MLIS S1-04

ONE YEAR MASTER'S DEGREE COURSE

(CBCS in Accordance with NEP 2020)

SEMESTER - I

DIGITAL LIBRARIES

LIBRARY AND INFORMATION SCIENCES

Guru Prasad Das School of Interdisciplinary and Vocational Studies
KRISHNA KANTA HANDIQUI STATE OPEN UNIVERSITY

Subject Experts

Prof. Sanjay Kumar Singh, DLIS, Gauhati University, Guwahati

Dr. Dipen Deka, DLIS, Gauhati University, Guwahati

Dr. Deepjyoti Kalita, DLIS, Cotton University, Guwahati

Course Co-ordinator: Komal Kumari, KKHSOU

SLM Preparation Team

UNITS	CONTRIBUTORS
1, 2, 3, 4, 13	Mr. Pranjit Kalita, KKHSOU
5, 6, 12, 14	Mr. Manash Kumar, Swahid Jadav Nath Homoeopathic Medical College & Hospital
9,10,11	Dr. Deepjyoti Kalita, Cotton University
7,8	Dr. Nishant Kashyap Ghatowar, Bikali College

Editorial Team

Content	:	Mr. Dibyajyoti Patgiri, KKHSOU
		Ms. Komal Kumari, KKHSOU
Structure, Format & Graphics	:	Pranjit Kalita, KKHSOU

August, 2026 ***ISBN: 978-93-7839-015-9***

This Self Learning Material (SLM) of the Krishna Kanta Handiqui State Open University is made available under a Creative Commons Attribution-Non Commercial-Share Alike 4.0 License (international): http://creativecommons.org/licenses/by-nc-sa/4.0/

Printed and published by Registrar on behalf of the Krishna Kanta Handiqui State Open University

The University acknowledges with thanks the financial support provided by the **Distance Education Bureau**, **UGC** for the preparation of this study material.

Headquarter : Patgaon, Rani Gate, Guwahati - 781017

City Office : Resham Nagar, Khanapara, Guwahati-781022; Website: www.kkhsou.in

CONTENTS

UNIT 1: INTRODUCTION TO DIGITAL LIBRARY 7-14

Definition, Historical Evolution of DL, Differences from Traditional/Virtual Libraries, Need of DL, Components of DL System, DL Services

UNIT 2: TECHNICAL INFRASTRUCTURE, ARCHITECTURE AND STANDARDS OF DL 15-26

Hardware, Software, Networking, Client-Server Models, DL Architecture; Components and Models, Cloud Computing and DL, Standards for Communication and Interoperability (Z39.50, OAI-PMH).

UNIT 3: COLLECTION DEVELOPMENT IN DL 27-34

Digital Collection; Types, Digitization and Its Benefits, E Resources: Selection Criteria and Evaluation, Licensing Aspects, Knowledge Organization in DL: Need, Challenges, Levels, Tools, Knowledge Organization in Traditional Vs Digital Library.

UNIT 4: PLANNING, IMPLEMENTATION, INFRASTRUCTURE AND HRM IN DL 35-45

Planning of DL; Information Needs, Quality in Plan, Security, IT Infrastructure, Digitization: Need, Determining Materials, Process, Tools, Access, Staffing and Budget, Key Guidelines for Creating DL System. Requirements And Feasibility, Human Resource Planning, IFLA/UNESCO Manifesto for DL.

UNIT 5: BROWSING AND SEARCH INTERFACE IN DL 46-54

Searching in DL, Browsing in DL, Emerging trends of searching and browsing in Digital Library

UNIT 6: SEMANTIC WEB AND LINKED DATA IN DL 55-66

Evaluation of Semantic Digital Library (SDL), Core services Provided by SDL Core Technologies and Architecture of SDL Example of SDL Initiatives, Application of Linked Data in DL: Benefits and Challenges

UNIT 7: DIGITAL PRESERVATION & CONSERVATION IN DIGITAL LIBRARY 67-77

Concept of Digital Preservation, Preservation Strategies in Digital Library, Challenges of Digital Preservation, Digital Library Conservation; Aspects of digital conservation, Role of Librarians in digital preservation and conservation, Metadata Standards for digital curation (PREMIS, METS) Digitization techniques, file formats, long term storage, Cloud Archiving and Blockchain

UNIT 8: OPEN-SOURCE TOOLKITS **78-86**

Comparative Studies Between Popular Open-Source Software for Digital Libraries: DSpace, GSDL, E-print, Omeka, Benefits of Open-Source Software Of DL, Popular Examples DLs Based on Open-Source Software: Indian and Global.

UNIT 9: REGISTRY AND ITEM MANAGEMENT IN DIGITAL LIBRARY **87-98**

Registries, Digital Library Registries, Functions of Digital Library Registries, Digital Library Handle System, DOI as a handle system, Integrating Author Identification services to DL/IL, Orcid Integration to DSpace

UNIT 10 CASE STUDY OF DL SOFTWARE: E PRINT **99-108**

E print software, Key features, Versions of E-Prints, System Requirements for E-Print installation, E-Print Workflow

UNIT 11: CASE STUDY OF DL SOFTWARE: GSDL **109-118**

Greenstone Digital Library Software (GSDL), Versions of Greenstone, System Requirements, Customization whereabouts in Greenstone2 and Greenstone3, Greenstone Librarian Interface (GLI), Advanced Management Features

UNIT 12: CASE STUDY OF POPULAR DL SOFTWARE: DSPACE **119-129**

Functional Overview: Data Model, Metadata, E-People, Authorization, Ingest Process and Workflow, Persistent URLs and Identifiers, Search and Browse, OAI Support, System Architecture, DL adopting DSpace

UNIT 13: DL EVALUATION AND MARKETING **130-137**

DL Evaluation: Concept And Need, Process of Evaluation, Criteria of Evaluation, Models of DL Evaluation, Marketing Of DL

UNIT 14: DL INITIATIVES AND RECENT TRENDS **138-148**

Popular Indian Initiatives: NDLI, E-Shodhsindhu, Vigyan Prasar Digital Library, NDBIAC, National Mission for Manuscripts, E-Shodhsindhu, TKDL, Kalasampada, ONOS; Global Initiatives: Project Gutenberg, NDLTD, NDSL, DLESE, ArXiv, Pandora, Europeana, Oxford Digital Library, Cambridge Digital Library, OAISTER.

INTRODUCTION

The course *"**Digital Libraries**"* is designed to provide an extensive understanding of the theoretical foundations, technological infrastructure, and managerial aspects essential for the development and maintenance of digital library systems. With the rapid advancement of information and communication technologies, libraries have transitioned from traditional repositories of physical materials to dynamic digital platforms that facilitate equitable access, retrieval, preservation, and dissemination of knowledge resources. This course aims to equip learners with the conceptual frameworks and practical competencies required to create and manage digital library environments in academic, research, and public domains.

The curriculum commences with an introduction to the definition and historical evolution of digital libraries, delineating their distinction from traditional and virtual libraries. Emphasis is placed on understanding the rationale, key components, architecture, and services of a digital library system. The technical dimensions, including hardware, software, network configuration, client-server architecture, and interoperability standards such as Z39.50 and OAI-PMH, are comprehensively addressed to familiarize learners with the underlying technological ecosystem.

Then we will learn about digital collection development, encompassing digitization processes, evaluation and selection of electronic resources, licensing considerations, and approaches to knowledge organization in digital environments. Planning and implementation processes are treated as critical components of digital library development, with attention to infrastructural requirements, budgeting, human resource management, information quality, and security measures. The course aligns with international guidelines, notably the IFLA/UNESCO Manifesto on Digital Libraries, to ensure adherence to global best practices.

Advanced topics include semantic web technologies, linked data applications, and the conceptualization of Semantic Digital Libraries (SDL). The modules on digital preservation and conservation address strategies, challenges, metadata standards such as PREMIS and METS, and emerging solutions including cloud archiving and blockchain-based preservation. Learners are also introduced to prominent open-source digital library software packages DSpace, Greenstone, E-Prints, and Omeka through comparative studies and detailed case analyses that highlight their architectures, workflows, and operational features.

The concluding sections encompass evaluation and marketing strategies for digital libraries, as well as a study of major national and international initiatives such as the National Digital Library of India (NDLI), e-ShodhSindhu, Project Gutenberg, Europeana, and ArXiv. Upon completion, learners will demonstrate a comprehensive understanding of the principles, technologies, and management strategies that underpin sustainable digital library systems, thereby preparing them to contribute effectively to the digital knowledge infrastructure in contemporary scholarly and institutional contexts.

UNIT 1: INTRODUCTION TO DIGITAL LIBRARY

1.1 Introduction

1.2 Learning Objectives

1.3 Definition

1.4 Historical Evolution of DL

1.5 Differences from Traditional/Virtual Libraries

1.6 Need of DL

1.7 Components of DL System

1.8 DL Services.

1.9 Answers to Check your Progress

1.10 Further Readings

1.11 Probable Questions

1.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- Learn about what digital library is, evolution, differences between traditional libraries, Digital, Virtual, and Hybrid libraries, understanding the unique features of each.
- Describe the essential components of a digital library system, such as the user interface, repository, and search mechanisms.

1.2 INTRODUCTION

We live in a time where there is an incredible explosion of information. The conventional library concept, which consists of a tangible collection of books and journals, is changing significantly. The Digital Library is a new paradigm in information storage and retrieval brought about by the quick development of computer and communication technology. This transition reflects a fundamental rethinking of how knowledge is arranged, conserved, and shared in the twenty-first century; it goes beyond a simple format change from print to pixels.

A digital collection of artifacts is only one aspect of a digital library. Digital content is produced, arranged, and made available to a worldwide user base in this intricately controlled environment. A digital library is a dynamic, interactive environment where

information is accessible around-the-clock, surpassing geographical and physical constraints, in contrast to the static shelves of a real library. It combines the flexibility and speed of contemporary technology with the strong organizational principles of classical librarianship to provide services ranging from straightforward document retrieval to intricate data analysis and individualized information delivery.

The necessity to manage the massive amount of information produced today, which frequently surpasses the physical capacity of conventional institutions, has led to the development of digital libraries. Furthermore, the public and academic community are calling for more immediate, remote access to information. By offering a smooth interface to a huge network of multimedia resources—text, audio, video, and data sets—that are accessible from a desktop computer or mobile device anywhere in the world, digital libraries respond to this demand.

We will examine the fundamental ideas of digital libraries in this unit. From the early, imaginative concepts of a “World Brain” to the complex, web-based systems of today, we shall chart their historical development. We will differentiate digital libraries from similar ideas such as virtual and hybrid libraries, look at the essential elements of a digital library system, and talk about the various services they provide. A thorough grasp of the reasons why digital libraries are crucial pieces of infrastructure for the contemporary information society will be provided by this excursion.

1.3 DEFINITION

It is difficult to define a “digital library” because different communities use the term in different ways.

For computer scientists, it is frequently seen as a distributed system of managed multimedia data or a sizable database.

Regarding Librarians: It is a digital extension of the collecting, organization, preservation, and access functions of a traditional library.

For publishers, it could just be an electronic journal collection or an online catalogue.

Clifford Lynch (1995) gave a thorough definition, saying that “a digital library is a system providing a community of users with coherent access to a large, organized repository of digital information and knowledge.”

The definition provided by Deegan and Tanner (2002) is similar: “A managed collection of information, with associated services, where the information is stored in

digital formats and accessible over a network."

Key Features of DL:

According to these definitions, a digital library has the following features:

- Managed collections are a carefully selected collection of digital items (text, video, and audio) with metadata rather than an unstructured network of links like the WWW.
- Associated Services: It offers browsing, search, and reference services that are comparable to those of a real library.
- Digital Format: Machine-readable formats are used to store information.
- Network Access: It can be accessed via an intranet or the internet.
- Community Focus: It caters to a certain user community.

1.4 HISTORICAL EVOLUTION OF DL

It took time for the idea of the digital library to emerge. It is the outcome of decades of innovative thinking and advancements in technology.

➢ **The 1930s–1960s Visionary Era:**

- In 1938, H.G. Wells proposed the "World Brain," an all-inclusive encyclopaedia.
- Vannevar Bush (1945): In his groundbreaking article "As We May Think," he discussed the "Memex," a hypothetical mechanical device that would hold documents, conversations, and books and enable users to connect them—a forerunner to hypertext.
- In his 1965 book "Libraries of the Future," J.C.R. Licklider predicted that computers would eventually take the role of physical libraries as the primary means of storing and retrieving knowledge.

➢ **The 1970s–1980s Automation Era:**

- Libraries started utilizing computers for housekeeping tasks including circulation and cataloguing (OPACs).
- The foundation for electronic information access was established by the introduction of CD-ROMs and online databases like DIALOG.

➢ **The 1990s–Present Digital Library Era:**

- Early 1990s: With the introduction of the World Wide Web, the phrase "Digital Library" became more widely used.

- Digital Library Initiatives (DLI): To create digital library technologies, the US National Science Foundation (NSF), NASA, and ARPA provided funding for extensive research initiatives in 1994.

Growth: Initiatives like the Gutenberg Project and the Internet Archive (1996) sped up the digitization of books in the public domain. Large-scale projects like the Digital Public Library of America (DPLA) and Google Books are available now.

1.5 DIFFERENCES FROM TRADITIONAL/VIRTUAL LIBRARIES

To comprehend their distinct identity, it is essential to differentiate digital libraries from other kinds of libraries.

1.5.1 TRADITIONAL VS. DIGITAL LIBRARY

Feature	Traditional Library	Digital Library
Collection	Physical (Print, Microfilm)	Digital (E-books, Audio, Video)
Storage	Physical Shelves (Space limited)	Electronic Servers (Huge capacity)
Access	Physical visit required; limited hours	Remote access (24/7); anywhere
Search	Card Catalogue / OPAC	Full-text search; Metadata search
Users	One user per item at a time	Multiple users per item simultaneously
Preservation		Conservation of physical material Migration and emulation of digital formats

1.5.2 DIGITAL VS VIRTUAL LIBRARY

Although both terms are frequently used interchangeably, they differ slightly.

Digital Library: Maintains, maintains, and conserves a targeted collection of digital items. The server is a “place” where content is stored. The Digital Library of India is one example.

Virtual Library: A virtual library is sometimes referred to as a “library without walls.” It doesn’t always contain the actual content. Rather, it serves as a gateway or portal, offering connections to dispersed resources situated on many servers worldwide. Although it may not directly oversee the content, it aggregates access. For instance, the WWW Virtual Library.

1.5.3 HYBRID LIBRARY

The majority of libraries nowadays are hybrids. They provide access to digital resources (e-journals, databases) in addition to maintaining a physical collection of books. They combine the digital and paper worlds.

1.6 NEED OF DL

The following compelling needs and benefits are driving the migration to digital libraries:

- **Information Explosion:** No single physical library could possibly purchase and house every publication due to the exponential explosion of knowledge. Large collections can be stored in a little amount of space thanks to digital storage.
- **Remote access:** It eliminates the requirement for users to physically visit a building. For distant learning and international research collaboration, information is provided to their desktop or mobile device.
- **24/7 Availability:** Digital libraries are "always on," serving patrons day and night, in contrast to traditional libraries that have set hours of operation.
- **Multiple Access**: Only one individual can utilize a popular physical book. Waiting lists are eliminated since hundreds of people can access a digital document at once.
- **Preservation:** By offering a superior digital replacement for everyday usage and shielding the original from deterioration, digitization aids in the preservation of rare and delicate items (manuscripts, antique maps).
- **Improved Search:** Unlike physical books, digital libraries enable deep text searching, which locates every instance of a word in a book.

1.7 COMPONENTS OF DL SYSTEM

Information is stored, managed, and retrieved by a number of interconnected parts that make up a functional digital library system.

1.7.1 INTERFACE (UI) FOR USERS

The digital library's "front door" is this. Users engage with the system through its web-based interface. A decent user interface (UI) should facilitate searching and browsing (by subject, author). It manages results display and digital object visualization (e.g., an image zooming tool or PDF viewer).

1.7.2 STORAGE SYSTEM (REPOSITORY)

The digital objects (files) and the metadata that goes with them are kept here, in the center of the digital library**.**

- **Digital Objects:** The real information (pictures, sounds, and text).
- **Metadata:** Title, Author, Date, and File Format are examples of metadata, which are details that characterize the content.

The integrity and security of these files are guaranteed by the repository.

1.7.3 HANDLE SYSTEM (SYSTEM IDENTIFIER)

Links can break in the digital world (Link Rot). Persistent identifiers, such as DOIs or Handles, are used by digital libraries to guarantee that a citation to a document is still valid even if the file is transferred to a new server.

1.7.4 SYSTEM FOR SEARCH AND RETRIEVAL

This part handles user inquiries and indexes the repository. It provides support for:

Searching particular fields, such as Author or Title, is known as metadata search.

- Full-Text Search: Looking at a document's whole contents.
- It ranks results according to relevancy using algorithms.

1.7.5 DRM (RIGHTS MANAGEMENT SYSTEM)

This component controls access privileges in accordance with copyright regulations. It allows actions (e.g., permitting viewing but not printing) and authenticates users (checking if they are authorized).

1.8 DL SERVICES

Digital libraries bring new, technologically advanced features while replicating many classic library services.

- **Search and retrieval:** The primary service is search and retrieval. To locate specific information, users can use straightforward keyword searches or intricate Boolean queries.
- **Browsing:** To replicate the experience of exploring actual stacks, users can go through collections using subject hierarchies or categories.

- **Reference Service:** To assist customers with complicated research questions, "Ask a Librarian" services are frequently offered via email, chat, or video conferencing.
- **Alerting Services (CAS/SDI):** When new information that interests them is published to the library, users can set up profiles to get email notifications (Alerts) (e.g., a new article on "Climate Change").
- **Personalization:** To organize their research, bookmark their favourite documents, and save searches, users can create "My Library" accounts.
- **Document Delivery:** Secure transmission of electronic documents to the user's device is known as document delivery.

CHECK YOUR PROGRESS

A. Fill In the Blanks

- The innovative author _____ first put up the idea of the "World Brain" in 1938, which was a forerunner to the concept of a digital library.
- A library that serves as a gateway or portal to dispersed resources housed on various servers—often without actually holding the content itself—is known as a _____ library.
- A digital library's _____ system makes use of persistent identifiers, such as DOIs, to guarantee that citations are still valid even if a file is transferred.
- The program known as _____ allows users to get automated email alerts whenever new materials that are relevant to their particular interests are uploaded to the library.

1.9 LET US SUM UP

The revolutionary idea of the Digital Library—a managed environment where data is kept in digital formats and accessible via a network—was presented in Unit 1. Digital libraries are structured repositories that offer curated collections and specialized services, surpassing the disarray of the World Wide Web. From the groundbreaking "Memex" of 1945 to the advanced, web-based systems of today, the unit documented their development.

We made a distinction between virtual libraries, which frequently serve as gateways rather than repositories, and traditional physical libraries, which emphasize digital access and storage. The proliferation of information, space limitations, and the necessity for remote access around-the-clock underscored the pressing need for digital libraries. Lastly,

we looked at the essential elements of a digital library system (search system, repository, and user interface) and the value-added services they offer, including full-text retrieval and tailored alerting, which empower users in the digital age.

✓ 1.10 ANSWERS TO CHECK YOUR PROGRESS

A. Fill In the Blanks

a. H.G. Wells, **b.** Virtual**, c.** Handle**, d.** Alerting Services

1.11 FURTHER READINGS

1. Arora, J. (n.d.). Digital Libraries. In *NDLI*. Retrieved December 16, 2025, from https://epgp.inflibnet.ac.in/epgpdata/uploads/epgp_content/S000021LI/P000038/M000344/ET/1483073244P8_M01.pdf
2. *Digital Libraries*. (n.d.). https://www.msuniv.ac.in/images/distance%20education/learning%20materials/ug%20pg%202023/pg%202021/Mlisc%202021/DLSE27%20-%20II%20Semester%20-%20Digital%20Libraries.pdf
3. *MODERN LIBRARY: AUTOMATED, DIGITAL ANDVIRTUAL*. (n.d.). https://nios.ac.in/media/documents/SrSecLibrary/LCh-003.pdf
4. *UNIT 7 DIGITAL LIBRARIES Structure*. (n.d.). https://egyankosh.ac.in/bitstream/123456789/25563/1/Unit-7.pdf

1.12 PROBABLE QUESTIONS

A. Short Answer Type Questions

a. Define Hybrid Library.

b. What is the Difference between a digital library and World Wide Web.

c. What are the components of Digital Library.

B. Long Answer Type Questions

a. Discuss about how the contemporary information world requires digital libraries. What particular issues do they resolve?

b. Describe a digital library system's architecture. Explain each significant component's purpose.

*** ***** ***

UNIT 2: TECHNICAL INFRASTRUCTURE, ARCHITECTURE AND STANDARDS OF DL

2.1. Learning Objectives

2.2. Introduction

2.3. Hardware, Software, Networking

2.4. Client-Server Models

2.5. DL Architecture; Components and Models

2.6. Cloud Computing and DL

2.7. Standards for Communication and Interoperability (Z39.50, OAI-PMH)

2.8. Let Us Sum Up

2.9. Answers to Check Your Progress

2.10. Further Readings

2.11. Probable Questions

2.1. LEARNING OBJECTIVES

After going through this unit, you will be able to,

· Explain the basic client-server models and architectural that organize these systems and list the necessary hardware and software components needed to construct a digital library.

· Explain about how cloud computing contributes to the scalability of modern libraries and show how communication standards such as OAI-PMH and Z39.50 allow for resource sharing and interoperability across various digital repositories.

2.2 INTRODUCTION

Building a digital library is similar to building a contemporary skyscraper in that it needs a solid foundation, a well-thought-out structure, and standard protocols to guarantee operation and safety. This “construction” in the context of digital libraries is mostly dependent on an intricate technical infrastructure. Digital libraries are characterized by their hardware capabilities, software effectiveness, and network connectivity, as opposed to traditional libraries, which are characterized by actual shelves and reading rooms.

The foundation of any digital library system is its technical infrastructure. It includes the actual devices (clients and servers) that store and show data, the software that controls

and retrieves this data, and the networks that link these elements to people all over the world. Furthermore, a digital library cannot function in a vacuum in order to be genuinely beneficial. Interoperability standards enable it to communicate with other libraries and systems.The “nuts and bolts” of digital libraries are covered in detail in this section. We will examine the client-server models that underpin these systems, the architectural frameworks that structure them, and the particular hardware and software needs. We will also look at how cloud computing is starting to provide scalable solutions and how important standards like OAI-PMH and Z39.50 enable various digital libraries to “talk” to one another. Anyone participating in the design, administration, or research of digital library systems must comprehend these technical facets.

2.3. HARDWARE, SOFTWARE, NETWORKING

A digital library’s hardware architecture falls into two primary categories: client-side components, which are what users utilize, and server-side components, which house the data.

2.3.1 HARDWARE

2.3.1.1 Server-side Hardware

The digital library’s main component is the server. Databases, digital content, and the software that powers library services are all hosted there.

· **Servers:** High-performance computers with strong processors and lots of RAM are called servers. A digital library may utilize different servers for different purposes, such as an object server to store the actual digital assets (PDFs, photos, and videos), a database server to manage metadata, and a web server to handle user traffic.

· **Storage Devices:** Large amounts of storage are necessary for digital libraries. RAID (Redundant Array of Independent Disks) technology, which distributes data across several drives for speed and data safety, is a popular solution. Libraries may use tape libraries or optical discs (DVD/Blu-ray) for long-term archiving.

· **Input Devices:** Libraries require top-notch scanners in order to digitize tangible resources. These include specialized book scanners (such as overhead planetary scanners) that can digitize delicate books without destroying their binding, as well as flatbed scanners for common documents.

2.3.1.2 Client-Side hardware

The gadgets people use to access the digital library are referred to here.

· **User Devices:** Personal computers (PCs), laptops, tablets, and increasingly smartphones are examples of user devices. These gadgets need to be able to view multimedia material and run contemporary web browsers.

· **Display:** Users who want to examine detailed photos or read lengthy texts should utilize high-resolution monitors.

2.3.2 SOFTWARE

It handles searches, oversees the collection, and shows the user the interface. Usually, a suite of complementary tools is utilized rather than a single software program that can handle everything.

2.3.2.1 Digital Library Management Software

This is the main program that combines different features. Among the examples are:

· **Greenstone:** An open-source software package used to create and share digital library collections.

· **DSpace:** It is frequently used to organize research output in academic institutional repositories.

· **EPrints:** Another well-liked open-source platform and is mostly used for scientific research archives.

2.3.2.2 Database Management System

The metadata, or catalogue information, is stored and arranged by the DBMS. The structured data that represents the digital objects is typically handled by relational database management systems (RDBMS), such as Oracle, MySQL, or PostgreSQL**.**

2.3.2.3 Search and Retrieval Software

This program handles user inquiries and indexes the content. It enables sophisticated Boolean queries and full-text searching. Due to its speed and adaptability, Apache Lucene or Solr are frequently used as the fundamental search engine in systems.

2.3.2.4 Image Processing and OCR

· **Image editing:** Scanned photos are cleaned up using programs like Adobe Photoshop or GIMP (cropping, color correction).

· **OCR:** Software such as ABBYY FineReader uses optical character recognition (OCR) to turn scanned text images into editable, searchable text, which makes the content

"readable" by the computer.

2.3.3 Networking

Remote access is made possible by the networking infrastructure that links the clients and servers.

- **Local Area Network (LAN):** Provides high-speed connectivity and internal administration for computers in the library building.
- **Internet connectivity:** In order to serve external consumers, a high-bandwidth internet connection is necessary. The amount of bandwidth needed varies depending on the type of media; streaming video requires a large amount of bandwidth, whereas text requires very little.
- **Network Device:** Network equipment include firewalls to defend servers against online attacks, switches to link devices, and routers to route data.

2.4. CLIENT-SERVER MODELS

The standard operating model for digital libraries and the Internet is the Client-Server Architecture. It assigns duties to the client, who requests the service, and the server, which provides the resource.

Client: The client is typically the web browser that the user is using, such as Chrome or Firefox. It makes a request to the server (such as a search query) and shows the results. Presentation logic and the User Interface (UI) are managed by the client.

Server: The request is processed by the server, a distant computer. After conducting a database search, it finds the pertinent digital object and returns it to the client. Processing logic and data management are handled by the server.

middleware: In order to facilitate communication between the client and server, particularly when multiple databases or operating systems are involved, "middleware" is frequently present. It guarantees that a Linux server can comprehend a query from a Windows client.

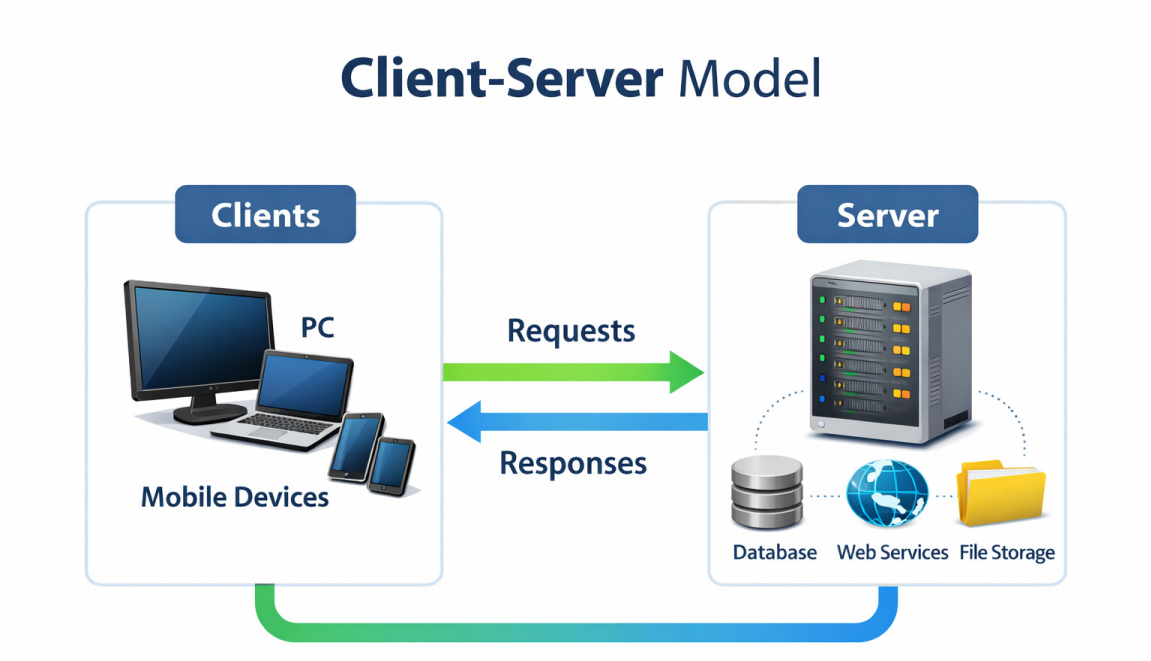


Figure 1: Client server Model

2.5. DL ARCHITECTURE; COMPONENTS AND MODELS

The conceptual design that specifies how the system's components work together is called digital library architecture.

2.5.1 KEY COMPONENTS

These five essential elements are present in most architectures:

- **User-Interface:** The front-end web portal where people browse and search.
- **Repository:** The safe place to keep metadata and digital items.
- **Handle System:** A method for giving digital things distinct, enduring identifiers (such as DOIs or Handles) so that links remain intact even when files are relocated.
- **Search System:** The data indexing and search query execution engine.
- **Rights Management System:** Regulates access according to user permissions and copyright (e.g., granting certain users view-only access).

2.5.2 ARCHITECTURE MODELS

DL design has been impacted by a number of theoretical models:

- **Kahn-Wilensky Architecture**

This is one of the fundamental frameworks for designing digital libraries, and it was first

put forth by Robert Kahn and Robert Wilensky in 1995. A distributed digital library system with a vast number of digital artifacts was intended to be supported by its design.

Fundamental Idea: The digital library is viewed by the architecture as a collection of dispersed "Digital Object Services." An object should be retrieved independently of where it is kept on the network since it essentially decouples the digital object (the content) from its location.

Key Contribution: The establishment of a naming convention for identifying and locating digital items is its most enduring contribution. In order to ensure that links to documents remain intact even if the file is transferred to a new server, it was stressed that each digital item must have a distinct, location-independent identity (a forerunner to the Handle System and DOIs used today).

- **Dienst Architecture and NCSTRL**

The Digital Library Research Group at Cornell University created the Dienst architecture, which is based on the ideas of federation and open architecture. The Networked Computer Science Technical Research Library (NCSTRL) is credited with implementing it.

Service Modules: Dienst divides the online library into four separate services that speak to each other using a predetermined protocol:

- **User Interface Service:** Provides a means of communication between users and the system.
- **Repository Service:** Digital documents are stored and made accessible through a repository service.
- **Indexing Service:** Gathers user queries and compares them with an index to provide pertinent results.
- **Collection Service:** Specifies the contents of the library and oversees how the other services communicate with one another.

Functionality: "Document Decomposition," a fundamental component of Dienst, enables a single document to be stored in many formats (such as ASCII, PostScript, and TIFF) and even divided into logical or physical components. Even though the data may be dispersed throughout more than 150 collaborating institutions, the system offers a unified display, allowing a user to perceive a single logical collection.

- **CRADDL (Cornell Reference Architecture for Distributed Digital Libraries)**

CRADDL (pronounced "cradle") is a Cornell University model that replaces previous versions. It moves the emphasis to an architecture that is built on components or services.

The "Lego" Approach: CRADDL specifies a collection of fundamental services that are interchangeable rather than a single, monolithic system.

- **Repository Service:** To store and deposit items.
- **Naming Service:** For the purpose of allocating URNs, or unique resource numbers.
- **Indexing Service:** To find items through queries.
- **Collection Service:** To combine items into useful clusters.
- **User Interface Gateway:** For communication between people.

Benefits: Libraries can quickly add additional, higher-level services, like automatic summarization or payment processing, thanks to its modular design, all without having to completely rebuild the system. It serves as an adaptable framework for the development of specialized library applications.

- **Federated Database Systems (FDBS)**

When independent libraries wish to collaborate while retaining their independence, they employ the Federated approach. The Networked Digital Library of Theses and Dissertations (NDLTD) and the National Science Digital Library (NSDL) are two examples.

Loosely-Coupled: This paradigm keeps the local database management systems used by participating libraries (components) intact. They are "loosely coupled" by means of a middleware layer that facilitates communication.

Interoperability Levels:

- **Federation**: Parties commit to rigorously adhering to common norms and procedures.
- **Harvesting**: Without complete integration, partners consent to make certain metadata available for shared services (such as a central search index).
- **Gathering:** The most basic level, in which a central system merely collects publicly available data from partners who might not actively collaborate using web crawlers.
- **CORBA (Common Object Request Broker Architecture)**

A distributed computing standard called CORBA makes it possible for software applications to collaborate regardless of where they are located or what language they are built in.

- **The mediator:** The Object Request Broker (ORB) is the middleware that CORBA employs. Consider the ORB to be a traffic cop and universal translator. When a client makes a request for a digital object, the ORB finds the appropriate item on a distant

server, routes the request, and transparently manages the communication details.

- **Platform Independence:** One of CORBA's main advantages is that it makes it possible to create a digital library with a variety of hardware and applications. Because CORBA manages the translation between a UNIX server and a Windows client, data requests from the latter can be made with ease.

- **Software Agents Architecture (UMDL)**

This cutting-edge approach, which is utilized by the University of Michigan Digital Library (UMDL), makes use of "Software Agents", self-governing programs that behave on behalf of users or collections.

Team of Robots: To accomplish tasks, the system's agents bargain with one another.

- **User Interface Agents (UIA):** The user is represented via User Interface Agents (UIA). After learning the user's preferences, they convert their questions into a format that the system can comprehend.

- **Collection Interface Agents (CIA**): Data is represented via Collection Interface Agents (CIA). They explain how to search a collection and what is contained inside it.

- **Mediation Agents (MA):** Managers are the Mediation Agents (MA). They route traffic while monitoring which agents are most qualified to respond to a given question.

Dynamic: The library system is intelligent and adaptable because to these agents' ability to "team up" dynamically to handle challenging issues.

- **OAIS (Open Archival Information System)**

OAIS emphasizes preservation, whereas many other models concentrate on access. For digital archiving, it is now the ISO standard (ISO 14721:2003).

- **Long-Term Focus:** It offers a structure for long-term information preservation, guaranteeing that it is comprehensible even as technology advances.

· **Preservation Metadata:** According to OAIS, "Preservation Description Information" (PDI) is essential and comprises the following:

- **Provenance:** The file's creator and its chain of custody.

- **Context:** The purpose of the file's creation and its relationships with other files.

- **Reference:** Unambiguous references to the material through unique identifiers (such as ISBNs).

- **Fixity:** Digital signatures or checksums to demonstrate that the file hasn't been corrupted or changed over time

2.6. CLOUD COMPUTING AND DL

Cloud computing encompasses the hardware and systems software in the data centres that offer those services as well as the applications that are supplied as services over the Internet. (Armbrust and others, 2009). Without needing institutions or individual users to invest in a substantial ICT infrastructure, it generates, manages, and uses information. Cloud computing-based digital libraries can make it easier for people to share computing and networking resources. To build up the repositories, individual institutions do not need to invest a lot of money in ICT infrastructure. Better access to research and academic literature is made possible by cloud-based services of open access repositories that connect with other specialized cloud research and content services. Additionally, amortizing the total cost of constructing the data centers for institutional repositories might lower the cost of open access digital library services. The economic and environmental sustainability of institutional repositories and open access materials can be greatly enhanced by cloud computing technology.

Benefits for digital library:

- **Cost Efficiency**: shifts IT spending from Capital Expenditure (CapEx - buying servers) to Operational Expenditure (OpEx - monthly subscription).
- **Elasticity:** If a library launches a popular new collection, traffic might spike. Cloud servers can automatically “scale up” to handle the load and “scale down” when traffic drops, ensuring the site doesn’t crash.
- **Disaster Recovery:** Cloud providers replicate data across multiple geographic locations. If the physical library is damaged by fire or flood, the digital collection remains safe in the cloud.

2.7. STANDARDS FOR COMMUNICATION AND INTEROPERABILITY (Z39.50, OAI-PMH)

The superpower of digital libraries is interoperability, which enables them to function as a single, massive worldwide network rather than as discrete islands. This link is based on two protocols: OAI-PMH (Harvesting) and Z39.50 (Search).

2.7.1 Z39.50

A well-established session-based protocol, Z39.50 (ISO 23950) is mostly utilized for

Federated Searching, which is the simultaneous search of several different live databases.

How It Operates: A "Client-Server" model, often referred to as "Origin" and "Target" is used.

- **Translation:** A user looks up "History of India" in the catalog of their local library. This query is converted into the standard Z39.50 format by the local system (Origin).
- **Broadcast:** This standardized query is simultaneously sent to five more remote libraries (Targets).
- **Retrieval:** After processing the query, the distant libraries return the results (bibliographic records).
- **Presentation:** The user sees a single, cohesive list of these external findings after the local system converts them back into its own format.

Key Features:

One of its main features is that it permits "semantic interoperability." Z39.50 ensures correct search results across various systems even if Library A uses the term "Author" while Library B uses "Creator," mapping them to a common attribute (e.g., Bib-1 Use Attribute 1003).

2.7.2 OAI-PMH

Z39.50 can be slow when querying hundreds of repositories at once, which is why OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) was created. OAI-PMH concentrates on metadata harvesting rather than live searches.

The "Data Provider" and "Service Provider" models are how it operates.

- **Data Provider (Repository):** Titles, authors, and abstracts are exposed by an institutional repository (such as one at a university) in a standard format, typically Dublin Core XML.
- **Service Provider (Harvester):** The repository is routinely asked, "What records have you added or changed since yesterday?" by a central search engine such as Google Scholar or OAIster.
- **Harvesting**: The harvester receives fresh metadata records from the repository.
- **Indexing:** Millions of entries from thousands of colleges are compiled into a vast central index by the harvester.

The six verbs OAI-PMH is extremely basic and operates using only six HTTP requests (verbs):

- **(Who are you?) Identify**
- **ListMetadataFormats: Which formats are supported?**
- **ListSets: What is the structure of your collection?**
- **ListIdentifiers (Display record IDs**
- **ListRecords (Please provide me with the real metadata)**
- **GetRecord (Provide me with a single record**

CHECK YOUR PROGRESS

A. Fill in the Blanks

a. Cornell University created the _____ architecture, a distributed digital library paradigm that divides the library into four distinct services: collection, indexing, repository, and user interface.

b. _____ is a standard protocol used for real-time federated searching across several distant library catalogs in the context of digital library interoperability.

c. An open-source digital library program called _____ is frequently used to build institutional repositories for the management and preservation of research results.

d. A low-barrier method for repository interoperability that enables service providers to gather metadata from data sources is the _____ Protocol for Metadata Harvesting.

2.8. LET US SUM UP

The strong technical underpinnings needed for contemporary digital libraries (DLs) were covered in Unit 2. We explored the essential hardware (servers, scanners, storage arrays) and software components (DL management systems like DSpace, RDBMS, and search engines like Lucene) that drive these systems. The unit detailed the pivotal Client-Server architecture, where users' browsers (clients) interact with powerful remote servers to retrieve information.

We analyzed various architectural models, including the Kahn-Wilensky framework (emphasizing unique identifiers), Dienst (distributed services), and Federated systems (connecting autonomous libraries). The unit also highlighted the transformative role of Cloud Computing (SaaS, IaaS, PaaS) in offering scalable, cost-effective storage and services. Finally, we discussed critical interoperability standards: Z39.50 for real-time cross-database searching and OAI-PMH for efficient metadata harvesting, ensuring that disparate digital libraries can function as a unified global network

✓ 2.9. ANSWERS TO CHECK YOUR PROGRESS

A. Fill in the Blanks:

a. Dienst, **b.** Z39.50, **c**. DSpace, **d.** OAI-PMH

🔍 2.10. FURTHER READINGS

Arora, J. (2025). *Components of a Digital Library – Digital Libraries*. Inflibnet.ac.in. https://ebooks.inflibnet.ac.in/lisp8/chapter/components-of-a-digital-library/

DIGITAL LIBRARIES: AN OVERVIEW Structure. (n.d.). IGNOU. Retrieved December 19, 2025, from https://egyankosh.ac.in/bitstream/123456789/25604/1/Unit-4.pdf

Digital Library Architecture – Digital Libraries. (2024). Inflibnet.ac.in. https://ebooks.inflibnet.ac.in/lisp8/chapter/digital-library-architecture/

Open Access and Digital Library – Digital Libraries. (2025). Inflibnet.ac.in. https://ebooks.inflibnet.ac.in/lisp8/chapter/open-access-and-digital-library/

Sanchati, R., & Kulkarni, G. (2011). *Double Blind Peer Reviewed International Research Journal Publisher: Global Journals Inc*, *11*. https://globaljournals.org/GJCST_Volume11/6-Cloud-Computing-in-Digital-and-University.pdf

? 2.11. PROBABLE QUESTIONS

A. Short Answer type questions

a. What role does the “Handle System” play in the construction of a digital library?

b. Contrast “Harvesting” with “Federation” when interoperability becomes closer.

c. Provide a quick explanation of “Infrastructure as a Service” (IaaS) in relation to online libraries.

d. What are the five key components of a Digital Library Architecture.

B. Long Answer Type Questions

a. Discuss about the Kahn-Wilensky digital library architecture. What are its fundamental components and primary contributions to the creation of digital libraries?

b. Describe how standards contribute to interoperability in digital libraries. Examine and contrast the OAI-PMH and Z39.50 protocols.

*** ***** ***

UNIT 3: COLLECTION DEVELOPMENT IN DL

UNIT STRUCTURE

3.1. Learning Objectives

3.2. Introduction

3.3. Digital Collection; Types, Digitization and Its Benefits

3.4. E Resources: Selection Criteria and Evaluation, Licensing Aspects

3.5. Knowledge Organization in DL: Need, Challenges, Levels, Tools

3.6. Knowledge Organization in Traditional Vs Digital Library

3.7. Let Us Sum Up

3.8. Answers to Check Your Progress

3.9. Further Readings

3.10 Probable Questions

3.1. LEARNING OBJECTIVES

After going through this unit, you will be able to,

- Classify the different categories of digital collections and comprehend the advantages of digitalization.
- Implement distinct criteria for the selection and assessment of electronic resources.
- Comprehend the licensing frameworks related to the procurement of digital content.
- Explain the necessity, obstacles, and tiers of knowledge organization inside digital libraries.
- Analyse and differentiate the structure of knowledge in conventional libraries compared to digital libraries.
- Identify the instruments employed for structuring knowledge in a digital context.

3.2 INTRODUCTION

Libraries have consistently been the core of academic and scientific institutions, functioning as the principal store of knowledge. In the digital era, the notion of a library has transformed from a tangible location containing books and journals to a virtual platform providing access to an extensive range of electronic material. This revolution has required a change in the development and management of collections. Collection creation in a digital library encompasses not only the procurement of physical things but also the

selection, licensing, and arrangement of digital resources. This lesson examines the complexities of constructing a digital collection, encompassing the many types of digital materials and the criteria employed for their selection and assessment. We will examine the essential elements of knowledge organization, which guarantees that these extensive digital libraries stay accessible and beneficial to users.

3.3. DIGITAL COLLECTION; TYPES, DIGITIZATION AND ITS BENEFITS

A digital library is not only an assemblage of content in electronic format; it is a system that offers a community of users' cohesive access to a vast, structured reservoir of information and expertise. The efficacy of a digital library mostly hinges on the caliber and pertinence of its digital collection.

3.3.1 TYPES OF DIGITAL RESOURCES

Digital collections can be classified into two primary categories based on their source. The initial category consists of "born-digital" resources. These items were initially produced in a digital format, including electronic journals, e-books, online databases, and websites. The second category comprises "digital surrogates," generated by transforming conventional physical resources (such as paper books, manuscripts, or pictures) into digital representations via a process known as digitization.

In an academic context, a digital collection may encompass a diverse array of resources:

- **E-journals and E-books:** These are digital formats of conventional serials and monographs, frequently equipped with sophisticated search functionalities.
- **Databases:** Organized assemblies of data, such bibliographic databases or full-text repositories, frequently available through the Internet or CD-ROM.
- **Theses and Dissertations:** Numerous institutions currently archive research outputs digitally to facilitate broader distribution.
- **Multimedia:** It encompasses digital audio, video segments, and graphics that can be incorporated into the educational process.

3.3.2 DIGITIZATION AND ITS BENEFITS

Digitization refers to the conversion of analog information into a digital format. This is an essential endeavour for libraries aiming to render their rare or distinctive physical collections accessible to a wider audience. The advantages of digitization are numerous:

- **Enhanced Accessibility:** Digital collections provide instantaneous access to high-demand products for multiple users concurrently, irrespective of their geographical location.
- **Preservation:** It serves as a preservation method by generating a surrogate for delicate or scarce originals, minimizing physical handling and so prolonging the lifespan of the original object.
- **Improved Searchability:** In contrast to traditional literature, digital collections allow for keyword searches, facilitating substantially faster and more accurate information retrieval.
- **Space Efficiency:** Digital resources eliminate the need for physical shelf space, presenting a significant benefit for libraries encountering spatial limitations.
- **Resource Sharing:** It enables the virtual consolidation of disparate collections, facilitating more efficient resource exchange among libraries.

3.4. E RESOURCES: SELECTION CRITERIA AND EVALUATION, LICENSING ASPECTS

Creating a digital collection necessitates stringent selection and assessment procedures. In contrast to traditional books, which have unequivocal ownership, e-resources can entail intricate access and license arrangements.

3.4.1 SELECTION CRITERIA AND EVALUATION

The selection of electronic resources is governed by criteria akin to those used to traditional materials, supplemented by extra technical and legal factors. Librarians must guarantee that the chosen materials align with the institutional objective and fulfil user requirements.

- **Content Quality:** The resource must deliver precise, credible, and impartial information. The credibility of the author or publisher serves as a crucial determinant of quality.
- **Relevance and Scope:** The content must be pertinent to the library's intended audience. Selectors must assess the comprehensiveness of coverage, timeliness of information, and whether the resource addresses a deficiency in the current collection.
- **Functionality and Accessibility:** Functionality is essential for electronic resources. Selectors must evaluate the user interface for usability, search functionality, and compatibility with current hardware and applications. They must also evaluate if the

resource facilitates remote access for off-campus users.

- **Cost and Sustainability:** The overall cost encompasses not only the membership charge but other concealed expenses such as personnel time for training, maintenance, and technical support. Librarians must evaluate if the resource provides adequate value relative to alternative goods.
- **Technical Support:** The provision of dependable technical support from the vendor is crucial for maintaining continuous access.

3.4.2 LICENSING ASPECTS

Licensing constitutes a unique aspect in digital collection development. In contrast to acquiring a physical book, libraries frequently "lease" access to digital literature via licensing agreements. Comprehending these models is essential:

- **Subscription versus Perpetual Access:** In a subscription arrangement, the library remits a periodic cost for access. Upon cessation of the subscription, access is typically forfeited. Conversely, a perpetual license entails a singular payment that confers onto the library enduring rights to utilize the information, akin to possessing a book.
- **Site licenses:** It permit all users within a designated physical location, such as a university campus, or a certain IP range to access the resource. This is typical for academic libraries.
- **Consortial Licensing:** Libraries frequently establish consortia to negotiate more favorable pricing and conditions with vendors. This joint bargaining power enables individual libraries to obtain more resources than they could independently afford.
- **Rights and Restrictions:** Licenses govern the permissible usage of the item. Librarians must verify that the license authorizes critical actions such as printing, downloading, and utilizing materials for interlibrary loans or electronic reserves.

3.5. KNOWLEDGE ORGANIZATION IN DL: NEED, CHALLENGES, LEVELS, TOOLS

Knowledge organization encompasses the processes of categorizing information for efficient retrieval. The complexity of a digital library arises from the extensive volume and diversity of digital objects.

3.5.1 NEED FOR KNOWLEDGE ORGANIZATION

The principal objective of knowledge organization is to facilitate efficient information retrieval for users. A digital library lacking a systematic system would devolve into a disorganized collection of files. Efficient organization allows users to navigate collections, identify specific papers, and comprehend the connections among various pieces of information. It converts a simple aggregation of files into a functional library.

3.5.2 CHALLENGES OF KNOWLEDGE ORGANIZATION

Structuring digital information poses distinct difficulties:

Interoperability: Digital libraries frequently encompass resources from diverse sources exhibiting multiple formats. Facilitating communication across these varied systems (semantic interoperability) is a significant challenge.

Granularity: In conventional libraries, the primary unit of organization is typically a book. In digital libraries, it may consist of an article, a chapter, an image, or a singular dataset. Determining the appropriate level for metadata application is challenging.

Volume and dynamism: The large volume of digital content and its rapid growth render conventional manual categorizing approaches impractical. Innovative, automated methodologies are frequently necessary.

3.5.3 LEVEL OF KNOWLEDGE ORGANIZATION

Information within a digital library is structured at various tiers:

- **Database Level:** Knowledge is organized into tables and fields within a database management system.
- **Document Structure:** Information within a document is systematically arranged into chapters, sections, and paragraphs. Markup languages such as XML facilitate the definition of this internal structure.
- **Metadata Level:** This refers to the descriptive information pertaining to a document, such as title, author, and date. Metadata is allocated to each digital asset to enhance discovery across networks.

3.5.4 TOOLS OF KNOWLEDGE ORGANIZATION

Digital libraries utilize specialized tools for organizational management.

- **Metadata Schemas:** Standards such as Dublin Core or MARC21 offer a uniform method for describing digital items. A metadata schema delineates the components (fields) utilized to characterize a resource.

- **Unique Object Identifiers**: To consistently access a digital thing, it requires a distinctive, enduring designation. Common identifiers encompass the Digital Object Identifier (DOI), the Handle System, and the Persistent Uniform Resource Locator (PURL). These guarantee that even if a file is transferred to a new server, the link to it remains operational.

- **Markup Languages**: Instruments such as XML (Extensible Markup Language) are employed for "descriptive markup." They designate sections of a document (e.g., <title>, <author>) instead than merely regulating its appearance, enabling computers to interpret the content more effectively.

3.6. KNOWLEDGE ORGANIZATION IN TRADITIONAL VS DIGITAL LIBRARY

The primary objective of arranging content is consistent to enhance accessibility yet the techniques and instruments vary considerably between traditional and digital libraries.

Classification versus Collections

Traditional libraries employ physical classification systems, such as the Dewey Decimal Classification (DDC), to organize volumes on shelves, thereby grouping related themes together. This facilitates physical browsing. In a digital library, geographical location is inconsequential. Knowledge is instead structured into "collections" or communities (e.g., the research output of a certain department). A singular digital object can concurrently belong to numerous collections without necessitating physical duplication.

Cataloguing versus Metadata

Conventional cataloguing depends on standardized codes such as AACR2 to generate card catalogs or OPAC entries. These entries frequently pertain solely to the description of the physical item (book). Digital libraries include metadata formats, such as Dublin Core, which offer more flexibility and can delineate many levels of granularity, including images, text, and video. Moreover, digital libraries employ many forms of metadata: descriptive (for finding), structural (for traversing intricate items), and administrative (for preservation and rights management), while traditional cataloging predominantly emphasizes descriptive components.

Identification and Retrieval

A user utilizes a “Call Number” (Class Number + Book Number) to find a book in a conventional library. In a digital library, retrieval is managed by unique identifiers such as DOIs or Handles. Moreover, whereas traditional libraries predominantly depend on designated subject headings for retrieval, digital libraries frequently facilitate full-text searching, enabling users to locate a document based on any term present in its text, rather than solely the cataloguer’s keywords.

Access Points

Conventional systems provide restricted access points (Author, Title, Subject). Digital libraries offer numerous access points. Users may search by file type, creation date, format, or specific technical metadata. The digital environment facilitates “semantic interoperability,” enabling use

CHECK YOUR PROGRESS

A. Fill in the Blanks

a. Resources that are originally developed in a digital format, such as e-books and webpages, are known as __resources.

b. The process of transferring analog information into a digital format to create a surrogate is termed _____.

c. A ____license entails a singular payment that confers upon the library enduring rights to utilize the information, akin to possessing a book.

d. In contrast to conventional libraries that employ call numbers, digital libraries utilize unique identifiers such as the _____ to search and retrieve specific digital objects.

3.7. LET US SUM UP

a.

b

3.8. ANSWERS TO CHECK YOUR PROGRESS

A. Fill in the Blanks

a. Born digital

b. Digitization

c. Perpetual

d. Digital Object Identifier

3.9. FURTHER READINGS

1. Hodge, G. (2000). *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. https://www.clir.org/wp-content/uploads/sites/6/pub91.pdf

2. Knowledge Organization in Digital Libraries – Digital Libraries. (2025). Inflibnet.ac.in. https://ebooks.inflibnet.ac.in/lisp8/chapter/knowledge-organization-in-digital-libraries/

3. SELECTION CRITERIA FOR TRADITIONAL AND ELECTRONIC RESOURCES. (n.d.). Retrieved December 25, 2025, from https://www.loc.gov/catdir/bibcontrol/selection_criteria.pdf

4. Swamy, H. M. C. (2005). *DIGITAL COLLECTIONS IN ACADEMIC LIBRARIES*. IR@INFLIBNET. https://ir.inflibnet.ac.in/server/api/core/bitstreams/be9a46c8-3e4e-4552-aabe-26e1b171c9e3/content

3.10 PROBABLE QUESTIONS

A. Short Answer Type Questions

a. Distinguish between “born-digital” resources and “digital surrogates.”

b. Define Perpetual license in case of digital library.

c. What is the role of metadata in digital library?

B. B. Long Answer Type Questions

a. Examine the difficulties of Knowledge Organization in Digital Libraries in contrast to Traditional Libraries.

b. Detail the criteria for selecting electronic materials, emphasizing the significance of licensing.

*** ***** ***

UNIT 4: PLANNING, IMPLEMENTATION, INFRASTRUCT-URE AND HRM IN DL

4.1 Learning Objectives

4.2 Introduction

4.3 Planning of DL; Information Needs, Quality in Plan, Security, IT Infrastructure

4.4 Digitization; Need, Determining Materials, Process, Tools, Access, Staffing and Budget

4.5 Key Guidelines for Creating DL System, Requirements and Feasibility

4.6 Human Resource Planning

4.7 IFLA/UNESCO Manifesto for DL

4.8 Let Us Sum Up

4.9 Answers to Check Your Progress

4.10 Further Readings

4.11 Probable Questions

4.1 LEARNING OBJECTIVES

After going through this unit, you will be able to,

- Comprehend the extensive planning procedure for a digital library, encompassing feasibility and requirements study.
- Ascertain the informational requirements of users and guarantee excellence in the strategizing of digital services.
- Explain the elements of IT infrastructure necessary for a digital library, encompassing hardware, software, and networking.
- Explain the digitization process, encompassing the selection of content, utilization of certain tools, and facilitation of access.

4.2 INTRODUCTION

The creation of a digital library (DL) is a multifaceted undertaking that goes beyond the simple procurement of equipment. It denotes a methodical procedure for establishing a structured store of knowledge, available without physical limitations, to cater to a designated user community. Planning and executing a digital library entail a comprehensive strategy that incorporates collection creation, technological infrastructure, financial management,

and human resource planning. Digital libraries markedly differ from traditional libraries in their storing, retrieval, and dissemination methods, necessitating a distinct array of tactics to guarantee sustainability and efficacy. This subject examines the pivotal stages of planning, the complexities of infrastructure and digitalization, and the fundamental importance of human resource management in establishing an effective digital library system.

4.3 PLANNING OF DL; INFORMATION NEEDS AND FEASIBILITY, SECURITY, IT INFRASTRUCTURE

Planning is the cornerstone for establishing a digital library. It is a methodical procedure that converts a concept into reality by identifying needs, evaluating feasibility, and delineating implementation options

4.3.1 INFORMATION NEEDS AND FEASIBILITY

Prior to initiating a digital library project, it is crucial to define its objective and significance. This entails recognizing the intended user demographic and comprehending their particular informational requirements. To delineate the project's scope, it is imperative to address inquiries such as

"Who will utilize it?",

"What categories of resources will it encompass?", and

"What methods will be employed for access?".

A feasibility study is essential to evaluate the project's viability, taking into account aspects such as document volume, user demand, and the availability of technical skills and financial resources.

4.3.2 QUALITY IN PLAN

An exemplary strategy is thorough and anticipatory. The document must specify the project's objectives, outputs, and priorities, ensuring alignment with the institution's strategic goals. Quality planning encompasses:

- **Collection Development**: Identifying resource kinds (born-digital versus digitized surrogates) and establishing selection criteria based on intellectual merit and user requirements.
- **Managerial Planning:** Organizing tasks, optimizing time management, and

tracking progress through methodologies such as PERT or SWOT analysis.

- **Financial Planning:** Assessing expenditures for hardware, software, digitization, and continuous operations, guaranteeing the budget is feasible and sustainable.

4.3.3 SECURITY

Security within a digital library framework is complex, encompassing the safeguarding of digital assets and the confidentiality of users.

- **Data Protection:** Digital assets must be safeguarded against inadvertent or deliberate alteration, loss, or theft. This entails resilient backup systems and stringent access controls.
- **Intellectual Property Rights (IPR):** Planning must consider copyright limitations and authorizations. Mechanisms for managing intellectual property rights, including authentication and permission, are crucial to guarantee that only authorized users can access exclusive content.
- **Cybersecurity:** Given the dependence on networks, digital libraries must be protected from cyber risks such as hacking and virus, hence maintaining the system's integrity and availability.

4.3.4 IT INFRASTRUCTURE

IT infrastructure constitutes the foundation of a digital library. It includes the hardware, software, and network elements required to store, manage, and distribute digital content.

- **Hardware:** This encompasses servers for content hosting, storage devices (such as RAID arrays or optical jukeboxes) for data preservation, scanners for digitization, and workstations for personnel and users.
- **Software:** Digital library software solutions, whether open-source or commercial, are necessary for collection management, metadata creation, and user interface provision. Examples encompass DSpace, EPrints, and Greenstone.
- **Networking:** A resilient network infrastructure, encompassing Local Area Networks (LAN) and high-speed Internet connectivity, is essential for facilitating remote access. Bandwidth requirements must be strategically managed to accommodate traffic, particularly for multimedia content.
- **Hosting:** Institutions must choose between self-hosting, necessitating substantial local infrastructure and experience, or utilizing cloud-based services that provide scalability

and maintenance assistance.

4.4 DIGITIZATION; NEED, DETERMINING MATERIALS, PROCESS, TOOLS, ACCESS, STAFFING AND BUDGET

Digitization is the process of transforming analogue information (such as books, manuscripts, and photographs) into digital representations (bits and bytes). It is a fundamental activity in the development of digital collections, separate from the acquisition of born-digital resources.

4.4.1 NEEDS FOR DIGITIZATION

The principal catalysts for digitization are:

- **Enhanced Accessibility:** It enables several people to concurrently access the same document from faraway locations, transcending physical limitations.
- **Preservation:** It aids in safeguarding rare and delicate things by producing high-quality replicas, hence minimizing the necessity to manipulate the originals.
- **Searchability:** Digitized text can be analysed using Optical Character Recognition (OCR), rendering the content entirely searchable and greatly augmenting its utility for research purposes.

4.4.2 DETERMINING MATERIALS

Not all entities require or are amenable to digitization. The selection process is a crucial phase determined by parameters including:

- **Intellectual Value:** Materials possessing significant scientific merit, distinctiveness, or demand are given precedence.
- **Physical Condition:** Things deemed fragile may be prioritized for preservation, whereas stable things may be deferred.
- **Copyright Status:** Public domain materials are more readily digitizable than copyrighted materials, which necessitate authorization.

4.4.3 PROCESS

The digitization process comprises numerous discrete phases:

- **Scanning:** Employing scanners (flatbed, drum, or overhead) or digital cameras to capture the image of a physical object. Resolution (dpi) and bit depth (color or grayscale)

are determinants of quality.

- **Image Processing:** Improving scanned images by cropping, de-skewing, and modifying contrast. Compression methods, such as lossless formats like TIFF and lossy formats like JPEG, are employed to regulate file sizes.
- **OCR (Optical Character Recognition):** OCR software, such as ABBYY FineReader, transforms photographs of textual documents into editable text, facilitating full-text search capabilities.
- **Metadata Development:** Allocating descriptive, structural, and administrative metadata to guarantee the digital entity is discoverable and manageable.
- **Storage and Retrieval:** Archiving files in a repository and connecting them to a user interface for accessibility.

Access:

Facilitating access entails not merely uploading files online but also guaranteeing their usability. This encompasses the creation of intuitive interfaces, compatibility with several file formats (PDF, HTML, JPEG), and the execution of search functionalities. Access policy must cover authentication for restricted materials and adherence to intellectual property rights legislation.

Staffing and Budget Allocation

Digitization requires substantial resources.

Staffing: Staffing necessitates a team with varied expertise, comprising project managers, scanning operators, metadata specialists, and IT personnel. Training is frequently necessary to enhance the competencies of current library personnel.

Budget: Expenses encompass capital expenditures for equipment (scanners, servers) and software, in addition to operational costs for personnel, training, and maintenance. Financial planning must consider the long-term expenses associated with data migration and storage to guarantee sustainability.

4.5 KEY GUIDELINES FOR CREATING DL SYSTEM

Establishing an effective digital library system necessitates compliance with established protocols and a comprehensive grasp of requirements and feasibility. These recommendations provide a framework for addressing the intricacies of technical, legal, and operational difficulties.

- **Define Clear Objectives:** The project must have a clearly defined mission, whether it is to preserve heritage, support research, or improve public access. This vision guides all subsequent decisions.
- **Adhere to Standards:** Use open standards for metadata (Dublin Core, MARC), file formats (TIFF, PDF/A), and interoperability (OAI-PMH). This ensures that the digital library can exchange data with other systems and remains accessible over time.
- **Prioritize User Needs:** The system design should be user-centric. Regular user assessment and feedback mechanisms should be integrated into the planning and implementation phases to ensure the system meets actual needs.
- **Ensure Sustainability:** Plan for the long term. This means securing ongoing funding, planning for technology refresh cycles (hardware/software upgrades), and ensuring digital preservation strategies are in place to prevent data loss.
- **Manage IPR and Ethics:** strictly adhere to copyright laws and ethical guidelines regarding data privacy and cultural sensitivity. Clear policies on access rights and reuse must be established.

Requirements and Feasibility analysis:

A thorough requirements study precedes the technological implementation. It encompasses:

- **Technical Specifications:** Assessing the requisite bandwidth, storage capacity (expandable to petabytes if necessary), and computational power.
- **Functional Requirements:** Specifying the system's capabilities searching, browsing, displaying, downloading, and restricting access according to user roles.
- **Feasibility Study:** This serves as a "reality assessment." It evaluates the feasibility of the proposed project in light of the constraints.

➢ **Technical Feasibility:** Whether the technology available and sufficiently developed?

➢ **Economic Viability:** Whether the financial allocation adequate for initial and recurring expenses?

➢ **Operational Feasibility:** Whether the institution equipped with the necessary personnel and organizational culture to implement this new system?

4.6 HUMAN RESOURCE PLANNING

The "human" component frequently serves as the paramount success determinant in

a digital library initiative. Establishing and sustaining a digital library necessitates a transformation in competencies and responsibilities relative to conventional librarianship.

Competencies, Skills and Responsibilities

A digital library team comprises multiple disciplines. Essential responsibilities encompass:

- **Project Managers:** To supervise the complete lifecycle, administer finances, and facilitate coordination among various teams.
- **Digital Librarians:** Tasked with selection, metadata development, and strategies for digital preservation. Technical Specialists: System administrators, software developers, and network engineers responsible for maintaining the IT infrastructure.
- **Digitization Personnel**: Technicians proficient in operating scanners, image processing software, and OCR technologies.
- **Legal and Intellectual Property Rights Experts:** To manage copyright permissions and licensing agreements. Personnel Acquisition and Development Organizations frequently encounter a deficiency in skills.

Human resource planning must tackle this by:

- **Training:** Offering ongoing professional development for current personnel in domains such as digital curation, metadata standards, and software utilization.
- **Recruitment:** Employing new personnel possessing particular technological competencies absent in the existing library workforce.
- **Outsourcing:** For particular, time-sensitive projects such as mass digitalization, it may be more practical to engage specialized contractors instead of employing temporary personnel, while quality control remains an internal obligation.

4.7 IFLA/UNESCO MANIFESTO FOR DL

- **Closing the Digital Divide**

The digital divide fundamentally represents an information split or difference. Bridging this gap is essential for attaining the “Millennium Development Goals (MDGs)” established by the United Nations. Access to information and communication facilitates health, education, cultural, and economic advancement. Information transmission facilitates lifelong learning and informed societal engagement. Equal access to cultural and scientific heritage constitutes a fundamental human right. Libraries provide tranquillity, ethical principles, education, and cultural comprehension. Digital library services bridge cultures

across geographical and socioeconomic divides.

- **The Concept of Digital Libraries**

A digital library is an online repository of verified digital assets. Digital items are generated or acquired and administered in accordance with globally recognized standards. Access is granted in a logical, sustainable, and service-oriented manner. Digital libraries provide an essential component of library services, employing technology to facilitate access to digital holdings. Collections are generated, managed, and accessed efficiently for specific populations. Collaborative digital libraries facilitate networks of public and academic libraries. Interoperability between partner systems is crucial. Digital libraries augment digital archives and preservation efforts.

- **Objectives of Digital Libraries**

The objective is to furnish immediate, organized, and authoritative access to informational resources. Digital libraries integrate information technology, education, and culture within library services.

Objectives of Digital Libraries. Facilitate digitization, accessibility, and preservation of cultural and scientific heritage. Facilitate access to informational resources while upholding intellectual property rights. Develop interoperable systems utilizing open standards. Advocate for the function of libraries in establishing uniform standards and optimal practices. Guarantee perpetual accessibility to digital resources. Connect digital libraries with high-velocity research and development networks. Leverage the convergence of communication media to produce and distribute digital content.

- **Content Generation, Accessibility, and Conservation**

Digital libraries necessitate content that is either digitized or inherently digital. Numerous nations have implemented national digitization initiatives, endorsed by IFLA. Digitization facilitates the establishment of virtual collections spanning continents. Digitization also facilitates the preservation of deteriorating original materials. Digitized products and born-digital resources necessitate long-term preservation. All digital library initiatives must incorporate preservation strategies sanctioned by relevant authorities. Digital libraries facilitate the comprehensive life cycle of information and knowledge. Interoperability and sustainability constitute fundamental principles. Open standards and protocols enhance global information dissemination.

- **Executing the Manifesto**

IFLA implores governments and organizations to acknowledge and endorse digital

libraries. Extensive digitization enhances accessibility to cultural and scientific resources. Legislation and financial assistance are crucial for sustainability. Long-term strategies should prioritize education, literacy, culture, and access to information. Bridging the digital gap encompasses the incorporation of information literacy into educational courses. There is a necessity for awareness that a significant amount of precious heritage remains undigitized. Libraries ought to partner with cultural and scientific heritage institutions. Consultation with rights holders and indigenous groups is essential. Digital libraries must guarantee equitable access, particularly for individuals with disabilities.

- **Digital Library Planning Requirements**

Availability of qualified personnel. Sufficient structures and amenities. Coordinated planning for libraries and archives. Sustainable funding mechanisms. Precise objective establishment. Conformity with national e-strategies as advised by the World Summit on the Information Society.

CHECK YOUR PROGRESS

A. Fill in the Blanks

a. Establishing and implementing a digital library necessitates a thorough strategy that encompasses collection development, technological infrastructure, financial oversight, and _____ planning.

b. A feasibility study assesses a digital library initiative by examining document volume, user demand, and the availability of _____ and financial resources.

c. Digital assets in a digital library must be safeguarded against manipulation, loss, or theft through robust backup systems and stringent ____ controls.

d. Optical Character Recognition (OCR) transforms scanned textual documents into ____ text to facilitate full-text searching.

4.8 LET US SUM UP

The unit explored the extensive domain of developing and executing a digital library. It commenced by recognizing that a digital library constitutes a sophisticated system necessitating careful planning, wherein identifying user requirements, guaranteeing quality, and doing feasibility analyses are essential. We examined the comprehensive IT infrastructure hardware, software, and networks that underpins these systems. Digitization was recognized as an essential procedure for improving access and preservation,

encompassing defined workflows from resource selection to optical character recognition and storage. We emphasized the importance of Human Resource Planning, asserting that proficient workers are the driving force behind any digital library. Ultimately, we discussed the IFLA/UNESCO Manifesto, which establishes the ethical and global framework for our endeavours, emphasizing that the primary objective of digital libraries is to ensure fair access to knowledge for all individuals.

✓ 4.9 ANSWERS TO CHECK YOUR PROGRESS

A. Fill in the blanks

a. Human Resource**, b.** Technical Skills**, c**. Access**, d.** Editable

4.10 FURTHER READINGS

1. *Digital Libraries Module: 10 Digitization Part - 1*. (n.d.). NDLI. Retrieved 2025, from https://epgp.inflibnet.ac.in/epgpdata/uploads/epgp_content/S000021LI/P000038/M001877/ET/1483076002P8_M10.pdf
2. *Digital Libraries Module: 11 Digitization Part - 2*. (n.d.). NDLI. Retrieved 2025, from https://epgp.inflibnet.ac.in/epgpdata/uploads/epgp_content/S000021LI/P000038/M001878/ET/1483076026P8_M11.pdf
3. *Digital Library Planning and Implementation – Digital Libraries*. (2024). Inflibnet.ac.in. https://ebooks.inflibnet.ac.in/lisp8/chapter/digital-library-planning-and-implementation/
4. *IFLA/UNESCO Manifesto for Digital Libraries*. (n.d.). https://www.ifla.org/wp-content/uploads/2019/05/assets/digital-libraries/documents/ifla-unesco-digital-libraries-manifesto.pdf
5. Module 2 ELECTRONIC RESOURCES. (n.d.). In *NIOS*. NIOS. https://nios.ac.in/media/documents/SrSecLibrary/LCh-008.pdf
6. *NDLI: Digital library planning and implemenation*. (2015). Iitkgp.ac.in. http://ndl.iitkgp.ac.in/he_document/inflibnet_epgp/inflibnet_epgp/IN__I_e_P_P__1_L_a_I_S__28899_P_0_D_I__29579_M_0_D_I_p_a_i__29595_29596
7. *UNIT 8 DIGITISATION CONCEPT AND NEED*. (n.d.). Egyankosh; IGNOU. Retrieved 2025, from https://egyankosh.ac.in/bitstream/123456789/26305/1/Unit-8.pdf

4.11 PROBABLE QUESTIONS

A. Short Answer type Questions

a. Define planning in the context of a digital library.

b. Discuss the importance of feasibility study before initiating a digital library project.

c. What is the role of security in a digital library system.

d. Discuss the types of professionals are required for digitization activities.

B. Long Answer Type Questions

a. Examine the essential planning elements necessary for the establishment of a digital library system.

b. Discuss IFLA/ UNESCO Manifesto for Digital Library.

*** ***** ***

UNIT 5: BROWSING AND SEARCH INTERFACE IN DL

UNIT STRUCTURE

5.1 Learning Objectives

5.2 Introduction

5.3 Searching in Digital Library

5.4 Browsing in Digital Library

5.5 Emerging Trends of Searching and Browsing in Digital Library

5.6 Let Us Sum Up

5.7 Answers To Check Your Progress

5.8 Further Readings

5.9 Probable Questions

5.1 LEARNING OUTCOME

After going this unit, learners will be to

- Know about the concept of Searching and Browsing.
- Know about the different types of search facilities available in digital library.

5.2 INTRODUCTION

Digital libraries play a crucial role in providing access to large and diverse collections of information. As the content continues to grow, effective mechanisms for information discovery have become very essential. Searching and browsing are two main approaches that enable the users to find and explore information within digital library systems.

Searching is typically query-driven and goal-oriented, allowing users to retrieve specific information. In contrast, browsing supports exploratory behavior by enabling users to navigate information through categories, links, and relationships. The integration of both approaches reflects the evolving nature of digital libraries, which have progressed from basic keyword retrieval systems to more advanced discovery tools incorporating metadata, faceted navigation, and semantic technologies.

Effective search and browsing are fundamental to information accessibility, helping users locate relevant resources with minimal barriers. They also significantly influence user satisfaction and engagement by supporting intuitive and flexible interactions.

5.3 SEARCHING IN DIGITAL LIBRARY

Searching is a primary mechanism for information retrieval in digital libraries, enabling users to find relevant resources by submitting specific queries to the system. Search-based information retrieval operates by matching user queries against indexed digital content, including metadata records and full-text documents, and presenting results ranked according to relevance. This approach is particularly effective when users clear about information need and seek specific documents or facts.

Search-based information retrieval is characterized by its query-driven and goal-oriented nature. Users must articulate their information needs using keywords, phrases, or structured expressions, which the system interprets through retrieval algorithms. The process is inherently interactive, as users often evaluate initial results and refine their queries to improve precision and recall. Relevance ranking, efficiency, and scalability are central features of search systems in digital libraries, given the size and diversity of digital collections.

Digital libraries typically support multiple types of searches to accommodate varying user expertise and information-seeking behaviors.

5.3.1 SIMPLE SEARCH

Simple search is the basic search mechanism in digital libraries. It enables users to retrieve information by entering one or more keywords into a single search box. The primary objective of simple search is to support quick and intuitive access to digital resources, particularly for non-expert users.

In digital library systems, simple search generally performs keyword-based retrieval across multiple fields such as title, author, subject, keywords, and often the full text of documents. The system automatically processes user queries using techniques such as stemming, stop-word removal, and relevance ranking, thereby reducing user effort.

Simple search interfaces are displayed on digital library homepages and follow a user-centered design approach. Platforms such as DSpace, Greenstone, and EPrints employ simple search as the default access point to their content. The search results are typically ranked according to relevance, enabling users to identify potentially useful documents quickly.

The major strengths of simple search include ease of use, rapid access to information,

and broad retrieval coverage. However, simple search has limitations, including low precision, lack of field-level control, and limited suitability for complex or systematic research tasks. Consequently, simple search is most effective when complemented by advanced search and browsing functionalities

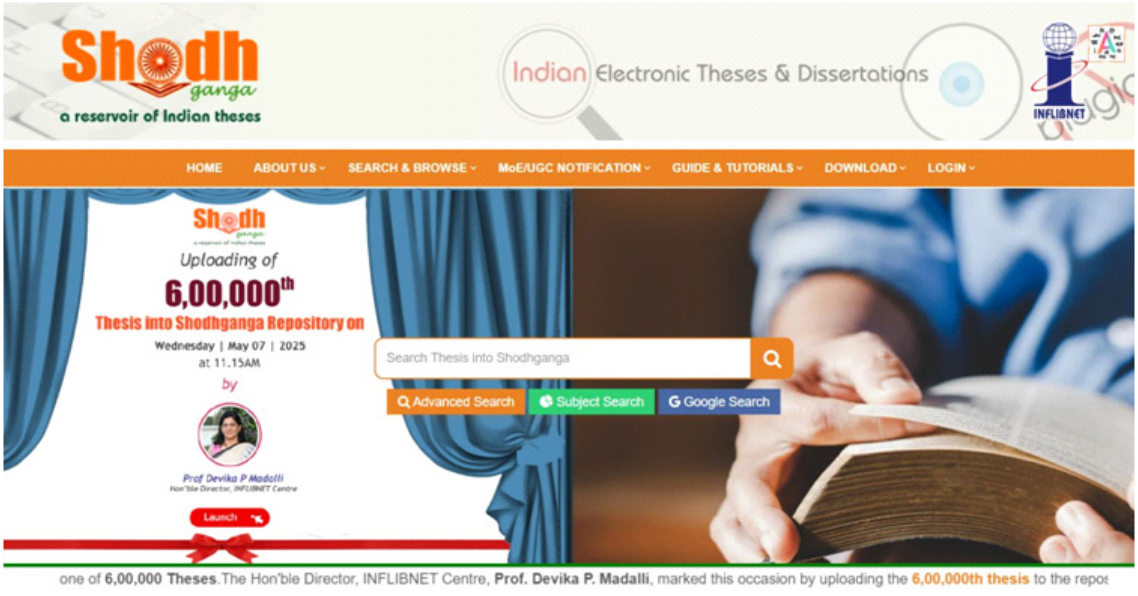


Figure 1: Simple/Basic search box in ShodhGanga (DSpace)

5.3.2 ADVANCED SEARCH

Advanced search is structured and feature-rich search mechanism that allows users to formulate precise and complex queries. Unlike simple search, advanced search provides greater control over the search process by enabling field-specific querying, Boolean logic, and multiple refinement options.

Advanced search interfaces typically allow users to search within specific metadata fields such as title, author, subject, abstract, publication date, and resource type. Boolean operators (AND, OR, NOT), phrase searching, truncation, and range searching are also generally supported. It allows users to narrow or broaden queries according to research needs. Many digital libraries also integrate filters and facets to further refine search results.

From an information retrieval perspective, advanced search improves precision by reducing ambiguity and limiting retrieval to relevant records. Digital library platforms such as DSpace, EPrints, and Greenstone implement advanced search using structured metadata and indexed fields, often supported by search engines such as Apache Solr.

Effective searching in digital libraries also depends on query formulation and refinement. Query formulation involves translating an information need into an appropriate set of

search terms, often requiring the selection of relevant keywords, synonyms, or controlled vocabulary terms. Query refinement is an iterative process in which users modify their initial queries based on system feedback and result analysis. Techniques such as adding or removing terms, applying field restrictions, and using filters contribute to improved retrieval performance. Together, query formulation and refinement support more accurate and efficient information discovery within digital library environments.

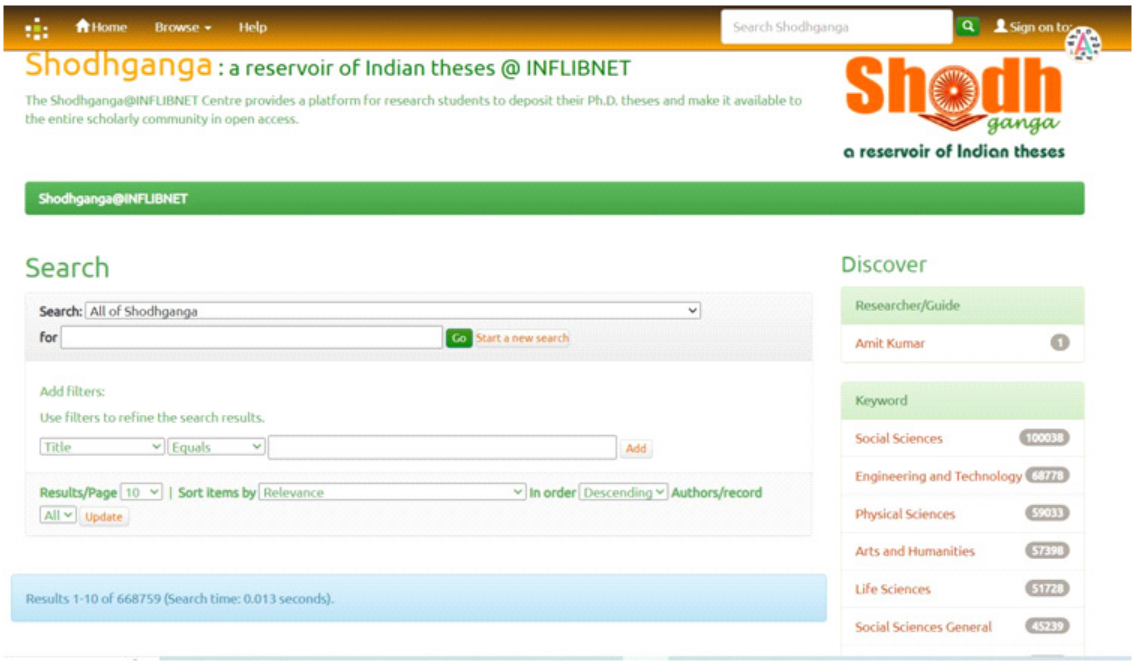


Figure 2: Advanced search interface in ShodhGanga (Dspace)

5.3.3 META SEARCH

Meta search in digital libraries refers to a search approach in which a single user query is simultaneously submitted to multiple information sources or digital libraries, and the retrieved results are aggregated and presented through a unified interface. Meta search enables cross-system discovery without requiring centralized storage of all content.

Meta search systems function by translating a user's query into formats compatible with different target databases and transmitting it to distributed search services. The results returned from these sources are then normalized, deduplicated, and ranked, before being displayed to the user. Common technologies supporting meta search in digital libraries include standard protocols such as Z39.50, SRU/SRW, and web-based APIs.

The primary advantage of meta search lies in its ability to provide comprehensive access to distributed digital collections, thereby enhancing resource visibility and saving

user time. However, meta search systems often face challenges related to response time, inconsistent metadata quality, limited control over relevance ranking.

5.4 BROWSING IN DIGITAL LIBRARY

Browsing is a fundamental information access mechanism in digital libraries that enables users to explore collections in a structured and systematic manner. Browsing supports exploratory information seeking and is particularly useful when users are uncertain about their information needs or unfamiliar with specific terminology.

In digital libraries, browsing is typically organized around logical and intellectual structures such as subject categories, classification schemes, author lists, titles, document types, communities and collections, or chronological arrangements. These structures are derived from metadata and reflect established bibliographic and knowledge organization principles. Digital library systems such as DSpace, Greenstone, and EPrints commonly provide browsing options based on author, title, subject, date of publication, and collection hierarchy.

Browsing interfaces facilitate gradual refinement of information through hierarchical navigation, allowing users to move from broad categories to more specific subsets. This approach enhances serendipitous discovery, as users may encounter relevant resources that would not be retrieved through keyword searching alone. Browsing is therefore particularly valuable in academic and cultural heritage digital libraries, where contextual understanding of collections is important.

From a usability perspective, browsing reduces cognitive load by presenting information in an organized and predictable manner. It complements search functionalities by offering an alternative access pathway, especially for novice users and interdisciplinary exploration. However, browsing may be less efficient for locating highly specific information.

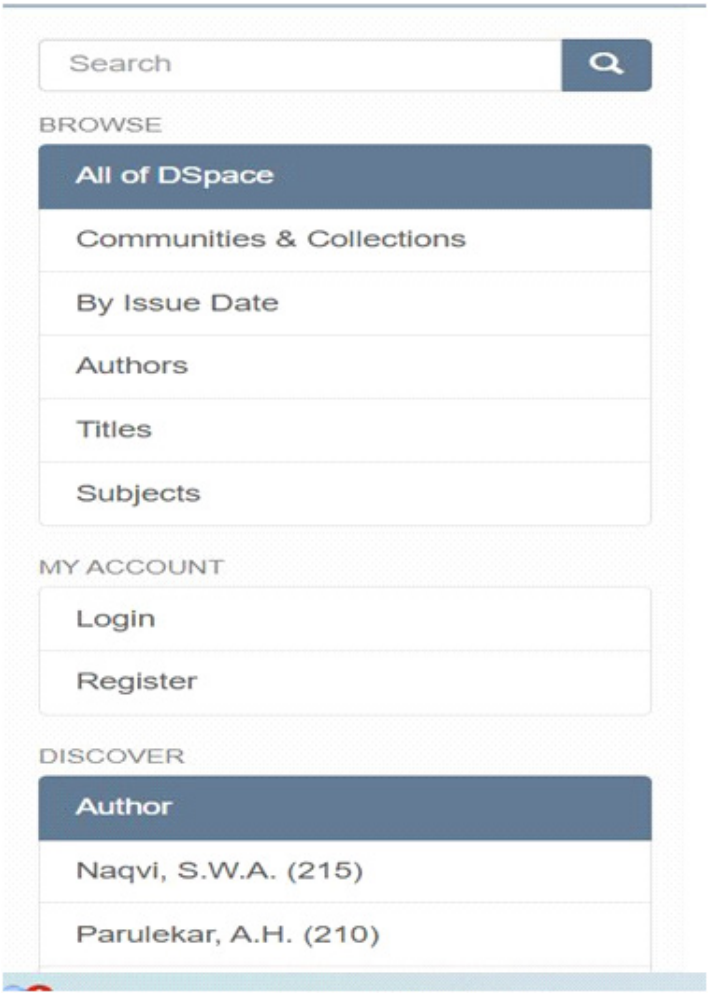


Figure 3: Browsing interface in DSpace Based Digital Library

5.5 EMERGING TRENDS OF SEARCHING AND BROWSING IN DIGITAL LIBRARY

The rapid advancement of information and communication technologies has significantly transformed search and browsing interfaces in digital libraries. Traditional keyword-based retrieval and static browsing structures are increasingly supplemented by intelligent, user-centric systems that enhance discoverability, relevance, and usability. Emerging trends in this area emphasize semantic understanding, artificial intelligence, natural language interaction, and personalized user experiences.

5.5.1 SEMANTIC AND CONCEPT-BASED SEARCH

Semantic and concept-based search represents a major shift from syntactic keyword matching to meaning-oriented information retrieval. This approach based on semantic

web technologies, ontologies, controlled vocabularies, and linked data to understand relationships among concepts, entities, and subjects. In digital libraries, semantic search enables users to retrieve relevant resources even when exact search terms do not match the indexed metadata or full text. By interpreting user intent and contextual meaning, semantic search improves precision and recall, particularly in multidisciplinary and research-oriented environments. Concept-based browsing further allows users to navigate knowledge structures through related terms and thematic relationships, supporting deeper exploration of digital collections.

5.5.2 AI-ASSISTED AND RECOMMENDATION-BASED DISCOVERY

Artificial intelligence has introduced advanced capabilities in search and browsing interfaces through machine learning and data-driven recommendation systems. AI-assisted discovery analyzes user behaviours, search history, and content similarity to suggest relevant documents, authors, or subjects. Recommendation-based interfaces, commonly used in scholarly digital libraries, support functions such as “related items,” “users also viewed,” and citation-based recommendations. These systems enhance serendipitous discovery and reduce information overload by guiding users toward contextually relevant resources. However, ethical considerations related to transparency, bias, and privacy remain important challenges in the deployment of AI-driven discovery tools.

5.5.3 VOICE AND NATURAL LANGUAGE SEARCH

Voice-enabled and natural language search interfaces reflect the growing adoption of conversational technologies and intelligent assistants. Unlike traditional query formulation, natural language search allows users to express information needs in everyday language. In digital libraries, this trend improves accessibility and inclusivity, particularly for users with limited technical skills or physical disabilities. Natural language processing techniques enable systems to interpret complex queries, questions, and contextual cues, thereby supporting more intuitive interaction with digital collections. While still emerging in academic libraries, voice-based search has significant potential for mobile and hands-free information access.

5.5.4 PERSONALIZATION AND ADAPTIVE INTERFACES

Personalization and adaptive interfaces aim to tailor search and browsing experiences

to individual user preferences, roles, and information needs. These systems dynamically adjust interface elements, search results, and navigation pathways based on user profiles, prior interactions, and contextual factors. In digital libraries, personalization enhances efficiency by prioritizing relevant content and reducing irrelevant results. Adaptive interfaces also support diverse user groups, such as students, researchers, and faculty members, by offering customized views and functionalities. Despite their benefits, personalized systems must balance usability with concerns related to data security and user autonomy.

CHECK YOUR PROGRESS

a) What is semantic search?

b) Name two protocols used in Meta search.

5.6 LET US SUM UP

- In conclusion, searching and browsing are complementary and crucial components of digital library access systems, each addressing distinct yet interrelated information-seeking behaviors. Searching facilitates precise and goal-oriented retrieval by allowing users to locate specific items or information through well-defined queries, supported by metadata, indexing, and relevance ranking mechanisms. In contrast, browsing enables exploratory interaction with digital collections, supporting contextual understanding, intellectual discovery, and serendipitous encounters through structured navigation and classification-based organization.
- When effectively integrated, searching and browsing provide a balanced and flexible access environment that accommodates all types of users and enhances overall usability. Collectively, these access mechanisms strengthen the capacity of digital libraries to function as robust and user-centered infrastructures for knowledge discovery, scholarly communication, and the dissemination of digital resources.

5.7 ANSWER TO CHECK PROGRESS

a) Semantic and concept-based search transcends keyword matching by emphasizing the significance of information. It comprehends relationships among concepts and entities through the utilization of semantic web technologies, ontologies, and linked data. This enables users to locate pertinent items in digital libraries, even in the absence of precise phrase matches. By interpreting user intent and context, it enhances search accuracy

and supports deeper exploration through concept-based browsing.

b) Z39.50, SRU/SRW, and web-based APIs.

5.8 FURTHER READINGS

1. Pradhan, D. (n.d.). *Search and browse interface in digital library*. In *Digital Libraries*. INFLIBNET Centre. Retrieved January 22, 2026, from https://ebooks.inflibnet.ac.in/lisp8/chapter/search-and-browse-interface-in-digital-library/
2. Smith, A. G. (2000). *Search features of digital libraries*. *Information Research, 5*(3), paper73. Retrieved January 22, 2026, from https://informationr.net/ir/5-3/paper73.html

5.9 PROBABLE QUESTIONS

A. Give short answer

1. Discuss difference between searching and browsing.
2. What is recommendation system.

B. Write short note

1. Boolean search.
2. Browsing.

C. Long Answer type Questions

1. Discuss different search options of content in context of digital library.

*** ***** ***

UNIT 6: SEMANTIC WEB AND LINKED DATA IN DL

UNIT STRUCTURE

6.1 Learning Outcome

6.2 Introduction

6.3 Evaluation Of Semantic Digital Library

6.4 Core Services Provided by Semantic Digital Library

6.5 Core Technologies and Architecture of the Semantic Digital Library

6.6 Example of Semantic Digital Library Initiatives

6.7 Application of Linked data in Digital library: Benefits and Challenges

6.8 Let Us Sum Up

6.9 Answers to Check Your Progress

6.10 Further Readings

6.11 Probable Questions

6.1 LEARNING OBJECTIVES

After going this unit, learners will be able to

- Know about the concept of the Semantic Digital Library as a natural progression in digital librarianship, addressing both theoretical and practical dimensions.
- Know about the various Semantic Digital Library initiative around the world.

6.2 INTRODUCTION

The recent transformation of libraries in this era has moved beyond the mere digitization of documents toward the development of intelligent, interoperable, and user-oriented information systems. Within this evolving landscape, the Semantic Digital Library (SDL) represents a crucial advancement that integrates digital library systems with Semantic Web technologies to enable meaning-based organization, retrieval, and knowledge discovery.

Traditional digital libraries predominantly rely on syntactic approaches to information retrieval, such as keyword matching and static metadata records. While effective for basic access, these approaches are limited in their ability to capture semantic relationships, contextual relevance, and conceptual linkages among information resources. As digital collections grow in scale and complexity, users increasingly require systems capable of

supporting advanced queries, interdisciplinary exploration, and intelligent information services. The Semantic Digital Library responds to these needs by embedding machine-understandable semantics into metadata, content, and services through technologies such as ontologies, RDF, OWL, linked data, and knowledge graphs.

6.3 EVALUATION OF SEMANTIC DIGITAL LIBRARY:

In traditional Digital Library collections are stored in digital formats and accessible by electronic devices over internet or LAN. Digital resources that can include text, images, audio, video, or other digital media. Objects can consist of digitized content, as well as born-digital. In addition to storing content, digital libraries provide means for organizing, searching, and retrieving the content contained in the collection.

Digital libraries have grown since the time of its origination in both numbers and volume of resources. Some of them are designed on top of operating systems such as Unix, Linux, or Windows. And many are implemented on top of relational database systems such as Oracle, SQL Server. Lately, the advent of open-source operating systems and database engines with enterprise level features and reliability has created a trend towards digital libraries implemented on top of Linux and MySQL.

With the explosive growth of information, Digital libraries have faced many issues like-

- Inconsistent & Incomplete Metadata: Different vocabularies, formats, and absence of uniform standards make resources difficult to find across systems.
- Siloed Information: Data is often isolated, making it difficult to link related items across different collections or platforms.
- Keyword-Based Search Limitations: Depends on exact word matches, missing relation among concepts and failing to understand user's intent.
- Poor Interoperability: Due to different data models, systems fail to share and understand data among them.

Semantic Digital Libraries (SDLs) address these issues by integrating traditional digital library services with Semantic Web technologies. It is an ecosystem where digital objects are not only stored as files but are described using formal, machine-processable metadata and ontologies.

Unlike traditional digital libraries that rely primarily on keyword-based indexing, a Semantic Digital Library enables machines to understand the meaning, context, and relationships between resources. This facilitates advanced knowledge discovery, seamless

interoperability between heterogeneous data sources, and more precise information retrieval for the end user.

6.4 CORE SERVICES PROVIDED BY SEMANTIC DIGITAL LIBRARY

- **Semantic Interoperability:** Uses ontologies to allow different systems to communicate and understand data consistently**.**
- **Contextual Search:** Develops retrieval strategies based on Semantic Web technology, elevating information retrieval to the semantic level. This enhancement significantly improves both precision and recall rates, making retrieval more intelligent and effective. By using semantic web technology SDLs understands the meaning of queries and resources, providing more relevant results by linking concepts (e.g., search query of "climate change impacts" yielding results of document discussing "global warming effects").
- **Automated Management:** Agents can automate tasks like metadata harvesting, quality checks, and resource organization, reducing manual labour.
- **Enhanced Discovery:** Offers smarter browsing, filtering, and personalized recommendations based on user profiles and interests.
- **Linked Data & Provenance:** Connects information objects, tracking their history, relationships, and origin.
- **Rich User Profiles:** Creates comprehensive, interoperable user profiles by integrating data from various sources, improving personalization.

6.5 CORE TECHNOLOGIES AND ARCHITECTURE OF THE SEMANTIC DIGITAL LIBRARY

Core Technology

By design, Semantic Digital Libraries transcend simple data storage to facilitate complex semantic representation, automated reasoning, and seamless interoperability. This evolution from static collections to intelligent knowledge environments has been fundamentally realized through the Linked Data paradigm and Semantic Web frameworks. Some of the core technologies are-

Uniform Resource Identifier (URI)

URIs provide a unique and unambiguous way to identify resources on the web. In the Semantic Web, not only documents but also real-world entities such as authors, books,

places, and concepts are identified using URIs

Resource Description Framework (RDF)

RDF is the fundamental data model of the Semantic Web. It represents information in the form of triples, consisting of:

- Subject
- Predicate
- Object

For example: *Book A – hasAuthor – Author B*. This simple structure enables flexible data representation and linking.

RDF Schema (RDFS)

RDFS provides a basic vocabulary for describing classes, properties, and hierarchical relationships among resources. It allows the creation of simple ontologies and supports inferencing such as class inheritance.

Ontologies and Knowledge Organization

An ontology is a formal, explicit specification of a shared conceptualization of a domain. Ontologies enable semantic interoperability by ensuring that different systems interpret data consistently.

Web Ontology Language (OWL)

OWL is a more expressive ontology language that allows detailed description of concepts, relationships, and constraints. It is used to model complex knowledge domains and supports advanced reasoning.

SPARQL

SPARQL is a query language designed to retrieve and manipulate data stored in RDF format. It plays a role similar to SQL in relational databases but is specifically tailored for graph-based data.

Linked Data

Linked Data refers to a set of best practices for publishing and interlinking structured data on the web in a way that it can be easily discovered and reused. Tim Berners-Lee proposed four fundamental principles of Linked Data:

- Use URIs to identify things.
- Use HTTP URIs so that people and machines can look them up.
- When a URI is looked up, provide useful information using standards such as

RDF and SPARQL.

- Include links to other related URIs to enable discovery of more data.

Linked Data transforms isolated data silos into a global, interconnected data space.

Architecture:

Kruk et.al proposed the architectural model of Semantic Digital Library drawn from analyses of systems like FEDORA, BRICKS, Greenstone, and JeromeDL, providing a blueprint for scalable, semantically enriched libraries. The architecture addresses dynamic, heterogeneous content and modern user expectations, through layered structures, vertical service groupings, and supporting infrastructures like SOA, P2P, and Grid. They presented the system architecture from three perspectives: top-down layered architecture, vertical architecture of core services, and stack of enabling infrastructures

1. Top-Down Layered Architecture Perspective

The top-down layered architecture defines the structural organization of a semantic digital library as a hierarchy of functional layers, each with clearly delineated responsibilities. This perspective emphasizes modularity and separation of concerns, ensuring that changes in one layer do not adversely affect others.

At the lowest level, the data source layer encapsulates heterogeneous repositories, including RDF triple stores for semantic metadata, conventional storage systems for digital objects, full-text indexes, and federated repositories. This layer is responsible solely for data persistence and retrieval.

Above this, the data abstraction layer provides a unified and technology-independent interface to underlying data sources. It abstracts differences among data models—such as RDF graphs, relational databases, and distributed indexes—allowing higher-level services to operate without direct dependency on specific storage technologies.

The data manipulation layer constitutes the core operational tier of the SDL. It implements essential business logic, including metadata management, semantic indexing, classification, ontology-based reasoning, and information retrieval operations. Semantic processing at this layer enables concept-based access and knowledge inference.

The data preparation layer performs validation, access control enforcement, and contextual adaptation of information prior to delivery. It ensures policy compliance, metadata consistency, and enrichment through integration with internal or external semantic services.

At the top, the user interface layer facilitates interaction between users and the system. It supports semantic search, faceted navigation, visualization of relationships, and adaptive

interfaces tailored to user profiles and contexts.

This layered perspective provides a systematic framework for embedding semantic technologies throughout the digital library lifecycle.

2. Vertical Architecture of Core Business Logic Services

The vertical architecture perspective focuses on the functional classification of services that operate across multiple layers of the SDL. Unlike the layered view, which is structural, this perspective is service-centric, emphasizing how business logic is delivered to users and administrators.

Four primary categories of services are identified:

Information Discovery Services enable user-facing functionality such as semantic search, browsing, faceted navigation, annotation, ranking, collaborative filtering, and context-aware discovery. These services leverage semantic relationships to improve precision and relevance.

Advanced Management Services support administrative and professional tasks, including ontology management, collection administration, user and role management, and workflow orchestration. These services are critical for maintaining semantic consistency and system governance.

Basic Services provide foundational operations required for system functionality, such as content ingestion, metadata creation, indexing, feature extraction, and access policy enforcement.

Interoperability Services facilitate integration with external systems and platforms through metadata harvesting, SPARQL endpoints, service mediation, and standard exchange protocols.

This perspective highlights that semantic digital libraries function as service-oriented ecosystems, where intelligent behaviour emerges from coordinated service execution rather than isolated components.

3. Stack of Enabling Infrastructure Technologies Perspective

The third perspective situates the semantic digital library within a supporting infrastructure stack that ensures system performance, scalability, and reliability. This perspective addresses non-functional requirements that are critical for large-scale semantic processing.

At the architectural level, Service-Oriented Architecture (SOA) principles govern the design of Semantic Digital Library services, promoting loose coupling, reusability, and

interoperability. RESTful API interfaces are particularly emphasized for efficient integration and service composition.

The Peer-to-Peer (P2P) infrastructure supports distributed storage and indexing across multiple nodes, enabling federated digital libraries. Semantic annotations enhance P2P efficiency through semantic overlay networks that optimize resource discovery.

The Grid computing infrastructure provides computational support for resource-intensive operations such as semantic reasoning, natural language processing, large-scale indexing, and multimedia feature extraction. By offloading heavy processing tasks, grid infrastructure ensures that advanced semantic services do not degrade system responsiveness.

6.6 EXAMPLE OF SEMANTIC DIGITAL LIBRARY INITIATIVES

The evolution of Semantic Digital Libraries has been shaped not only by theoretical models but also by experimental and operational systems that demonstrate the practical application of Semantic Web technologies in digital library environments. Such two initiatives are discussed below-

JeromeDL:

JeromeDL is a fully semantic digital library system designed to support collaborative knowledge creation and intelligent information retrieval. It integrates Semantic Web technologies directly into the core digital library workflow, rather than treating semantics as an auxiliary feature.

JeromeDL employs ontology-based metadata modelling, using RDF and OWL to represent bibliographic entities, user profiles, annotations, and relationships among resources. Metadata is stored in semantic repositories (triple stores), enabling reasoning and inference across collections.

The system supports:

Semantic annotations of documents,

Ontology-driven classification,

Integration of bibliographic and user-generated metadata.

Services and Functionality:

JeromeDL provides advanced semantic services, including:

Concept-based and relationship-aware search,

Collaborative annotation and tagging,

Recommendation services based on semantic similarity and user behavior.

Technical Significance:

JeromeDL is significant because it demonstrates a native semantic architecture, where metadata, users, and services are all represented as interlinked semantic entities. It serves as a reference implementation for research-oriented semantic digital libraries.

BRICKS Digital Library Infrastructure

BRICKS (Building Resources for Integrated Cultural Knowledge Services) is a European Project which aims at enabling integrated access to distributed resources in the cultural heritage domain. BRICKS is a distributed semantic digital library infrastructure designed to integrate heterogeneous digital library systems across institutional boundaries. Unlike JeromeDL, BRICKS focuses primarily on interoperability and federation.

BRICKS adopts a semantic mediation layer that enables mapping between different metadata schemas and domain ontologies. Instead of enforcing a single metadata standard, it allows multiple vocabularies to coexist and interoperate through semantic mappings.

Key features:

- Ontology alignment and schema mediation,
- Semantic service discovery,
- Cross-collection search and access.
- Architectural Approach

The system is implemented using a service-oriented architecture, where semantic services are loosely coupled and exposed through standardized interfaces. Peer-to-peer mechanisms support distributed indexing and retrieval.

Technical Significance:

BRICKS illustrates how semantic technologies can be applied at an infrastructure level to support large-scale, distributed digital library ecosystems. It is particularly relevant for national and international digital library networks.

6.7 APPLICATION OF LINKED DATA IN DIGITAL LIBRARY: BENEFITS AND CHALLENGES

Linked data is utilized in digital libraries to transform isolated information into web-accessible, interconnected entities that facilitate enhanced discovery and reuse.

Fundamental applications

- Distributing catalogues as RDF graphs utilizing URIs for works, authors, subjects, locations, and institutions, facilitating machine-actionable connections among data.
- Substituting or augmenting MARC with RDF-based frameworks like FRBR, FRAD, RDA, and BIBFRAME, enabling bibliographic and authority data to integrate into the Semantic Web.
- Facilitating extensive aggregation platforms: Europeana, DPLA, data.bnf.fr, VIAF, and WorldCat present their digital collections and authority files as linked data for inter-repository search and reutilization.
- Facilitating entity-centric discovery layers (e.g., Share-VDE, OCLC initiatives) that aggregate records based on works, creators, and subjects, rather than merely compiling title lists.
- Integrating repositories and CMS: DSpace and EPrints export items as RDF (N-Triples, RDF/XML, RDFa), while Drupal and WordPress use RDF modules and plugins (e.g., WordLift) to embed linked data in digital library portals.
- Integrating library data with external knowledge repositories like DBpedia, Wikidata, and schema.org to improve authority control, contextual relevance, and overall visibility in web searches.

Benefits.

- 24/7 web access to premium, curated library data, enhancing the exposure and reutilization of collections beyond local OPAC interfaces.
- Semantic enrichment of discovery through faceted browsing, related-works navigation, and links between works, expressions, manifestations, items, authors, subjects, places, events, and user annotations.
- Interoperability among institutions and systems: initiatives like Europeana, DPLA, data.bnf.fr, VIAF, and WorldCat linked data demonstrate how RDF/XML-based frameworks facilitate cross-repository aggregation and reutilization of metadata.
- Minimization of redundancy and isolated descriptions, as various representations of the same thing can be consolidated using common URIs and vocabularies.

Challenges:

- The reliance on MARC, with its inflexible, record-centric framework, is ill-suited for Web standards, rendering the transition to FRBR/BIBFRAME or other RDF models technically challenging.
- Insufficient comprehensive tools and methods for converting current catalogues

(MARC to RDF), harmonizing disparate schemas, and sustaining persistent URIs at scale.

- Need for substantial technical infrastructure: Web servers, digital preservation systems, SPARQL endpoints, and staff skilled in RDF, RDFS, OWL, SPARQL, and ontology design.
- Legal and organizational impediments, encompassing copyright, licensing, privacy, data ownership, and ambiguous laws about the commercial versus non-commercial reutilization of linked data.
- The presence of multilingual and heterogeneous data, along with the proliferation of vocabularies and ontologies, complicates schema selection, mapping, and worldwide interoperability.
- Notwithstanding these challenges, both assessments contend that linked data is strategically vital for libraries to maintain the visibility, interoperability, and relevance of their metadata in a Web 3.0 context.

CHECK YOUR PROGRESS

a) What is meant by a Semantic Digital Library?

b) What is linked data? Discuss its application in library systems.

6.8 LET US SUM UP

- The Semantic Digital Library represents a significant conceptual and technological advancement in the evolution of digital librarianship. By integrating Semantic Web technologies with established principles of knowledge organization, Semantic Digital Libraries transcend the limitations of traditional keyword-based systems and enable meaning-driven information representation, discovery, and reuse.
- Through ontology-based metadata, linked data frameworks, and machine-interpretable knowledge structures, Semantic Digital Libraries transform digital libraries from passive information repositories into intelligent knowledge environments.

6.9 ANSWER TO CHECK PROGRESS

a) A semantic digital library is an ecosystem where digital objects are stored not merely as files, but are accompanied by formal, machine-readable metadata and ontologies. This enables machines to comprehend the meaning, context, and interrelations of

resources, thereby facilitating advanced knowledge discovery, interoperability, and accurate information retrieval.

b) Linked Data refers to a set of best practices for publishing and interlinking structured data on the web in a way that it can be easily discovered and reused. Application includes:

i. Distributing catalogues as RDF graphs utilizing URIs for works, authors, subjects, locations, and institutions, facilitating machine-actionable connections among data.

ii. Substituting or augmenting MARC with RDF-based frameworks like FRBR, FRAD, RDA, and BIBFRAME, enabling bibliographic and authority data to integrate into the Semantic Web.

6.10 FURTHER READING

1. Kruk, S. R., & McDaniel, B. (Eds.). (2009). *Semantic digital libraries* (1st ed.). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-85434-0
2. Pradhan, D., & Patel, Y. (n.d.). *Digital library and semantic web*. In *Digital Libraries*. INFLIBNET Centre. Retrieved from https://ebooks.inflibnet.ac.in/lisp8/chapter/digital-library-and-semantic-web/
3. Behera, S. K., & Maharana, B. (n.d.). Libraries of the future. In *AN OVERVIEW OF LINKED DATA AND ITS APPLICATION IN LIBRARIES* (pp. 238–242). https://dspace.nitrkl.ac.in/dspace/bitstream/2080/4139/1/2023_MANLIBNET_SKBehera_AnOverview.pdf
4. *Client challenge*. (n.d.). https://www.scribd.com/document/606757450/Application-of-Linked-Data-Technologies-in-Digital-Libraries
5. Raza, M. (2019). Application of linked data technologies in digital libraries: a review of literature. *Library Hi Tech News*, *36*(3), 9–12. https://doi.org/10.1108/lhtn-10-2018-0067

6.11 PROBABLE QUESTIONS

1. What role do ontologies play in Semantic Digital Libraries?
2. Define linked data in the context of digital libraries.
3. Discuss the core technologies used in Semantic Digital Libraries.
4. Explain the metadata perspective of a Semantic Digital Library.
5. Describe any one Semantic Digital Library system (JeromeDL or BRICKS).

6. Discuss the architecture of a Semantic Digital Library, highlighting its layered structure.
7. Critically examine the role of Semantic Digital Libraries in enhancing information discovery and interoperability in modern digital libraries.

*** ***** ***

UNIT 7: DIGITAL PRESERVATION & CONSERVATION IN DIGITAL LIBRARY

UNIT STRUCTURE

7.1 Learning Objectives

7.2 Introduction

7.3 Concept of Digital Preservation

7.4 Preservation Strategies in Digital Libraries

7.5 Challenges of Digital Preservation

7.6 Digital Library Conservation

7.7 Aspects of Digital Conservation

7.8 Role of Librarians in Digital Preservation and Conservation

7.9 Metadata Standards for Digital Curation (PREMIS, METS)

7.10 Digitization Techniques

7.11 File Formats

7.12 Long-Term Storage

7.13 Cloud Archiving and Blockchain

7.14 Let Us Sum Up

7.15 Answers to Check Your Progress

7.16 Further Readings

7.17 Probable Questions

7.1 LEARNING OBJECTIVES

After this unit, learner will be able to,

- Understand the concept of digital preservation and conservation.
- Identify preservation strategies used in digital libraries.
- Explain the role of librarians and metadata standards in digital curation.
- Describe digitization techniques, file formats, and long-term storage solutions.

7.2 INTRODUCTION

Digital preservation and conservation are essential functions of modern digital libraries. They ensure that digital resources such as e-books, journals, theses, images, and

multimedia remain accessible, authentic, and usable over a long period of time. Due to rapid technological changes and risks of data loss, libraries must adopt systematic methods, standards, and tools to protect digital information for present and future users.

Digital preservation and conservation aims at ensuring long-term access to digital information resources. With the rapid growth of born-digital and digitized materials such as e-books, e-journals, research data, multimedia files, and institutional repositories, libraries face significant challenges related to technological obsolescence, data degradation, and format incompatibility. Digital preservation aims to ensure that digital information remains usable, authentic, accurate, and accessible for a long period of time. It involves the use of proper planning, policies, and technologies to manage digital materials effectively. Digital conservation focuses on protecting digital content from damage, loss, or unauthorized changes by using reliable storage systems, regular backups, and appropriate security measures. Together, digital preservation and conservation help libraries protect important intellectual and cultural resources and support continuous learning and research activities. As most information is now created and stored in digital form, effective preservation practices are essential to maintain knowledge resources and make them available to future generations of students and researchers

7.3 CONCEPT OF DIGITAL PRESERVATION

Digital preservation is the concept of ensuring long-term access to digital information by protecting it from technological obsolescence, media failure, and data loss. It is based on traditional preservation principles combined with modern digital technologies and standards. Libraries, archives, books, and records contain valuable information that must be preserved in durable and sustainable forms. This need has led to the development of digital preservation practices.

The first step in digital preservation is digitization, which involves converting printed or physical documents into electronic formats. Once digitized, these documents can be stored on servers and accessed by multiple users at the same time through the internet. Materials such as books, manuscripts, photographs, maps, and letters can be digitized for wider access and long-term use.

Digital preservation also includes the management of born-digital materials, such as e-journals, databases, websites, and multimedia resources, as well as digitized materials like scanned books and manuscripts. Preservation activities include format migration,

emulation, regular data backups, and the use of standard file formats, metadata creation, and integrity checks. Trusted digital repositories are used to ensure the authenticity, accessibility, and sustainability of digital resources.

Digital preservation is not a one-time task but a continuous process that requires proper policies, skilled professionals, and appropriate technologies. In digital libraries, it plays a vital role in preserving intellectual and cultural heritage and supporting education and research for present and future users.

7.4 PRESERVATION STRATEGIES IN DIGITAL LIBRARIES

Digital preservation strategies aim to ensure the long-term maintenance of digital content and continued access to it. No single strategy is suitable for all types of digital materials, so a combination of approaches is often required. Based on the lifespan of digital objects, preservation can be short-term, medium-term, or long-term. United Nations Educational, Scientific and Cultural Organization (UNESCO) groups digital preservation strategies into four broad categories.

- **Short-term strategies** include bit-stream copying, refreshing, replication, and backward compatibility, which mainly protect data from immediate loss due to hardware or media failure.
- **Medium- to long-term strategies** focus on maintaining usability over time through migration, emulation, canonicalization, and providing viewers at the point of access to address technological obsolescence.
- **Investment strategies** involve planning at the time of creation or archiving, such as restricting file formats, relying on open standards, data abstraction, and software re-engineering.
- **Alternative strategies** include analogue backups and digital archaeology for data recovery.

Since digital technologies change rapidly, effective digital preservation depends on using multiple strategies together to ensure authenticity, integrity, and long-term accessibility of digital resources.

7.5 CHALLENGES OF DIGITAL PRESERVATION

Digital preservation faces several challenges that affect the long-term accessibility of digital resources.

- **Technological Obsolescence** – Rapid changes in hardware, software, file formats, and standards make digital materials inaccessible over time.
- **Continuous Migration** – Digital resources must be regularly transferred to new formats and systems to keep them usable.
- **Fragility of Digital Media** – Storage media like CDs, DVDs, and magnetic tapes deteriorate quickly due to environmental factors.
- **Short Life Span of Storage Media** – Digital media have a much shorter lifespan compared to printed materials.
- **Dynamic Nature of Digital Content** – Websites, databases, and interactive resources change frequently, making preservation complex.
- **Lack of Policies and Legislation** – Inadequate legal frameworks and preservation policies hinder systematic preservation efforts.
- **Lack of Awareness and Skilled Professionals** – Limited knowledge and expertise affect effective digital preservation.
- **Insufficient Collaboration and Partnerships** – Poor coordination among institutions weakens long-term preservation.
- **Disaster Planning and Recovery Issues** – Absence of proper disaster management can result in permanent data loss.
- **Limited Financial Resources** – Insufficient funding makes it difficult for libraries to invest in advanced preservation technologies, infrastructure, and skilled personnel.

These challenges require careful planning, continuous monitoring, and institutional commitment to ensure effective digital preservation in libraries.

7.6 DIGITAL LIBRARY CONSERVATION

Digital library conservation refers to the protection and maintenance of digital resources to prevent loss, damage, or unauthorized alteration. It focuses on ensuring the safety and stability of digital materials such as e-books, e-journals, databases, images, and multimedia content.

Conservation in digital libraries involves the use of secure storage systems, regular data backups, virus protection, and access control mechanisms. It also includes monitoring digital files to detect errors or corruption and taking corrective measures when needed.

Unlike traditional conservation, digital conservation requires continuous attention due to rapid technological changes and security risks. Effective digital library conservation

supports digital preservation by maintaining the integrity and reliability of digital resources. By applying proper conservation practices, digital libraries can ensure uninterrupted access to information and protect valuable knowledge resources for students, researchers, and future generations.

7.7 ASPECTS OF DIGITAL CONSERVATION

Digital conservation involves several important aspects that help protect and maintain digital resources in libraries.

- One key aspect is *secure storage*, which ensures that digital files are stored in safe and reliable systems.
- *Data backup* is another essential aspect, allowing recovery of information in case of data loss or system failure.
- *Access control* and *security measures* help prevent unauthorized use, modification, or deletion of digital content.
- *Regular monitoring* and *integrity checks* are conducted to identify file corruption or errors at an early stage.
- The use of *standard file formats* supports long-term usability and reduces the risk of obsolescence.
- *Proper documentation and metadata* also play an important role in managing and conserving digital materials effectively.

Together, these aspects of digital conservation help maintain the safety, accuracy, and availability of digital resources, ensuring continuous access for users in digital libraries.

7.8 ROLE OF LIBRARIANS IN DIGITAL PRESERVATION AND CONSERVATION

Librarians play a crucial role in digital preservation and conservation by ensuring the long-term accessibility and protection of digital resources. They are responsible for developing and implementing preservation policies and guidelines for managing digital collections effectively. This includes selecting appropriate file formats, organizing metadata, and maintaining proper documentation of digital materials to support their authenticity and usability over time.

Librarians also oversee secure storage systems, conduct regular data backups, and implement access control measures to protect digital content from loss, corruption, or

unauthorized use. Monitoring the integrity and authenticity of digital resources is an ongoing task, which ensures that users can trust and rely on these materials for research, education, and other purposes.

In addition, librarians coordinate digitization activities, converting physical resources into digital formats for wider accessibility, and provide training to users on the proper use of digital resources. By continuously updating their technical skills and collaborating with information technology professionals, librarians help digital libraries tackle technological challenges effectively. Through these efforts, they ensure that digital knowledge resources remain reliable, usable, and accessible for present and future generations.

7.9 METADATA STANDARDS FOR DIGITAL CURATION (PREMIS, METS)

Digital curation refers to the active management of digital data throughout its life cycle, from creation and organization to long-term preservation and access. It ensures that digital resources remain usable, reliable, and meaningful over time. Digital libraries curate a wide range of data such as documents, images, research data, and multimedia files. To manage these resources effectively, metadata standards are essential. Metadata provides structured information about digital objects, helping in their identification, description, management, preservation, and retrieval. Without standardized metadata, digital resources become difficult to manage, preserve, and access over the long term.

PREMIS (PREservation Metadata: Implementation Strategies) is an international metadata standard specifically designed for digital preservation. It records essential information needed to maintain the authenticity, integrity, and usability of digital objects. PREMIS metadata includes details about digital objects, preservation actions, technical environment, and rights information. It helps libraries track changes, document preservation activities, and ensure long-term access.

The ***METS (Metadata Encoding and Transmission Standard)*** schema is a standard format used to encode descriptive, administrative, and structural metadata for digital objects in a digital library. It is expressed using the Extensible Markup Language (XML) schema language developed by the World Wide Web Consortium (W3C). The METS standard is maintained by the METS Board in collaboration with the Network Development and MARC Standards Office of the Library of Congress and was originally initiated by the Digital Library Federation (DLF). METS is widely used in digital libraries to manage complex

digital objects and support preservation and access.

7.10 DIGITIZATION TECHNIQUES

Digitization techniques refer to the methods used to convert physical materials into digital format for storage, access, and preservation. Commonly digitized materials include books, manuscripts, photographs, maps, and audio-visual resources.

- One important technique is *scanning*, where flatbed or overhead scanners are used to capture high-quality images of printed documents.
- For rare or fragile materials, *non-contact scanners* are preferred to prevent physical damage.
- *Optical Character Recognition (OCR)* is used to convert scanned text images into machine-readable and searchable text.
- For photographs and artworks, *high-resolution digital cameras* are often used.
- *Audio and video materials* are digitized using specialized equipment to convert analog signals into digital files.

Proper resolution, file format selection, and quality control are essential in all digitization processes. Effective digitization techniques improve access, support preservation, and help digital libraries manage and share information efficiently.

7.11 FILE FORMATS

File formats define how digital information is stored and organized in a computer system. In digital libraries, the selection of appropriate file formats is important for long-term access and preservation.

- Common text formats include .pdf, .docx, and .txt,
- While images are stored in formats such as .png, .jpeg, and .tiff
- Audio files use formats like .mp3, and .wav, and video files are stored in .mp4 or .avi formats.
- Compressed archives like .zip are also used for efficient storage and easy transfer of multiple files.

For preservation purposes, libraries prefer open and standard formats because they are more stable and widely supported. Choosing suitable file formats helps ensure compatibility, reduces the risk of obsolescence, and supports effective digital preservation.

7.12 LONG-TERM STORAGE

Long-term storage in digital libraries involves preserving digital resources for decades while ensuring their integrity, accessibility, and usability. Libraries use a combination of storage media such as

- Hard Disk Drives (HDDs) for high capacity,
- Solid-State Drives (SSDs) for speed and durability, and
- Magnetic Tapes (LTO) for large, infrequently accessed archives.

Optical media like CDs, DVDs, or Blu-ray can be used for stable offline storage, while cloud storage services (e.g., AWS Glacier, Azure Archive) provide off-site redundancy and disaster recovery.

Key strategies include the 3-2-1 rule (three copies, on two types of media, with one offsite), format migration to prevent obsolescence (e.g., WordStar to PDF/A), and regular checksums and integrity validation. Digital preservation systems like LOCKSS, Archivematica, or Preservica manage formats, metadata, and access over time.

Challenges include media degradation and technological obsolescence, which can make files inaccessible. Therefore, digital libraries adopt a layered storage strategy, combining multiple media types with proactive preservation practices to ensure enduring access to valuable digital resources.

7.13 CLOUD ARCHIVING AND BLOCKCHAIN

Cloud archiving is the practice of storing digital data on remote, cloud-based servers for long-term preservation, accessibility, and management. It provides scalable, cost-effective storage, easy access, and secure backup of digital resources, helping libraries and organizations maintain large collections without relying on physical infrastructure. Examples of cloud archiving include AWS Glacier (Amazon Web Services), Microsoft Azure Archive Storage, and Google Cloud Storage (Nearline & Coldline). However, since cloud data often resides on centralized servers, it can create a single point of failure if proper backups are not maintained.

Blockchain technology is a shared and unchangeable (immutable) digital ledger that records transactions across many computers (nodes) in a network. It creates a secure, transparent, and decentralized database. Data is organized into “blocks,” which are cryptographically linked in a chronological “chain,” making the system tamper-proof. This

technology is ideal for tracking assets, enabling cryptocurrencies, and establishing trust without intermediaries like banks. A blockchain was introduced in 2008 by an individual or group using the pseudonym Satoshi Nakamoto as the public, distributed ledger for Bitcoin transactions. This innovation built upon earlier research and concepts developed by Stuart Haber, W. Scott Stornetta, and Dave Bayer. Examples include Evernym/Sovrin for identity management, KodakOne for copyright protection of digital images, Artory for tracking provenance of artworks, and IPFS (InterPlanetary File System) with blockchain for secure decentralized file storage.

Combining cloud archiving with blockchain enhances digital preservation. Digital items are first digitized and stored in the cloud for easy access. A digital fingerprint or hash of each file, along with metadata like author and date, is recorded on the blockchain. When users access the files, the cloud provides the content, while the blockchain verifies the integrity and authenticity of the resources. This integration ensures secure, reliable, and long-term preservation of digital collections.

CHECK YOUR PROGRESS

A. Fill in the Blanks

a. Digital preservation aims to ensure the long-term _____, authenticity, and usability of digital information resources.

b. _____ metadata standard is specifically designed to support digital preservation by recording preservation actions and technical environment details.

c. The _____ rule suggests keeping three copies of digital data on two different types of media, with one copy stored offsite.

7.14 LET US SUM UP

Digital preservation and conservation are crucial for guaranteeing sustained access to library resources in the face of swift technological advancements. Preservation emphasizes retaining usability against obsolescence with tactics like as migration and emulation, while conservation ensures data integrity through backups and security measures. Librarians play a vital role in this process, overseeing digitization, storage, and metadata standards such as PREMIS and METS. Notwithstanding obstacles like finance and media instability, libraries utilize various storage methods—including cloud archiving and blockchain to verify and safeguard intellectual legacy. These collaborative endeavours ensure that

knowledge remains accessible, authentic, and applicable for future generations of scholars and learners.

✓ 7.15 ANSWERS TO CHECK YOUR PROGRESS

a. (accessibility), b. (PREMIS), c. (3-2-1)

7.16 FURTHER READINGS

1. Beagrie, N., Charlesworth, A., & Miller, P. (2014). *How cloud storage can address the needs of public archives in the UK* (1st ed.). The National Archives. https://cdn.nationalarchives.gov.uk/documents/archives/cloud-storage-guidance.pdf
2. Caplan, P. (2008). *Foundations and standards.* In *Library Technology Reports* (Vol. 44, No. 2, pp. 14–18). ALA TechSource. https://doi.org/10.5860/ltr.44n2
3. Caplan, P. (2008). *Preservation practices.* In *Library Technology Reports* (Vol. 44, No. 2, pp. 10–13). ALA TechSource. https://doi.org/10.5860/ltr.44n2
4. Caplan, P. (2008). *Repository applications.* In *Library Technology Reports* (Vol. 44, No. 2, pp. 26–32). ALA TechSource. https://doi.org/10.5860/ltr.44n2
5. Caplan, P. (2008). *What is digital preservation?* In *Library Technology Reports* (Vol. 44, No. 2, pp. 7–9). ALA TechSource. https://doi.org/10.5860/ltr.44n2
6. Library of Congress. (2015). *PREMIS data dictionary for preservation metadata* (Version 3.0). https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf
7. Library of Congress. (2025). *METS 2 guide.* https://www.loc.gov/standards/mets/mets2.html
8. National Library of Australia. (2003). *Guidelines for the preservation of digital heritage* (Document code CI.2003/WS/3). UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000130071
9. Najar, J. K., & Wani, J. A. (2019). *Digital preservation: An overview.* Library Philosophy and Practice (e-journal), Article 2989. https://digitalcommons.unl.edu/libphilprac/2989
10. Panda, S. K., Jena, A. K., Swain, S. K., & Satpathy, S. C. (2021). *Blockchain technology: Applications and challenges.* Springer.

7.17 PROBABLE QUESTIONS

a. Short Answer Type Questions:

i. What is digital preservation? Explain its importance in digital libraries.

ii. Mention any four challenges of digital preservation in libraries.

b. Long Answer Type Questions:

i. Discuss the major digital preservation strategies adopted by digital libraries. Explain short-term, medium-term, and long-term strategies with suitable examples.

ii. Explain the role of librarians and metadata standards (PREMIS and METS) in digital preservation and digital curation.

*** ***** ***

UNIT 8: OPEN-SOURCE TOOLKITS

UNIT STRUCTURE

8.1 Learning Objectives

8.2 Introduction

8.3 Comparative Studies Between Popular Open-Source Software for Digital Libraries: DSpace, GSDL, E-print, Omeka

8.4 Benefits of Open-Source Software Of DL

8.5 Popular Examples DLs Based on Open-Source Software: Indian and Global

8.6 Let Us Sum Up

8.7 Answers to Check Your Progress

8.8 Further Readings

8.9 Probable Questions

8.1 LEARNING OBJECTIVES

After this unit, learner will be able to

- Learn about different kinds of open-source digital library software
- Get a comparative idea between these open-source DL software.

8.2 INTRODUCTION

Digital libraries rely on robust software platforms to manage, preserve, and provide access to digital content. Among the popular open-source solutions, DSpace, Greenstone Digital Library (GSDL), EPrints, and Omeka are widely used by academic institutions, research organizations, and cultural heritage institutions. Each software has its unique features, strengths, and limitations, making it suitable for specific digital library requirements.

DSpace

DSpace is a free, open-source repository software widely used to create open access institutional repositories for scholarly and published digital content. Unlike general content or document management systems, DSpace is designed specifically as a digital archive, emphasizing long-term storage, access, and preservation of digital resources. Globally, over 3,000 repositories are listed in the DSpace Registry.

The DSpace project was initiated in July 2000 as part of the HP-MIT alliance. The

project was given $1.8 million USD by HP over two years to build a digital archive for MIT that would handle the 10,000 articles produced by MIT authors annually. The first public version was released in November 2002 through collaboration between MIT and HP Labs. Governance evolved through the DSpace Federation, the DSpace Foundation, and later DuraSpace, which merged with LYRASIS in 2019. Today, LYRASIS provides leadership to the DSpace community.

Technologically, DSpace is built using Java web applications, stores metadata in PostgreSQL or Oracle databases, and uses Apache Solr for faceted search and browsing. It supports multiple digital formats, granular access control, UTF-8, and interfaces in 22 languages. DSpace runs on Linux, Unix, Windows, and macOS, with Linux being most common. Notable repositories globally include DSpace@MIT, Apollo - University of Cambridge Repository, Digital Access to Scholarship at Harvard, WHO Institutional Repository for Information Sharing, and the World Bank- Open Knowledge Repository. Some prominent DSpace repositories in India include Shodhganga (INFLIBNET Centre), DSpace@IIT Delhi, DSpace@IIT Bombay, DSpace@IISc Bangalore, and DSpace@JNU.

Greenstone Digital Library (GSDL)

Greenstone is an open-source, multilingual software suite for building and distributing digital library collections, developed by the New Zealand Digital Library Project at the University of Waikato in collaboration with UNESCO and the Human Info NGO in 1996. It enables users to organize information and publish it online or on removable media such as DVDs and USB drives. Released under the GNU General Public License, Greenstone is freely available and customizable.

The software is designed to empower universities, libraries, and public institutions to create their own digital libraries, facilitating easier access to information. It supports information dissemination and acquisition in education, science, and culture, particularly in developing countries. Greenstone aims to promote open access to knowledge, enabling communities and institutions to share information widely, enhance digital literacy, and place valuable content in the public domain, contributing to global education and cultural development.

EPrints

EPrints is an open-source digital repository platform developed at the University of Southampton in 2000 and maintained by the EPrints Community, led by EPrints Services. It is primarily used as an Open Access repository to promote and share academic outputs

through a discoverable web presence. EPrints can also store research data, educational resources, and other specialized content. It is adaptable for institution-wide use, for specific outputs like thesis, or for journals and conferences.

The platform is highly flexible, allowing customizations in branding, data schema, workflows, metadata presentation, and integrations with external systems such as CRIS or institutional websites. Technically, EPrints runs on Linux, uses a MySQL database, and is served via Apache, making it lightweight, flexible, and easy to integrate. A vibrant user and developer community actively supports its use and guides future development, ensuring continuous improvement and adoption in academic institutions worldwide.

Omeka

Omeka (Omeka Classic) is a free, open-source content management system for online digital collections, designed to allow users to publish and exhibit cultural heritage objects. Unlike traditional repository software like DSpace or Fedora, Omeka is lightweight, focuses on display, and uses the unqualified Dublin Core metadata standard. Functionality can be extended with themes and plugins.

Omeka is used by institutions such as the Newberry Library, USA, small museums, historical societies, and the Missouri School of Journalism, USA for sharing archives, including 38,000 photographs from the Pictures of the Year International contest. Originally developed by the Roy Rosenzweig Center for History and New Media at George Mason University in 2008, it has received recognition from the Andrew Mellon Foundation and is used in teaching curation.

Since 2016, Omeka has been maintained by the non-profit Corporation for Digital Scholarship. In 2017, Omeka S was released for institutional use, supporting multiple sites drawing from a common resource pool, while Omeka Classic continues to serve individual projects and educators.

8.3 COMPARATIVE STUDIES BETWEEN POPULAR OPEN-SOURCE SOFTWARE FOR DIGITAL LIBRARIES: DSPACE, GSDL, E-PRINT, OMEKA

Open-source digital library software enables institutions to **manage, preserve, and share digital collections** efficiently without high costs. Among the most popular platforms are **DSpace, Greenstone Digital Library Software (GSDL), EPrints, and Omeka.** Each software differs in **purpose, target users, technical requirements, metadata support, and content types**, making it important to compare their features before selection. Some

are optimized for **institutional repositories** (DSpace, EPrints), while others focus on **multimedia collections or exhibitions** (GSDL, Omeka).

The following table provides a **comprehensive comparison**, including their **technical aspects, key features, strengths, limitations, and examples** of real-world deployments.

Feature	DSpace	GSDL (Greenstone)	EPrints	Omeka
Developer	MIT & HP Labs; now LYRASIS	University of Waikato	University of Southampton	Roy Rosenzweig Center for History & New Media
License	BSD	GNU General Public License (GPL)	GNU General Public License (GPL)	GNU General Public License (GPL)
Primary Use	Institutional repositories, theses, research outputs	Digital library collections, multimedia	Open access repositories, academic outputs	Digital exhibitions, cultural heritage collections
Target Users	Universities, research institutions	Libraries, NGOs, communities	Universities, research institutions	Museums, educators, small institutions
Programming Language	Java	C++, Perl	Perl	PHP
Database	PostgreSQL, Oracle	Built-in or external	MySQL	MySQL
Operating System	Linux, Windows, macOS, Unix	Cross-platform (Linux, macOS and Microsoft Windows)	Linux/Unix	Linux, Windows
Metadata Standards	Dublin Core	Dublin Core	Dublin Core	Dublin Core (unqualified)
Web Interface	Yes	Yes	Yes	Yes
Content Types	Text, image, audio, video	Text, image, audio, video	Text, research data, educational resources	Images, text, multimedia
Key Features	Long-term preservation, SOLR search, multi-language, access control	Multilingual, offline access, flexible collection organization	Flexible workflows, customizable schema, CRIS integration	Lightweight, exhibition-focused, themes & plugins, multiple sites (Omeka S)
Strengths	Scalable, preservation-focused, widely used	Easy to use, multilingual, supports removable media	Flexible, good for open access, strong community	User-friendly, ideal for digital storytelling & exhibits
Limitations	Complex installation & maintenance	Outdated interface, limited modern UI	Less suited for exhibitions	Not ideal for large repositories
Examples	Shodhganga (INFLIBNET), DSpace@MIT, DSpace@Harvard, DSpace@Cambridge	Manioc by the University of the Antilles and Guyane; University of Waikato	ePrints Soton (University of Southampton Institutional Research Repository); Caltech thesis repository (California Institute of Technology (Caltech) (USA)	Missouri School of Journalism archives, Newberry Library collections, small museums exhibitions

8.4 BENEFITS OF OPEN-SOURCE SOFTWARE OF DL

Open-source software (OSS) plays a vital role in the development and management of

digital libraries. It offers several advantages that make it suitable for academic institutions, especially libraries with limited financial resources. The major benefits are explained below in simple and exam-oriented points:

1. Cost-effectiveness

Open-source digital library software such as DSpace, Greenstone, and EPrints is free to use. Libraries do not need to pay license or subscription fees, which reduces overall operational costs.

2. Freedom to Customize

The source code is openly available, allowing libraries to modify the software according to their local needs, language requirements, and institutional policies.

3. Interoperability and Standards Compliance

Most open-source DL software follows international standards such as OAI-PMH, Dublin Core, and MARC, enabling easy data sharing and interoperability with other systems.

4. Community Support and Collaboration

OSS is supported by a global community of developers and users. Libraries benefit from shared knowledge, documentation, discussion forums, and regular updates.

5. Scalability and Flexibility

Open-source DL software can be used by small libraries and scaled up for large academic institutions without major changes or extra costs.

6. Security and Reliability

Since the source code is open, security issues can be quickly identified and fixed by the community, making the software reliable and transparent.

7. Long-term Sustainability

Libraries are not dependent on a single vendor. This avoids vendor lock-in and ensures long-term access to digital content.

8. Support for Open Access

Open-source DL software strongly supports open access initiatives by enabling free and wide dissemination of scholarly content.

Open-source software empowers digital libraries by providing affordable, flexible, and sustainable solutions, making it ideal for academic and research libraries.

8.5 POPULAR EXAMPLES DLS BASED ON OPEN-SOURCE SOFTWARE: INDIAN AND GLOBAL

Open-source software has become the backbone of many digital library (DL) initiatives worldwide. Its flexibility, low cost, and strong support for open access have encouraged academic institutions, governments, and research organizations to adopt open-source platforms such as DSpace, EPrints, Greenstone, and other community-driven systems. The following sections highlight popular Indian and global digital libraries developed using open-source software.

Indian Digital Libraries Based on Open-Source Software

- **Shodhganga**

Shodhganga is a national-level digital repository in India, established in 2011 and managed by the Information and Library Network Centre (INFLIBNET Centre) under the University Grants Commission (UGC). It is developed using the open-source DSpace platform and provides open access to theses and dissertations submitted by Indian universities. Shodhganga plays a crucial role in preventing duplication of research and promoting academic transparency by making doctoral research widely accessible.

- **National Digital Library of India (NDLI)**

The National Digital Library of India (NDLI) is a flagship project of the Ministry of Education, India, established in 2015 and developed and maintained by IIT Kharagpur. NDLI uses a customized open-source architecture largely based on DSpace. It aggregates millions of digital learning resources, including books, articles, videos, and theses, catering to users ranging from school students to researchers. Its open-source foundation allows scalability and integration with diverse content providers.

- **Institutional Repositories of IITs and Universities**

Many premier institutions in India, such as IIT Bombay, IIT Delhi, and IIT Madras, use DSpace to develop their institutional repositories. These repositories provide access to faculty publications, conference papers, technical reports, and theses. The use of open-source software ensures long-term preservation of institutional knowledge and supports open access mandates.

- **ePrints@IISc**

ePrints@IISc is the institutional repository of the Indian Institute of Science, Bengaluru, India, established in 2007 and developed using the open-source EPrints software. It

provides global access to research publications produced by IISc scholars and promotes the dissemination of high-quality Indian research output.

Global Digital Libraries Based on Open-Source Software

- **MIT DSpace Repository (USA)**

The MIT DSpace Repository in the USA was established in 2002. Developed using DSpace, it is one of the earliest and most influential institutional repositories in the world, providing open access to the scholarly output of the Massachusetts Institute of Technology. This repository has served as a model for many academic institutions globally.

- **arXiv.org**

arXiv.org is a globally recognized open-access digital repository for preprints in physics, mathematics, computer science, and related fields. It was established in 1991 in the United States. Built on open-source software, arXiv enables rapid dissemination of research findings prior to formal publication, significantly accelerating scholarly communication.

- **PubMed Central (PMC)**

PubMed Central is a free digital archive of biomedical and life sciences literature, established in **2000** in the **United States**, and maintained by the U.S. National Institutes of Health. It is based on open-source repository frameworks and provides permanent access to peer-reviewed research articles, supporting open science initiatives worldwide.

- **Europeana**

Europeana is a large-scale European digital library, established in **2008** in the **Netherlands**. It provides access to millions of cultural heritage items such as books, manuscripts, images, and audiovisual materials. Developed using open-source technologies, Europeana supports interoperability among museums, libraries, and archives across Europe.

- **Greenstone Digital Library (New Zealand)**

The Greenstone Digital Library, developed by the University of Waikato in **New Zealand**, was established in **1997**. It is both an open-source digital library system and a working digital library, widely used globally for building and managing digital collections, especially in developing countries.

These Indian and global examples clearly demonstrate that open-source software provides a sustainable, scalable, and cost-effective foundation for developing digital libraries. By supporting open access and interoperability, open-source DLs contribute significantly to global knowledge sharing and academic advancement.

CHECK YOUR PROGRESS

A. Fill in the Blanks

a. Shodhganga, a national-level digital repository in India, was established in _____ and is managed by the _____ under the University Grants Commission (UGC).

b. The National Digital Library of India (NDLI) was established in _____, developed and maintained by _____, and uses a customized open-source architecture largely based on _____.

8.6 LET US SUM UP

Digital libraries depend on reliable open-source software such as DSpace, Greenstone, EPrints, and Omeka for the efficient management and preservation of collections. DSpace functions as the benchmark for institutional repositories, Greenstone facilitates decentralized dissemination, EPrints is dedicated to open-access research, and Omeka emphasizes visual cultural displays. These platforms offer essential advantages including customization, interoperability, and liberation from vendor lock-in. Their worldwide influence is seen in significant initiatives like as India's Shodhganga and the National Digital Library, which utilize these tools to democratize access to scholarly and cultural information.

8.7 ANSWERS TO CHECK YOUR PROGRESS

a. UGC**, b.** 2018**,** IIT Kharagpur

8.8 FURTHER READINGS

1. Roy Rosenzweig Center for History and New Media. (2011). *User guide for librarians: Instructions for building a website in Omeka.net for sharing special collections.* George Mason University. https://info.omeka.net/wp-content/uploads/2011/11/user_guide_for_librarians.pdf

2. Smith, M., Barton, M., Bass, M., Branschofsky, M., McClellan, G., Stuve, D., Tansley, R., & Walker, J. H. (2003, January). *DSpace: An open source dynamic digital repository. D-Lib Magazine, 9*(1). https://www.dlib.org/dlib/january03/smith/01smith.html

3. Tansley, R., & Harnad, S. (2000, October). *Eprints.org software for creating institutional and individual open archives. D-Lib Magazine, 10*(6). https://doi.org/

10.1045/october2000-inbrief

4. Witten, I. H. (2003, March). *Examples of practical digital libraries: Collections built internationally using Greenstone. D-Lib Magazine.* http://dlib.org/dlib/march03/witten/03witten.html

5. Witten, I. H., & Bainbridge, D. (2003). *How to build a digital library.* Morgan Kaufman Publishers.

6. Witten, I. H., Boddie, S., & Thompson, J. (2006, March). *Greenstone digital library: User's guide* (gsdl-2.70). https://files.greenstone.org/manuals/gsdl2/en/html/User_en_all.html

8.9 PROBABLE QUESTIONS

A. Short Answer Questions

a. Name any two key features of DSpace that make it suitable for institutional repositories.

b. Mention any two institutions or repositories that use Omeka for digital exhibitions.

c. State two benefits of using open-source software for digital libraries.

B. Long Answer Questions

i. Explain the main differences between DSpace, Greenstone, EPrints, and Omeka in terms of purpose, target users, and key features. Support your answer with examples.

ii. Discuss the role of open-source software in promoting open access and long-term preservation in digital libraries.

iii. Describe at least three major Indian and three major global digital libraries developed using open-source software, including their year of establishment and country.

*** ***** ***

UNIT 9: REGISTRY AND ITEM MANAGEMENT IN DIGITAL LIBRARY

UNIT STRUCTURE

9.1 Learning Objectives

9.2 Introduction

9.3 Registries

9.4 Digital Registries

9.5 Functions of Digital library registries:

9.6 Digital Library handle System

9.7 Digital Object Identifier (DOI) as a handle system

9.8 Integrating Author identification services to Digital Libraries/Institutional Repositories

9.9 Let Us Sum Up

9.10 Answers to Check Your Progress

9.11 Further Readings

9.12 Probable Questions

9.1 LEARNING OBJECTIVES

After this unit, learners will be able to,

- Know about registries, types, digital registries and its functions
- Digital Library Handle system, DOI as handle system.

9.2 INTRODUCTION

This subject examines the essential architecture that converts digital repositories from basic storage facilities into internationally accessible and dependable library systems. We commence by delineating the notion of a “registry”—an authoritative database utilized for monitoring official records—and scrutinize its particular implementation within the library sector via platforms such as ROAR (Registry of Open Access Repositories) and OpenDOAR (Directory of Open Access Repositories). You will understand how these registries operate by indexing repositories rather than storing information, thereby improving visibility, standardizing metadata, and enabling worldwide research aggregation.

Moreover, the section tackles the technical difficulty of guaranteeing sustained access to digital materials. We will examine the Handle System and Digital Object IDs (DOI),

which offer persistent IDs that mitigate “link rot” and guarantee the accessibility of digital items despite changes in their locations. We conclude by examining the incorporation of author identity services such as ORCID, which are vital for automating metadata processes, upholding authority control, and guaranteeing accurate attribution in the academic record. Upon completion of this unit, you will comprehend the backend systems that render digital libraries interoperable, permanent, and reliable.

9.3 REGISTRIES

Cambridge Dictionary defines registry as “a place where official records are kept” while Collin’s dictionary categorises it as a noun describing it as “a collection of all the official records relating to something, or the place where they are kept.” The term registry broadly refers to an authoritative database or official record-keeping system used to track specific types of information over time. Registries come handy as a central source while people are in search of authenticated information on any domain of human knowledge. They are essential tools for organization, research, and system configuration across various industries.

Primary functions of registries are

- Centralization: They consolidate information that would otherwise be scattered, making it easier to search and manage.
- Longitudinal Tracking: Unlike simple lists, many registries (especially medical ones) track data over long periods to observe changes and patterns.
- Research & Policy: Data from registries often drive scientific discoveries and public health or safety regulations.
- Verification: They serve as legal proof of status, such as a ship’s nationality or a professional’s license.

There are registries for various types-

- **Computing (Windows Registry):** A central, hierarchical database in the Windows operating system. It stores critical configuration settings for hardware, software, and user preferences. Accessing it (via regedit) is usually for advanced troubleshooting, as incorrect changes can cause system failure.
- **Medical & Patient Registries:** Organized systems that use observational methods to collect uniform data on individuals with a specific disease or condition.
- **Disease Registries:** Track conditions like cancer or diabetes to monitor trends

and outcomes.

- **Product Registries:** Monitor the safety and effectiveness of medical devices or drugs.
- **Clinical Trial Registries:** Connect willing volunteers with researchers for medical studies.
- **Government & Civil Registries:** Official lists managed by authorities to maintain public records, such as:

Vital Records: Registries for births, deaths, and marriages.

Property/Land Registries: Official records of land ownership.

Motor Vehicle Registries: Databases for vehicle registration and licensing.

- **Commercial & Personal Registries:**

Gift Registries: Retailer-managed lists for events like weddings or birthdays.

Domain Name Registries: Databases of all domain names registered under specific top-level domains like .com or .org

9.4 DIGITAL LIBRARY (DL) REGISTRIES

The movement of development of institutional repositories or digital libraries have started along with the open access movement. The problem of serial crisis made it impossible for many institutions to subscribe to article of major scientific publishers and encouraged institutions to create repositories so that the visibility of their scientific output can be increased and scientific discoveries reaches the right audience. The time from 1990s to 2000s have seen the emergence of many open-source digital library software packages like Dspace, ePrints, Greenstone etc and it helped the cause of librarians to build digital libraries at institutional levels. This growth of DL and IRs across institutions have encouraged a demand for having a central registry to showcase the data of various Institutional Repositories across the world. Therefore, a DL registry can be defined as a platform or service used to organize, index, and provide access to digital repositories across the globe. In simple terms a digital library registry is used to register, list, describe, and help discover digital libraries, rather than to store the actual digital content. Some popular registries are-

- **ROAR (Registry of Open Access Repositories) https://roar.eprints.org/**

ROAR is hosted at the University of Southampton, UK and is made possible by funding from the JISC. It was launched in 2003. Originally known as the Institutional Archives

Registry, it was designed to index the creation, location, and growth of open-access institutional repositories worldwide. It was subsequently renamed ROAR in 2006. As of 2026, more than 4000 repositories around are globe are part of ROAR. It provides information such as repository name, country, software used, and growth statistics, helping track the development of open-access repositories. ROAR also provides specific institutional id which now a days are very much essential to curate institutional publication data.

- **OpenDOAR (Directory of Open Access Repositories) https://opendoar.ac.uk/**

OpenDOAR was developed and is maintained by the University of Nottingham in 2005, in collaboration with Lund University. It was initially funded by a partnership including the Open Society Institute (OSI), Jisc (UK Joint Information Systems Committee), SPARC Europe, and the Consortium of Research Libraries (CURL). Currently it is maintained by SHERPA services. Till 2026, OpenDOAR has more than 5900 member institutions.

- **OpenAIRE (https://explore.openaire.eu/)**

IT is a European open-access infrastructure that registers repositories and supports open science by linking publications, datasets, and research projects funded by the EU. OpenAIRE acts as a central aggregator and infrastructure, linking thousands of institutional, thematic, and data repositories to create a “Global Open Science Graph”. It harvests metadata for literature, data, and software, requiring repositories to be OpenAIRE-compliant to ensure visibility, with Zenodo acting as its catch-all, CERN-hosted repository. The OpenAIRE initiative, including its underlying repository infrastructure, began in 2009-2010 as a project to support Open Access in Europe. While the infrastructure developed from earlier projects, OpenAIRE was established as a formal legal entity in 2018 to provide a permanent structure for open scholarly communication.

9.5 FUNCTIONS OF DIGITAL LIBRARY REGISTRIES

Registration and Identification of Digital Libraries:

The primary function of a digital library registry is to register and uniquely identify digital libraries and repositories. Institutions submit basic information such as the repository name, hosting organization, URL, and repository type. Example: ROAR (Registry of Open Access Repositories) registers institutional repositories like IIT Delhi’s repository and assigns them a unique registry entry.

Metadata Collection and Standardization:

Digital library registries collect descriptive metadata about repositories, including subject coverage, content types, languages, and software platforms used. This metadata is usually standardized to ensure consistency and comparability. Example: OpenDOAR records metadata such as whether a repository contains theses, articles, datasets, or conference papers, and whether it uses software like DSpace or EPrints.

Discovery and Visibility Enhancement

Registries improve the visibility and discoverability of digital libraries by making repository information searchable through a single interface. This helps researchers, librarians, and aggregators locate relevant repositories efficiently. Example: Researchers searching OpenDOAR can filter repositories by country (India), subject (Library Science), or content type (theses).

Support for Interoperability and Harvesting

Digital library registries indicate whether repositories comply with interoperability standards such as OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting). This enables service providers to harvest metadata from multiple repositories. Example: BASE and CORE rely on registry information from OpenDOAR to identify OAI-PMH compliant repositories for metadata harvesting.

Quality Control and Repository Evaluation

Registries often apply quality criteria before listing repositories, such as checking openness, accessibility, and technical standards. This helps users identify reliable and functional repositories. Example: OpenDOAR manually reviews repositories before inclusion, ensuring that only genuinely open-access repositories are listed.

Policy and Open Access Compliance Monitoring

Some registries record information about open access policies, embargo periods, and licensing practices. This helps funding agencies and institutions monitor compliance with open-access mandates. Example: Funding bodies can use registry data to verify whether an institution maintains an open-access repository for public-funded research.

Statistical and Analytical Support

Digital library registries provide statistical data such as the number of repositories by country, growth trends, and software usage. This data supports research, planning, and policy formulation. Example: ROAR provides statistics showing the growth of institutional repositories in India over time.

Promotion of Open Access and Scholarly Communication

By listing and promoting digital libraries, registries encourage open access publishing and strengthen the global scholarly communication ecosystem. Example: OpenDOAR and ROAR play a key role in promoting open-access repositories worldwide by making them visible to search engines and indexing services.

9.6 DIGITAL LIBRARY HANDLE SYSTEM

Digital library handles refer to persistent identifiers from the CNRI (Corporation for National Research Initiatives) Handle System, designed for stable access to digital objects in libraries and repositories. The Handle System assigns unique, location-independent identifiers (handles) to digital content like articles, datasets, and web resources. It uses a distributed network of global and local servers to resolve handles to current locations, even if files move. Developed by CNRI (and now managed by the DONA Foundation), it resolves opaque, unique handles into current locations (URIs), ensuring, for example, that a DOI never breaks. It handles digital object identification, resolution, and management via a network of local and global servers.

Syntax of a handle:

A handle has the form:

n/d

Where n is a naming authority and d is an arbitrary string. The string d is unique, for that naming authority. The following are examples of valid handles:

berkeley.cs/1994.12.05.23.42.12;7 (the part before / is the naming authority)

cnri.dlib.papers/tn95-137

All alphanumeric and certain other ASCII characters are allowed. Text is case insensitive. Within n, a period (.) has a special significance.

For identification in Internet applications, a handle may be preceded by “hdl:”, “hdl://”, “x-hdl:”, or “x-hdl://” as in the following example: hdl:cnri.dlib/tn95-01 (Arms & Ely, 1995)

Typical example of an item handle in Dspace DL software may look like this- http://hdl.handle.net/123456789/100. Here the portion “……handle.net” represents the server address of the repository, while the part “/123456789/100” is generated by the internal handle system of Dspace. This handle system is analogues to many other digital library software systems like ePrints and GSDL. In the DSpace system architecture, the Handle System lies primarily within the Business Logic Layer (often referred to as the Core Service Layer). It acts as a service that interacts closely with the content management API to

assign and manage persistent identifiers for communities, collections, and items.

The various types and components of the Handle System include:

Global Handle Server (G)

The global handle server is a single, distributed system that is publicly accessible, highly secure, and fault-tolerant. Its primary roles include:

Long-term Persistence: It stores handle records for items requiring very long-term retention, such as legal records or copyright registrations.

System Registry: It stores handles for all naming authorities and local handle servers.

Public Resolution: It is a public service that any client can query to resolve any handle. If it does not store a specific handle itself, it provides the address of the primary handle server for the relevant naming authority.

Local Handle Servers

Local handle servers are optional components that work in conjunction with the global server to resolve handles locally.

Local Autonomy: They provide organizations with increased control over their handles and distribute the computing load away from the global server.

Access Control: Unlike the global server, local handle servers are not public; administrators can set permissions (currently based on IP addresses) to restrict who can resolve handles stored there.

Redundancy: By storing handles on both local and global servers, organizations can use them to back up one another.

Caching Handle Servers

Caching servers are used to improve performance and minimize the load on global and local servers.

Performance: They provide rapid resolution by storing recently accessed handle records.

Configuration: Caches can be integrated into individual client software or set up as shared servers for a group of clients. The system architecture recommends no more than two levels of caching to keep the resolution process efficient.

Proxy Servers

Proxy servers act as intermediaries that allow standard Web browsers (like Mosaic) to resolve handles without needing specialized client software.

URL Resolution: If a handle can be resolved into one or more URLs, the proxy server returns a URL to the client browser.

Standardization: The proxy server developed by the CNRI is designed to be compatible with other Web community proxy standards.

Naming Authorities

While not a “server” type, naming authorities are the organizational entities authorized to create and manage handles.

Hierarchy: They follow a hierarchical structure (e.g., berkeley.cs), where high-level names are issued by the global naming authority.

Primary Handle Server (P): Every naming authority is associated with a primary handle server. When a new handle is created, it is either stored in this primary server or an indirect handle (a forwarding address) is placed there to point to the actual storage location.

9.7 DIGITAL OBJECT IDENTIFIER (DOI) AS A HANDLE SYSTEM

The Digital Object Identifier (DOI) is a persistent identifier system used to uniquely identify and reliably locate digital objects such as journal articles, books, datasets, reports, and multimedia content. Technically, the DOI system is an application of the Handle System, which provides the underlying infrastructure for persistent identification and resolution.

The Handle System, developed by the Corporation for National Research Initiatives (CNRI), is a distributed system designed to assign, manage, and resolve persistent identifiers (handles) for digital objects.

A DOI is a specialized form of a handle that follows the syntax:

oi:prefix/suffix

or

https://doi.org/prefix/suffix

The prefix identifies the registrant (publisher or organization).

The suffix uniquely identifies the digital object.

Example: 10.1000/182

Here, 10 indicates the DOI namespace within the Handle System, 1000 identifies the registrant, and 182 identifies the specific object.

How DOI Works as a Handle

When a user clicks or enters a DOI (e.g., https://doi.org/10.1016/j.ipm.2023.103207), the Handle System resolver redirects the request to the current URL of the digital object. Even if the object’s web location changes, the DOI remains the same, ensuring persistent access.

Key Functions of DOI as a Handle System

- Persistent Identification: Provides a permanent, unique identifier independent of physical location.
- Resolution Mechanism: Uses Handle System servers to resolve DOIs to current URLs.
- Metadata Association: Links rich metadata (author, title, publisher, rights) with the identifier.
- Interoperability: Works across publishers, repositories, and digital library systems.
- Long-term Access: Prevents link rot in scholarly communication.

9.8 INTEGRATING AUTHOR IDENTIFICATION SERVICES TO DIGITAL LIBRARIES/INSTITUTIONAL REPOSITORIES

Integrating authors identification services like Orcid, ResearcherID into digital library software enhances researcher identification, automates metadata workflows, and improves discoverability of scholarly outputs. DSpace and EPrints are popular open-source digital library software with built-in ORCID support, allowing users to import records, authenticate via ORCID iD, and push metadata like publications back to ORCID profiles. Integration helps in

- Researchers authenticate their ORCID ID during submission, allowing automatic linking of works to their ORCID record.
- ORCID acts as an authority control tool for author names. Libraries can embed ORCID iDs in metadata fields (e.g., Dublin Core, MODS), improving consistency across systems.
- With researcher consent, institutional libraries can push and pull data between the repository and ORCID, reducing manual data entry and errors.
- integration helps libraries track compliance with open-access mandates and institutional research policies by reliably linking outputs to authors.

9.9 ORCID INTEGRATION TO DSPACE

Prerequisites

Institutions must join ORCID as a member to obtain Member API credentials (Client ID and Secret) for full read/write capabilities; Public API suffices for basic authentication.

Activate required DSpace features like Researcher Profiles, Configurable Entities

(Person, Publication, Project), and the ORCID synchronization queue in

local.cfg.

Install DSpace 7.3+ where native bidirectional support exists via ORCID API v3.0.

Configuration Steps

Set

orcid.authentication-enabled=true and

orcid.synchronization-enabled=true

in local.cfg, along with API endpoints and scopes for read/write on biography, works, and funding.

Map DSpace metadata to ORCID fields in

orcid.cfg (e.g., person.firstName to ORCID given-names, dc.type to ORCID work-types via converters).

Enable event consumers like orcidqueue for automatic nightly syncs, with manual queues for user approval of pushes.

User Workflow and Maintenance

Users connect profiles via ORCID login during deposit, choosing sync preferences (e.g., biographical data, publications as Works).

Validate data pre-push against ORCID controlled lists (e.g., work types, funding details) to avoid errors; log failures for admins.

Display ORCID iDs on item pages for certification compliance, test thoroughly in staging, and monitor via ORCID's integration recipes.

CHECK YOUR PROGRESS

a. A registry is an authoritative _____ used for official record keeping.

b. ROAR stands for _____ of Open Access Repositories.

c. OpenDOAR is maintained by the _____ University.

d. Digital library registries do not store actual content but provide _____ and discovery of repositories.

e. The Handle System was developed by _____.

f. A handle consists of two parts: naming authority and _____.

g. DOI is an application of the _____ System.

9.9 LET US SUM UP

Registries serve as authoritative databases essential for the centralization, tracking, and verification of information across diverse fields, from computing and medicine to government records. In the specialized realm of digital libraries, these registries—such as ROAR, OpenDOAR, and OpenAIRE—act as vital platforms for organizing and indexing global repositories, thereby enhancing the visibility and discovery of scientific output in response to the "serial crisis" and the open access movement.

The long-term reliability of these digital ecosystems depends heavily on persistent identification. The Handle System and its application, the Digital Object Identifier (DOI), provide stable, location-independent identifiers that prevent "link rot" and ensure consistent access to digital objects even if their web location changes. These systems are supported by a complex architecture of global, local, and proxy servers designed for fault tolerance and efficient resolution.

Integration of author identification services like ORCID into digital library software (such as DSpace or EPrints) represents a critical step in automating metadata workflows. By linking unique researcher IDs directly to publications, institutions can maintain authority control, reduce manual data entry errors, and effectively monitor compliance with open-access mandates. Together, these management tools and identification systems form the infrastructure necessary for a transparent, interoperable, and permanent global scholarly communication network

9.10 ANSWERS TO CHECK YOUR PROGRESS

a.

b.

9.11 FURTHER READINGS

1. Arms, W., & Ely, D. (1995, June 23). *The handle system: A technical overview*. Corporation for National Research Initiatives. Available at: https://www.cnri.reston.va.us/home/cstr/handle-overview.html
2. Deka, D., (2006) *The role of open source software building through Institutional Repository*. 4th Convention PLANNER, Mizoram University, Aizawl, 09-10 November, 2006. INFLIBNET Centre, Ahmedabad. 121-127.

3. Frumkin, J. (2006). The need for a digital library service registry. *OCLC Systems & Services: International Digital Library Perspectives, 22*(1), 23–25. https://doi.org/10.1108/10650750610640766

4. Garcia-Vera V., Roig-Vila, R., Ferrandez, P. G. and Marti P., (2015) How can directories of open access repositories improve the reuse of learning objects in building engineering? *International Journal of Continuing Engineering Education and Life-Long Learning. 25*(3) 274–92. DOI:10.1504/IJCEELL.2015.072664.

5. Powell, J. E., Black, K., & Collins, L. M. (2011). A semantic registry for digital library collections and services. *D-Lib Magazine, 17*(11/12). https://doi.org/10.1045/november2011-powell

6. Registry of Open Access Repositories. (n.d.). About the registry. University of Southampton. http://roar.eprints.org/

9.12 PROBABLE QUESTIONS

a. Define a registry and explain its importance in digital library systems.

b. Discuss the role of digital library registries in enhancing the visibility and discoverability of institutional repositories.

c. Explain the functions of ROAR and OpenDOAR as digital library registries.

d. Describe the Handle System and explain its significance in digital libraries.

e. Explain the structure and components of the Handle System, highlighting global and local handle servers.

f. Discuss DOI as a persistent identifier and explain how it functions as a handle system.

g. Explain the importance of integrating author identification services like ORCID into digital libraries and institutional repositories.

*** ***** ***

UNIT 10: CASE STUDY OF DL SOFTWARE: E PRINT

UNIT STRUCTURE

10.1 Learning Objectives

10.2 Introduction

10.3 E-Print Software

10.4 Key Features

10.5 Versions of E-prints

10.6 System requirements for E prints installation

10.7 E-Prints workflow

10.8 Let Us Sum Up

10.9 Answers to Check Your Progress

10.10 Further Readings

10.11 Probable Questions

10.1 LEARNING OBJECTIVES

After this unit, learners will be able to

- Know about E print DL software in details
- Know about its features, versions, requirements workflow

10.2 INTRODUCTION

E-Prints is a groundbreaking open-access repository software, recognized as the first professional platform expressly engineered to enable the establishment of high-quality, OAI-compliant institutional repositories. E-Prints, developed at the University of Southampton's School of Electronics and Computer Science in the early 2000s and released under the GNU General Public License, has facilitated the establishment of digital repositories at numerous institutions globally, enabling the collection, preservation, and dissemination of scholarly outputs including pre-prints, post-prints, theses, and research data.

This lesson presents a detailed case study of EPrints software, analyzing its technical architecture, fundamental features, and practical application. This exploration will examine how EPrints facilitates the "Green Road" to open-access publishing, allowing universities to internationally display their research output while ensuring complete interoperability

via OAI-PMH compliance. This article examines the progression of EPrints from its first 1.0 release to the contemporary 3.4 series, evaluating the software's development into a stable, plugin-oriented system that harmonizes flexibility and reliability.

The key technical aspects addressed encompass EPrints Perl-based architecture, its multilingual support functionalities, extensive metadata management features, and role-based content workflows that guarantee quality control via structured submission and moderation procedures. Through the analysis of system requirements, workflow processes, and practical applications, one will acquire valuable insights into the deployment and management of EPrints repositories inside academic and institutional environments. This case study illustrates how open-source software may proficiently tackle the obstacles of academic communication, offering institutions tailored, economical solutions for enduring digital preservation and worldwide research visibility.

10.3 E-PRINTS SOFTWARE

ePrints is the world's first professional software platform specifically designed to facilitate the creation of high-quality, OAI-compliant institutional repositories. Developed at the University of Southampton, England, in early 2000, it is released as free, open-source software under the GNU General Public License. The software is intended to help research communities collect, preserve, and disseminate their intellectual output—such as pre-prints, post-prints, theses, and scientific data in a digital format for easy retrieval.

Key aspects of e-Prints include:

- **Open Access Focus:** It is specifically focused on Open Access (OA) functionality and supports the "Green Road" channel of OA publishing, which allows institutions to showcase their scholarly literature to the global community.
- **Technical Foundation:** The platform is primarily written in Perl and operates using a traditional open-source technology stack, including the Apache web server and MySQL for metadata storage.
- **Versatility:** While its default configuration is optimized for research papers, EPrints is an extensible content management system that can archive any digital content, including images, audio, video, and research data.
- **Interoperability**: It is fully compliant with the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). This allows metadata from EPrints repositories to be programmatically harvested by centralized search engines, creating a global network of

cross-searchable research information

10.4 KEY FEATURES

10.4.1 LANGUAGE SUPPORT

- **Unicode and UTF-8 Compliance:** EPrints is fully Unicode compliant, which allows it to process, store, and display almost all the major languages of the world that have scripts. It uses UTF-8 encoding internally, ensuring that metadata and documents in native languages are interpreted correctly.
- **Multilingual User Interfaces:** The software supports the creation of user interfaces in multiple languages. This is achieved by duplicating and translating static XML files and phrase files (e.g., translating phrases-en.xml to phrases-hi.xml for a Hindi interface).
- **Localized Tools:** For languages with unique scripts, E-Prints can integrate virtual keyboard interfaces (e.g., Java-based keyboards for Hindi or Kannada) to assist users who do not have localized hardware.

10.4.2 FILE FORMATS SUPPORT

- **Diverse Digital Content:** E-Prints is built to manage “anything that can be stored digitally”. While its default configuration is optimized for research papers, it can easily be used for images, scientific data, audio, and video archives.
- **Specific Supported Formats:** The system functions with a wide variety of file types, including PDF, HTML, JPEG, TIFF, MP3, AVI, MS Word, and PowerPoint.
- **Storage and Thumbnails:** Metadata is stored in a relational database (MySQL), while the actual digital files are stored on the OS file-system. The software can automatically generate thumbnail previews for documents and images upon upload.

10.4.3 METADATA SUPPORT

- **Customizable Schema:** Metadata fields are highly user-defined. Administrators can specify which metadata fields are held for each item, which are mandatory, and how they appear on submission forms.
- **Standards and Interoperability**: EPrints is natively OAI-PMH compliant, allowing metadata to be programmatically harvested by global search engines and centralized harvesters. It supports various metadata formats, including Dublin Core, METS, MODS,

and MPEG-21 DIDL.

- **Bulk Import and Export:** It supports importing and exporting data in several popular formats such as BibTeX, PubMed XML, EndNote, and Reference Manager.
- **Quality Control**: E-Prints utilizes authority files to prevent duplicate entries and maintain high metadata uniformity.

10.4.4 CONTENT MANAGEMENT SUPPORT

- **Role-Based Management:** The system delegates responsibilities through specific user roles:
- **Author:** Can deposit and manage their own documents.
- **Editor/Moderator:** Reviews submissions for authenticity and metadata accuracy before they move to the "live" archive.
- **Administrator:** Controls all back-end options, including server settings and interface appearance.
- **Submission Workflow:** E-Prints uses a multi-stage workflow where documents move from a temporary buffer (submission) to a review phase (moderation) before becoming publicly accessible.
- **Access Controls**: It supports multi-tier access control, allowing documents to be restricted to "Anyone" (public), "Registered users only" (institutional members), or "Archive staff only".
- **Discovery and Retrieval:** Users can discover content through simple and advanced search or by browsing metadata fields like subject, year, author, or document type.
- **Engagement and Updates:** The platform supports Web 2.0 functionalities such as RSS feeds and email alerts to keep researchers updated on new additions to the repository.

10.5 VERSIONS OF E-PRINTS

EPrints, an open-source digital repository software originating from the University of Southampton, saw its first releases in the early 2000s following the 1999 Santa Fe OAI meeting. The evolution of EPrints software over the last two decades reflects a progression from foundational open-access standards to a modern, plugin-based architecture focused on stability.

10.5.1 THE FOUNDATIONAL YEARS (2000–2002)

Development accelerated in 2000 with the release of EPrints 1.0, which introduced support for OAI 0.2. This was followed by Version 1.1 in January 2001, which updated the software to support OAI 1.0. The transition to EPrints 2.0 (Olive) in early 2002 was a major milestone, introducing OAI 2.0 and subscription features. During this period, the software also began incorporating more specialized roles, such as the addition of subject editors in Version 2.2 (Pumpkin).

10.5.2 MATURITY AND THE TRANSITION TO VERSION 3 (2004–2011)

Following the refined 2.x series, which included numerous patches up to 2.3.13, the software underwent a significant architectural shift with the start of EPrints 3 in late 2006. This era was defined by a series of stable releases that improved the core system:

Version 3.1 (September 2008)

Version 3.2 (March 2010)

Version 3.3 (September 2011)

10.5.3 THE MODERN ERA: THE 3.4 SERIES (2018–2024)

After a period of further development, the 3.4 series launched in May 2018. This series represents the most recent major phase of the software's evolution, characterized by a high frequency of iterative updates. Between 2018 and late 2024, the software moved from version 3.4.0 to 3.4.7.

The focus of these modern releases has shifted toward stability, the integration of plugins, and the implementation of modern features to ensure the software remains relevant in the current digital repository landscape. As of September 2024, Version 3.4.7 stands as the latest release in this two-decade-long history of development.

10.6 SYSTEM REQUIREMENTS FOR E PRINTS INSTALLATION

10.6.1 HARDWARE REQUIREMENTS

The sources recommend the following minimum hardware specifications for a professional installation:

Processor: An Intel P4 processor or equivalent is recommended, though the software

can run on lower configurations.

Memory (RAM): A minimum of 512 MB RAM is suggested for stability, though experimental setups have been run on as little as 256 MB.

Storage: At least 40 GB of hard disk space is recommended, but the actual requirement depends heavily on the size and type of the digital collection being archived.

Performance Scaling: As the number of users and the size of repositories grow, hardware must be scaled up to meet higher performance demands.

10.6.2 SOFTWARE REQUIREMENTS

The software environment for EPrints is built on several “open source” foundations, often referred to as a traditional technology stack.

Operating Systems: EPrints is primarily recommended for UNIX-like operating systems, specifically various Linux distributions such as Fedora, Ubuntu, Debian, RHEL, CentOS, and Suse, as well as Solaris. It is also compatible with Windows XP and later versions.

Web Server: The software requires the Apache web server (version 2.0 or later). It must include the mod_perl module, which significantly increases the performance of the Perl scripts used by the system.

Database Management: MySQL (version 5.0 recommended) is used to store metadata for records and users. Some sources note that PostgreSQL or Oracle may also be supported depending on the version.

Programming Language: EPrints is written in Perl, and several specific Perl modules are required, including XML-LibXML, XML-Parser, MIME-Lite, and Term::ReadKey.

10.6.3 SUPPORT UTILITIES

Full-text Indexing: Tools like Xpdf and antiword are used to facilitate the indexing of documents.

Rendering and Processing: ImageMagick and tetex-latex are utilized for rendering equations and generating document previews.

Standard Commands: Basic Linux utilities such as wget, tar, and unzip are necessary for the installation process.

10.7 E-PRINTS WORKFLOW

10.7.1 USER REGISTRATION AND ACCESS

Account Creation: To submit documents, authors must first create a repository account. This typically involves providing required information and completing an email activation procedure.

Role-Based Access: The workflow is governed by specific user roles:

Authors/Student Users: Responsible for depositing their own work.

Editors/Moderators: Responsible for reviewing submissions for metadata accuracy and authenticity.

Administrators: Manage the overall system configuration, including metadata schemas and user rights.

10.7.2 THE SUBMISSION (DEPOSIT) PROCESS

Authorized users (authors or creators) initiate the workflow by depositing documents through a web-based interface.

Step-by-Step Submission: Users are guided through the process one step at a time.

Metadata Entry: Users provide information such as document type (e.g., article, thesis, book chapter), title, author, and date via a web form. Features like “autocomplete-as-you-type” are available to speed up data entry.

File Upload: An electronic copy of the document is uploaded. EPrints supports diverse formats including PDF, HTML, JPEG, and MP3. Upon upload, the system automatically generates thumbnail previews.

Access Rights: The depositor must assign access levels: “Anyone” (public), “Registered users only,” or “Archive staff only”.

Buffer Stage: Once submitted, documents do not go live immediately; instead, they are placed in a temporary storage buffer.

10.7.3 THE REVIEW (MODERATION) PROCESS

In this stage, editors or reviewers manage the submissions held in the buffer.

Quality Control: Reviewers check for authenticity, validate the quality of metadata, and ensure the documents meet institutional policies.

Format Verification: Librarians may check for online readability and convert documents to standardized formats if necessary.

Editorial Actions: The moderator can take one of several actions:

Accept: Move the document to the “Live Archive” for public view.

Return: Send the document back to the author with comments for revision.

Reject/Destroy: Delete the document if it fails to satisfy archive policies.

10.7.4 LIVE ARCHIVE AND RETRIEVAL

Once a document is moderated to the Live Archive, it becomes discoverable by end-users.

Indexing: The system processes the metadata and full text to allow retrieval through search engines like Google Scholar.

Discovery Mechanisms: Users can retrieve documents through simple or advanced searches, or by browsing fields such as subject, year, or author.

Updates: Users can stay updated on the latest additions through RSS feeds and email alerts.

10.7.5 TECHNICAL FLOW BEHIND THE WORKFLOW

From a system architecture perspective, the user interacts with the web browser, which communicates with the web server (Apache). The server invokes Perl scripts to process requests, which consult the MySQL database for metadata and configuration files before delivering results back to the user

More about the e prints workflow can be experienced at this demo practice site of e-Prints: https://tryme.demo.ePrints-hosting.org/

CHECK YOUR PROGRESS

a) EPrints was developed at the University of _____ in England.

b) EPrints software is released under the _____ General Public License.

c) The primary protocol used by EPrints for metadata harvesting is _____.

d) EPrints is mainly written in the _____ programming language.

e) Metadata in EPrints is stored using the _____ database management system.

10.8 LET US SUM UP

E-Prints is a mature, open-source digital repository software specifically designed to support Open Access and the long-term management of institutional scholarly output. Developed at the University of Southampton, it provides a robust, OAI-PMH–compliant platform for collecting, preserving, and disseminating research materials in multiple formats and languages. Its strengths lie in flexible metadata customization, strong interoperability, role-based content management, and a well-defined submission and moderation workflow that ensures quality control. Over two decades of development, EPrints has evolved into a stable, plugin-based system with regular updates, making it suitable for institutions of varying sizes. Overall, EPrints remains a reliable and widely adopted solution for building sustainable institutional repositories and enhancing global visibility of research output.

10.9 ANSWERS TO CHECK YOUR PROGRESS

a) Southampton

b) GNU

c) OAI-PMH

d) Perl

e) MySQL

10.10 FURTHERREADINGS

1. Dalkýran, Ö., Aker, Ý., Öztemiz, S., Taþkýn, Z., & Tunç, S. K. (2014). Usability testing of digital libraries: The experience of EPrints. *Procedia - Social and Behavioral Sciences*, *147*, 535–543.
2. EIFL. (2011, May). *EPrints - Introducing EPrints 3 software* [Webinar]. Contributors: Nixon, W. J., Peternel, M., & Zakaryan, T.
3. Kamble, V. T., Raj, H., & Sangeeta. (2012). Open source library management and digital library software. *DESIDOC Journal of Library & Information Technology*, *32*(5), 388–392.
4. Ram, S. (n.d.). *Case study: Eprints*. Digital Libraries.
5. Ridwan, Rohana, & Sa'adiyah, H. (2024). Analysis of the implementation of EPrints-based digital libraries. *JIPI (Jurnal Ilmu Perpustakaan dan Informasi)*, *9*(2), 338–345.

6. Sahu, R. R., & Karadia, A. (n.d.). *Comparative study of an open source digital library software: DSpace, Greenstone and EPrint*.

7. Sponsler, E., & Van de Velde, E. F. (2001, July 27). *Eprints.org software: A review*. California Institute of Technology.

8. Tramboo, S., Humma, Shafi, S. M., & Gul, S. (2012). A study on the open source digital library software's: Special reference to DSpace, EPrints and Greenstone. *International Journal of Computer Applications*, *59*(16), 42–47.

9. Vaidya, N. N., & Jayakanth, F. (2006, February 2–4). *Setting up a multilingual information repository: A case study with Eprints.org software* [Paper presentation]. 4th International Convention CALIBER-2006, Gulbarga, India

10.11 PROBABLE QUESTIONS

1. Explain the concept of Open Access and discuss how EPrints supports the Green Road to Open Access publishing.
2. Describe the key features of EPrints software with reference to language support and file format compatibility.
3. Discuss the role of metadata in EPrints and explain how it ensures interoperability and quality control.
4. Trace the evolution of EPrints software from its early versions to the modern 3.4 series.
5. Explain the system requirements for installing EPrints, highlighting both hardware and software components.
6. Describe the EPrints submission and moderation workflow in detail.
7. Discuss the role-based access control mechanism in EPrints and its importance in repository management.

*** ***** ***

UNIT 11: CASE STUDY OF DL SOFTWARE: GSDL

UNIT STRUCTURE

11.1 Learning Objectives

11.2 Introduction

11.3 Greenstone Digital Library Software (GSDL)

11.4 Versions of Greenstone

11.5 System Requirements

11.6 Customization whereabouts in Greenstone2 and Greenstone3

11.7 Greenstone Librarian Interface (GLI)

11.8 Advanced Management Features

11.9 Let Us Sum Up

11.10 Answers to Check Your Progress

11.11 Further readings

11.12 Probable Questions

11.1 LEARNING OBJECTIVES

After this unit, learners will be able to

- Know about Greenstone digital library software
- Know about system requirements, customization, GLI etc.

11.2 INTRODUCTION

Greenstone is an all-inclusive, open-source software suite for building and sharing digital library collections. It was created by the University of Waikato's New Zealand Digital Library Project in partnership with UNESCO and enables organizations to arrange a variety of data, including text, audio, and video, into searchable, metadata-driven resources. Greenstone, which is released under the GNU General Public License, focuses on meeting the needs of developing nations and promoting knowledge in the public domain. Strong multilingual support and compatibility with Windows, Linux, and macOS make it an adaptable tool for libraries and colleges to efficiently create and distribute their own digital repository.

11.3 GREENSTONE DIGITAL LIBRARY SOFTWARE (GSDL)

Greenstone is a comprehensive, open-source software suite designed for building

and distributing digital library collections. It is not a digital library itself, but rather a specialized tool for organizing information and publishing it as a fully searchable, metadata-driven digital resource. Developed by the New Zealand Digital Library Project at the University of Waikato in cooperation with UNESCO and the Human Info NGO, the software is issued under the terms of the GNU General Public License. In 1998, the New Zealand Digital Library Project released the first version of the software on CD-ROM.

The primary objective of Greenstone is to empower institutions such as universities, libraries, and public service organizations to construct their own digital repositories. It is particularly focused on sharing information in the public domain and supporting the needs of developing countries.

Greenstone is characterized by several key functional and technical features:

Core Interfaces: The software provides two distinct interactive interfaces: the Reader interface, which end-users use to access the library via a web browser, and the Greenstone Librarian Interface (GLI). The GLI is a graphical tool used to manage the entire lifecycle of a collection, from gathering and enriching material with metadata to designing searching and browsing facilities.

Document and Metadata Support: It can handle collections containing thousands or millions of documents, including text (PDF, Word, HTML), images, audio, and video. It supports standard metadata sets such as Dublin Core, MARC, and CDS/ISIS. 2.2 Multilingual Nature: One of its unique strengths is its high degree of multilingual support; the reader's interface is available in over 60 languages. It is Unicode-compliant, allowing it to process and display documents in any language consistently.

Interoperability: Greenstone is highly interoperable, supporting contemporary standards such as OAI-PMH for metadata harvesting and the Z39.50 protocol. The system can ingest or export collections in METS and Dublin Core formats.

Flexible Distribution: Collections may be disseminated via the Internet or published on removable media such as CD-ROMs, DVDs, and USB flash drives for use in offline environments.

Platform Compatibility: The software is a multiplatform application, compatible with all versions of Windows, Unix/Linux, and Mac OS X.

11.4 VERSIONS OF GREENSTONE

The development history of Greenstone is intriguing. Initially released on CD-ROM in

1998, UNESCO began distribution in French and Spanish alongside English in 2003. Currently, UNESCO has released two major versions of the Greenstone software, Greenstone2 and Greenstone3, along with their subsequent minor versions. While both versions share the core objective of enabling institutions to construct digital libraries, they exhibit significant differences in their technical architecture and flexibility.

11.4.1 Greenstone 2 is the legacy iteration, originally developed in C++.

11.4.1.1 Architecture and Language: Greenstone2 was created in 2000 and is implemented in C++. Due to the absence of standardized web technologies at the time of its inception, it relies significantly on custom formats and syntax.

11.4.1.2 Workflow: It functions predominantly as a batch-mode process, indicating that collections are not constructed interactively. Incremental collection assembly in this version was typically facilitated through a now-defunct web interface depositor.

11.4.1.3 Structure: It employs a singular "collect" directory and a sole run-time Common Gateway Interface program.

11.4.1.4 Current Status: It is presently in maintenance mode, with developers issuing periodic maintenance releases for existing users.

11.4.2 Greenstone 3: It is the recommended version for download and active development originally released in 2017, written in Java.

11.4.2.1 Architecture and Language: Greenstone3 is written in Java and leverages modern web standards such as XML Transforms (XSLT) for presentation and the Java Authentication and Authorization Service (JASS) for security.

11.4.2.1 Design and Modularity: It is structured as a network of independent modules that communicate via XML. This modularity simplifies the addition of new services and allows the system to run in a distributed fashion across multiple servers.

11.4.2.3 Dynamic Features: Unlike the batch-oriented Greenstone2, Greenstone3 allows documents to be dynamically added to collections and provides library management facilities directly within the web browser interface.

11.4.2.4 Customization: It is highly customizable through the use of XSLT templates and jQuery-based themes, which allow non-technical users to change the library's appearance without coding knowledge.

11.4.3 Compatibility and Comparison

11.4.3.1 Backwards Compatibility: Greenstone3 is fully backwards compatible with Greenstone2.The Greenstone Librarian Interface (GLI) features the capability to import

and execute legacy Greenstone2 collections without alterations.

11.4.3.2 User Interface Modes: Greenstone2's GLI traditionally provided four user modes (Assistant Librarian, Librarian, Library System Specialist, and Expert), while Greenstone3 has streamlined these into three modes (Assistant, Librarian, and Expert).

11.4.3.3 Platform Support: Both versions are multiplatform, compatible with all iterations of Windows, Unix/Linux, and Mac OS X. They exhibit substantial interoperability with standards such as OAI-PMH, METS, and Z39.50.

11.5 SYSTEM REQUIREMENTS

The installation of the Greenstone Digital Library (GSDL) necessitates a combination of fundamental hardware and specific software dependencies that differ slightly by operating system. As of 2026, the most recent stable release is Greenstone 3.12.

11.5.1 Software Requirements Operating Systems: Greenstone is cross-platform and compatible with Windows (Win 10/11), Linux (recommended Ubuntu 20.04+), and macOS (Monterey 12.7.6 to Sonoma 14.7.6).

Java Runtime Environment (JRE): Essential for the Greenstone Librarian Interface (GLI). Version 8 or higher is required.

Perl: Necessary for constructing and processing collections. It is generally pre-installed on Unix/Linux but must be installed on Windows (often included with the GSDL installer).

Web Browser: Required for the Reader interface. Modern browsers such as Chrome, Firefox, or Edge are standard; older versions necessitated Netscape 4.0 or Internet Explorer 4.0+.

11.5.2 Web Server: Local Library: Utilizes a compact integrated server (no additional software required). Web Library: Necessitates an external server such as Apache (for Greenstone 2) or Tomcat (for Greenstone 3).

11.5.3 Optional Dependencies:

ImageMagick: Essential for processing image-centric collections (thumbnails, covers).

Ghostscript: Necessary for managing and converting PDF or PostScript files.

11.5.4 Hardware Requirements: Greenstone is engineered to operate on minimal infrastructure, ensuring high accessibility.

Processor: Minimum Pentium 3.0 GHz or equivalent.

Memory (RAM): Minimum 512 MB; however, 1 GB or more is advisable for extensive collections.

Disk Space: Binary Installation: Approximately 40–50 MB. Source Code (Compilation): Approximately 90–160 MB. Storage: Additional space is necessary for the digital collections (e.g., 10 GB+ recommended for institutional use).

Version Selection:

Greenstone 3: The currently recommended version. It is written in Java and uses modern web technologies like XML and XSLT.

Greenstone 2: The legacy C++ version, still maintained for existing users and compatibility.

For the most current downloads and documentation, visit the official Greenstone website. (https://files.greenstone.org/manuals/gsdl2/en/html/Chapter_installation_procedure.htm).

11.6. CUSTOMIZATION WHEREABOUTS IN GREENSTONE2 AND GREENSTONE3

11.6.1 Programming and Communication: Greenstone2 was written in C++ and functions as a batch-mode process where a single run-time Common Gateway Interface (CGI) program handles requests. In contrast, Greenstone3 is written in Java and is structured as a network of independent modules that communicate using XML, allowing for a distributed system that can be spread across multiple servers.

11.6.2 Dynamic Capabilities: A significant customization advantage of Greenstone3 is its ability to dynamically add documents to collections and provide library management facilities directly within the web browser interface. Greenstone2 generally necessitated the reconstruction of collections in batch mode to incorporate modifications.

11.6.3 Configuration Files: Customization in Greenstone2 is primarily focused on the collect.cfg and build.cfg files. In contrast, Greenstone3 employs XML-compliant files such as Config.xml and buildConfig.xml, which are more consistent with contemporary web standards.

11.6.4 Templates vs. Macros: The visual presentation of Greenstone2 is determined by macro files (.dm) and CSS. Greenstone3 supersedes this custom macro system with XSLT templates, facilitating more precise control over the transformation of XML output from the library into HTML for the browser. 5.5 Themes for Non-Technical Users: Greenstone3 introduced jQuery-based themes, enabling users to modify colour schemes and layouts without any proficiency in CSS, HTML, or XSLT. Greenstone2 does not possess the advanced “point-and-click” aesthetic customization feature.

11.6.5 Format Statements: Both iterations utilize format statements to regulate the presentation of documents and search outcomes. However, Greenstone3 employs an XML-based GSF syntax, which the system converts into XSLT, while Greenstone2 utilizes a specific custom format.

11.6.6 Security Standards: Greenstone3 integrates the Java Authentication and Authorization Service (JASS) for security and authentication purposes. In contrast, Greenstone2 depends on its internal administrative functions to oversee user groups and document security.

11.6.7 User Levels: The Greenstone Librarian Interface (GLI) in Greenstone2 traditionally provided four user modes (Assistant Librarian, Librarian, Library Systems Specialist, and Expert). Greenstone3 has streamlined these into three modes: Assistant, Librarian, and Expert.

5.9 Remote Management: Greenstone3 provides built-in library management facilities accessible through the browser after logging in, which is considered a significant advancement in administrative customization over Greenstone2.

11.7 GREENSTONE LIBRARIAN INTERFACE (GLI)

The Greenstone Librarian Interface (GLI) is a central, Java-based graphical tool designed to manage the entire lifecycle of a digital library collection. It serves as the primary workspace for collection developers, allowing them to gather content, assign metadata, design searching and browsing structures, and build the final collection.

The GLI is organized into a series of functional panels that generally follow the workflow of collection building:

11.7.1 Download: This panel enables users to retrieve documents from the Internet using various protocols, including Web (HTTP/FTP), MediaWiki, OAI-PMH, Z39.50, and SRW.

11.7.2 Gather: In this section, developers explore their local file space and drag-and-drop files into the collection area. The GLI preserves any existing metadata when documents are imported from other Greenstone collections.

11.7.3 Enrich: Users manually add or edit metadata for individual documents, groups of files, or entire folders. Metadata assigned to a folder is automatically inherited by all documents within it, which expedites the organization of large collections.

11.7.4 Design: This panel is critical for determining how users interact with the library;

it is used to configure document plugins for processing different file types and to set up search indexes (using engines like MG, MGPP, or Lucene) and browsing classifiers.

11.7.5 Create: This is where the physical building of the collection takes place. The interface provides visual feedback through a progress bar and allows expert users to monitor the output of underlying Perl scripts to ensure the process is running correctly.

11.7.6 Format: Developers use this panel to customize the user interface and define how document content and search results are displayed on the web pages generated “on the fly” by system.

User Expertise Levels- The GLI accommodates four distinct user proficiency tiers:

- Assistant Librarian: Able to execute fundamental tasks, such as appending documents/metadata to current collections or duplicating an existing collection’s framework.
- Librarian (Default): Proficient in creating new collections by modifying existing templates, incorporating new document types, and configuring search/browse functionalities.
- Library Systems Specialist: Equipped for advanced customizations, including the application of regular expressions for collection segmentation.
- Expert: Possesses comprehensive access to all functionalities, including the capability to schedule builds and scrutinize detailed system logs.

11.8 ADVANCED MANAGEMENT FEATURES

The GLI encompasses specialized tools such as the Greenstone Editor for Metadata Sets (GEMS), which facilitates the creation or modification of custom metadata schemas. In collaborative settings, the Remote Librarian Interface (ClientGLI) enables users to edit collections hosted on a remote server without necessitating a local Greenstone installation. To preserve data integrity, the system employs a collection locking mechanism (via a .lck file) to ensure that only one individual can edit a specific collection at any given time. Additionally, while Greenstone3 represents a significant technical overhaul, it incorporates a feature within the GLI to import legacy Greenstone2 collections for contemporary use. Greenstone Demo Site for practice: GSDL operates a demo site that permits users to explore its various services. https://www.greenstone.org/greenstone3/library

CHECK YOUR PROGRESS

A. Fill in the Blanks

a) Greenstone Digital Library Software was developed by the New Zealand Digital Library Project at the -----____.

b) Greenstone software is distributed under the _____ General Public License.

c) The graphical tool used to manage the entire lifecycle of a Greenstone collection is called _____.

11.9 LET US SUM UP

Greenstone Digital Library Software (GSDL) is a robust, open-source solution for building, managing, and disseminating digital library collections. Designed with a strong focus on accessibility, interoperability, and support for developing countries, it enables institutions to organize diverse digital resources using international metadata standards and multilingual interfaces. The evolution from Greenstone2 to Greenstone3 reflects a major technological advancement, with Greenstone3 offering modular architecture, dynamic collection management, enhanced customization, and modern web standards while maintaining backward compatibility. The Greenstone Librarian Interface (GLI) plays a central role by providing an intuitive, workflow-based environment for collection development and administration. Overall, GSDL remains a flexible, scalable, and cost-effective digital library platform suitable for academic, cultural, and public information institutions worldwide.

11.10 ANSWERS TO CHECK YOUR PROGRESS

A, Fill in the Blanks

a) University of Waikato.

b) GPL

c) Greenstone Librarian Interface

11.11 FURTHER READINGS

1. Bainbridge, D., & Witten, I. H. (2004, June 7–11). Greenstone digital library software: Current research. In *Proceedings of the 2004 Joint Conference on Digital Libraries (JCDL '04)* (pp. 63–69). Tucson, AZ: ACM.
2. Biswas, G., & Paul, D. (2010). An evaluative study on the open source digital library softwares for institutional repository: Special reference to Dspace and greenstone digital library. *International Journal of Library and Information Science*, 38–62.

3. Greenstone Digital Library. (2015, June 1). *Factsheet :: Greenstone Digital Library Software*. Retrieved from http://www.greenstone.org/factsheet
4. Greenstone Wiki. (2023, March 13). *en:beginner:customization*. Retrieved from http://wiki.greenstone.org
5. Greenstone Wiki. (2023, March 13). *en:beginner:install_basic*. Retrieved from http://wiki.greenstone.org
6. Greenstone Wiki. (2023, March 13). *en:beginner:overview_of_greenstone*. Retrieved from http://wiki.greenstone.org
7. Greenstone Wiki. (2023, March 13*). en:user_advanced:remote_greenstone*. Retrieved from http://wiki.greenstone.org
8. Kumar, S. (2013). Case studies: Greenstone digital library software. In C. K. Ramaiah (Ed.), *Electronic Resources Management in Libraries.*
9. Parichi, R., & Nisha, F. (2015). Greenstone Digital Library Management System – a functional review based on selected criteria. *Library Hi Tech News, 32*(10), 16–21. https://doi.org/10.1108/LHTN-06-2015-0041
10. Singh, R. (2020). *Greenstone Digital Library Software*. Lucknow: University of Lucknow.
11. Tramullas, J., Sánchez-Casabón, A.-I., & Garrido-Picazo, P. (2013). An Evaluation Based on the Digital Library User: An Experience with Greenstone Software. *Procedia - Social and Behavioral Sciences, 73*, 167–174.
12. Witten, I. H. (n.d.). *The development and usage of the Greenstone digital library software*. Hamilton, New Zealand: University of Waikato.
13. Witten, I. H., & Bainbridge, D. (2007, January). *A brief history of the Greenstone Digital Library Software.* Hamilton, New Zealand: University of Waikato.
14. Witten, I. H., & Bainbridge, D. (n.d*.). Creating Digital Library Collections with Greenstone*. Hamilton, New Zealand: New Zealand Digital Library Project, University of Waikato

11.12 PROBABLE QUESTIONS

a) Explain the objectives and key features of Greenstone Digital Library Software (GSDL).

b) Discuss the role and functions of the Greenstone Librarian Interface (GLI) in building digital library collections.

c) Compare Greenstone2 and Greenstone3 with reference to architecture, workflow, and customization features.

d) Describe the system requirements and platform compatibility of Greenstone Digital Library Software.

e) Explain the customization features available in Greenstone3 and how they differ from Greenstone2.

f) Discuss the interoperability and multilingual features of Greenstone Digital Library Software.

*** ***** ***

UNIT 12: CASE STUDY OF POPULAR DL SOFTWARE: DSPACE

UNIT STRUCTURE

12.1 Learning Objectives

12.2 Introduction

12.3 Functional Overview

12.4 Digital Library adopting DSpace

12.5 Let Us Sum Up

12.6 Answers to Check Your Progress

12.7 Further readings

12.8 Probable Questions

12.1 LEARNING OBJECTIVES

After going this unit, learners will be able to

- Know about the various features and functionality of DSpace.
- Know about the Information Model and Architecture of DSpace.

12.2 INTRODUCTION

DSpace is one of the most widely used open-source digital library and institutional repository software platforms worldwide. DSpace is designed to capture, store, index, preserve, and disseminate digital content. Originally developed in 2002 through a collaboration between the Massachusetts Institute of Technology (MIT) and Hewlett-Packard (HP), DSpace was conceived to address the long-term preservation and access challenges associated with institutional intellectual output. Over time, it has evolved into a mature, community-driven platform maintained by LYRASIS, with continuous contributions from an international network of developers, librarians, and information professionals. The latest stable DSpace version is DSpace 8.3, released December 17, 2025, with DSpace 9.2 also available, offering new features and support for newer platforms.

12.3 FUNCTIONAL OVERVIEW

DSpace functions as a centralized institutional repository, enabling various communities

like departments, centers, schools, to maintain distinct sections within the platform. Community members can upload content easily through an intuitive web interface, or administrators can use the batch importer for large-scale content. Communities can designate 'gatekeepers' to review and refine submissions before they go live in the repository.

The system indexes the accompanying metadata and grants access based on community-defined privileges. To launch a practical service quickly, DSpace adopted a 'breadth-first' development approach: it implemented core repository features in straightforward ways, allowing refinements while the system runs in production

DSpace serves as a centralized platform for institutional repositories. Communities such as labs, centers, schools, or departments each get dedicated spaces. Members deposit items via a user-friendly web interface or through batch import for bulk uploads. Gatekeepers, appointed by communities, review and edit submissions before approval. (Tansley et al., 2003)

Once ingested, DSpace indexes the metadata and enforces access controls set by the community. Developed using a 'breadth-first' strategy, it prioritizes basic functionality across all key areas, enabling iterative improvements on a live system.

The functional aspects of DSpace can be summarized as follows:

12.3.1 DATA MODEL

DSpace employs a hierarchical, container-based data model to organize digital content, reflecting institutional structures while enabling flexible preservation and access. At the highest level, the repository is divided into communities, which may be further organized into sub-communities to represent entities such as colleges, departments, research centres, or laboratories. Each community contains one or more collections, which serve as logical groupings of related digital content. A single collection may be associated with more than one community, thereby supporting interdisciplinary or cross-organizational representation.

Collections comprise items, which constitute the fundamental archival units within DSpace. Each item is formally owned by a single collection, ensuring clear administrative responsibility and provenance. However, to enhance discoverability and reuse, an item may also be mapped to additional collections without duplication. Items are further organized internally into files and its associated metadata. The structure is summarised in Figure 1.

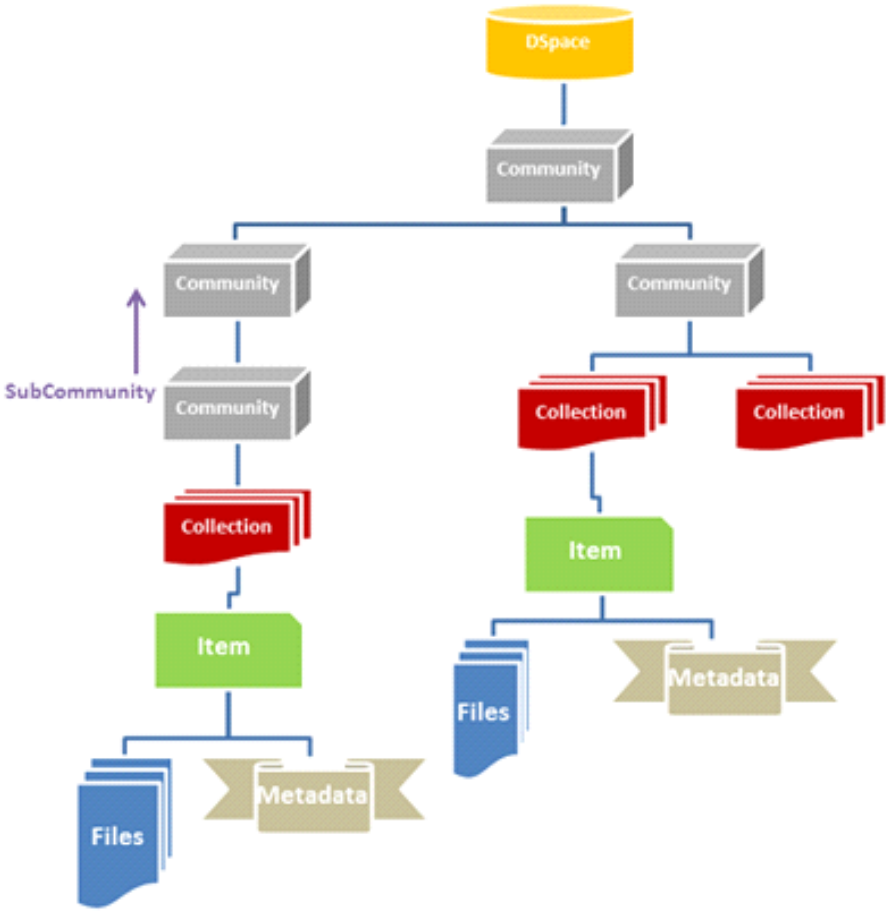


Figure 1: Data Model

12.3.2 METADATA

DSpace captures three primary metadata categories for archived content: descriptive, administrative, and structural.

- **Descriptive Metadata**

Every Item requires one qualified Dublin Core (QDC) record. Institutions can customize the element/qualifier registry, though search and submission interfaces require manual updates. Additional formats like MARC records store as serialized bitstreams; communities and collections hold basic name/prose metadata in the DBMS.

- **Administrative Metadata**

DSpace stores administrative metadata including preservation details, provenance records, and authorization policies primarily within its relational DBMS schema. Provenance information in prose form embeds directly into Dublin Core records, while select attributes like bitstream sizes and MIME types duplicate there for external accessibility, such as

through OAI-PMH.

- **Structural Metadata**

This category defines item presentation strategies and relationships among components, such as page sequencing in a thesis composed of TIFF images where each represents a single page. Current capabilities remain basic: bitstreams organize into bundles (with optional primary bitstreams, e.g., the lead HTML file for browser display), and each carries a unique sequence ID for persistent identification within items. Advanced structural data can serialize into bitstreams, though DSpace lacks native parsing in this version.

12.3.3 E-PEOPLE

Authenticated users or “e-people” perform actions like submissions, subscriptions, or administration. DSpace groups e-people for streamlined management.

DSpace stores these details per e-people:

- Email address
- First and last names
- Encrypted authentication data (e.g., passwords)
- Subscribed collections for new-item alerts
- Self-registration flag (auto-created vs. imported from personnel DB)

E-people join groups to simplify administrator tasks when assigning authorization policies.

12.3.4 AUTHORIZATION

DSpace has a granular and flexible authorization system. DSpace’s authorization system is based on associating actions with objects and the lists of EPeople who can perform them. The associations are called Resource Policies, and the lists of EPeople are called Groups. There are two built-in groups: ‘Administrators’, who can do anything in a site, and ‘Anonymous’, which is a list that contains all users.

To perform any action on a DSpace object, a user must be granted the appropriate permission. Permissions in DSpace are non-transitive, meaning that authorization granted at one level does not automatically extend to related objects.

The authorization system recognizes a defined set of actions, such as READ, WRITE, ADD, and REMOVE. Notably, there is no explicit DELETE action within DSpace. Instead, object deletion is achieved indirectly. To remove an object (such as an item) from the

repository, a user must possess REMOVE permission on all parent objects that contain it (for example, the owning collection).

12.3.5 INGEST PROCESS AND WORKFLOW

DSpace's ingest process systematically ingests digital objects into its repository, ensuring data integrity, metadata standardization, and workflow governance. This process supports both individual and bulk submissions, transforming raw content into searchable, preserved items through validation, review, and dissemination.

Depending on the submission policies defined for a particular collection, the deposit of an item may trigger a workflow process. This process typically involves one or more designated human reviewers who are responsible for examining the submission to verify its suitability, accuracy, and compliance with collection guidelines. Workflow configurations are established at the time a collection is created and are applied to all items submitted to that collection.

When a submission is completed through either the batch item importer or the web-based submission interface, the system initiates the next stage of the ingest process, which may involve workflow review or direct item installation. At this point, a provenance statement is recorded within the Dublin Core metadata, documenting key details of the submission, including the filenames and checksums of the associated content files. Similarly, each transition within the workflow results in the addition of a corresponding provenance entry. These provenance records collectively enable the tracking of changes and actions applied to an item from the moment of submission onward.

Once the workflow process is successfully completed, the InProgressSubmission object is processed by an item installer, which transforms it into a fully archived and publicly accessible item within DSpace.

The workflow associated with a DSpace collection may consist of up to three sequential steps, each designed to introduce varying levels of review and control before an item is archived.

In workflow step one (review), designated reviewers examine the submission and may either accept it for inclusion or reject it.

Workflow step two (edit) allows authorized users to review and edit the metadata supplied by the submitter, although they are not permitted to modify the submitted files. At this stage, the submission may again be accepted or rejected.

Workflow step three (final edit) also permits metadata editing without altering the files; however, at this final stage the reviewer must either commit the item to the archive or halt the process, as rejection is no longer permitted.

Each workflow step may be assigned to a specific e-person group. If no group is assigned to a particular step, that step is skipped. In cases where no e-person groups are assigned to any of the workflow steps, submissions to the collection bypass the workflow entirely and are installed directly into the archive.

Operationally, the workflow proceeds as follows: once a collection receives a submission, DSpace checks whether a group has been assigned to workflow step one. If so, that step is activated and the relevant group is notified; if not, the system advances to the next step. The same logic applies to workflow steps two and three, which are executed only if corresponding groups are defined.

When a workflow step is initiated, the associated task is placed into a task pool for the assigned group. A single member of the group claims the task from the pool, at which point it is removed to prevent multiple users from unknowingly performing the same review simultaneously.

If a submission is rejected, the reviewer must provide a reason, which is communicated to the submitter via email.

This structured, group-based workflow model ensures controlled review, prevents duplication of effort, and maintains consistency and accountability in the management of digital content within DSpace repositories.

12.3.6 PERSISTENT URLS AND IDENTIFIERS

DSpace generates a persistent identifier for every community, collection, and item stored within the system. In order to ensure long-term persistence, these identifiers must be independent of both physical storage location and internal system structure. For this purpose, DSpace adopts the CNRI Handle System, a globally recognized infrastructure for persistent identification.

Within DSpace, Handles serve primarily as globally unique identifiers for repository objects. Each DSpace installation is required to obtain a unique Handle prefix from CNRI, ensuring that identifiers created within one repository do not conflict with those generated elsewhere. The combination of this prefix with a locally assigned suffix guarantees global uniqueness.

A significant advantage of the Handle System is its global resolution infrastructure. An end-user may enter a Handle into any Handle-aware service, such as a web browser, and be redirected to the object identified by that Handle. To support this functionality, each DSpace site operates a local Handle server, which receives resolution requests and redirects users to the appropriate community, collection, or item. All necessary software components for running this service are included within the DSpace distribution.

Handles may be expressed in two equivalent forms:

hdl:1721.123/4567

http://hdl.handle.net/1721.123/4567

Both representations refer to the same identifier. While the first form is useful as a compact identifier, the second form is more practical for end-users, as it can be resolved directly by standard web browsers. Consequently, DSpace displays Handles using the HTTP-based format to maximize usability and accessibility.

In addition to Handles for higher-level objects, DSpace provides persistent identifiers for bitstreams. These identifiers are described as “persistent” in a limited sense, as they remain valid only within the context of a given DSpace installation. If content is moved to another server or organizational environment, these identifiers may no longer resolve correctly.

Each bitstream is assigned a sequence identifier, which is unique within the scope of its parent item. This sequence identifier is used to construct a bitstream URL of the following form:

DSpace-URL/bitstream/Handle/sequence-ID/filename

12.3.7 SEARCH AND BROWSE

Users can discover content in DSpace in multiple ways. Including:

- Keyword Search
- Browsing through different access point (Author, Date)
- External Reference (Web search, Handle)

DSpace offers a rich and responsive search experience. Its indexing and search module provides a straightforward application programming interface (API) that supports the indexing of newly added content, regeneration of indexes when required, and execution of searches across the entire repository or within specific communities or collections. Internally, this functionality is powered by the Java-based open-source search engine

Apache Lucene, which provides advanced features such as fielded searching, stop-word removal, stemming, and incremental indexing. These capabilities allow new content to be added to the index without necessitating a complete reindexing of the repository.

In addition to search, browsing serves as an equally important discovery mechanism in DSpace. Browsing enables users to explore repository content by navigating through predefined indices, such as item titles, issue dates, or authors, in order to identify materials of interest. The browse subsystem offers a simple API that allows the system to present selected segments of an index based on user input. Browsing can be applied globally across the repository or restricted to a particular community or collection, thereby supporting both broad exploration and focused discovery.

12.3.8 OAI SUPPORT

The Open Archives Initiative (OAI) has developed a standardized protocol for metadata harvesting, enabling repositories and service providers to collect metadata from multiple distributed sources. The harvested metadata can be aggregated and reused for the purpose of centralized searching, indexing, linking, and analytics.

DSpace implements this protocol by exposing Dublin Core metadata for all items that are publicly accessible without authentication. In addition to item-level metadata, DSpace makes its collection structure available through the OAI protocol's sets mechanism, thereby allowing harvesters to selectively retrieve metadata from specific communities or collections. This functionality is implemented using OCLC's open-source OAICat framework.

Furthermore, DSpace can be configured to disseminate metadata in formats other than Dublin Core by utilizing appropriate metadata crosswalk plugins, enabling support for schemas such as MODS and others as required by service providers.

Using record status, DSpace's OAI service is capable of exposing deletion information for withdrawn items, ensuring that harvesters are informed when content has been removed from public view. However, metadata for items that have been permanently expunged from the system is not exposed. DSpace also supports the use of OAI-PMH resumption tokens, which facilitate the efficient harvesting of large result sets by allowing metadata to be retrieved in manageable segments.

12.3.9 SYSTEM ARCHITECTURE

The latest stable version of DSpace (DSpace 8.3) adopts a modern, API-driven, service-oriented architecture that clearly separates the user interface from backend services, building on the architectural redesign introduced in DSpace 7. This approach enhances scalability, interoperability, and maintainability.

At a high level, DSpace 8 follows a three-tier architecture:

- User Interface (Angular)
- Application Services (REST API)
- Storage Layer (Database + File Storage)

The User Interface layer is implemented as an Angular-based single-page application that handles content discovery, submission, workflow interaction, and administration. It communicates exclusively with the backend through RESTful APIs, allowing the UI to be customized or replaced independently.

The Application and Service layer is built using Spring Boot and exposes core repository functionality via a comprehensive REST API. This layer implements business logic such as workflow management, authorization, metadata processing, persistent identifier assignment (Handles and DOIs), and interoperability services.

The Storage layer manages repository data. Metadata and system information are stored in a relational database (typically PostgreSQL), while digital files (bitstreams) are stored in the file system or object storage. File integrity is ensured through checksums and format registries.

For discovery, DSpace 8 integrates Apache Solr, providing full-text search, faceted browsing, and usage statistics. Authentication and authorization are handled through a pluggable framework supporting local accounts, LDAP, Shibboleth, and OAuth-based identity providers.

12.4 DIGITAL LIBRARY ADOPTING DSPACE

Some prominent institute adopting DSpace for digital library plateform are

- Shodhganga (INFLIBNET) https://shodhganga.inflibnet.ac.in/
- eGyanKosh (IGNOU) https://egyankosh.ac.in/
- National Institute of Oceanography https://drs.nio.res.in/drs/

CHECK YOUR PROGRESS

a) What is descriptive metadata?

b) What is handle system.

12.5 LET US SUM UP

DSpace is a well adopted digital library software because of its powerful, free, open-source platform for creating digital repositories, offering robust features for capturing, preserving, and providing open access to diverse digital content (text, media, datasets). Its large, active community, customizability, support for open standards, and flexibility to manage various file formats make it a trusted, cost-effective, and widely adopted solution for academic, government, and research institutions globally. In essence, DSpace provides a comprehensive, reliable, and community-driven solution for institutions to manage their digital assets, ensuring they remain accessible and preserved for the long term.

12.6 FURTHER READINGS

1. Tansley, R., Bass, M., Stuve, D., Branschofsky, M., Chudnov, D., McClellan, G., & Smith, M. (2003). The DSpace institutional digital repository system: Current functionality. In *Proceedings of the 2003 Joint Conference on Digital Libraries* (pp. 87–97). IEEE.
2. LYRASIS. (n.d.). *DSpace 9.x documentation* https://wiki.lyrasis.org/display/DSDOC9x

12.7 ANSWER TO CHECK YOUR PROGRESS

a) Descriptive metadata is defined as the intellectual content of a digital resource to help its discovery, identification, and selection. It covers key elements such as the title, author, abstract, keywords, and publication date, working like a digital library catalog card. This type of metadata is essential for allowing users to effectively search, browse, and retrieve relevant information from large collections

b) It is defined as a secure, distributed and open protocol system for assigning, managing, and resolving persistent identifier for digital and physical objects or resources available on internet.

12.8 PROBABLE QUESTIONS

3. Who developed DSpace?
4. Who is e-people in context of DSpace?
5. Discuss the Dublin Core Metadata.
6. Describe data model of DSpace.
7. Discuss the architecture of DSpace.

*** ***** ***

UNIT 13: DL EVALUATION AND MARKETING

UNIT STRUCTURE

13.1 Learning Objectives

13.2 Introduction

13.3 Concept And Need

13.4 Process of Evaluation

13.5 Criteria of Evaluation

13.6 Models of DL Evaluation, Marketing Of DL

13.7 Let Us Sum Up

13.8 Answers to Check Your Progress

13.9 Further Readings

13.10 Probable Questions

13.1 LEARNING OBJECTIVES

After this unit, learner will be able to

- Explain the idea of digital library assessment and elucidate its necessity.
- Outline the sequential procedure for assessing a digital library, encompassing planning through to analysis.
- Determine the principal criteria employed to evaluate digital libraries, including usability and system performance.
- Examine prominent evaluation models such as the DELOS model and DigiQUAL.
- Elucidate the notion of marketing within a library framework and enumerate techniques to enhance the promotion of digital services.

13.2 INTRODUCTION

Digital libraries are intricate systems that amalgamate technology, information resources, and human users to enable knowledge creation and dissemination. Similar to conventional libraries, digital libraries necessitate periodic evaluation to ensure they are achieving their objectives and successfully serving their customers. Nonetheless, due to their dependence on swiftly evolving technology and frequent service to remote consumers, the evaluation methodologies are unique and complex. This lesson examines the criteria for assessing the quality and success of a digital library through systematic evaluation. It also tackles a

vital although frequently neglected facet of library administration: marketing. A digital collection, regardless of its quality, is ineffective if prospective users are unaware of its existence or how to utilize it. By integrating assessment methodologies with efficient marketing, library administrators may guarantee their systems remain pertinent, sustainable, and extensively utilized.

13.3 CONCEPT AND NEED

Evaluation within digital libraries refers to the methodical assessment of the system's operation, performance, and value, along with its offerings. It is not solely a technical assessment to determine the functionality of the software. Rather, it is a comprehensive evaluation that assesses the system's effectiveness in achieving its objectives. This entails evaluating several attributes, including effectiveness, which assesses whether the system fulfils its intended purpose, and efficiency, which examines the resources and work necessary to attain those outcomes. Evaluation is a research-oriented process that gathers data such as user comments or usage metrics to assess the library's quality.

The principal objective of evaluation is to ascertain the proper functioning and utility of the digital library for its community. Digital libraries necessitate substantial investment in financial resources, trained personnel, and advanced technology. Administrators and funders require evidence that these resources are utilized judiciously and that the library is sustainable in the long term. Evaluation supplies the requisite proof to substantiate ongoing financing or to solicit budget augmentations.

Moreover, assessment is crucial for enhancement. It assists in identifying issues that users may encounter, such as challenges in locating documents or sluggish download speeds. By comprehending these concerns, library managers can make informed decisions regarding software upgrades, policy modifications, or interface enhancements. In the absence of review, a digital library risks becoming obsolete or irrelevant, as management would be unaware of the continued necessity of the content or whether the technology has grown outdated. It effectively connects the library's offerings with the users' actual need.

13.4 PROCESS OF EVALUATION

The evaluation of a digital library is a systematic cycle that commences well in advance of data collection. It necessitates meticulous planning to guarantee that the outcomes are

both valid and beneficial for decision-making.

1. Requirement Identification

The initial step is to explicitly articulate the purpose of the evaluation. The assessors must ascertain the particular activities of the library that require assessment and the inquiries that necessitate resolution. The objective may be to ascertain the reasons for the decline in user traffic or to evaluate the functionality of a new search feature. This phase entails determining the intended audience for the evaluation report, including library administration or external funding organizations.

2. Viability Assessment

Prior to initiating a comprehensive review, a feasibility study is typically performed to ascertain the practicality of the planned evaluation plan. This entails verifying whether the library possesses the funds, time, and resources to conduct the study. It guarantees that the assessment does not expend more resources than its value warrants. This step may include evaluate technical limitations, such as the accessibility of server logs for analysis.

3. Strategic Planning and Methodological Selection

Upon clarifying the needs, a comprehensive plan is produced. This entails choosing the particular techniques for data collection. Prevalent techniques encompass surveys to collect user feedback, observation to monitor user interactions with the system, and transaction log analysis to examine usage patterns documented by the server. The plan delineates a timeframe and allocates duties to personnel**.**

4. Data Acquisition and Examination

This is the implementation phase during which the selected methods are executed. Upon selection, a survey is disseminated to users. If log analysis is selected, the data is taken from the system. depicts the progression from data acquisition to analysis. The raw data is subsequently processed and analyzed to identify trends, patterns, or specific issues. Analysis may indicate that users often terminate searches after yielding no results, suggesting an issue with the search engine or the indexing of the collection.

5. Reporting and Decision-Making

The concluding stage involves analysing the results and consolidating them into a report for stakeholders. This report supplies the requisite evidence to implement policy modifications, rectify technological issues, or revise the library's collection development strategy. The objective is to transform data into implementable enhancements that augment the digital library's worth.

13.5 CRITERIA OF EVALUATION

Assessors employ defined standards to evaluate the quality of a digital library. These criteria serve as benchmarks for evaluating the library's performance.

- **Functionality**

Usability is a key criterion as it directly influences the user's experience. It assesses the user's ease of learning and utilizing the technology. Essential elements of usability encompass interface design, navigation simplicity, and the efficiency with which a user may accomplish activities such as locating a certain document. An effective system is intuitive, allowing a novice user to comprehend it without substantial training. Errors must be minimized, and the system should provide swift recovery for users in the event of a mistake.

- **Quality of Collection**

The essence of every library is in its collection. The evaluation criteria for the collection encompass the documents' relevance to the user population, the breadth and depth of the subjects, and the timeliness of the material. Evaluators assess the quality of digital items, including the legibility of scanned photographs and the searchability of text through Optical Character Recognition (OCR). They evaluate the metadata to guarantee that resources are appropriately defined and readily accessible.

- **System Efficiency**

This technical criterion evaluates the performance of the software and hardware. Key criteria encompass retrieval speed (the rapidity of search result display), dependability (the frequency of system failures or downtime), and scalability (the system's capacity to accommodate increased users or data in the future). Two fundamental concepts in information retrieval are typically assessed: precision (the proportion of relevant things retrieved to the total items retrieved) and recall (the proportion of relevant items retrieved to the total relevant items in the database).

- **Quality of Service**

In addition to material and software, digital libraries function as service organizations. This category assesses the assistance offered to users. It encompasses the accessibility of assistance features, the promptness of staff in addressing email inquiries, and the existence of community elements such as news updates or discussion forums. Models such as DigiQUAL emphasize this dimension, evaluating if users perceive their specific

demands as fulfilled and if they have confidence in the library's services.

- **Expenditure and Sustainability**

Ultimately, economic criteria are employed to assess the cost-effectiveness of the digital library. This entails evaluating the expenses associated with content licensing, server maintenance, and personnel compensation in relation to the advantages afforded to the community. Sustainability assesses if the library possesses a feasible long-term strategy to safeguard its digital assets and maintain operations despite any alterations in financing sources.

13.6 MODELS OF DL EVALUATION

Models of Deep Learning Evaluation

Researchers have created many approaches to standardize evaluation by establishing a framework for measurement criteria.

DELOS Reference Model: This extensive framework delineates a digital library via six principal concepts: Content, User, Functionality, Quality, Policy, and Architecture. It is especially beneficial as it encompasses not only the technological system but also the regulations regulating it and the quality standards for each component. It assists assessors in examining the library from all essential perspectives, encompassing both the data and the governing regulations.

5S Model: This formal framework delineates digital libraries through five abstract concepts: **Streams** (content categories such as text and video), **Structures** (information organization), **Spaces** (logical and presentation perspectives), **Scenarios** (services and user interactions), and **Societies** (user communities). It offers a stringent mathematical framework for characterizing digital libraries and is frequently employed to officially delineate quality indicators, so assuring that concepts such as "completeness" and "consistency" are quantitatively assessed.

DigiQUAL: Derived from the renowned LibQUAL+ instrument utilized in traditional libraries, DigiQUAL aims to assess service quality in digital contexts. It emphasizes user perceptions and anticipations. It employs surveys to inquire about users' experiences with the library's website, collection, and services, assisting managers in identifying discrepancies between user expectations and the library's actual offerings.

13.7 MARKETING OF DL

Digital library marketing is the management process that identifies, anticipates, and fulfills user needs in a profitable or efficient manner. Numerous librarians erroneously perceive marketing solely as selling or promoting. It encompasses the complete process of developing a desired service and ensuring user awareness of it.

The Concept and Necessity of Marketing: Digital libraries function within a competitive information landscape primarily influenced by search engines such as Google. If consumers are oblivious to the high-quality, validated resources within a digital library, the library fails to fulfill its objective. Marketing enhances exposure, use, and user engagement. It transforms the library's perception from a static repository into a dynamic centre of knowledge.

Marketing Strategies (The 4 Ps): Marketing strategies frequently employ the adapted "4 Ps" framework for libraries.

Product: The digital library, encompassing its collections and services. The product must exhibit superior quality and fulfil user requirements.

Price: Although frequently complimentary to users, the "cost" may encompass the user's time and work. Marketing emphasizes the advantages of time efficiency gained by utilizing structured, dependable library resources in contrast to disorganized web searches.

Place: This denotes distribution. In digital libraries, the "place" refers to the website or mobile application. It must be available around the clock from any location.

Promotion: This encompasses all methods via which the library engages with users. Strategies encompass email newsletters, social media initiatives, backlinks from other institutional websites, and online tutorials.

Promotion Strategies: Successful promotion depends on engaging consumers in their current context. Libraries utilize social media platforms to disseminate narratives regarding distinctive artifacts within their collections. They utilize Search Engine Optimization (SEO) to guarantee that library information is visible in general web searches. User education initiatives, like webinars and online guides, serve as effective marketing instruments by instructing users on how to derive value from the library, thereby enhancing their loyalty and utilization.

CHECK YOUR PROGRESS

A. Fill in the Blanks

a. The DELOS Reference Model outlines a digital library founded on six fundamental concepts: Content, User, Functionality, Quality, Policy, and _____.

b. In the 5S framework of digital libraries, the five conceptual elements are Streams, Structures, Spaces, Scenarios, and _____.

c. DigiQUAL is a service quality evaluation instrument derived from the _____ protocol utilized in conventional libraries.

d. The marketing mix method commonly utilized in libraries is referred to as the "4 Ps," representing _____ and ____.

13.8 LET US SUM UP

This unit addressed the fundamental techniques of assessment and promotion for digital libraries. Evaluation is a methodical examination employed to ascertain whether a digital library is achieving its aims successfully and efficiently. The procedure entails identifying requirements, strategizing, gathering data using techniques such as surveys and log analysis, and utilizing that data to make informed decisions. We examined essential evaluation factors, encompassing usability, collection quality, system performance, and service quality. Numerous models, including DELOS and DigiQUAL, offer frameworks to facilitate these evaluations. The unit emphasized that evaluation should be integrated with marketing. Marketing guarantees the visibility and utilization of important resources within a digital library by the intended audience. By using techniques such as the 4Ps and utilizing contemporary promotional tools, libraries can maintain their significance in the digital era.

13.9 ANSWERS TO CHECK YOUR PROGRESS

A. Fill in the blanks

a. Management, **b.** Societies, **c.** LibQUAL+, **d.** Product, Price, Place, Promotion

13.10 FURTHER READINGS

1. Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., Klas, C.-P., Kovács, L., Landoni, M., Micsik, A., Papatheodorou, C., Peters, C., & Sølvberg, I. (2007). Evaluation of digital libraries. *International Journal on Digital Libraries*, *8*(1),

21–38. https://doi.org/10.1007/s00799-007-0011-z

2. Khoo, M., & MacDonald, C. (2011). An Organizational Model for Digital Library Evaluation. *Research and Advanced Technology for Digital Libraries*, 329–340. https://doi.org/10.1007/978-3-642-24469-8_34
3. *NDLI: Digital library : marketing and promotions*. (2026). Iitkgp.ac.in. http://ndl.iitkgp.ac.in/he_document/inflibnet_epgp/inflibnet_epgp/IN__I_e_P_P__1_L_a_I_S__28899_P_0_D_I__29579_M_1_D_I_m_a_p__29623_29624
4. *NDLI: Evaluation and case studies*. (2026). Iitkgp.ac.in. http://ndl.iitkgp.ac.in/he_document/inflibnet_epgp/inflibnet_epgp/IN__I_e_P_P__1_L_a_I_S__28899_P_0_D_I__29579_M_2_E_a_c_s__29613_29614
5. Pandya, M. (2025). *Evaluation of Digital Library – Digital Libraries*. Inflibnet.ac.in. https://ebooks.inflibnet.ac.in/lisp8/chapter/evaluation-of-digital-library/
6. Prasad, A. (n.d.). *Digital Library Software and Evaluation Criteria*. Retrieved January 3, 2026, from https://egyankosh.ac.in/bitstream/123456789/61854/1/Digital%20Library%20Software%20%26%20Evaluation%20Criteria.pdf
7. Wankhede, K. A. (2025). *Digital Library Evaluation*. *13*, 2320–2882. https://www.ijcrt.org/papers/IJCRT2503233.pdf

13.11 PROBABLE QUESTIONS

A. Short Answer Type Questions

a. What is the main objective of performing a feasibility study prior to evaluating a digital library?

b. Elucidate the distinction between “precision” and “recall” as metrics for assessing system efficacy.

c. Enumerate three separate methodologies employed for data collection in the assessment of a digital library.

B. Long Answer Type Questions

a. Discuss the DELOS Reference Model for digital libraries, outlining its six fundamental elements and clarifying its role in facilitating thorough examination.

b. Examine the significance of marketing for digital libraries and elucidate how the “4 Ps” of marketing (Product, Price, Place, Promotion) may be utilized in this context to enhance user engagement.

*** ***** ***

UNIT 14: POPULAR DIGITAL LIBRARY INITIATIVES

UNIT STRUCTURE

14.1 Learning Objectives

14.2 Introduction

14.3 Indian Initiatives

14.4 Global Initiatives

14.5 Let Us Sum Up

14.6 Answers to Check Your Progress

14.7 Further Readings

14.8 Probable Questions

14.1 LEARNING OBJECTIVES

After this unit, learner will be able to

- Identify and explain significant digital library projects in India, including the National Digital Library of India (NDLI), Vigyan Prasar, and the National Mission for Manuscripts.
- Examine significant international digital library initiatives, such as Project Gutenberg, the Networked Digital Library of Theses and Dissertations (NDLTD), and Europeana.

14.2 INTRODUCTION

The emergence of digital technologies has transformed the methods of information collection, preservation, and dissemination. Digital libraries have become an essential element of the global information infrastructure, surpassed geographical limitations and offered immediate access to an extensive repository of human knowledge. The digital library movement in India has gained considerable speed in recent decades, propelled by governmental initiatives and the necessity to conserve the nation's extensive cultural heritage and scientific contributions. Likewise, international initiatives have established benchmarks in digitizing standards, open access, and collaborative resource sharing.This unit offers a comprehensive examination of prominent digital library efforts, both in India and internationally. It examines the aims, collections, and influence of these projects, emphasizing their contribution to democratizing information access. These programs

illustrate the varied applications of digital library technologies, ranging from the preservation of ancient manuscripts to the provision of access to advanced scientific research. The unit is organized to initially present key Indian initiatives, thereafter providing an in-depth examination of notable global projects that have influenced the digital landscape.

14.3 POPULAR INDIAN INITIATIVES

The digital library landscape in India emphasizes the preservation of its extensive cultural legacy and the advancement of scientific knowledge. Government agencies, frequently in partnership with university institutions, have led these initiatives.

14.3.1 NDLI

The National Digital Library of India (NDLI) is a premier initiative within the National Mission on Education through Information and Communication Technology (NMEICT). Created by the Indian Institute of Technology Kharagpur, NDLI functions as a virtual repository of educational resources using a unified search interface. It is intended for learners of all proficiency levels, ranging from school pupils to lifelong learners, and accommodates various languages and material formats, including books, articles, videos, and simulations.

NDLI functions as an aggregator, collecting metadata from many educational institutions and digital repositories globally and within India. It offers unrestricted access to an extensive array of educational resources, rendering it an essential asset for students and scholars. The platform includes a mobile application, hence improving accessibility. NDLI is essential in closing the digital divide in education by consolidating various resources into a single platform.

14.3.2 VIGYAN PRASAR DIGITAL LIBRARY

Vigyan Prasar, an autonomous entity within the Department of Science and Technology, launched its digital library to disseminate knowledge of science and technology to the public. The library has a collection of notable scientific publications by Vigyan Prasar, encompassing books, periodicals, and audio-visual materials. The major objective is to disseminate scientific reasoning and information by providing these resources in a publicly accessible digital format.

The digital collection comprises digitized full-text editions of books in English, Hindi,

and various other Indian languages. It additionally archives Dream 2047, a renowned science magazine, from its first issue. Vigyan Prasar greatly enhances science communication and education in India by offering open access to scientific literature. The platform guarantees that significant scientific knowledge disseminates to a broader audience, surpassing the constraints of print dissemination.

14.3.3 NATIONAL DATABANK ON INDIAN ART AND CULTURE (NDBIAC)

The National Databank on Indian Art and Culture (NDBIAC) was established by the Indira Gandhi National Centre for the Arts (IGNCA) as a pilot initiative in partnership with the Archaeological Survey of India (ASI). The project seeks to improve the accessibility of Indian cultural resources through digital technologies. It is as a repository for digitized photographs, audiovisual materials, and writings pertaining to India's extensive artistic and cultural heritage.

NDBIAC offers access to virtual tours of archaeological sites, past editions of the ASI magazine Indian Archaeology - A Review, and rare texts in Indic languages such as Sanskrit and Hindi. The project emphasizes the significance of digital preservation in protecting India's tangible and intangible cultural heritage. Through the digitization of these resources, NDBIAC guarantees their accessibility to scholars and the public, thereby promoting a more profound comprehension of Indian history and culture.

14.3.4 NATIONAL MISSION FOR MANUSCRIPTS

Initiated in 2003 by the Ministry of Tourism and Culture, the National Mission for Manuscripts (NMM) seeks to discover and save India's extensive collection of manuscripts. India harbors around five million manuscripts, establishing it as one of the most extensive collections of handwritten information globally. The mission's mandate encompasses the documentation, conservation, and digitization of ancient manuscripts to avert their degradation and disappearance. NMM operates a nationwide database named Kritisampada, which offers information on manuscripts accessible throughout the country. A network of Manuscript Resource Centres (MRCs) and Manuscript Conservation Centres (MCCs) has been formed to locate and preserve endangered manuscripts. The objective additionally emphasizes the digitization of certain manuscripts to enhance accessibility for scholars and researchers. NMM is crucial in safeguarding India's intellectual heritage for future generations.

14.3.5 E-SHODHSINDHU

E-Shodhsindhu was established by the amalgamation of three prominent consortia: UGC-INFONET Digital Library Consortium, NLIST, and INDEST-AICTE Consortium. This effort is undertaken by the Ministry of Education, Government of India, and implemented by the INFLIBNET Centre. The consortium offers both current and archival access to over 10,000 essential and peer-reviewed journals, along with several bibliographic, reference, and factual databases.

The principal aim of E-Shodhsindhu is to provide access to high-quality electronic resources, encompassing full-text, bibliographic, and factual databases, for academic institutions at a reduced subscription cost. It seeks to close the digital divide and advance toward an information-abundant society. The consortium negotiates tariffs and access terms to guarantee that member institutions, such as universities and colleges, obtain cheap access to academic literature.

14.3.6 TRADITIONAL KNOWLEDGE DIGITAL LIBRARY

The Traditional Knowledge Digital Library (TKDL) is an innovative project designed to safeguard India's traditional medical knowledge against bio-piracy and unjust patenting practices. This is a joint initiative of the Council of Scientific and Industrial Research (CSIR) and the Department of Ayurveda, Yoga, Naturopathy, Unani, Siddha, and Homoeopathy (AYUSH). TKDL catalogs traditional knowledge found in public domain writings and translates it into five worldwide languages: English, German, French, Japanese, and Spanish.

By providing this information to patent examiners globally, TKDL inhibits the issuance of patents for pre-existing conventional knowledge. The database functions as a protective measure, guaranteeing that India's historic knowledge is acknowledged as previous art. The Traditional Knowledge Digital Library (TKDL) exemplifies a framework for nations aiming to safeguard their traditional knowledge and illustrates the efficacy of digital libraries in the administration of intellectual property rights.

14.3.7 KALÂSAMPADÂ

Kalâsampadâ, referred to as the Digital Library: Resources of Indian Cultural Heritage (DL-RICH), is a notable endeavour by the Indira Gandhi National Centre for the Arts (IGNCA). The project seeks to establish an extensive database of cultural heritage assets,

encompassing manuscripts, rare pictures, paintings, sculptures, and audiovisual items. It employs sophisticated digital technology to conserve and showcase these riches in an interactive multimedia manner. The digital library offers access to more than 272,000 manuscripts, several rare volumes, and extensive audio and video recordings totalling hundreds of hours. It has an intuitive interface that enables scholars to investigate India's varied cultural traditions. Kalâsampadâ has been acknowledged for its outstanding execution of e-governance in the cultural domain. By incorporating many media forms, it provides a comprehensive perspective on India's artistic and cultural heritage.

14.3.8 ONOS

On 15 August 2022, the Prime Minister of India, in his presentation from the Red Fort, emphasized the vital significance of Research and Development during the Amrit Kaal and proclaimed the rallying cry of "Jai Anusandhan," therefore strengthening the national emphasis on research. The National Education Policy 2020 has acknowledged research as an essential component for attaining excellence in education and fostering national growth. In accordance with the objectives of Atmanirbhar Bharat and Viksit Bharat@2047, the Government of India has sanctioned the One Nation One Subscription (ONOS) initiative to ensure universal access to high-impact international scholarly journals and research publications for students, faculty, and researchers affiliated with central and state higher education institutions and central government research and development organizations. ONOS is acquiring national licenses from 30 prominent worldwide publishers, granting access to their journals for around 18 million beneficiaries across several fields, including individuals in tier-2 and tier-3 cities, thus promoting both fundamental and multidisciplinary research. The initiative commenced on 1 January 2025, with Phase I sanctioned for the period 2025 to 2027. Access is managed by the INFLIBNET Centre in Gandhinagar via a consolidated ONOS portal, while the ANRF conducts periodic assessments of usage and publication output. Before ONOS, access to subscription journals was disjointed and administered through various library consortia affiliated with different ministries and departments, in addition to individual institutional subscriptions. ONOS represents a timely and transformative advancement in enhancing India's global research ecosystem by streamlining and democratizing access to scholarly knowledge.

14.4 GLOBAL INITIATIVES

14.4.1 PROJECT GUTENBERG

Project Gutenberg, established in 1971 by Michael Hart, is the earliest digital library, aimed at promoting the creation and dissemination of eBooks. It emphasizes the digitization of public domain works, rendering them accessible to the public at no cost. The initiative primarily functions through the contributions of volunteers who scan, proofread, and format texts.

As of 2024, Project Gutenberg provides more than 70,000 complimentary eBooks in multiple formats, including plain text, HTML, ePub, and Kindle. The collection predominantly comprises literature from the Western cultural tradition, while also encompassing works in several other languages. Project Gutenberg's principles of open access and volunteerism have motivated numerous additional digital library initiatives. It continues to be an essential resource for readers and academics globally, guaranteeing the accessibility of classic literature in the digital era.

14.4.2 NDLTD

The Networked Digital Library of Theses and Dissertations (NDLTD) is a global entity committed to advancing the adoption, creation, utilization, dissemination, and preservation of electronic theses and dissertations (ETDs). Founded in 1996, NDLTD assists colleges and libraries in the execution of ETD projects. It seeks to enhance graduate education by enabling students to utilize digital technology for more effective dissemination of their findings.

Member institutions of NDLTD, comprising numerous universities worldwide, provide metadata to a centralized unified catalog. This enables researchers to search and access millions of theses and dissertations through a unified interface. The association conducts annual conferences to examine best practices and technology innovations in ETDs. NDLTD significantly contributes to the academic communication environment by enabling the international dissemination of graduate research.

14.4.3 NSDL

The National Science Digital Library (NSDL) was created by the National Science Foundation (NSF) in the United States to promote education in science, technology,

engineering, and mathematics (STEM). Established in 2000, NSDL functions as an extensive platform for superior educational resources catering to learners of all ages. It consolidates content from various sources, including universities, museums, and professional organizations.

NSDL provides an extensive array of resources, including lesson plans, databases, visualizations, and interactive simulations. It additionally offers tools and services to assist educators and learners in the optimal use of these resources. The library is structured into several collections and paths designed for particular audiences and subject domains. NSDL constitutes a substantial investment in digital educational infrastructure, with the objective of enhancing STEM literacy and the quality of education.

14.4.4 DLESE

The Digital Library for Earth System Education (DLESE) is a community-driven initiative financed by the National Science Foundation. It emphasizes granting access to superior educational resources pertaining to the Earth system, encompassing geology, geography, and environmental science. DLESE was created to facilitate comprehensive educational reform and enhance the quality of Earth-related teaching and learning.

The library offers a diverse collection of resources, including maps, databases, photos, and lesson plans, all evaluated for scientific precision and educational merit. DLESE functions on a community-oriented paradigm, wherein users provide resources and engage in the evaluation process. It additionally provides services to assist educators in incorporating digital resources into their curriculum. DLESE demonstrates the capability of subject-specific digital libraries to assist specialized educational communities.

14.4.5 ARXIV

ArXiv is a prominent open-access repository for academic articles in physics, mathematics, computer science, and related fields. Established in 1991 by Paul Ginsparg at Los Alamos National Laboratory, it is currently maintained by Cornell University. ArXiv enables researchers to self-archive their preprints and postprints, so aiding the swift dissemination of scientific discoveries.

The repository presently contains over two million articles and acquires thousands of new submissions monthly. It has evolved into a crucial instrument for researchers in the physical and computational sciences, frequently functioning as the principal means of

conveying new findings. The popularity of ArXiv has prompted the adoption of the "preprint server" paradigm in several disciplines, including biology (bioRxiv) and social sciences (SocArXiv). It illustrates the efficacy of subject-specific repositories in expediting scientific discovery.

14.4.6 PANDORA – AUSTRALIA'S WEB ARCHIVE

PANDORA (Preserving and Accessing Networked Documentary Resources of Australia) was founded by the National Library of Australia in 1996 to preserve Australian online publications. It was among the inaugural web archiving initiatives globally. PANDORA archives specific websites and online materials to safeguard Australia's digital documentary legacy for future generations.

The archive comprises a varied assortment of items, including government publications, academic periodicals, and transient websites that document cultural events. In 2019, PANDORA integrated with the broader Australian Web Archive, accessible via the Trove discovery service. The research underscores the significance of selective web archiving in safeguarding the history of the internet and digital culture. It functions as a cooperative initiative among state libraries and various cultural institutions around Australia.

14.4.7 EUROPEANA

Europeana is an online site initiated by the European Union in 2008 to facilitate access to the digital cultural heritage of Europe. It consolidates metadata from numerous archives, libraries, and museums throughout the continent. Europeana allows users to access millions of digital artifacts, encompassing artworks, books, movies, and audio, from a unified platform.

The initiative advocates for the reutilization of digital culture in education, research, and creative sectors. It collaborates with cultural institutions to standardize metadata and enhance the quality of digital assets. Europeana additionally endorses open licensing frameworks, promoting the unrestricted utilization of public domain resources. Europeana promotes a unified European identity and enables transnational cultural interchange by linking various collections.

14.4.8 OXFORD DIGITAL LIBRARY

The Oxford Digital Library (ODL) serves as the primary digitization service at the Bodleian

Libraries, University of Oxford. It seeks to establish a substantial repository of digital resources from the university's extensive and historic holdings. ODL supplies the technical framework and standards for digitalization initiatives throughout the university's libraries. Notable initiatives within ODL encompass the digitalization of early English texts, medieval manuscripts, and 19th-century literary works. The library partners with various entities, including Google and the Vatican Library, to execute extensive digitization initiatives. The Polonsky Foundation Digitization Project is a significant collaboration aimed at digitizing Hebrew and Greek manuscripts. ODL guarantees the preservation and accessibility of Oxford's academic treasures for researchers globally, thereby augmenting the university's impact on international study.

14.4.9 THE CAMBRIDGE DIGITAL LIBRARY

The Cambridge Digital Library (CUDL) was created by Cambridge University Library to provide online access to its vast contents. Initiated with an emphasis on "Foundations of Faith" and "Foundations of Science," the library features superior digital images of manuscripts and rare books. Notable collections comprise the manuscripts of Isaac Newton, Charles Darwin, and important religious books.

CUDL underscores the significance of academic context by offering transcriptions and comprehensive information in conjunction with digital photographs. It seeks to benefit both the scholarly community and the general populace. The site persists in its expansion, incorporating new collections that encompass board games and World War I archive. By making its archives accessible to the public, Cambridge University Library fosters novel opportunities for scholarship and public interaction with history and science.

14.4.10 OAISTER

OAIster is a collective inventory of open access digital resources, initially created at the University of Michigan and currently administered by OCLC. It collects metadata from global open access repositories utilizing the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). OAIster offers a searchable index of millions of digital resources, encompassing theses, technical reports, and image collections.

OAIster is one of the largest collections of open access content, comprising over 50 million records from numerous authors. It is crucial in rendering "deep web" academic information accessible. By incorporating these resources into the WorldCat system, OCLC guarantees that

open access materials are shown alongside conventional library collections. OAIster exemplifies the efficacy of interoperability standards in linking diverse digital repositories.

CHECK YOUR PROGRESS

A. Fill in the Blanks

a. The National Digital Library of India (NDLI) is an initiative under the National Mission on Education through Information and Communication Technology (NMEICT) and is produced by _____

b. The National Mission for Manuscripts oversees a national database of manuscripts referred to as_____

c. ____ is a community-driven digital library initiative financed by the National Science Foundation, dedicated to offering educational materials pertaining to the Earth system.

d. The "One Nation One Subscription" (ONOS) concept seeks to provide a unified national license for access to ____ and databases for all educational and research institutions in India.

14.5 LET US SUM UP

This unit has provided a comprehensive overview of popular digital library initiatives in India and across the globe. Indian programs such as NDLI and the National Mission for Manuscripts are effectively fulfilling the dual objectives of education and preservation. These initiatives utilize technology to render India's extensive knowledge traditions available to a broader audience. Likewise, worldwide initiatives such as Project Gutenberg, Europeana, and ArXiv exemplify the transformative potential of digital libraries in democratizing information access and promoting international collaboration. These initiatives exemplify the varied and dynamic functions of digital libraries in the 21st century, from safeguarding the transient web to enhancing scientific collaboration. As technology progresses, these platforms will certainly assume a more significant role in influencing the future of education and research.

14.6 ANSWERS TO CHECK YOUR PROGRESS

a. IIT Kharagpur**, b. K**ritisampada**, c.** Digital Library for Earth System Education (DLESE)**, d.** Scientific Journals

14.7 FURTHER READINGS

1. *Cambridge Digital Library ¯ University of Cambridge*. (n.d.). Cudl.lib.cam.ac.uk. https://cudl.lib.cam.ac.uk/
2. *Digital Libraries*. (2026). Iitkgp.ac.in. http://ndl.iitkgp.ac.in/he_document/inflibnet_epgp/inflibnet_epgp/library_and_information_science_38_1887_23193_E_Text
3. *Digital Library intiatives in UK*. (n.d.). NDLI. Retrieved 2026, from https://epgp.inflibnet.ac.in/epgpdata/uploads/epgp_content/S000021LI/P000038/M001875/ET/1483076376P8_M26.pdf
4. *Digital Library Project in USA*. (n.d.). NDLI. Retrieved 2026, from https://epgp.inflibnet.ac.in/epgpdata/uploads/epgp_content/S000021LI/P000038/M001874/ET/1483076353P8_M25.pdf
5. Europeana. (n.d.). *Discover inspiring European cultural heritage*. Www.europeana.eu. https://www.europeana.eu/en
6. India. (2022). *ONOS | India*. Onos.gov.in. https://www.onos.gov.in/about
7. Kannan P. (n.d.). *Major Digital Library Projects: Rest of the World*. NDLI. Retrieved 2026, from https://epgp.inflibnet.ac.in/epgpdata/uploads/epgp_content/S000021LI/P000038/M001873/ET/1483076397P8_M27.pdf
8. *National Digital Library of India*. (2015). Iitkgp.ac.in. https://ndl.iitkgp.ac.in/
9. *NDLI: E-Text: Digital libraries*. (n.d.). Iitkgp.ac.in. Retrieved January 3, 2026, from http://ndl.iitkgp.ac.in/he_document/inflibnet_epgp/inflibnet_epgp/library_and_information_science_38_1888_24390_E_Text
10. *OAIster: Catalog of open access resources*. (2020, November 10). OCLC. https://www.oclc.org/en/oaister.html
11. *Oxford Digital Library*. (2026). Ox.ac.uk. https://digital.humanities.ox.ac.uk/project/oxford-digital-library

14.8 PROBABLE QUESTIONS

1. Discuss in brief the digital library initiatives in India.
2. Discuss the digital Library initiatives in global scenario.
3. Write briefly about ONOS.
4. What is e shodhganga?
5. What is Project Gutenberg?

*** ***** ***