, whether items will be migrated to new file formats, if the repository will ensure the continued readability and usability of its contents (dealing with software and hardware obsolescence), or who will take it back up and if there will be a backup strategy in place. What systems are to be put in place to ensure that the contents are preserved for the long term? Will some types of contents be prioritized over others for preservation? Will data files be migrated to new file formats where necessary to preserve access to their intellectual contents? In the event that the repository is closed, will the data be transferred to another appropriate archive? Will the repository regularly back up its files according to current best practices? In the policy of preservation, it should be clear what file formats (see Unit 3) will be accepted for preservation in the repository (word.doc, Adobe.pdf, etc). There are clear differences between file formats because a file format that is good for access today may not be a format that is easy to migrate. On the other hand, a format that is easy to migrate may not be easy to read. A set of recommendations may be liste here on the basis of the best-practice guidelines:

- Items will be retained indefinitely in a repository;
- Items may be migrated to new file formats (preferably open formats) when necessary;
- Storing objects in formats that can be migrated as per the latest technology;
- To access un-migrated formats, software emulations will be provided.
- Repositories may keep more than one copy of a submitted file for the purposes of security, backup, and preservation.
- Repository will perform regular backups of its files in accordance with industry best practices, and
- The database will be transferred to another appropriate archive as the repository is closed down.

8.3.3 People

Members or E-people of an IDR include submitters, metadata editors, reviewers, and final approvers. But the most important question is: who is eligible to submit? Major policy decisions will be needed related to this submission policy. In the submission policy, it is

important to define who will be able to submit content. A very important decision in the submission policy is whether or not an institutional repository will provide any assistance with the submissions. And in the case of mediated submissions, what will be the extent of the submission? Or what is the workflow for submission? As per the OpenDOAR (2012) database, 80% of repositories have not defined submission policies. Generally, only registered or authorized users can submit objects to an IDR. Only a few experts suggested using a system of "*mediated deposits"* to assist the contributors in the submission process. In general, it is found that most of the IDRs allow authors to archive their items, and the authority assists the author in the deposition process, if required.

8.3.4 Workflow

A repository workflow is a breakdown of the administrative tasks involved. There are several types of workflow in a typical repository, depending upon the types of documents and software used. Generally, these include workflows to manage user registration and administration; workflows to manage authorization and permissions within the repository; and various administrative workflows to allow for maintenance and software updates. However, the most significant workflow focuses on the submission process (see Figure 8.3).

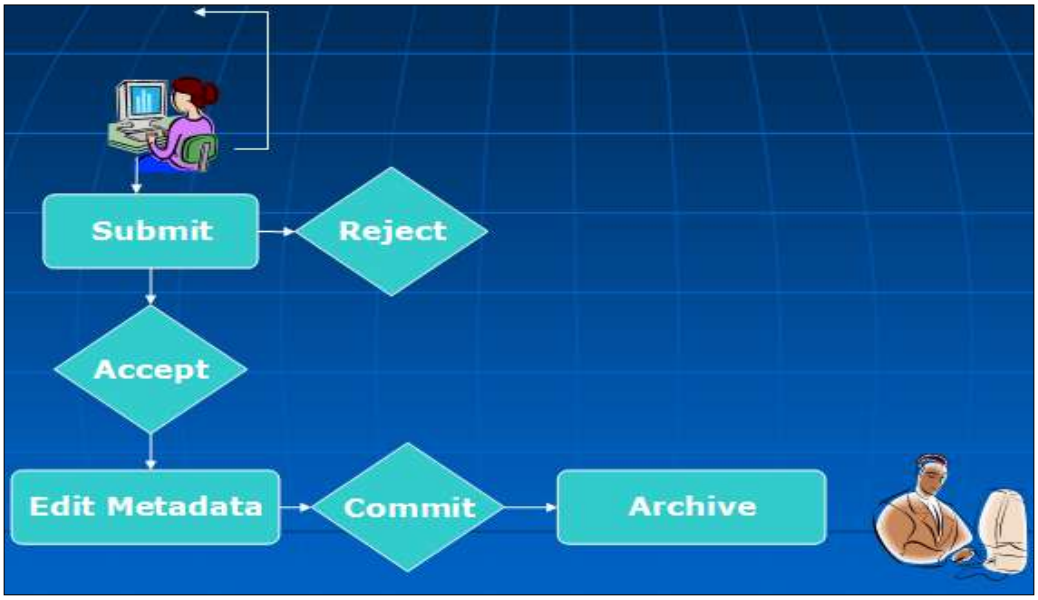

Figure 8.3: Submission workflow in an IDR

- Workflow 1: Accept/Reject Step: This step is used to allow a user to simply accept an item, or reject it. If they reject it, they can give a reason which will be emailed to the submitter. The item will appear back in the submitter's workspace, if it is rejected;
- Workflow 2: Accept/Reject/Edit Metadata Step: This step is used to allow E-person to either accept or reject an item, and edit its metadata. If they reject it, they can give a reason which will be emailed to the submitter. The item will appear back in the submitter's workspace if it is rejected. But they cannot change the submitted files. They can accept submission for inclusion, or reject submission; and
- Workflow 3: Edit Metadata Step: This step is used to allow the user to edit the metadata. This might be done to correct the metadata, or to improve it. But they cannot change the submitted files. They must then commit to archive; may not reject submission.

8.3.5 Core services

There are many potential services and benefits that an IDR can extend, and these services exist at various levels, from the individual researcher to the university as a whole. Experts listed the following main services of an IDR: scholarly communication, education, e-publishing, collection management, long term preservation, institutional prestige, knowledge management and research assessment exercises. These services support three main service areas for an institution - to increase global visibility, to preserve knowledge object as produced, and to provide free access to the institution's scholarships. However, the services and the corresponding benefits may be listed here from different points of view:

- IDRs have potentially significant benefits for institutions if they are integrated holistically into university frameworks. The most prominent reason is the increase in visibility and impact of research output;
- IDRs may become the new university presses and local peer review and quality control will evolve into full scale publishing ventures;
- Teaching and learning can be supported by links to IDR content from virtual learning environments (VLEs) and the library catalogue;
- IDRs are a practical, cost effective, and strategic means for institutions to build partnerships with their faculty to advance scholarly communication;
- IDRs are beneficial for all the stakeholders, including publishers, editors and authors

as they can substantially increase their impact and the impact factor for the source journals;

- IDRs facilitate immediate access to research and scholarship and maximize the potential research impact of archived publications;
- IDRs may archive supplementary material such as supporting evidence and data, interim reports and draft versions of papers may also be stored;
- IDRs can produce hit counts on papers, personalized publication lists and citation analyses alongside article-level metrics;
- IDRs offer advantages to both 'academics-as-authors' and 'academics-as-readers' - the same system that facilitates the dissemination of academics' own work also enables them to gain access to the work of others; and
- Finally, as scholarship is shared, society at large is benefited as IDRs - provide access to the world's research; provide local access to global research; ensures long-term preservation of institutes' academic output; and can accommodate increased volume of research output (no page limits, can accept large data-sets, 'null-results', etc.).

A summary of benefits have been illustrated in Table 8.2

Table 8.2: Benefits of IDRs

Stakeholders	Benefits
Institutions	Increases visibility and prestige; acts as an advertisement to funding sources, potential new faculty, raising the institutional profile, total intellectual output, teaching and learning, supporting institutional record keeping, cost savings, unique place of resources.
Users	Dissemination and impact, IR content, feedback and commentary, added value services, personal and promotional uses, networked information.
Researchers	Provide a central archive of their work, increase the dissemination and impact of their research, more control over their work.
Society	Provide access to the world's research; ensures long-term preservation of institutes' academic output.

8.4 DEVELOPING IDR: POLICY ISSUES

This section explores the critical parts of developing an IDR: everyday work, the services, the usefulness of the services in practice, and the quality of the IDR. Some think IDR has achieved more than they need, and some believe it has not yet met the desired level. This discussion looked at what has been done, what can be learned from the earlier studies, and where the gaps are.

A. Existing policy

A successful IDR cannot be developed without giving serious consideration to its overall structure and design. This overall structure and governance can be initiated by developing IDR policies. Repository policies need to have clear explanations and examples, but in case of IDRs in India, several policy issues are missing and have not been considered and need to be developed in the line of global recommendations.

B. Advocacy and promotion

The success of the IDR system depends much on the users' participation and voluntary involvement with the system. There should be a public relations and branding policy for IDR resources. But initiatives in India lack focus on how IDR will increase in quantity and quality in order to become competitive with other providers. Another issue, training and documentation (for end users, authors, and administrators), has to be considered seriously.

C. Legal framework and licensing model

IDRs face few legal challenges in their early stages, but they may face significant challenges in developing a critical mass of journal papers written by researchers and faculty members and commercially published. There are three parties involved in this system: the author, the institution, and the publisher. It is not possible to have a single, straightforward licensing policy as different publishers have different licenses. It would therefore be desirable to have a customizable licensing system for items submitted to the repository. Studies do not have a licensing policy for the legal issues in regard to uploading and accessing the documents. Even studies do not propose any licensing models for IDR System rights management.

D. Access control and rights management

The IDR system must have mechanisms to restrict access to the information when OA is premature or otherwise not desirable. It is also desirable to ensure that every knowledge object archived is having required OA license (CreativeCommons or other licensing of similar nature).

E. Preservation and curation

The advocacies do not discuss any specific preservation strategy or backup strategy. Only a few studies discuss file preservation. There is no study of in-house preservation systems, and these services are still under development. There are two technical issues: the first is ensuring that the physical item remains intact at the bitstream level, and the second is ensuring that the digital object remains understandable. These issues have not been properly discussed. This challenge remains a long way from being solved. A coordinated strategy like the OAIS (Open Archival Information Systems) reference model (see Unit 7 section 7.5), the defacto standard for digital archive architecture, is required to ensure long-term preservation of the IDR contents.

F. Content quality

Especially for the quality of the IDR contents, only a few studies have been done in this area. Validation and verification of the data are important factors, and there are two levels to this task: the metadata about a particular scholarly work and the organisational and contextual information about the work. Studies do not clearly explain how quality is assured or who will measure the contents' quality. So, further studies are urgently needed in this area.

G. Model

Developing a sustainable IDR needs funding and business models; however, in India, to date, we don’t have a business model at the national level that could be followed in developing IDRs.

8.5. DEVELOPING IDR: TECHNICAL ISSUES

The following technical issues are required to be considered in developing and maintaining an IDR as per the global standards, best practices and guidelines:

A. Content

Content is king for every retrieval system. IDRs are no exception. There are many decisions associated with the content policy. There will be different descriptive requirements, preservation issues, and workflow patterns associated with different types of documents. Consideration should be given to the types of checks and who is likely to perform them to ensure the quality and authenticity of the content available through an IDR.

B. Resource organization and management

For any IDR system, data standardization tools like standard lists, code lists, or vocabulary control devices need to be followed, but initiatives in India do not propose common

standards, methods, tools, or techniques for the organization of resources in an IDR. As a result, services are not based on an internationally agreed-upon data standard.

C. Metadata Schema

All repositories are using DC metadata as generic metadata standard but initiatives in India do not recommend any domain-specific metadata schema suitable for different types of objects. It would therefore be desirable to have a flexible policy for metadata schemas that will be able to combine the different requirements for each of the possible content domains to produce a metadata set to be collected that is exactly appropriate. Another issue, authority control, is a critical area in metadata submission. But this issue has not been given due importance in developing IDRs in India.

D. Indexing services and standards

Another key issue is developing common platforms to make IDR interoperable with other systems in order to import and export resources. Users aren't interested in browsing each IDR; there is a need to integrate a resource discovery and search tool (like BASE or CORE) that can be built using the OAI-PMH metadata harvesting from different IDRs in a centralized index.

E. Resource identification for perpetual access

Identifying and locating online contents is a key issue in repository systems for granting long-term access. IDRs need to assure the permanence of object names in the repository. It provides access to the contents on the same timeline and shows the relationship with other items. Each object should have a unique and persistent identifier such as CNRI, handle, DOI, etc.

F. User Interface

There are hardly any studies that have been done in this area. Any IDR system demands customizable, accessible web service interfaces so that the repository can participate in distributed application systems. The development of multilingual interfaces and mechanisms to access multimedia learning objects has been ignored in the IDR initiatives.

G. Multilinguality

In a multilingual country like India provisions for multilingual search and access to multilingual content is a must. IDRs need to support the Universal Character Set (UCS) of Unicode. Additional problems and challenges are related to: data management and representation of information, interoperability (linking between systems), etc. Multilingual document indexing is also challenging because each language has different characteristics

and rules.

H. Standard compliance

As the repository field is a new and developing one, some parts of the system have more agreed-upon standards than others. Some standards, such as OAI-PMH, are part of the software and well developed. The standards for domain-specific metadata are still under development. All systems are able to provide Dublin Core output, but this is not descriptive enough to be able to ensure complete interoperability and transfer of data between different systems. The system should emerging interoperability standards like OAI/ORE for compound digital objects, SUSHE or PIRUS for usage statistics, DRIVER and OpenAIRE to support network-level interoperability and so on.

I. Software

IDRs are part of open knowledge movement and therefore mostly use open source software and open standards for developing such systems. But there is no uniformity in using software for the IDR system. Every IDR software does not meet global standards. Moreover, sometimes they need to be customized to meet local requirements. Apart from supporting all the open standards in the IDR domain, an IDR software must be compliant with the emerging standards like REST/API based interaction, JSON-formatted data exchange and so on.

8.6. ACCESS SERVICES FOR INSTITUTIONAL REPOSITORIES

One of the major issues with the open-access institutional repositories is their distributed nature. They are distributed unevenly across the globe. For example, a query to OpenDOAR (see section 8.6.1) shows that there are presently around 6,000 repositories distributed in more than 100 countries (see Figure 8.4). It makes it difficult for researchers to retrieve distributed knowledge resources. Fortunately, a few universities and academic consortiums came forward to provide solutions to this issue with an array of services. These services may be studied under three groups: a) pathfinder services; b) search services; and c) data services.

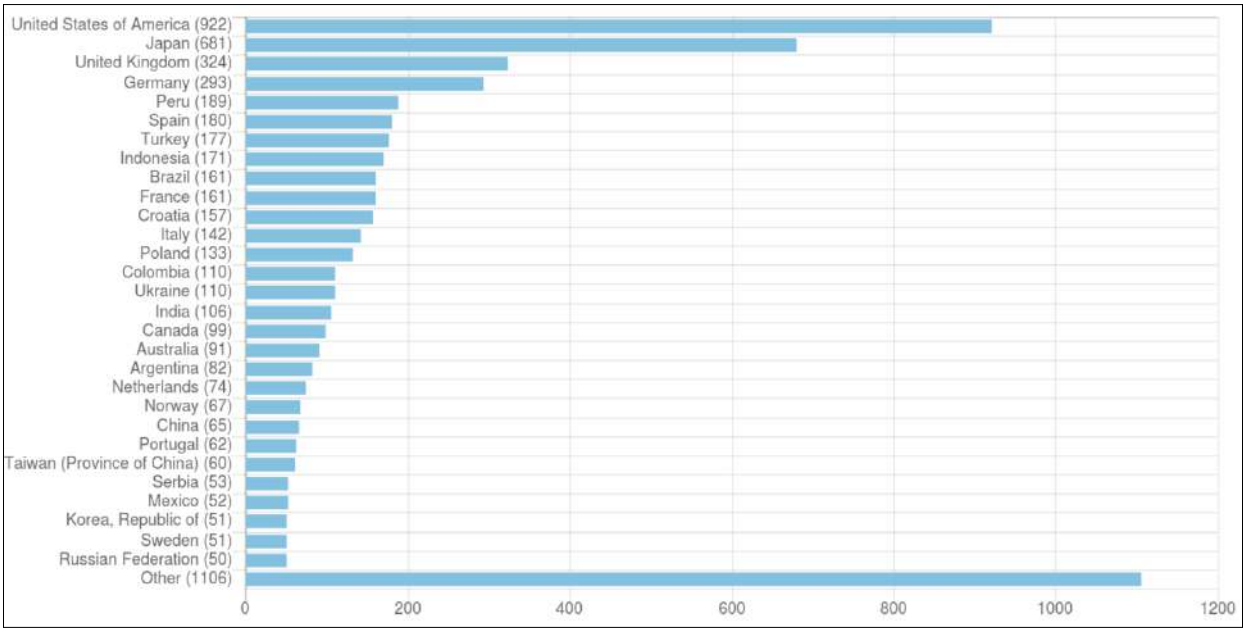

Figure 8.4: Repositories by countries

8.6.1 Pathfinder Services for Institutional Repositories

As a student of library science you already know that the most important sources for information professionals are the tertiary sources of information. These are also called meta tools and are mainly directories. The first ever such a path finder service or a directory was developed by the University of Southampton and it is known as ROAR (Registry of Open Access Repositories – http://roar.eprints.org/). The other one, initiated by University of Nottingham and Lund University, funded by OSI, Jisc, SPARC Europe and CURL, is known as OpenDOAR (Open Directory of Open Access Repositories – https://v2.sherpa.ac.uk/opendoar/). Both of these services allow browsing, searching and filtering of global OA repositories. Additionally, OpenDOAR allows a few statistical services along with data visualization facilities. ROAR presently lists around 4725 OA repositories and OpenDOAR includes 5980 OA repositories (as on 30.11.2022). Both of these services allow REST/API based data fetching for local services. An example for record of a data repository from OpenDOAR is given in Table 8.3.

Table 8.3: A sample repository record in OpenDOAR

Repository Name	AIJR Preprints [English]

Repository Type	Aggregating
Contact Email	preprints@aijr.org
Repository URL	https://preprints.aijr.org/index.php/ap/preprints
OAI-PMH URL	http://preprints.aijr.org/index.php/ap/oai
Software Name	Other (OPS)
Content Types	Journal Articles
Subjects	Science,Technology,Engineering,Mathematics,Health and Medicine, Arts,Humanities,Social Sciences
Organisation Name	AIJR Publisher [English]
Organisation URL	https://www.aijr.in
Country	India
Metadata policy	https://preprints.aijr.org/index.php/ap/ethics
Data policy	https://preprints.aijr.org/index.php/ap/ethics
Content plocicy	https://preprints.aijr.org/index.php/ap/ethics
Preservation policy	https://preprints.aijr.org/index.php/ap/ethics
ID	9935
Date Created	21 September 2020 07:54:00 UTC
Last Modified	12 January 2022 15:36:33 UTC
URI	https://v2.sherpa.ac.uk/id/repository/9935

8.6.2 Search Services for Institutional Repositories

The pathfinder services, as discussed in the previous section, allow browsing and searching at the repository level but not at the content level. A few global search services are presently allowing search and browsing of institutional repositories (the "green path" of OA) at the content level. The most important in this array of search services is BASE (Bielefeld Academic Search Engine), an initiative of the Bielefeld University Library in Bielefeld, Germany. It indexes, apart from journal articles (gold path), theses, book chapters, etc., the repositories listed in OpenDOAR, ROAR, re3data, or Open Archives. It is based on open source tools: the VuFind discovery service, the Apache Solr text retrieval engine, and open standards-based metadata harvesting (OAI and PMH). It presently includes 310,625,858

resources (310 million) from 10,269 content providers. It indexes 260 sources from India that include OA journals, OA repositories, and OA theses. BASE has a very sophisticated search interface where end users can filter results by a number of options like field-level search, range search, Boolean join, and filtering by document types, license type, language, year, and so on. Moreover, it allows browsing of OA resources by DDC class number, document type, etc. The display of retrieved results is also library-like, with a rich set of information elements (see Figure 8.5). It provides an array of personalized search services like search history, RSS feed-based alerting, a save citation option, sharing, record export, and so on.

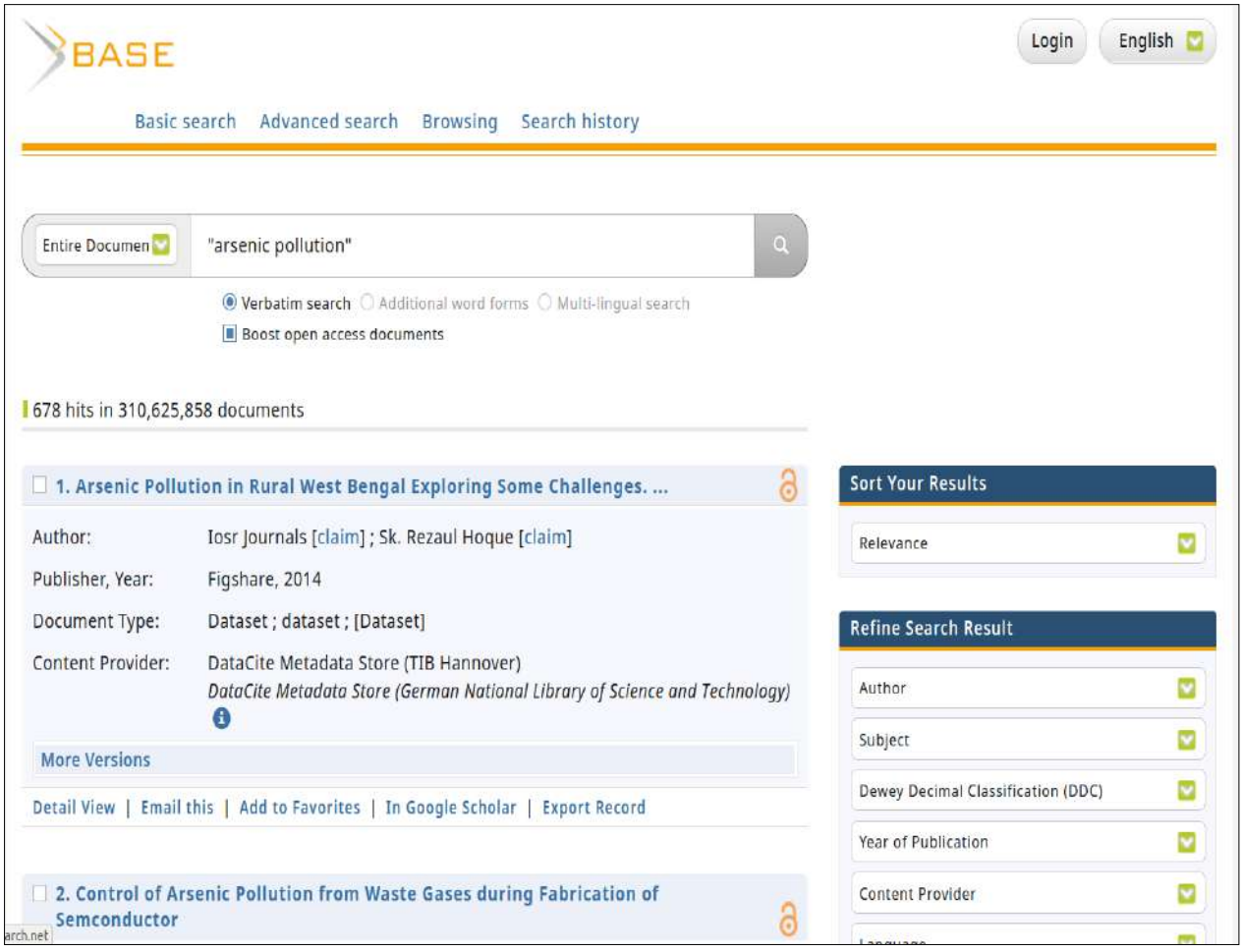

Figure 8.5: Retrieval interface of BASE

Apart from BASE, the following search services also index institutional repositories at the content level:

- **Lens** (https://www.lens.org/): With information from Microsoft Academic, PubMed,

Crossref, OpenAlex, UnPaywall open access information, CORE full text, and linkages to ORCID, Lens provides access to more than 249 million scholarly records. For the first time, the entire scholarly citation graph is made available as a free public resource. A considerable portion of these resources are coming from the green path of OA.

- **Dimensions** (https://app.dimensions.ai/discover/publication): More than 100 million publications are available in Dimensions, including preprints, postprints, conference proceedings, monographs, book chapters, and articles that have been published in scientific journals. All papers have links to financing, publications, patents, clinical trials, and policy documents that contextualize them. Additionally, you can examine profiles for researchers, institutions, sponsors, and associated categories.
- **Semantic Scholar** (https://www.semanticscholar.org/): For the benefit of the entire world's academic community, Semantic Scholar offers free, AI-driven search and discovery tools including OA resources, covering content of the OA repositories. It indexes more than 200 million scholarly works obtained from web crawls, publisher partnerships, and data sources. Research articles published in all categories are listed in the Semantic Scholar Academic Graph (S2AG) Dataset and APIs as an accessible JSON archive.

8.6.3 Data Repositories

A new generation of repositories known as “data repositories” are emerging across the world. It allows researchers to upload datasets related to their research so that other researchers can use and explore the datasets without repeating the same set of activities. These datasets are generally made available as open OA resources under the Open Data Commons Open Database License (ODbL – https://opendatacommons.org/licenses/odbl/). Some repositories line Zenodo (developed by CERN – https://zenodo.org/) allows archiving textual content and datasets with the provison for linking research reports with the corresponding resaerch dataests. Many journal publishers advise resaerchers/contributors to pick a data repository that generates a persistent identifier, preferably a Digital Object Identifier (DOI), and that has developed a thorough preservation strategy to guarantee the data is kept safe forever. Some of generic data repositories that are quite well known to researchers across the world is listed here for your ready reference:

- 4TU.ResearchData - https://data.4tu.nl/info/en/
- ANDS contributing repositories - https://researchdata.ands.org.au/contributors
- Dryad Digital Repository - https://datadryad.org/

- Figshare - https://figshare.com/
- Harvard Dataverse - http://dataverse.harvard.edu/
- Mendeley Data - https://data.mendeley.com/
- Open Science Framework - http://osf.io/
- Science Data Bank - https://www.scidb.cn/en
- Zenodo - https://zenodo.org/
- Code Ocean (with code) - https://codeocean.com/

There are two serach services for data repositories that can help a researcher to find out suitable subject-based data repositories for archiving research datasets (see Figure 8.6) – 1) FAIRsharing (https://fairsharing.org/) and re3data.org (https://www.re3data.org/).

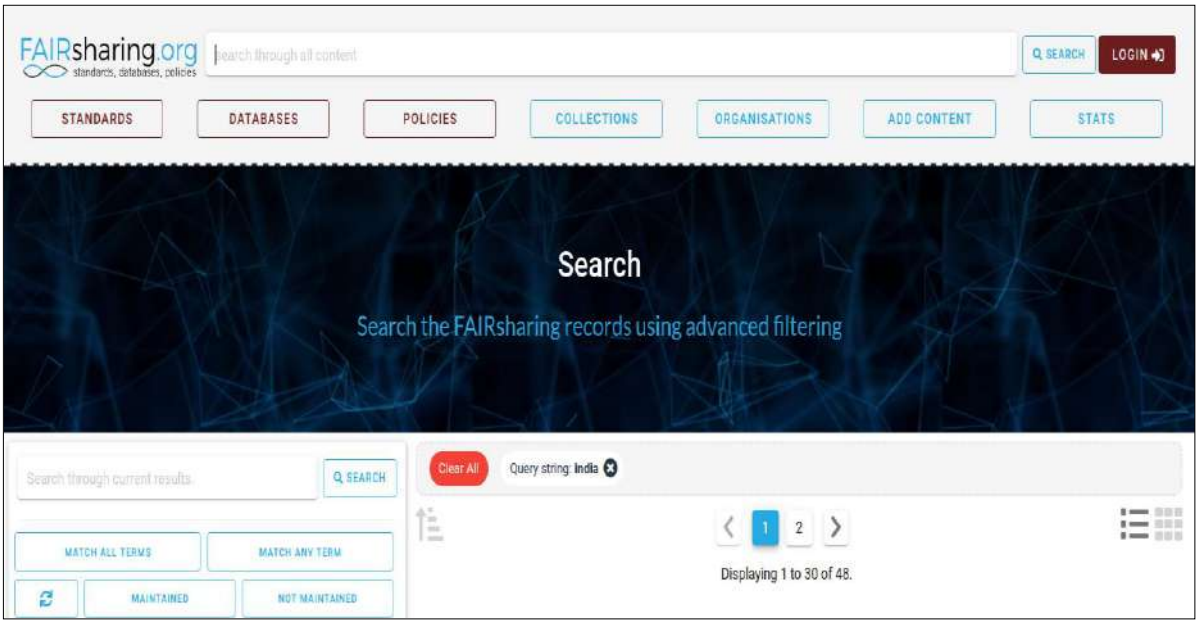

Figure 8.6: Search interface of FAIRsharing

8.7 INSTITUTIONAL REPOSITORY MOVEMENT IN INDIA

The web-enabled, distributed IDR system is a socio-technical concept. It is a multi-faceted domain. It is not all about information storage and retrieval, but rather the organisation and management of resources and bibliographic items to support open access to knowledge. India is having one of the largest chain of institutions in the world with 1000+ universities, 10,000+ stand-alone institutions, 40,000+ colleges but as on November 30th

2022 OpenDOAR lists only 106 institutional repositories and ROAR includes 135 repositories with the highest contribution from Indian Institute of Astrophysics (6525 records). However, the current adoption levels of OARs are pleasingly high. The two major registries of OARs (viz. OpenDOAR & ROAR) show that adoption in India is already very high, teh country has started taking the advantages that open access confers. The area is in rapid development, and OpenDOAR statistics show a steep incline from around 2005 to the present in the availability of both open archives and open records (Figure 8.7).

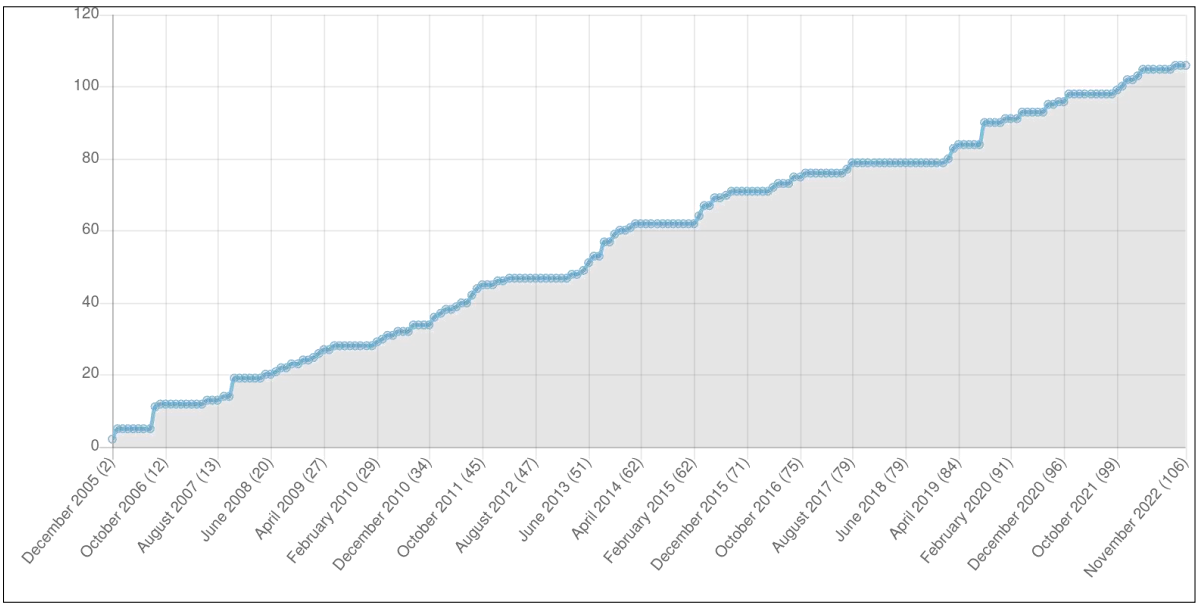

Figure 8.7: Growth of IDRs in India (source OpenDOAR)

But the situation is fast changing, and a closer look reveals a shining picture. Our country has been the most prominent partner in the field of the open knowledge movement. Our professionals and working librarians are trying to build up the necessary information infrastructure, which is essential for open access development, and our experts are trying to establish digital libraries and institutional repositories with open source software (OSS). India, in its own way, made its contribution by developing and launching more than 100 repositories as on date. Several agencies, learned societies, and professional bodies have come forward and established IDRs on their own. A nation-wide movement has started, and government organizations like the University Grants Commission and the National Knowledge Commission have recommended OA for publicly funded research. In one word, "open access" to information is the realization of Ranganathan’s Five Laws of Library Science in the Internet world. In April 2005, there were only two (2) institutional repositories

in India (Figure 8.X), and this number had climbed to more than hundred (100) in 2022 (November 30th, 2022), with an average increase of 98%. Based on worldwide numbers in both OpenDOAR and ROAR databases, this growth has placed India as the seventh leading nation in IDR development. Much of this success is undoubtedly connected to government support as well as sponsorship from MoE (the Ministry of Education), UGC, NKC, and other professional agencies (e.g., the Information and Library Network (INFLIBNET)) and research institutes. As a result, a silent revolution is taking place rapidly in India in the area of repository development. A subject-wise analysis (one repository may have many subject categories) shows that Science & Technology topped the list, followed by Social Sciences and Health & Medicine (Figure 8.8).

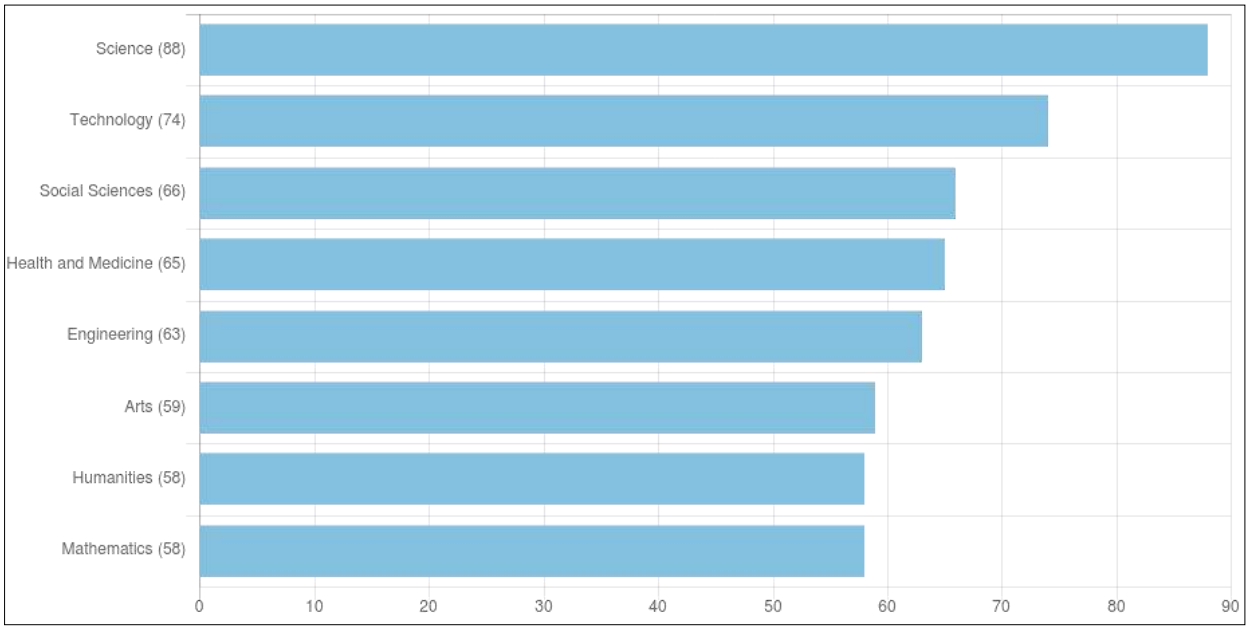

Figure 8.8: Subject-wise division of IDRs in India (source OpenDOAR)

8.8. SUMMARY

IDR is an important component of the open access movement. Digital library systems of different types, along with IDR, are making the scholarly communication process open for all. It is a viable alternative to the present toll-based journal access system. The role of a librarian is very important in developing an institutional repository. In general, the question

of who will manage the IDR may arise. Or where will it be located? An in-depth study of the global IDRs reveals that these information entities are mostly run by the host institution's library (or information services). There are a number of reasons cited as to why the library is the appropriate locus for the leadership of such projects. Libraries are an essential component of a nation’s information infrastructure, and their staff play a significant role in the development of institutional repositories. "Linking people to resources" has been the task of information specialists for many years. The ARL SPEC Kit survey results reveal that libraries play a critical role in initiating, planning, and implementing IDRs all over the world. This unit discusses all the issues related to IDR, ranging from policies to practices, software to content quality, standard compliance to promotion and advocacy, and so on.

8.9 CHECK YOUR PROGRESS

Q. 1: What is ROAR?	
A	Registry of Open Access Repositories
B	Register of Open Access Repositories
C	Registry of One Access Repositories
D	Registry of Open Access Repository

Q. 2: What is FAIRsharing?	
A	Open Access repository
B	Search engine for OA repositories
C	Search service for data repositories
D	Search engine for web resources

Q. 3: Which organization is responsible for developing BASE?	
A	Bielefeld University Library, Italy
B	Bielefeld University Library, Germany

C	MIT, US
D	INFLIBNET, India

Q. 4: Match the followings:			
a	ODbL	i	A service to know OA policy of publishers
b	Lens	ii	A standard for usage statistics
c	SUSHI	iii	A search service for academic resources
d	Sherpa/RoMEO	iv	Open license for datasets
Code:			
A	a – i; b – ii; c – iii; d - iv		
B	a – i; b – ii; c – iv; d - iii		
C	a – iii; b – ii; c – iv; d - i		
D	a – iv; b – iii; c – ii; d - i		

Q. 5: Match the followings:			
a	OpenDOAR	i	University of Nottingham
b	ROAR	ii	University of Southampton
c	Zenodo	iii	CERN
d	Dataverse	iv	Harvard University
Code:			
A	a – i; b – ii; c – iii; d - iv		
B	a – i; b – ii; c – iv; d - iii		
C	a – iv; b – iii; c – i; d - ii		
D	a – iv; b – iii; c – ii; d - i		

Answer keys: Q 1: A ; Q. 2: C ; Q. 3: B ; Q. 4: D ; Q. 5: A

8.10 KEYWORDS

- **OAI-ORE** (Open Archives Initiative – Object Reuse and Exchange): is an interoperability standard for compound digital objects,
- **OpenAIRE** (Open Access Infrastructure Research for Europe): provides guidelines and

standards to integrate OA repositories and OA journals.

- **ORCID** (Open Researcher & Contributor ID): is an open international initiative to provide a registry of unique researcher identifiers at global scale.
- **PersID:** supports persistent identification of knowledge objects through an international infrastructure and knowledge base.
- **PIRUS** (Publishers and Institutional Repository Usage Statistics): is a code of practice for managing usage data and is considered as open international standard in usage data for DL objects.
- **RDF** (Resource Description Framework): is a standard model for web-based data interchange.
- **SURE** (Statistics on the Usage of Repositories): aims to coordinate and aggregate usage data from repositories in Netherlands.

  SUSHI (Standardized Usage Statistics Harvesting Initiative): is a protocol designed for the transmission and sharing of COUNTER-compliant usage data from DL service providers.

8.11 QUESTIONS FOR SELF STUDY

1. Define IDR. How does it differ from a digital library?
2. Write a short note on IDR movement in India.
3. Write a plan for developing an IDR for KSOU considering policy issues and technical issues.
4. Discuss the role of IDRs in open access movement.
5. Compare retrieval features of BASE, Lens and Semantic Scholar as OA search services.

8.12 REFERENCES

Anuradha, K. T. (2005). Design and development of institutional repositories: A case study. *The International Information & Library Review,* 37(3), 169-178.

Arunachalam, S. (2004a, March). India's March Towards Open Access. In *Science and Development Network*. Retrieved February 10, 2011, from http://www.scidev.net/en/opinions/indias-march-towards-open-access.html

Arunachalam, S. (2004b). Open Access and the Developing World. *The National Medical Journal of India*, 17(6), 289-291.

Arunachalam, S. (2005, September). India moving ahead with open access. *Access*, 54. Retrieved June 10, 2011, from http://www.aardvarknet.info/access/

Crow, Raym. (2002). The Case for Institutional Repositories: A SPARC Position Paper. (Washington, DC: Scholarly Publishing & Academic Resources Coalition). Available from <http://www.arl.org/sparc/IR/IR_Final_Release_102.pdf>.

Phelps, Charles E. (1998) "Achieving Maximal Value from Digital Technologies in Scholarly Communication." In *The Proceedings of the 133rd Annual Meeting of theAssociation of Research Libraries*. Available from <http://www.arl.org/arl/proceedings/ 133/phelps.html>.

Pinfield, Stephen (2001). "How Do Physicists Use an E-Print Archive?" *D-Lib Magazine* 7 (12).

Pinfield, Stephen, Mike Gardner, and John MacColl (2002) "Setting up an institutional eprint archive" Ariadne 31. Available from <http://www.ariadne.ac.uk /issue31 /eprintarchives/intro.html>.

Rusbridge, Chris, and William J. Nixon (2001) "Setting up an institutional ePrints archive—what is involved?" Unpublished paper, UKOLN Meeting (July 11, 2001). Available from <http://www.lib.gla.ac.uk/eprintsglasgow.html>.

Van de Sompel, Herbert and Patrick Hochstenbach (1999) "Reference Linking in a Hybrid Library Environment." *D-Lib Magazine* 5 (4).

Wyly, Brendan J. (1998). "Competition in Scholarly Publishing? What Profits Reveal." *ARL* 200. Available from <www.arl.org/newsltr/200/wyly.html>