CERTIFICATE COURSE ON DIGITAL LIBRARY

Paper II

E Resources, Knowledge and Research Data Management in Digital Environment

Guru Prasad Das School of Interdisciplinary and Vocational Studies

KRISHNA KANTA HANDIQUI STATE OPEN UNIVERSITY

Subject Experts

Prof. Sanjay Kumar Singh, Gauhati University, Guwahati

Dr. Dipen Deka, Gauhati University, Guwahati

Dr. Deepjyoti Kalita, Cotton University, Guwahati

Course Co-ordinator: Dibyajyoti Patgiri, Pranjit Kalita, KKHSOU

SLM Preparation Team

UNITS	**CONTRIBUTORS**
8	Ms Sikha Das, Librarian, Dudhnoi College
10	Mr. Prasanna Kumar Konch, Librarian, Barpeta Polytechnique
11	Ms. Astha Borah, Librarian, Lanka Mahavidyalaya
1, 2,3 ,4,5,6, 7,9,12,13,14	Mr. Pranjit Kalita, KKHSOU

Editorial Team Mr. Dibyajyoti Patgiri, KKHSOU
Content Editor: Mr. Manash Kumar, Swahid Jadav Nath Homoeopathic Medical College & Hospital

Structure, Format & Graphics: Pranjit Kalita, KKHSOU

CONTENTS

Unit 1 E -resources in Libraries

Definition, concept and need of e-resource, Evolution from traditional print to electronic collections, Types, e- resource aggregators, Advantages and challenges of resource.

Unit 2 DRM with E resources

Digital right management, Intellectual property right, Evaluation, Subscription models, Pricing models, Renewal issues of collection development of e resources.

Unit 3 E resources and consortia management

Modes of subscription of e resources by libraries, Library consortia: concept, Advantage, issues and challenges of consortia approach to e resource subscription

Unit 4 ERM in libraries

E resource Management; Concept, need, Infrastructural Needs for ERM, ERM Challenges, Advantage and Limitations of ERM, ERAMS, needs, capabilities, basics elements, Popular ERMS

UNIT 5 Digital Library Technologies and Platforms

Concepts of Digital Libraries and Institutional Repositories, Open-Source Digital Repository Software, Digital Collection Development, Metadata Creation, Indexing

Unit 6 Open-Access

Concept of Open access; need and importance in DL system, Role of open access in research visibility and data sharing, Management and integration of open-access resources in digital environments.

Unit 7 Open Access Resources and platforms

Types of Open Access, Open access Resources, Open access Platforms, Major international and national repositories

Unit 8 Organization of Knowledge

Concept of Knowledge Organization, Library Classification: Concept, Cataloguing

Unit 9 Bibliographic Control and Subject heading

Bibliographic Control, Concept of Indexing: Type, Computerized Indexing, Subject Headings

Unit 10 Metadata

Metadata: Concept, Types and Functionalities, Metadata Standards: MARC, Dublin Core (DC), METS, etc. Protocol For Metadata Harvesting (PMH), Open Archives Initiatives (OAI Model)

Unit 11 Data in Library

DIKW Model: Concept of Data, information, knowledge and wisdom, Type of data and data resources, Concept of Big data in Libraries and Information Cntre, Open data: Characteristics and standards, , Linked data in library

Unit 12 Data, storage, security and data repositories

Data storage, data protection, security, sharing and ethics, Data repository: types, functions, resources and software for data repositories

Unit 13 Research Data Management

Data Management: data collection, data collection tools, Concept of research data management, RDM software and services: Free and Open sources

Unit 14 Emerging trends in Library data management

Artificial intelligence, machine learning, blockchain, and cloud solutions in data management, Data privacy, protection and cyber-security trends, new data governance strategies for libraries

INTRODUCTION

The course *"E-Resources, Knowledge and Research Data Management in Digital Environment"* aims to provide students with a comprehensive understanding of the principles, technologies, and practices involved in managing electronic information resources, organizing knowledge, and handling research data in the contemporary digital landscape. In the era of information proliferation and digital scholarship, libraries and information centres play a pivotal role in facilitating seamless access to, and preservation of, electronic resources and research data essential for academic and scientific advancement. This course has been designed to develop conceptual clarity and technical competence in the acquisition, organization, and management of digital content and data across diverse knowledge domains.

The course begins by introducing the concept, need, and evolution of electronic resources, tracing their transition from traditional print formats to digital repositories and subscription-based e-resource ecosystems. It explores the types, advantages, challenges, and the role of e-resource aggregators in modern library services. Building on this foundation, the course examines Digital Rights Management (DRM), Intellectual Property Rights (IPR), and key economic models governing e-resources, such as subscription, pricing, and renewal models. Students are further introduced to consortia-based resource sharing and collaborative subscription approaches that enhance information access and cost efficiency.

Subsequent units emphasize the concepts and infrastructural requirements of Electronic Resource Management (ERM) systems in libraries, including their capabilities, functions, and limitations. Learners will explore different ERM tools and software for managing the lifecycle of e-resources. The course further extends to digital library technologies, institutional repositories, and metadata-based organization of digital collections with standard schemas such as MARC, Dublin Core, METS, and OAI-PMH.

A significant component of the course is dedicated to open access and its transformative role in promoting research visibility, transparency, and collaborative scholarship. Students examine open-access movements, platforms, and major repositories at national and international levels. The course also addresses the organization of knowledge through classification, cataloguing, indexing, and subject heading systems that facilitate effective bibliographic control and retrieval in digital environments.

Advanced modules introduce the concept of research data management (RDM), its stages, tools, and ethical aspects of data handling. Learners gain insights into the DIKW (Data–Information–Knowledge–Wisdom) framework, big data analytics in libraries, and the management of open and linked data. Topics such as data storage, security, repositories, and governance frameworks are explored in depth, highlighting the growing importance of data protection, privacy, and sustainability.

The course concludes by discussing emerging trends in library data management, including artificial intelligence, machine learning, blockchain, cloud technologies, and contemporary cybersecurity protocols.

Unit 1 E -resources in Libraries

UNIT STRUCTURE

1.1 Learning Objectives
1.2 Introduction
1.3 Definition,
1.4 Need of e-resource
1.5 Evolution from traditional print to electronic collections,
1.6 Types and e- resource aggregators,
1.7 Advantages and challenges of e- resource.
1.8 Let Us Sum Up
1.9 Answers to Check Your Progress
1.10 Further Readings
1.11 Probable Questions

1.1 Learning Objectives

After this unit, learner will be able to,

- Describe what electronic resources are and why they are important to the current library system.
- Determine the different requirements and elements influencing the use of e-resources in public and academic libraries.
- Examine how library collections have changed over time, moving from print to digital and electronic media.
- Classify the electronic resources, including databases, e-books, and e-journals, and explain the function of aggregators of e-resources.

1.2 Introduction

Information generation, storage, and distribution have undergone a paradigm shift in the twenty-first century. Libraries, which were once thought of as peaceful repositories of tangible books and manuscripts, are now vibrant hubs for digital knowledge. The landscape of library services has changed significantly since the introduction of the internet and the quick development of information and communication technology (ICT). Information is no longer restricted to a building's actual shelves; instead, it is freely available on international networks and may be accessed at any time. The idea of "electronic resources," or "e-resources," which are now the foundation of contemporary research and education, is introduced in this course. E-resources include everything from sophisticated databases and multimedia archives to online journals and e-books. They signify the transition from "ownership" of hard copies to "access" of digital content, which has many advantages like round-the-clock availability, multi-user access, and robust search capabilities. Nevertheless, this shift also presents new difficulties,

such as the digital divide, expensive subscriptions, and long-term preservation concerns. The definition and types of e-resources will be covered in this course, along with the importance of aggregators in organizing this massive volume of data and the development of libraries from print-based establishments to hybrid information hubs. In order to comprehend how libraries are adjusting to satisfy the evolving information demands of patrons in the digital age, we will also examine the benefits and drawbacks of this change.

1.3 Definition

"E-resource" is an acronym for electronic resource. It describes any content or information source that has to be accessed via a computer or other electronic device. Items that are accessible in digital formats as opposed to the conventional physical print media, such as books and periodicals, are considered resources in the context of libraries. A material that needs computer mediation to access its content and make it useful is referred to as an electronic resource. This broad definition includes offline resources like CD-ROMs, which are physically housed in libraries but require a computer to operate, as well as online resources that may be accessed over the internet. The phrase refers to all of the information items that a library offers via a computer network, according to glossaries for libraries and information technology. This means that an item is deemed an e-resource if a library provides a service that allows patrons to read a book, an article, or view an image on a computer screen. These materials can be print resources that have been scanned and converted to electronic files, or they can be "born digital," which means they were first produced in a digital format. The According to a different viewpoint, e-resources are files that typically contain information that can be stored as electrical impulses. These are typically made available online or saved on a computer. They signify the transition from real, tangible objects to network-transmittable, immaterial digital data. Online databases, e-journals, e-books, e-reports, e-theses, and even webpages are examples of electronic resources. The Additionally, these resources' ownership paradigm differs from that of conventional print publications. When a book is bought in a traditional library, the actual copy is permanently owned by the library. However, the library does not actually "own" the content of many electronic resources, such as databases and online journals. Rather, the library purchases a license that gives its patrons access privileges for a predetermined amount of time. Because it alters the relationship between the library and the content, this distinction is essential to defining e-resources. The library only acts as a conduit for its patrons to access the information, which frequently sits on a server that is owned by a publisher or vendor. The In a traditional library context, electronic resources are useful research tools that supplement print-based resources. They offer a way to obtain information that could otherwise be limited because of cost or geographic limitations. Libraries may provide comprehensive links to more relevant material by utilizing digital formats, which makes the research process more dynamic and connected.

1.3 Need of e-resource

The rapid improvements in information and communication technology have led to a major increase in the requirement for e-resources in libraries. Library patrons' evolving information-seeking habits are the main factor driving this demand. Students, professors, and researchers

are among the modern users who favor rapid access to knowledge. Without needing to physically visit a library building, they want to be able to locate, read, and utilize materials with the touch of a mouse. One of the main reasons that e-resource adoption is required is this convenience.

The need for multi-access is one of the most important needs that e-resources address. A single physical book can only be used by one individual at a time in a typical library. Other users have to wait if a popular book is checked out. By enabling numerous users to access the same material concurrently from various locations, e-resources address this issue. By removing the obstacles of physical distance and library opening hours, this round-the-clock accessibility guarantees that information is available whenever the user needs it.

Another important consideration is speed. With e-resources, users can quickly search a large collection of materials for particular keywords or subjects. Compared to looking through actual card catalogue or scanning printed book index pages, this is substantially faster. It is very easy for users to extract, integrate, and cross-search information from other publications. For researchers who must sort through vast volumes of literature to locate pertinent data, this efficiency is essential.

The need for functionality that print media cannot provide is another factor driving the need for e-resources. Multimedia components including audio, video, and animation can be included in electronic formats. As a result, learning becomes more dynamic and interesting. For instance, an e-journal allows a medical student to see a video of a surgical procedure in addition to the text description, something that a print textbook cannot do.

Electronic collections are also highly needed in libraries due to space limits. The substantial shelf space needed for physical books and journals is sometimes scarce and costly to maintain. Since e-resources are kept on servers, the library hardly needs any physical storage space for them. This makes it possible for libraries to offer access to a far greater range of resources than their actual physical space could possibly accommodate.

In addition, libraries frequently experience budget cuts, and the price of print publications has been increasing. Institutions can combine their resources and make a greater variety of items available than they could separately by using e-resources and library consortia. Because of this budgetary reality, e-resources are an essential way to keep a collection current and thorough.

1.4 Evolution from traditional print to electronic collections

An important development in the history of information management is the shift from traditional print libraries to electronic collections. Libraries have been physical reservoirs of knowledge for millennia, using manual cataloguing systems, printed volumes, and manuscripts. Users had to physically visit the library, go through card catalogs, and pick the documents from the shelf. The actual object and the library building as the exclusive point of access were the focal points of this paradigm. The digitization of library services marked the start of the transformation process. This entailed enhancing traditional tasks using digital technology. Online Public Access Catalogs (OPACs) and the automation of library catalogs were among the first actions. Although this facilitated book searches, the majority of the

content was still available in print. Libraries started to change from being static repositories to interactive digital information centers as technology developed, especially with the introduction of the internet and high-speed networks. Electronic journals that are accessible online first appeared in the 1990s. At first, consumers had to download these from FTP sites or receive them as email attachments. Electronic publishing advanced with the creation of the World Wide Web in 1993, enabling the direct transmission of journals with graphic capabilities via web browsers. The widespread digitalization of pre-existing print collections carried on the evolution. Public domain texts began to be digitized and made freely available for download through initiatives like Project Gutenberg. Commercial publishers and aggregators started offering thousands of books and journals online after this. With the help of e-resources, libraries have transitioned from a "just-in-case" approach, where they purchased books in case someone needed them, to a "just-in-time" model, where access is available whenever needed. The function of the librarian has also changed as a result of this change. Librarians in the traditional era concentrated on maintaining tangible collections and assisting patrons in navigating the physical area. These days, the position has changed to that of a "digital librarian" or "information manager." To preserve and oversee digital collections, professionals today require expertise in web design, server architecture, and digital rights management. They help consumers who might be overwhelmed by the amount of information available online by acting as navigators in the digital realm. A shift in user behavior is another indicator of the progression. People now use search engines and online databases like Google Scholar instead of books and reference stacks. They anticipate smooth text, picture, and sound integration. In order to satisfy contemporary needs, libraries have embraced hybrid models, keeping physical collections for patrons who value the tactile experience while making significant investments in digital resources. This two-pronged strategy guarantees that libraries continue to be important cultural centers and entry points to the worldwide digital information network.

1.6 Types of E- resources

To meet diverse information needs, electronic resources are available in a variety of formats. The following are the main kinds found in libraries:

E-journals

These are periodicals that are published online or in electronic format. These could be periodicals that are exclusively published digitally or electronic versions of well-known print journals. Because they offer the most recent information and are accessible from any location, e-journals are essential for scholars. They frequently contain extra material that isn't included in the print version, such as data sets or films.

E-Books

A book-length digital release including both text and images is called an e-book. It is designed to be read on computers and other electronic devices, such as e-readers and tablets. The digital version of printed books, e-books have capabilities like search functionality, font size customization, and portability. With only one device, consumers can carry a library of thousands of books.

Electronic Database:

A database is a structured set of documents. These are typically collections of conference papers, book chapters, and journal articles that are organized for easy searching in a library setting.

- **Bibliographic databases**: These do not have the complete text, but they do include references to published works, including the author, title, and abstract. They assist researchers in locating pertinent publications.
- **Full-text databases:** These offer the documents' entire text. The full article or chapter is available for users to view, save, and print. JSTOR and ScienceDirect are two examples.

E-theses and Dissertations:

These are digital copies of the research projects that students have turned in for their degrees. ETDs are important resources for original research. Researchers may now access thousands of doctoral theses that were previously difficult to locate in print versions thanks to digital libraries like Shodhganga in India.

E-reports:

These are ad hoc or recurring publications that contain information in tabular or narrative form. University annual reports and government planning commission reports are two examples. These are frequently posted for public access on institutional websites.

Institutional Repositories:

These are online repositories created by academic institutions or research centers to compile and present the scholarly works of their staff and students. Preprints of articles, lecture notes, and other research materials might be included.

Electronic Conferences:

These platforms, which are often referred to as electronic forums or discussion groups, enable online topic discussion among scholars. They are crucial for figuring out what novel subjects are being discussed in a field prior to official publication.

Courseware and Tutorials:

These are online learning resources made specifically for e-learning. They offer self-paced learning modules, interactive tutorials, and lecture notes.

E-resources aggregators:

A company or service that offers digital access to the material of numerous serials and other resources from various publishers is known as an aggregator. They serve as a liaison between the libraries that subscribe to the content and the publishers who provide it.

Role of E resources aggregators:

Libraries face a challenging task in the electronic age when it comes to handling subscriptions to thousands of distinct periodicals from hundreds of different publishers. By compiling information from several sources and making it accessible via a single searchable interface, aggregators make this process simpler. They lessen the administrative load on library employees by managing the technical facets of access, licensing, and billing. They offer a "one-stop-shop" where consumers may look through a wide variety of resources.

Types of E resources aggregators:

Although they might not be the owners of the content rights, hosting aggregators offer the technological infrastructure needed to host electronic journals. Ovid and HighWire Press are two examples.

- **Full-Text Aggregators:** These services provide access to articles from a wide range of periodicals in their entirety. Typically, they lease the content once a year. ProQuest and EBSCO are two examples.
- **Gateways:** These serve as entry points to externally hosted information. Gateway services are frequently offered by subscription agencies.
- **Subject-based Aggregators:** These concentrate on gathering information about a particular field, such law, business, or medical.
- **Multidisciplinary Aggregators:** These are helpful for broad academic libraries and cover a variety of topics.

Examples of some aggregators are:

- **EBSCO:** A significant supplier of e-books, e-journals, and research databases.
- **JSTOR:** Itis a non-profit organization that offers a digitized collection of scholarly publications, dating back to their initial issues.
- **Credo Reference:** An aggregator dedicated solely to reference materials, such as encyclopedias and dictionaries.

As they give libraries an affordable means of granting access to a vast amount of content, aggregators are crucial. They can negotiate a lower total cost than if they subscribed to each newspaper separately.

Libraries must be mindful, though, since aggregators occasionally alter their journal listings or coverage years without much warning, which reduces the library's control over its collection.

1.7 Advantages and challenges of e- resource.

Advantages of E-resources:

Both libraries and their patrons gain a great deal from the transition to electronic resources.

- **Accessibility and Convenience:** The ability to access e-resources from any location in the world at any time is the biggest benefit. To conduct research, users do not have

to go to the library in person. Those who live in remote areas or are distance learners will especially benefit from this. Access provides unparalleled convenience and is available every day of the week.

- **Searchability:** E-resources provide strong search features. Users can conduct keyword searches across thousands of documents at once or inside a single document. Compared to searching through printed books or indexes by hand, this saves a ton of time. It enables non-sequential reading, in which the user can go straight to the pertinent portion.
- **Multi-User Access:** Multiple users can access a single electronic resource simultaneously. No one else can read a book that has been borrowed from a print library. A complete class can view the same article for an assignment at the same time using e-resources.
- **Publication Speed:** Before the print edition is printed and disseminated, articles and journal issues frequently surface online. This guarantees users instant access to the most recent findings and data.
- **Links and Multimedia:** Videos, audio snippets, and animations are examples of multimedia content that can be supported by electronic forms and enhance the educational process. Additionally, hypertext linkages create a knowledge web that print cannot match by enabling users to click directly to similar articles, citations, or external websites quickly.
- **Space Saving:** A lack of physical space is a persistent problem for libraries. Shelf space is no longer necessary thanks to e-resources. A server can house a huge collection of thousands of journals, freeing up physical space in the library for study spaces or other purposes.

Challenges of E resources:

Despite all of the advantages, e-resources come with several drawbacks.

- **Infrastructure and Cost:** Users and libraries require the proper infrastructure, such as computers, tablets, and dependable internet connectivity, in order to access e-resources. The cost of buying and maintaining this gear and software might be high. Furthermore, libraries frequently have limited funding, and some e-resources have extremely costly subscription fees.
- **Digital Divide:** Access to technology is not universal. The "digital divide" is the difference between people with and without access to computers and the internet. Users are disadvantaged and unable to take advantage of the library's electronic collection if they lack these resources or the necessary expertise.
- **Preservation and Archiving:** The physical copy of print works belongs to the library. The library still has the back issues it paid for even if a subscription is cancelled. The library frequently only licenses access to e-resources. The library can also lose access to all previous materials if the subscription expires. Because file formats and storage medium vary over time, ensuring the long-term preservation of digital data is also a technical problem.

- **Health Problems:** Spending a lot of time reading on a computer screen can be tiresome and bad for your eyes. It can lead to physical discomfort and eye strain, which is less of a problem with printed paper.
- **Copyright and Licensing:** It's difficult to manage e-resource licenses. Regarding who can access the content, how much can be downloaded, and if it can be shared, each publisher may have various guidelines. Digital Rights Management (DRM) software frequently limits the use of e-books by limiting the number of pages that may be printed, and libraries must carefully negotiate these restrictions.



Check Your Progress

A. Fill in the Blanks

a. An _____ is described as any material or information source that requires computer mediation in order to access its content and make it useful.
b. _____serve as go-betweens for publishers and libraries, compiling information from many sources to give users access to a single searchable interface.
c. Electronic resources offer_____ access, allowing several users to read the same document at once, in contrast to traditional books that only permit single-user access.
d. The library paradigm has changed from "ownership" of physical copies to ____ of digital information through licensing agreements as a result of the shift to e-resources.

1.8 Let Us Sum Up

In this unit, we have explored the concept of e-resources and their growing significance in modern libraries. We defined e-resources as materials that require computer access, ranging from e-journals and e-books to online databases. The need for these resources is driven by the demand for instant, 24/7 access to information by multiple users simultaneously, as well as the need to save physical space in libraries. We traced the evolution from traditional print libraries to digital collections, noting how technology has transformed the role of librarians and the habits of users. We also examined the various types of e-resources available and the vital role played by aggregators in organizing and providing access to this vast amount of content. Finally, we weighed the advantages, such as searchability and remote access, against the challenges, including high costs, the digital divide, and preservation issues. Understanding these aspects is crucial for effectively managing and utilizing library resources in the digital age.

1.9 Answers to Check Your Progress

a. E-resources**, b.** Aggregators, **c.** Simultaneous**, d.** Access

1.10 Further Readings

1. Ganthale, D., Lemdeo, S., Mahavidyalay, P., Tah-Kuhi, M., & Dist-Nagpur. (2020). DIGITAL LIBRARIES AND THEIR IMPACT ON TRADITIONAL LIBRARIES. *International Journal of Management*, *10*. https://www.ijmra.us/project%20doc/2020/IJMIE_DECEMBER2020/IJMIE10Dec20-DS.pdf
2. Joshi, Y. (n.d.). *Types of Electronic Resources and its Usage*. https://andjournal.in/wp-content/uploads/2019/11/2019a_22.pdf
3. *MODULE -2: ELECTRONIC RESOURCES*. (n.d.). https://nios.ac.in/media/documents/SrSecLibrary/LCh-008.pdf
4. Sathyabhama. (n.d.). *Electronic resources: Advantages and disadvantages for academic library users*. *6*(10), 1097–1100. https://www.allresearchjournal.com/archives/2020/vol6issue10/PartP/8-8-56-279.pdf

1.12 Probable Questions

A. Short Answer Type Questions

1. E-resource aggregators: what are they? Describe their various forms and the advantages they provide to libraries.
2. Describe electronic resources and explain how they differ from traditional print resources.
3. Write about the different requirements that have caused e-resources to be widely used in academic libraries.

B. Long Answer Type Questions:

1. Explain how libraries have changed from being traditional archives to becoming digital information hubs.
2. Describe the benefits and difficulties of implementing e-resources in libraries.
3. Give examples to illustrate the differences between full-text and bibliographic databases.
4. Write a brief essay discussing institutional repositories' function in the digital era.

Unit 2 DRM with E-resources

Unit Structure

2.1.Learning Objectives
2.2.Introduction
2.3.Digital Right Management
2.4.Intellectual Property Rights
2.5.Evaluation
2.6.Subscription Models
2.7.Pricing Models
2.8.Let Us Sum Up
2.9.Answers to Check Your Progress
2.10. Further Readings
2.11. Model Questions

2.1 Learning Objectives

After This unit, learner will be able to,

- Know about the concept of Digital Rights Management
- Learn about IPR and its aspects
- Evaluation, subscriptions related aspects of resources and various issues related to Renewal of e resources for library collection development

2.2 Introduction

The transition from physical to digital collections has drastically changed how materials are obtained, maintained, and accessible in today's libraries and information centers. E-journals, e-books, full-text databases, and multimedia resources are examples of electronic resources, or e-resources, which are materials that need computer access, whether via a personal computer, mainframe, or mobile device. The acquisition of e-resources frequently entails a complicated procedure of licensing and access management rather than straightforward ownership, in contrast to traditional print materials that are bought and owned outright. This change calls for a thorough grasp of how to manage the rights attached to these digital assets in order to safeguard the interests of publishers and artists while ensuring that they are still available to authorized users.

Intellectual property rights (IPR) and digital rights management (DRM) are essential components of this management procedure. While advocating for the broadest possible access for their customers, libraries must negotiate legal frameworks that safeguard the moral and financial rights of creators. In order to guarantee smooth service delivery, the management of these rights is not only required by law. For example, the user experience and the library's

capacity to efficiently serve its community are directly impacted by the mode of access, whether it is through IP authentication or login credentials. Additionally, the complexity of these agreements rises as libraries join consortia to pool their purchasing power, necessitating that librarians be skilled at negotiating conditions that prioritize institutional needs over constrictive publisher regulations.

The protection of intellectual property, digital rights management procedures, and the stringent evaluation procedures necessary to choose top-notch resources are the main topics of this unit's exploration of the crucial aspects of e-resource management. It also explores the financial issues of managing e-resources, looking at different pricing and subscription structures that libraries use to make the most of their tight budgets. By comprehending these ideas, library workers may make sure they are offering their patrons useful, affordable, and legally compliant resources.

2.3. Digital Right Management

In the context of library services, digital rights management (DRM) refers to the tools and procedures used to regulate who can access and use digital content. DRM is more widely focused with making sure that access to licensed resources is safely and effectively managed for authorized users in a library system, even if the phrase frequently conjures up restrictive technologies that prohibit duplicating. Managing who can access and utilize e-resources is a major operational challenge because they are usually accessed locally through library networks or remotely via the internet.

A fundamental aspect of DRM in libraries is the identification and administration of "Authorized Users." These are usually described in license agreements as individuals, such as faculty, staff, and students, who have a current, verified association with the subscribing institution. Furthermore, this definition frequently includes "walk-in users" visitors who physically visit the library as long as they use resources from within the actual building. Authentication mechanisms are used in the technical implementation of DRM to confirm these users. IP (Internet Protocol) authentication is the most widely used technique in academic libraries. As long as the user is utilizing a computer that is linked to the institution's network or a proxy server, this strategy enables users to access resources without the requirement for unique usernames and passwords.

Accessibility and security must be balanced in order for DRM systems to be effective. Access can be severely hampered by restrictive DRM that demands specific software or individual login information for each user. A vendor's insistence on password-based access, for instance, creates problems with password distribution and control, which may result in illegal use or, on the other hand, lock out authorized users who forget their login information. In order to extend the library's services beyond its physical walls, libraries employ DRM systems that meet industry standards like IP filtering and proxy access, which enable remote use access from home or off-campus locations.

The administration of usage rights, including printing, downloading, and interlibrary lending (ILL), is another facet of DRM. In order to stay within "fair use" or "fair dealing" clauses,

license agreements frequently stipulate how much content can be downloaded or printed by a single user. To stop systematic database harvesting, DRM solutions may place restrictions on how many pages a user may download in a single session. DRM protocols must also take into account the library's requirement to share resources with other libraries. In order to prevent the conventional library practice of resource sharing from being undermined by the transition to digital, licenses should ideally allow the use of secure document transfer methods for interlibrary loans.

2.4. Intellectual Property Rights

The legal rights that people or organizations have over their creative works are known as intellectual property rights, or IPRs. IPR is the cornerstone that enables producers, publishers, and creators to profit from their work and manages its distribution in the field of e-resources. In general, innovations, literary and artistic works, designs, and symbols utilized in trade are all considered forms of intellectual property (IP). Because it gives creators credit and financial incentives, protecting these works is crucial to promoting innovation and creativity.

IPR is a wide notion that encompasses two primary categories: copyright and individual property, which is frequently industrial property. Trademarks, industrial designs, and innovations (patents) are examples of individual property. Conversely, literary and artistic creations including novels, poetry, movies, music, and architectural designs are protected by copyright. Since copyright regulates the text-based and multimedia resources that comprise the majority of digital collections, it is the most relevant type of intellectual property for libraries.

Intellectual Property Rights Types:

- **Copyrights:** The owner of a literary or artistic work is granted the exclusive right to control its publication, dissemination, and adaptation. This is known as copyright. A concept's expression, not the idea itself, is protected by copyright. Copyright rules guarantee that, with the exception of some "fair use" or "fair dealing" clauses, digital content that libraries subscribe to cannot be duplicated or disseminated without authorization.
- **Trademarks:** These are words, symbols, or graphics that are used to identify a business or item. Trademarks aid in setting one company's products and services apart from another. Major publications' and database providers' logos, for instance, are protected trademarks. The information providers' reputation and brand identification are safeguarded by these marks.
- **Patents:** For a set amount of time, usually 20 years, patents provide the creator of an invention the sole authority to manage its creation, use, or sale. Patents are pertinent to the software and technology used to access library collections, even though they are less common in the content of these collections.
- **Industrial Design Rights:** These safeguard an object's non-utilitarian visual design. This encompasses an article's color, pattern, or shape.

IPR aims to strike a balance between the public interest and the interests of creators. IPR rules promote the creation of new knowledge by giving monopoly rights for a finite period of time. But in the digital age, problems like copyright violations and internet piracy have grown to be major obstacles. Because digital materials are so easy to duplicate and distribute, IPR must be strictly enforced, which frequently results in the complicated licensing agreements that libraries have to deal with.

2.5. Evaluation

The selection and evaluation of e-resources is a more complex process than that for traditional print materials. A selector cannot make a decision to acquire an e-resource in isolation; it requires consultation with various departments, including technical systems, acquisitions, and service delivery, to ensure the resource is suitable for the institution. The evaluation process involves a thorough assessment of multiple criteria, which can be broadly categorized into content, technical requirements, functionality, and vendor support.

Content Evaluation: The caliber and applicability of a resource's content are its main requirements. E-resources must complement the institution's curriculum and research objectives. Selectors must assess if the material is current, peer-reviewed, and comes from a reliable source in addition to adding depth to the current collection. The compatibility of the electronic version with any print equivalent is particularly important for e-resources. Additionally, it is important to confirm whether the resource provides "completeness of content," which refers to accurately reproducing the original publication's text and photos without omission.

- **Technical Requirements:** The resource's technical viability must be evaluated prior to acquisition. This includes figuring out the authentication options, compatibility with current hardware and software, and the access mechanism (local vs. remote). Browser compatibility is a crucial technical consideration; the resource should work with common web browsers and not require proprietary. Additionally, usability is affected by the content's format (HTML, PDF, XML, and SGML). Even though PDF preserves the original visual layout, XML is frequently used for huge documents because it offers greater information retrieval and lifespan.
- **Functionality and Reliability:** One of the main factors influencing a resource's worth is how easy it is to use. The interface should be intuitive, easy to navigate, and user-friendly, offering features like context-sensitive help and online tutorials. Strong search and retrieval capabilities are required, including full-text, Boolean, and keyword searches in addition to sophisticated features like truncation and relevance rating. The system's uptime is referred to as reliability; providers should promise round-the-clock availability with little unplanned downtime. To provide easy access, the resource should also work effectively with other library systems, such as link resolvers and discovery tools.
- **Vendor Support:** An important aspect of the assessment is the vendor connection. Libraries should assess the level of support provided, including the

availability of trial periods to test the product before purchase. Vendors ought to provide training resources and respond quickly to technical problems. The availability of statistical reporting is an additional crucial component. Vendors must provide usage statistics that comply with standards like COUNTER (Counting Online Usage of Networked Electronic Resources), which allows libraries to assess the cost-effectiveness of the subscription over time.

- **Open Access and Overlap Analysis:** In recent years, the evaluation process has also begun to include an analysis of open access (OA) content. Tools like CloudSource allow libraries to perform overlap analysis, identifying how much of a paid subscription database consists of content that is freely available as open access. Making decisions that will save money need this data. If a significant portion of a subscription database is available openly, the library might consider discontinuing that subscription to avoid paying for content that is free. This data-guided approach helps in optimizing the budget without compromising access to essential research

2.6. Subscription Models

Libraries are frequently offered a range of subscription alternatives, and there is no standard model for the packaging and sale of electronic materials. Choosing the arrangement that provides the institution with the best value and access terms requires an understanding of various types.

Purchase vs. Subscription:

E-resources are frequently rented, in contrast to print books, which are bought once. Some models, however, enable the purchase of perpetual access to a particular collection or year of content. This is frequently paired with an annual maintenance or access cost. Journal back-files and e-book collections frequently experience this. The regular subscription model, on the other hand, only grants access for the time of the contract; unless special post-termination rights have been negotiated, access to the content is lost in the event that the subscription is cancelled.

Bundled Packages and the Big Deal:

One common concept is the Big Deal, in which a publisher charges a single fee for access to all of their periodicals. Although this significantly expands the selection of titles that users can choose from, it may ultimately prove to be costly and rigid. Libraries frequently have to pay for a lot of titles that are rarely used. With some libraries choosing to unbundle packages in order to subscribe to only the most pertinent items, this strategy is coming under more and more scrutiny.

Consortia Pricing:

Libraries frequently form consortia to negotiate collective licensing in an effort to reduce excessive prices. In order to obtain better conditions and lower rates than individual libraries could obtain on their own, consortia pricing makes use of the purchasing power of several institutions. This strategy often entails a single license that covers all member libraries, with

fees allocated according to each member's size or usage. It makes negotiating easier and gives more people access to resources.

Pay-Per-View and Rental Models:

Libraries may choose to use a pay-per-view model, in which a fee is paid for each article or document accessed, for materials that are only seldom needed. For specialized publications with minimal usage, this can be economical. Libraries can also rent access to a collection for a brief time using rental models. These arrangements offer flexibility, but due to heavy usage, they must be closely watched to make sure expenses don't surpass the cost of a regular subscription.

Print and Electronic:

Libraries must keep the print copy in order to access the online version since some publishers link the electronic subscription to the print subscription. As libraries adopt e-only policies to reduce processing expenses and physical space, this paradigm is losing favor. Libraries ought to provide "unbundled" pricing, in which the print and electronic versions can be bought separately.

2.7. Pricing Models

E-resource pricing is complicated and differs greatly between providers. Although there isn't a single standard, expenses are determined using a number of common metrics. These models need to be thoroughly examined by selectors to make sure they fit the library's spending plan and usage trends.

FTE (Full-Time Equivalent) Pricing:

This is one of the most common pricing models for academic libraries. The cost is calculated based on the total number of full-time students and faculty at the institution. While this accounts for the potential audience size, it can be disadvantageous for large universities if the resource is specialized and only relevant to a small department. In such cases, libraries should negotiate for pricing based on the FTE of the specific department rather than the entire institution.

Concurrent User Pricing:

In this model, the price is determined by the number of users who can access the resource simultaneously. Licenses may be for a single user, a small group (e.g., 5 concurrent users), or unlimited users. Unlimited user licenses are preferred for high-demand resources to prevent users from being turned away when all licenses are in use. However, for specialized or low-use resources, a limited concurrent user license can be a significant cost-saving measure.

Usage-Based Pricing:

Some models base the price on the actual usage of the resource, such as the number of searches or downloads performed in the previous year. While this can align costs with value, it makes budgeting difficult as costs can fluctuate unpredictably from year to year. A variation of this is the Patron-Driven Acquisition (PDA) or Demand-Driven Acquisition (DDA) model for e-books, where the library loads records into the catalog, and a purchase is triggered only after a user accesses the book a certain number of times.

Tiered Pricing:

Publishers often categorize institutions into tiers based on their type (e.g., university, college, high school) or size. Each tier has a set price for the resource. This is a simplified version of FTE pricing and attempts to account for the varying ability of different types of institutions to pay.

Introductory and Multi-Year Deals:

Vendors may offer low introductory prices to encourage new subscriptions. Libraries must be cautious of these, as the price may jump significantly after the initial period. Multi-year deals with fixed price caps are beneficial as they protect the library from annual inflation increases, which typically range from 5% to 7%, allowing for more predictable budgeting.

2.8. Renewal Issues of Collection Development of e resources

When renewing subscriptions to electronic resources, libraries encounter a number of interrelated difficulties that span administrative, technical, financial, and skill-based aspects.

- **Challenges with Budget and Pricing**

One of the main challenges is financial limitations, since library budgets are getting less while subscription costs are rising. Libraries must weigh user needs against available resources while determining whether to continue subscriptions, which makes price negotiation a major concern. The decision to renew necessitates careful budgeting for the purchase of e-resources, especially when publishers raise prices with little warning. Discounts may be available for multi-year memberships, but long-term commitments carry risks in the event that budgets shrink or usage habits shift**.**

- **Workflow and Administrative Problems**

The intricate administrative routines involved in renewal procedures put a burden on library employees. Libraries must assess a number of variables at once, including changes to licensing terms and conditions, user happiness, publisher service quality, consumption statistics, and access agreement adjustments. Libraries must coordinate communications, keep track of renewal deadlines, and make sure the right notification protocols are followed due to the constant challenges associated with vendor and publisher management. Though it doesn't always happen consistently, vendors should ideally alert libraries at least two months prior to renewal dates. Notification timelines are crucial.

- **Issues with Licensing and Access**

Complex license agreements pose significant renewal problems. Libraries must document holdings for perpetual access rights in both new and existing licenses, negotiate or accept term changes, and comprehend automatic renewal clauses (sometimes known as evergreen clauses) that can call for clear termination notices. Continuity during renewal periods is further complicated by platform migrations, IP configuration changes, and access authentication problems. Without updated spreadsheets recording these changes, it becomes impossible to retain accurate holdings information when journals switch platforms or publishers merge.

- **Technical Capability and Professional Skills**

The inability of library employees to handle e-resource renewals professionally is a significant issue that has been noted in several researches. Evaluation capabilities use data interpretation, and efficient vendor and publisher communication are all impacted by this skill gap. These problems are made worse by a staffing shortage because renewal efforts compete with other library duties like administration, collection development, digitization, and cataloging. Many libraries find it difficult to retain the specific knowledge required for technical infrastructure requirements, such as software compatibility, system maintenance, and integration with library management systems.

- **Usage Assessment Challenges**

Methodological difficulties arise when evaluating the usefulness of e-resources for renewal decisions. Libraries are required to examine consumption of data; however cost-per-use computations are unreliable due to differences in data quality between publishers and platforms. It can be challenging to consistently collect user feedback in order to assess if resources satisfy institutional needs. Evaluation procedures are made more difficult by traditional Integrated Library Systems' incapacity to completely support e-resource management.

- **Coordination of Consortiums**

Decisions regarding renewals need more coordination for libraries that are part of consortium subscriptions. Prior to renewal, the consortium must obtain certification from each library, which adds time constraints and administrative burden. What is already a difficult process is made more complex by managing resources at both the consortium and individual library levels.



Check Your Progress

a) What is IPR?
b) What are the differences between copyright and trademark?
c) What is FTE?

2.9. Let Us Sum Up

Managing electronic resources is a complex issue that lies at the nexus of library science, technology, and law. It starts with the strong protection of rights via Intellectual Property Rights (IPR) and Digital Rights Management (DRM) frameworks, which guarantee that content is safe but still available to authorized users. The assessment of these resources is a thorough procedure that considers technological viability, interface usability, and vendor support dependability in addition to content quality. Additionally, libraries have to negotiate a complicated marketplace of subscription and pricing arrangements, from consortium agreements and packaged Big Deals to FTE-based and usage-based pricing. Librarians who are proficient in these areas will be able to curate digital collections that are economical, compliant with the law, and, most importantly, extremely helpful to their user communities.

2.10. Answers to Check Your Progress

a) IPR is the legal rights that people or organizations have over their innovations, including inventions, literary and creative works, and symbols. For a set amount of time, they provide creators complete control over how their works are used.

b)While trademarks protect names, symbols, and logos used in commerce to differentiate one company's goods or services from another, copyright protects original literary and artistic works (such as books and music) against unapproved copying.

c) FTE pricing is determined by the aggregate count of full-time personnel and students at an institution, irrespective of the actual utilization of the resource. Concurrent User price is determined by the maximum number of users permitted to utilize the resource simultaneously.

2.11. Further Readings

1. Austin, C. (2024). Evaluating E-Resources Subscriptions: Creating Cost Savings with CloudSource+. *Technical Services Quarterly*, 1–13. https://doi.org/10.1080/07317131.2024.2352981
2. IGNOU. (n.d.). INTELLECTUAL PROPERTY RIGHTS . In *Egyankosh*. Retrieved 2025, from https://www.egyankosh.ac.in/bitstream/123456789/21348/1/Unit-1.pdf
3. Johnson, S., Evensen, O., Gelfand, J., Lammers, G., Sipe, L., Zilper, N., Fronty, J., Hafner, J., Mansfield, J., & Schmolling, R. (2012). *Key Issues for E-Resource Collection Development: A Guide for Libraries Acquisition and Collection Development Section*. https://www.ifla.org/wp-content/uploads/2019/05/assets/acquisition-collection-development/publications/electronic-resource-guide-en.pdf
4. Subramanya, S. R., & Yi, B. K. (2006). Digital rights management. *IEEE Potentials*, *25*(2), 31–34. https://doi.org/10.1109/MP.2006.1649008

2.12. Probable Questions

a) Examine the significance of Digital Rights Management (DRM) in academic libraries and the difficulties posed by stringent DRM rules.
b) Detail the several criteria a librarian must assess when appraising a new electronic database for subscription.
c) Analyze the similarities and differences between the "Big Deal" subscription business and the "Print Plus Electronic" model. What are the merits and demerits of each?
d) Discuss the impact of Intellectual Property Rights (IPR) on the administration and dissemination of electronic resources within a digital library context.
e) Conduct a critical analysis of the effects of various pricing models (Full-Time Equivalent, Concurrent User, and Pay-Per-View) on a library's financial resources and collection growth plan.

Unit 3 E resources and consortia management

UNIT STRUCTRURE

3.1 Learning Objectives
3.2 Introduction
3.3 Modes of subscription of e resources by libraries
3.4 Library consortia: concept, Advantage
3.5 Issues and challenges of consortia approach to e resource subscription
3.6 Let Us Sum Up
3.7 Answers to check Your Progress
3.8 Further Readings
3.9 Probable Questions

3.1 Learning Objectives

After this unit, learner will be able to

- know the nature of electronic resources and why libraries need to plan carefully for their selection and subscription
- explain the concept and advantages of library consortia in the context of e resource access, and discuss key issues and challenges that consortia and member libraries face in managing shared subscriptions.

3.2 Introduction

As users anticipate online access to journals, books, databases, and multimedia materials at any time and from any location, electronic resources have become essential to library services in academic, research, and many special libraries. Electronic journals, e-books, full-text and bibliographic databases, reference materials, conference proceedings, standards, patents, and numerous other digital information products provided by publishers and aggregators are examples of e-resources. Their use is controlled by licensing rather than the customary ownership of physical copies, and they are typically accessed across networks via web interfaces. Particularly in science, technology, and medicine, where journal costs are high, it has become challenging for individual institutions to subscribe to all the information their customers need due to the growing cost of electronic resources and diminishing or stagnant library budgets. Libraries have responded by adopting various subscription models and increasingly forming consortiums, which allow groupings of institutions to share access, negotiate better prices, and employ financial and technical resources more effectively. Although consortium-based strategies have demonstrated definite benefits, they also present problems and difficulties that need for cautious handling, including governance, equitable cost sharing, resource selection, and long-term sustainability**.**

3.3 Modes of subscription of e resources by libraries

Different arrangements are used by libraries to access electronic resources, reflecting variations in publisher availability, technical capabilities, finances, and local demands. It is useful to consider subscription methods, or the primary ways that libraries or collections of libraries pay for and gain access to databases, electronic journals, and other resources. Since each contract has implications for cost, flexibility, local control, and cooperation, several common approaches can be discovered and compared, even though the specifics of each may vary.

Direct individual subscriptions between a particular library and a publisher or aggregator are one fundamental method; these are frequently for a specific collection of journals, a database, or a content bundle. In this mode, the library uses its own funds to pay the subscription and negotiates terms like price, the number of concurrent users, archival rights, interlibrary loan permissions, and off-campus access. This provides a high degree of local control and enables selection to meet local needs, but it may also result in duplication of subscriptions among institutions in the same region and greater costs per title, particularly for smaller institutions with less negotiating strength.

Consortial subscription is another significant method in which a collection of libraries works together to bargain for and finance each member's access to electronic resources. In this instance, the consortium negotiates deals with publishers on behalf of participating universities, frequently obtaining reduced costs per institution, increased access rights, or extra services like training and usage data. Member libraries share access to the licensed content, typically via IP-based authentication or federated identity systems, and make financial contributions in accordance with predetermined formulas that may take into account variables like size, type of institution, or prior print expenditure.

Hybrid or mixed models are also frequently used, particularly when individual libraries add more titles or databases through direct subscriptions while national or regional consortia supply a core set of electronic resources. This enables libraries to customize collections to their specific fields or research strengths while still taking use of the economies of scale provided by consortia. A complex yet adaptable subscription environment that necessitates careful administration and coordination between consortia and individual libraries is created by some publishers' tiered models, which allow consortium members to choose more modules or back-files at discounted prices.

3.4 Library consortia: concept, Advantage

A library consortium is a collection of libraries that have decided to collaborate formally or informally to accomplish shared objectives that would be more costly or challenging to accomplish separately, such as resource sharing, training, and technical cooperation. Depending on their goals, consortiums can be established at the institutional, regional, national, or subject-based levels and may consist of public, academic, research, and special libraries.

Consortia offer a forum for group bargaining with publishers and vendors from the standpoint of electronic resources. The consortium has more negotiating strength since it represents the aggregate prospective usage and spending of its members rather than numerous libraries negotiating independently. In addition to securing more flexible license conditions, including expanded access rights, greater archive provisions, or shared usage statistics; this frequently leads to better price, such as discounted rates or advantageous multi-year agreements. In contrast to individual subscriptions, Indian consortia like UGC-INFONET, INDEST-AICTE, and others have made it possible for member universities to access a vast number of publications and databases at substantially lower costs per institution.

3.4.1 Benefits of Partnerships for Subscription to E Resources

There are a number of related methods to characterize the benefits of consortium-based techniques. The first is economic advantage, which is sometimes referred to as cost effectiveness. Consortiums can improve the amount and variety of information available to each member library without causing a corresponding rise in local costs by combining demand and purchasing power to negotiate lower subscription fees, longer access to larger collections, or both. For smaller and less well-funded schools that might not otherwise be able to afford pricey publications or databases in vital subject areas, this is especially crucial.

Second, consortia facilitate more equitable and extensive information access amongst institutions. Users at smaller colleges, distant campuses, or specialized institutes can access the same high-quality content as users at major universities when a consortium licenses resources for all members. This helps to close information gaps and supports regional or national goals for research and education development. Many consortiums use institution-wide authentication for access, which expands the reach of e-resources by enabling users to connect from school networks and occasionally from off-campus locations via proxy or federated systems.

Third, beyond short-term cost savings, consortiums foster collaboration and capacity building. Training sessions, user awareness campaigns, cooperative technical assistance, and the sharing of best practices are examples of shared activities that boost member libraries' skill development and enhance local e-resource management. In order to facilitate better informed decision making, consortiums can help choose resources that best meet community needs, arrange trials of new goods, and assess usage patterns collectively.

Fourth, by promoting standardization and interoperability in areas like metadata, authentication systems, access methods, and usage statistics, consortia can improve user experience and streamline management. Member libraries gain from more stable technical environments and more uniform data when consortium agreements mandate that manufacturers provide specific access protocols or give standardized use reports. Longer term, consortia may also be able to influence policy talks at the national or industry level regarding digital infrastructure, copyright exceptions, and open access, providing libraries a collective voice that would not be feasible on an individual basis.

3.5 Issues and challenges of consortia approach to e resource subscription

Consortia based approaches have clear advantages, but they also face important issues and challenges that must be identified and managed carefully. These difficulties can have an impact on each member library as well as the consortium as a whole and are caused by organizational, financial, technical, and legal issues. Understanding these issues is essential for planning sustainable consortia and for making realistic decisions about membership, participation and long-term commitments.

Funding and cost sharing provide a significant obstacle. Because institutions vary in size, subject coverage, finances, and actual levels of use, it can be difficult to decide how to divide the overall cost among members, even when a consortium negotiates cheaper pricing overall. A cost-sharing formula must be adopted by consortia, such as equal shares, shares based on the number of students or faculty, or shares linked to prior print expenditures. However, these formulas may be viewed as unfair, particularly if usage statistics reveal significant disparities or if smaller institutions feel overburdened and larger institutions believe they are subsidizing others. Although they are not always simple to attain, budgetary transparency and a readiness to evaluate and modify cost-sharing arrangements over time are crucial.

Another area of complexity is decision-making and governance. The selection of resources, subject priorities, publishers or aggregators to negotiate with, and how to react to changes in platforms, prices, or products are all decisions that consortiums must make. In instance, when the consortium attempts to cover several disciplines rather than concentrating on depth in a few areas, some members may feel that the selected resources do not meet their local needs. Reaching consensus among multiple institutions can be slow and may require concessions. These problems can be managed with the establishment of governance structures including steering committees, working groups, and explicit guidelines for representation and voting; but these institutions also increase administrative complexity and require time and effort from member libraries.

Consortial subscription success is also impacted by access and technical issues. The ease with which users can access the e-resources licensed by the consortium is impacted by the disparities across member institutions' local infrastructure, which includes network bandwidth, computer hardware, authentication systems, and library administration software. Some libraries may lack sufficient technical staff to set up and maintain access through proxy servers, federated identity systems or link resolvers, which can lead to underuse of the resources that have been paid for. Effective access necessitates additional resources and coordination, but it also needs ensuring that all members fulfil basic technical standards and provide shared technical support or training.

Legal and license related issues present further challenges. When institutions have varied interests or risk tolerances, it can be challenging to ensure that consortium agreements conform with national laws and institutional norms while defining license terms that are acceptable to all members. Libraries' ability to support teaching and research may be limited by publishers' restrictions on uses like interlibrary lending, inclusion in course packs, text and data mining,

or walk-in user access. It can be difficult to negotiate licenses that preserve publishers' interests while allowing appropriate academic usage, and existing agreements may be disrupted by mergers, acquisitions, or changes in publishing policies. In order for libraries to understand what happens to their access in the event that subscriptions are cancelled or the consortium's structure changes, long-term access rights, such as archival or perpetual access, must also be clearly defined.

Another set of issues that consortia need to deal with is assessment and sustainability. Consortia need to assess regularly whether the subscribed e resources still match members' needs and whether the costs are justified by levels of usage and the benefits obtained. Reliable usage statistics, suitable analytical tools, and established protocols for adding new resources, removing underutilized ones, and adjusting cost-sharing agreements in response to evolving conditions are all necessary for this. The durability of consortial agreements may be threatened by external economic factors such as inflation, currency fluctuations, and institutional budget cuts. These circumstances may necessitate renegotiations, coverage reductions, or changes in member participation. Consortium leadership will continue to be responsible for managing these changes while preserving member libraries' mutual trust, collaboration, and feeling of benefit.



Check Your Progress

a) In what ways do library consortia enhance the economic and accessibility dimensions of electronic resource supply for member institutions?
b) What are the reasons for libraries transitioning from primarily depending on individual subscriptions to increased utilization of consortial and mixed models for electronic resource access?

3.6 Let Us Sum Up

The unit has addressed e-resources and consortium management as a crucial component of contemporary library practice. It has articulated those electronic resources, such as journals, books, and databases, have become integral to library services while also posing financial and technical hurdles for specific institutions. Various subscription modalities were delineated, encompassing direct individual subscriptions, consortia agreements, and hybrid models that integrate shared and local access, each possessing distinct advantages and drawbacks.

The unit has explained the notion of library consortia and emphasized the benefits of consortia strategies for e-resource subscriptions, particularly regarding cost efficiency, broader and more equitable access, collaborative training, and standardization. Simultaneously, it has analyzed critical concerns and obstacles confronting consortia, including cost sharing and funding, governance and selection, technological infrastructure, licensing terms, and long-term

viability. Comprehending the advantages and challenges of consortia enables librarians and policymakers to devise pragmatic ways for enhancing e-resource accessibility in a collaborative and sustainable fashion**.**

3.7 Answers to check Your Progress

a) Library consortia enhance economic factors by leveraging collective bargaining power to obtain discounted prices and expanded access packages, while also improving access by providing high-quality electronic resources to all member institutions, including smaller or remote ones, through shared licenses and networked authentication.

b) Libraries have transitioned to consortial and mixed models due to the high per-title costs and limited coverage associated with individual subscriptions. Consortia can aggregate demand and negotiate more favorable prices and terms, enabling more institutions to access a broader array of electronic resources within budgetary constraints.

3.8 Further Readings

1. *Access to E-Resources: TIFR Libraries' Consortia Approach*. (2026). Ajist.co. https://ajist.co/index.php/ajist/article/view/95/4685
2. Chari, S. N., Kumari, A. R., & Ravinder, D. (n.d.). Consortia based Subscription of E resources: Library trends in Sustainability. In *Contemporary Trends in Librarianship*. https://www.researchgate.net/publication/368667556_Consortia_based_subscription_to_e-resources_library_trends_towards_sustainable_growth/citation/download
3. Chaurasia, N. (n.d.). *E-Resources and Consortia Management*. https://library.iitd.ac.in/arpit_2020-2021/Week%203%20-%20Module%206%20-%20E-Resources%20and%20Consortia%20Management.pdf
4. *Indian Digital Library in Engineering Science and Technology (INDEST) Consortium: Consortia-Based Subscription to Electronic R*. (2026). Arizona.edu. https://repository.arizona.edu/bitstream/handle/10150/105608/?sequence=2
5. Kaushik, S., Somvir, & Ranjana. (2011). ACCESS OF E-RESOURCES THROUGH CONSORTIUM. *Nternational Conference on "Convergence of Libraries, Archives and Museum*. https://www.researchgate.net/publication/327859575_Access_of_E-Resources_through_Consortium/citation/download

6. Kumar, D. (n.d.). Effective use of E-Resources Consortia for Indian Universities: New Challenges. In *International Refereed Journal in Library and Information Science*. Retrieved January 24, 2026, from https://sijlis.weebly.com/uploads/3/9/3/7/39370233/2015_narasi_5.pdf
7. *LIBRARY NETWORKS AND CONSORTIA*. (n.d.). https://egyankosh.ac.in/bitstream/123456789/11304/3/Unit-14.pdf
8. Pal, J. K., & Das, P. K. (2024). *Progress Through Partnership: Consortia Based e-Resource Subscription Initiatives in India*. ISI Digital Commons. https://digitalcommons.isical.ac.in/conf-articles/365/
9. Srivastava , J. P., & Verma, V. K. (n.d.). *Library Consortia: Issues and Challenges with Special Reference to INDEST-AICTE Consortium*. 9th Convention PLANNER-2014. https://ir.inflibnet.ac.in/server/api/core/bitstreams/d8ae35eb-b1b7-41c0-b2fa-65d9dc589a6b/content
10. Suresh, M., & Natarajan, R. (2021). Use of E-Resources Through Consortia With Special Reference to College Libraries. *Research Anthology on Collaboration, Digital Services, and Resource Management for the Sustainability of Libraries*, 907–925. https://doi.org/10.4018/978-1-7998-8051-6.ch050
11. Vilochanan, T. S., & Chinnasamy, Dr. K. (2020). *Electronic Resources through Consortia in Higher Education: An Asset to the Users of Indian Academic Libraries*. Ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3599145

3.9 Probable Questions

a) Identify two primary distinctions between individual and consortial subscription models for electronic resources.

b) In what ways may consortia assist smaller institutions in enhancing their access to scholarly electronic resources?

c) What technical issues can impede the successful utilization of consortium-subscribed electronic resources by member libraries?
d) Explain the primary subscription modalities for electronic resources employed by libraries and examine the impact of each modality on cost, accessibility, and local governance.

e) Define a library consortium and elaborate on the benefits that consortia provide for e-resource subscriptions, supported with examples.

f) Examine the principal concerns and obstacles related to consortium-based methods for e-resource subscription and provide measures that consortia and member libraries might implement to address these challenges.

Unit 4 ERM in libraries

4.1 Learning Objectives
4.2 Introduction
4.3 E resource Management; Concept and need
4.4 Technical needs for ERM,
4.5 Advantage and limitations of ERM,
4.6 Key principles of ERAMS
4.7 Proprietary and open access, ERM software.
4.8 Global Scenario
4.9 Let Us Sum
4.10 Answers to Check Your Progress
4.11 Further Readings
4.12 Probable Questions

4.1 Learning Objectives

4.2 Introduction

Electronic Resource Management (ERM) refers to the management of electronic resources, including e-journals, e-books, online databases, datasets, streaming media, CD-ROMs, and software, throughout their entire life cycle, from investigation and selection to acquisition, licensing, access, evaluation, renewal or cancellation, and finally to preservation. It is an extension of collection management, as it involves technical, legal, financial, and administrative tasks, because electronic resources require constant monitoring of licenses, authentication systems, usage statistics, and platform or format changes. The primary objective of ERM is to ensure seamless and stable access, license compliance, optimal use of financial resources, and evidence-based development of electronic collections. The major responsibilities of ERM include selecting resources based on user needs, licensing and managing contracts with vendors and publishers, setting up and maintaining access (IP ranges, remote access, link management), and periodically evaluating resources using usage statistics and cost-benefit analysis for renewal or cancellation decisions.

4.3 E resource Management: Concept and need

Electronic Resource Management (ERM) is the management of electronic information resources by libraries and other information providing institutions throughout the entire life cycle of the resources. ERM involves the processes, methods, policies, and software applications used by librarians and information professionals to manage the effective selection, use, evaluation, and renewal or cancellation of electronic resources. The electronic resources managed through ERM include electronic journals, e-books, online databases, datasets, streaming media, CD-ROMs, and computer software. Unlike print

materials, electronic resources need constant monitoring of licenses, access, authentication, usage statistics, and compatibility. Thus, ERM is more than a traditional collection management process, as it combines technical, legal, financial, and administrative tasks.
ERM is a systematic approach to managing the complexity of electronic resources for seamless access, license compliance, optimal use of financial resources, and sound development of electronic collections.

The major tasks involved in ERM include:

- Selection of suitable electronic resources based on user needs
- Licensing and negotiations with vendors and publishers
- Access management, including authentication and link management
- Constant evaluation based on usage statistics and cost-benefit analysis
- Renewal, alteration, or cancellation of subscriptions

For these tasks, libraries can use Electronic Resource Management Systems (ERMS). ERMS are software applications that assist libraries in managing licenses, vendors, costs, access, and usage data in a centralized manner. Libraries can access ERM solutions in three ways:

- By Purchasing commercial ERM software from third-party vendors,
- Choosing open-source ERM software, or
- Creating in-house ERM workflows tailored to institutional needs.

The development of ERM is directly associated with the development of electronic resources during the Information Age. As libraries began to migrate from print to electronic resources in the latter part of the twentieth century, the need for traditional methods of management became obsolete. This resulted in the development of formalized ERM practices and standards, including those developed by the Electronic Resources Management Initiative (ERMI) and later by standardization bodies.

4.4 Infrastructural Needs for ERM

- Servers and workstations

Servers must be provided to run ERM/ILS/LSP software, proxy servers or remote access software, authentication servers, and local databases or repositories that facilitate e-resource discovery and access.

Sufficient numbers of networked workstations for staff and users are necessary for acquisition, cataloging, troubleshooting, and user access, with periodic upgrades, backup systems, power protection, and secure storage.

- Faster Internet facility

Fast, reliable Internet connectivity (wired and campus-wide Wi-Fi) is essential since almost all e-resource operations like discovery, access, usage statistics, license checking are dependent on continuous online connections to external servers and platforms.

Sufficient bandwidth, redundancy, and quality-of-service management are required to cope with peak usage, remote access, and large-scale downloading of e-books, e-journals, and databases without timeouts and service disruptions.

- Dynamic and forceful websites

A dynamic library website or web portal serves as the first point of entry to e-resources, combining OPAC/discovery systems, A-Z lists, database menus, subject guides, and authentication links in an easily accessible manner.
The website should enable frequent content updates, responsive design, linkage with ERM systems and link resolvers, and effective communication on access options, trials, downtime, and usage policies.

- Trained staff members

Effective ERM implementation needs staff members trained in licensing, metadata, access management, troubleshooting, usage analysis, and working with ERM/ILS/LSP tools, in addition to general ICT skills.
Continuous staff training and development through workshops, refreshers, and hands-on training are required to ensure that librarians are up-to-date with new platforms, standards, and business models to support users effectively.

4.5 ERM Challenges

Electronic Resource Management (ERM) is confronted with a number of closely related challenges: lifecycle, format diversity, system constraints, and governance models such as consortia versus individual subscriptions.

- Complex lifecycle

E-resources go through a number of complex iterative cycles (trial, selection, licensing, activation, access, usage analysis, renewal/cancellation, archiving), and each cycle requires different data, participants, and applications.

The processes are non-linear: packages evolve, journals move from one platform to another, licenses change, and coverage dates or access privileges vary constantly, making it hard to keep track of accurate and up-to-date information throughout the whole lifecycle.

- Availability in different forms and formats

Today, libraries are responsible for managing e-journals, e-books, databases, A&I sources, multimedia and open access materials, each with its own specific access mechanisms, metadata formats, and system requirements.

This diversity makes description, discovery, authentication, and statistics harvesting more difficult, as there is no standard or schema that can easily accommodate all these formats and the practices of the providers.

- Dynamic nature of e-resources

Electronic resources are very dynamic. The platforms keep changing, the URLs and authentication processes change from time to time, the title lists and dates of coverage are constantly being updated, and new business models emerge, such as evidence-based and transformative agreements.

The libraries have to keep track of the changes, troubleshoot access issues, and keep the knowledge bases up to date, which is time-consuming and prone to broken links, incorrect holdings, and user dissatisfaction.

- ILS and LSP not able to manage fully

The ILS systems were meant for a more stable and item-based print environment and therefore do not have any built-in support for licenses, packages, entitlements, and complex access rights, which are typical of e-resources.

Even the current Library Services Platforms (LSPs) have difficulties in supporting the full range of ERM functions; there is still a lack of perfect interoperability with other ERM systems, link resolvers, and knowledge bases, which causes duplication of data and fragmentation of workflows.

- Consortia vs. individual approach

While consortial purchasing may offer greater access, more favorable licensing, and shared costs, it also diminishes local control, commits libraries to “big deals,” and sometimes conflicts with the individual institutional agenda.

The complexity of managing e-resources at both the consortial and individual institutional level is increased by overlap in packages, diverse user needs, divergent usage patterns, and shared management.

4.6 Advantage and limitations of ERM

Some of the benefits and advantages of ERMS include:

- Workflow and staff efficiency

ERMS assist in time-saving and efficient workflows by providing a centralized source of information on licenses, holdings, access, and contacts, eliminating the need for searching through various files and systems.

They eliminate duplicity of work by enabling various departments (acquisition, cataloging, systems, and reference) to access the same updated information rather than working with separate spreadsheets or databases.

- User experience and access

ERMS enable faster searching through OPACs/discovery systems and link resolvers, allowing users to quickly find full-text and correct coverage information. Improved title lists and access links mean fewer broken links and access problems, resulting in a smoother user experience.

- Collection development and cost control

ERMS can help identify and remove costly overlap between packages and subscriptions, particularly when various bundles include the same titles. ERMS offer usage data and cost-per-use information that enable informed collection development, including cancelling unused resources and reallocating funds to high-value content for efficient use of the library budget.

- Handling scale and complexity

ERMS facilitate the increasingly large and complex nature of e-resources by providing functionality for bulk updates, package management, and knowledge base synchronization. They facilitate the setup and maintenance of e-resources (authentication information, URLs, dates of coverage, comments) with reduced manual labour and potential for error when changes occur in platforms or access arrangements.

- Handling consortia

In a consortial setting, ERMS facilitate the management of shared packages, licenses, and entitlements for multiple institutions in a standardized manner. They facilitate collective decision-making, transparent cost-sharing and equivalent usage analysis for member libraries, thus enhancing the overall effectiveness of the consortium.

Limitation of ERM:

- Format and collection-related limitations

The diversity of digital collections, such as images, high-resolution digital images, live video streaming, audio files, interactive content, and datasets, often outstrips the capabilities of ERMS, resulting in an incomplete description, access, or preservation of some formats. Compared to print media, electronic resources have a potentially less stable life span, depending on file formats, storage media, platforms, and vendors' viability, whereas paper media will last much longer under reasonable preservation conditions.

- Cost and resource requirements

Proprietary ERMS systems are often costly to acquire, implement, and maintain, and the total cost of ownership (licensing, hosting, customization, and integration) can be prohibitively expensive for academic libraries with limited budgets. New technologies and

ERMS system implementations demand financial, temporal, and experiential investments to fully leverage their potential, including system upgrades, staff training, and periodic system reconfiguration.

- Human resource and employment concerns

The ERM system demands a technically competent and knowledgeable workforce with IT, metadata, licensing, and problem-solving expertise, which is often in short supply in many institutions and regions. A perceived increase in unemployment or redundancy for some traditional library staff is likely due to highly automated ERM workflows, which can generate resistance and fear among library staff.

- Time-consuming processes and learning curve

The integration of electronic resources into ERM is a time-consuming process, including data migration, knowledge base clean-up, holdings linking, access setup, and constant updates, requiring extensive staff time. New technology and interfaces introduce a high learning curve; it may take a considerable amount of time for both staff and users to feel comfortable with electronic interfaces, replacing the familiar physical browsing, scanning, and copying with electronic searching, selecting, downloading, storing, and printing.

- Health and usability issues

Overreliance on screen reading and accessing e-resources may result in health issues such as eye strain, headaches, and discomfort from glare or prolonged screen use. Users may find computer-based reading to be less fluent and less easy to scan compared to printed texts; turning pages, marking, and visually scanning are still more intuitive and efficient for most users in print than in electronic form.

- Interface, device, and software dependency

Electronic texts and systems are not always user-friendly, particularly when interfaces are not well designed, making activities such as rapid page turning, deep visual scanning, or detailed annotation more difficult than in paper form.

Many electronic resources need special devices or particular hardware/software configurations (a specific operating system, reader, plug-in, or app), and some are tied to proprietary formats that may become obsolete or inaccessible over time.

4.7 ERAMS, needs, capabilities, basic elements, Popular ERMS

Electronic Resource Management Systems (ERMS) offer a systematic, workflow-based approach to managing the entire lifecycle of electronic resources, which cannot be effectively managed by conventional ILS and LSPs.

Need for ERMS

- The lifecycle of e-resources (investigation, selection, acquisition, licensing, access, evaluation, renewal/cancellation, and preservation) is complex and is presently managed by a variety of disparate tools such as paper-based files, spreadsheets, staff web pages, small databases, A-Z lists, and email.
- A significant portion of the library budget is being spent on e-resources, but there are deficiencies in the management of ER workflows in ILS/LSPs, and they are unable to provide any meaningful, consolidated data without an ERMS.
- The management of digital collections has become a significant source of worry and confusion, and more than 30 different tools are being used simultaneously, and hence a flexible and holistic solution is required to save time and streamline work.

The essential core capabilities of ERMS are:

- The ability to handle custom workflows and internal/external communication for all stages of the lifecycle.
- The ability to track renewals and cancellations, including dates, notices, and decision history.
- The ability to upload and store license files and documents in a structured and retrievable manner.
- Interoperability with other systems (ILS/LSP, link resolver, discovery, A-Z lists) and the ability to export and import data.
- Ease of use and navigation, preferably through a web-based interface accessible to multiple staff members.
- The ability to handle different levels of user permissions (roles) and to store administrative information such as contacts, credentials, and platform information.

Basic data elements of ERMS:

The data is organized into a number of basic element types by the presentation:

- Descriptive data: titles, packages, publishers/aggregators, platforms, identifiers, coverage, and other bibliographic/resource description information.
- Licensing data: license texts, terms of use, usage clauses, ILL, course packs, remote access, perpetual access, jurisdiction, DRM, and renewal terms.
- Financial data: subscription cost, fees (platform, hosting), discounts, fund codes, cost allocations, and data required for cost-per-use calculations.
- Administrative and support data: vendor and publisher contacts, support information, problem logs, downtime information, communication history, and internal comments.
- Access data: authentication schemes (IP addresses, remote access arrangements), URLs, admin module logins, discovery/A-Z links, and data required for holdings lists and access problems.

The basic data elements provide a means for an ERMS to record not only “what we subscribe to,” but also “under what terms, at what cost, on which platform, for whom, and with what performance.”

Popular ERMS:

The source briefly discusses a few key ERMS tools in order to inform users of the options available.

- EBSCO ERM Essentials
    - Vendor: EBSCO.
    - Key features: single-click access to e-resource information, report generation, trial management, renewal facilitation, workflow management, customizable fields, and auto-populated data fields.
    - The focus is on minimizing labor, managing information in one place, and facilitating collection management through reporting capabilities.
- 360 Resource Manager
    - Vendor: ProQuest (now Ex Libris/Clarivate).
    - Key features: integration and management of subscription information, enhanced reporting and analysis, sharing or inheriting resources and metadata, simplified acquisition and licensing processes, facilitating improved collection management decisions, and relatively quick implementation and utilization of an ERM.

    It is marketed as a full-featured, SaaS-like solution tightly integrated with link resolver and knowledge base services.

- CORAL (Open Source)

    Nature: Open-source ERMS, not proprietary to any one company.

    Modules: Resources, Licensing, Organisation, Usage Statistics, and Management, mirroring the key functional areas of ERM.

    CORAL shows that the functionality of ERMS can be provided by modular, community-driven software development with a focus on resource information, licensing, organizational data, and usage analysis.



Check Your Progress

a) Electronic Resource Management (ERM) involves the entire _____ of electronic resources, from selection to preservation.
b) A specific _____ is needed because conventional ILS and LSPs do not adequately support ER workflows.
c) In ERM, usage statistics and _____ analysis are critical for making renewal or cancellation decisions.
d) The basic data components of ERMS are descriptive, licensing, financial, administrative and support, and _____ data.

4.8 **Let Us Sum**

Electronic Resource Management (ERM) provides a systematic, lifecycle approach to managing e-journals, e-books, databases, and other electronic resources, coordinating selection, licensing, access, evaluation, renewal/cancellation, and preservation as a unified set of processes. As e-resources are dynamic, diverse, and license-dependent, ERM must necessarily consider technical, legal, financial, and administrative aspects all at once, going well beyond the print-centric management of collections. Specialized Electronic Resource Management Systems (ERMS) have thus become a necessity: they provide centralized access to descriptive, licensing, financial, administrative, and access information, enable customized workflows and communication, and integrate with ILS/LSPs, discovery systems, and link resolvers. When properly implemented with adequate infrastructure and qualified staff, ERM and ERMS make it possible for libraries to ensure seamless access, remain within license terms, provide useful usage and cost-per-use data, eliminate duplication and overlap, maximize budgets, and facilitate informed, evidence-based development of electronic collections in the face of daunting technical and administrative challenges.

4.9 **Answers to Check Your Progress**

a) lifecycle, b) Electronic Resource Management System (ERMS),

c)cost-benefit, d) access

4.10 Further Readings

1. Dutta, A. (2023). Management of E-resources in Academic Libraries: Some observations. *International Journal of Research in Library Science*, *9*(4), 162–172. https://doi.org/10.26761/ijrls.9.4.2023.1715
2. Hiremath, M. B. K., & Kenchakkanavar, A. Y. (2023). *Electronic Resource Management in Libraries*. *12*(3), 713–720. https://www.researchgate.net/publication/369561351_Electronic_Resource_Management_in_Libraries
3. Imre, A., Hartnett, E., & Hiatt, C. D. (2013). CORAL: Implementing an Open-Source ERM System. *The Serials Librarian*, *64*(1-4), 224–234. https://doi.org/10.1080/0361526x.2013.760414
4. John Paul Anbu K, Sanjay Kataria, & Shri Ram. (2013). Dynamics of Managing Electronic Resources: Electronic Resource Management System (ERMS) Initiatives. *DESIDOC Journal of Library & Information Technology*, *33*(4). https://www.researchgate.net/publication/260165509_Dynamics_of_Managing_Electronic_Resources_Electronic_Resource_Management_System_ERMS_Initiatives

4.11 Probable Questions

1. Explain the concept of Electronic Resource Management (ERM) and describe any four key tasks involved in ERM.
2. Why are conventional ILS and LSPs not sufficient for managing the entire ERM process?
3. Discussany four key functionalities of an Electronic Resource Management System (ERMS).
4. What are the basic data elements maintained in an ERMS? Provide one example each for the two categories.
5. Explain the requirement of ERMS in academic libraries, highlighting how ERMS overcome the challenges of managing the complex e-resource life cycle and the limitations of the current systems and tools.
6. Explain in detail the strengths and weaknesses of ERM, specifically in relation to workflow, user interface, collection development, and cost management.

UNIT 5 DIGITAL LIBRARY TECHNOLOGIES AND PLATFORMS

UNIT STRUCTURE

5.1 Learning Objectives
5.2 Introduction
5.3 Concepts of Digital Libraries and Institutional Repositories
5.4 Open-Source Digital Repository Software,
5.5 Digital Collection Development
5.6 Metadata Creation.
5.7 Let Us Sum UP
5.8 Answers to Check Your Progress
5.9 Further Readings
5.10 Probable Questions

5.1 Learning Objectives

After this unit, learners will be able to

- Know about digital library and institutional repositories
- Know about open-source DL software, meta creation

5.2 Introduction

The way information is stored, retrieved, and disseminated has undergone a profound upheaval with the arrival of digital technologies. Traditional libraries, which for centuries have been the custodians of tangible books and manuscripts, are today developing into hybrid and totally digital enterprises. The necessity to conserve intellectual output for future generations, the expansion of born-digital content, and the growing desire for remote access to information are the driving forces behind this change.

Digital libraries represent a huge leap in information management. They are managed, structured environments that combine technology, services, and content rather than just collections of electronic information. Without being limited by time or physical area, they let users to access a wide range of information, including text, photos, audio, and video. Alongside digital libraries, the concept of the Institutional Repository (IR) has evolved as a significant component of the scholarly ecosystem. An IR serves as a digital archive for the intellectual work produced by a university or research institution, ensuring that this information is captured, conserved, and made accessible to the worldwide community.

5.3 Concepts of Digital Libraries and Institutional Repositories

The term digital library is frequently used in a broad sense, sometimes including anything from an individual's personal file collection to the entire Internet. It has a more precise definition in

the area of library and information science, though. An organization that provides the means, including specialized personnel, to choose, organize, enable intellectual access to, interpret, distribute, maintain the integrity of, and guarantee the longevity of collections of digital works is known as a digital library.

A digital library is a controlled environment, in contrast to the Internet, which is an unstructured mass of information. It applies the fundamental tenets of traditional librarianship selection, organization, and service to digital content. Ranganathan's five rules of library science still hold true in this situation: digital libraries are dynamic, evolving creatures, they must link each user to the material they require (Every reader his book), and they exist to serve users (Documents or files in digital library usually consists of a combination of "digitized" converted from analog formats, such as scanned historical manuscripts and "born-digital", produced initially in digital form, such as e-journals). The following are the main attributes of a digital library:

- Digital Collection: Digital format is used for the main resources.
- Organization: To make retrieval easier, the content is arranged according to standard techniques rather than at random.
- Access: It offers services that let customers locate and utilize the data, frequently from a distance.
- Preservation: There is a commitment to keeping the digital assets usable over the long term.

5.3.2 Institutional repositories

One particular kind of digital library system is an Institutional Repository (IR). It is described as a collection of services that a university or research organization provides to its community members for the administration and distribution of digital content produced by the organization and its constituents.

An IR is focused domestically, but a generic digital library may gather materials from a variety of other sources (such as subscribed journals or purchased e-books). Capturing the host institution's intellectual output is its main goal. Faculty research articles, theses and dissertations, conference proceedings, technical reports, and datasets fall under this category.

The idea that an institution should take care of its own intellectual property is the foundation of the IR. By gathering this work in a central, open-access repository, the institution fulfils multiple goals:

- Visibility: It makes the most of its research's worldwide exposure and influence.
- Access: Anyone with an internet connection can read the research thanks to the removal of access obstacles.
- Preservation: It ensures that these digital materials are preserved for the long term, shielding them from digital obsolescence.

- Prestige: It presents to the world the institution's academic excellence and output.

The institutional repository and the digital library have a strong link. The IR is frequently a service offered by the digital library. The library helps authors deposit their work, maintains the software, and generates the metadata. As a result, the IR is frequently viewed as a particular collection or subset of the larger digital library infrastructure.

5.4 Open-Source Digital Repository Software

Specialized software is required to create an institutional repository or digital library. This software oversees the storage of digital files, the development of metadata, the user interface for searching and browsing, and the underlying preservation mechanisms. Although there are commercial solutions, open-source software is preferred by many academic institutions. Open-source platforms are maintained by vibrant international communities of developers and libraries, are free to obtain, and provide local modification.

Four of the most well-known open-source digital repository software platforms are covered here.

DSpace:

Perhaps the most popular institutional repository program worldwide is DSpace. It was first created by Hewlett-Packard (HP) and the MIT Libraries, and it was published in 2002. An organization's research material can be captured, stored, indexed, preserved, and redistributed in digital formats using DSpace.

Key Features:

- Structure: The manner that DSpace arranges its information is similar to how a university is structured. It makes use of Collections (which can relate to particular output kinds like Theses or Research Papers) and Communities (which can relate to departments or schools).
- Workflow: A variety of steps in the submission process are possible thanks to its adaptable workflow system. For instance, a student may turn in a thesis, which needs to be approved by a librarian before it can be made public.
- File Support: Text, photos, videos, databases, and other digital files are all accepted. In order to assist with preservation planning, it also makes an effort to identify the file format of each incoming file.
- Interoperability: The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is completely compliant with DSpace. This implies that search engines such as Google Scholar may easily locate content by harvesting metadata from a DSpace store.

EPrints

EPrints, the first software specifically designed for open access repositories, was created at the University of Southampton. It was developed especially to assist the "self-archiving" movement, in which scholars deposit their own works to make them publicly accessible.
Key features

- Emphasis on Research: Journal articles, conference papers, and technical reports are among the research publications that EPrints is most suited for handling. For making Eprint archives, it works quite well.
- Ease of Use: One of its design objectives is to encourage authors to contribute to the repository by making the submission procedure as easy as feasible.
- Flexibility: EPrints is quite adaptable and may be set up to manage several kinds of collections, such research data or multimedia assets, even if it has a default setup for research articles.
- OAI Compliance: EPrints is completely OAI-compliant, just like DSpace, guaranteeing that the research stored in it is discoverable globally and compatible with other systems.

Fedora

DSpace and EPrints are not the same as Fedora. It has no pre-made user interface and is not an "out-of-the-box" solution. Rather, it is a strong digital object repository architecture that offers the fundamental framework for libraries to develop a wide range of services. Cornell University and the University of Virginia collaborated on its development.

Key Features

- Flexibility: Fedora is incredibly flexible. A book with hundreds of page images and a transcription file are examples of complicated digital things that it can store.
- Scalability: It can manage enormous collections and intricate object relationships.
- Service-Oriented: Fedora enables repositories to link things to behaviors or services. When a user asks it, it can, for instance, dynamically convert a high-resolution image into a thumbnail.
- Use Case: Fedora is frequently selected by big libraries that need to create intricate digital libraries or preservation systems that go beyond typical research papers since it takes more technical know-how to construct a user interface on top of it.

Greenstone:

Greenstone is a software suite for building and distributing digital library collections. It is produced by the New Zealand Digital Library Project at the University of Waikato, in cooperation with UNESCO and Human Info NGO. is a software suite for building and distributing digital library collections. It is produced by the New Zealand Digital Library Project at the University of Waikato, in cooperation with UNESCO and Human Info NGO.

Key Features:

- Distribution: Greenstone's capacity to publish digital libraries on removable media, such as CD-ROMs or DVDs, in addition to serving them online, is a distinctive feature. Because of this, it is especially helpful in underdeveloped nations or places with poor internet access.

- Multilingual: It provides good support for different languages, making it a popular choice for worldwide projects
- Support for Metadata: Greenstone supports a number of metadata standards, such as METS (Metadata Encoding and Transmission Standard), and can harvest documents and collections using OAI-PMH.

5.5 Digital Collection Development

Building software is only one aspect of creating a digital library; the collection is at its core. The process of locating, choosing, obtaining, and arranging digital content is known as digital collection development. Although it adds new challenges with regard to technology, copyright, and licensing, it is very comparable to traditional collection development.

5.5.1 Identification and Selection

Determining what should be included in a digital collection is the first step in creating one. Digital material can be elusive, in contrast to traditional publishing, when books and journals are distributed through well-established routes**.**

- **Identification**: Librarians must identify sources of born-digital content (such databases, e-reports, or faculty websites) and analog assets appropriate for digitization. Because these materials frequently lack a central bibliography, this approach can be labor intensive.
- **Selection Criteria:** Not everything can or ought to be included, just like with print books. Libraries need to use strict selection standards. Important elements consist of:
    - Relevance: Does the information meet user demands and the institution's mission?
    - Content Quality: Is the data reliable and worthwhile?
    - Technical Viability: Are the file formats supported by the library? Is the cost of upkeep and storage sustainable?
    - Copyright: Is it permissible for the library to make this content available online? This is frequently the biggest obstacle to the development of digital collections.

5.5.2 Digitalization

Digitization is the link to the digital library for works that are only available in physical form, like rare books, theses, or old photos. The process of converting analog signals such as light bouncing off a page into digital signals bits and bytes is known as digitization.

There are various steps in the digitizing process:

- Choosing which tangible objects are worth the cost of scanning in order to digitize them. User demand, the requirement for preservation (to minimize handling of delicate originals), or the desire to display a distinctive collection are frequently the driving forces behind this.

- Preparation: Physical things must be checked for damage. Pages may need to be cleaned and bindings treated gently.
- Scanning/Capture: Using scanners or digital cameras to create digital images. Decisions must be made about resolution (DPI), color depth, and file format (e.g., TIFF for preservation masters, JPEG for web access).
- Processing: Image cropping, straightening, and color correction are examples of processing. Optical Character Recognition (OCR) is frequently used to transform text images into searchable text for text documents.

5.5.3 Electronic Resource Acquisition and Licensing

A large amount of the collection of a contemporary digital library is licensed from outside publishers rather than being produced or digitized locally (e.g., e-journals, e-books, databases).

Licensing: Unlike buying a physical book, libraries license access to these digital materials. This entails drafting legal agreements that specify who can access the content (such as on-campus versus remote users), how it can be utilized (such as course packs or interlibrary lending), and for how long.

Pricing Models: Libraries must deal with a variety of pricing models, including pay-per-view, annual subscriptions, and perpetual access payments, which require a one-time payment for continuous rights. Consortia (groups of libraries) often deal together to acquire better pricing.

5.6 Metadata Creation

In order for users to locate digital items, they must be characterized after they have been chosen or generated. We refer to this description as metadata, which means data about data. Metadata is the catalog item in a digital library that describes an object to both the user and the computer system. A digital object without enough metadata is akin to an unlabelled book misplaced in a warehouse.

5.6.1 Metadata Types

For digital libraries to operate well, they need a variety of different kinds of metadata:

- Descriptive metadata: Traditional cataloging is comparable to descriptive metadata. To assist users in finding the material, it explains its intellectual substance. Typical components consist of:

Title

Author/Creator

Topic/Keywords

Description/Abstract

Publication date:

- Administrative metadata: This offers details to aid in resource management. It consists of:
    - Technical metadata includes information on the file itself, including its size, resolution, program used to create it, and format (e.g., PDF, JPEG). This is critical for long-term preservation.
    - Rights Metadata: Details regarding license conditions, copyright status, and access limitations (e.g., "Open Access," "Campus Use Only").
- Structural Metadata: The relationship between the components of a digital entity is described by structural metadata. For instance, structural information connects the 100 individual image files (one for each page) that make up a digital book in the right order so the user can turn the page practically. The complex thing is bound together by it.

To ensure that digital libraries can talk to each other and share records, they must use standard formats.

- METS (Metadata Encoding and Transmission Standard): A complex standard used to wrap up all types of metadata (descriptive, administrative, and structural) and the digital content itself into a single XML package. It is often used for transferring objects between repositories.

- PREMIS (Preservation Metadata: Implementation Strategies): The international standard for preservation metadata, defining exactly what information a repository needs to keep to ensure a file remains usable in the future.

- Dublin Core:

  Dublin Core is a metadata standard that describes physical and digital resources to make it easier to find and manage them online. Often referred to as the digital library card catalog for the web, it is intended to be easy to use, succinct, and expandable.

  Standardization and Origin

  Origin: Developed in 1995 during a workshop held by OCLC and NCSA in Dublin, Ohio (thus the name "Dublin Core").

  Standards: Currently published as ISO 15836-1:2017 (Part 1: Core elements) and ISO 15836-2:2019 (Part 2: DCMI Properties and classes), it is an international standard. Another name for it is ANSI/NISO Z39.85.

  The 15 Essential Components (Basic Dublin Core)

  The 15 broad and general elements (DCMES) defined by the standard enable cross-domain resource description:

Element	Description
1. Title	The name given to the resource.
2. Creator	The entity primarily responsible for making the content (e.g., author).
3. Subject	The topic of the content (keywords, phrases, classification codes).
4.Description	An account of the content (abstract, table of contents).
5. Publisher	The entity responsible for making the resource available.
6.Contributor	An entity responsible for making contributions to the resource.
7. Date	A point or period of time associated with an event in the lifecycle of the resource.
8. Type	The nature or genre of the content (e.g., "home page", "novel").
9. Format	The physical or digital manifestation (e.g., MIME type, size).
10. Identifier	An unambiguous reference to the resource within a given context (e.g., URL, ISBN, DOI).
11. Source	A related resource from which the described resource is derived.
12. Language	The language of the resource.
13. Relation	A related resource.

Element	Description
14. Coverage	The spatial or temporal topic of the resource (location, era).
15. Rights	Information about rights held in and over the resource (copyright, licenses).

Important Features

- Interoperability: Makes it easier for records to be shared between various systems.
- Extensibility: Compatible with different metadata standards.
- Simplicity: Made to be understood and used by both experts and non-catalogers.



Check Your Progress

i. What is metadata?
ii. What are institutional repositories?
iii. What are some popular examples of DL software?

5.7 **Let Us Sum Up:**

The technological foundation of contemporary academic information systems has been examined in this section. We discovered that institutional repositories and digital libraries are complimentary systems for managing, preserving, and granting access to digital content. Institutional repositories particularly concentrate on capturing and exhibiting the intellectual output of their host institutions, whereas digital libraries play a broad role in gathering varied resources.

Popular Open-Source Software platforms that enable these systems are also discussed here.. The most well-liked options for institutional repositories are DSpace and EPrints, which provide strong management capabilities for research papers. While Greenstone is unique in its capacity to distribute libraries on portable media, Fedora offers a flexible architecture for intricate digital collections.

5.8 Answers to Check Your Progress

i. Data about data

ii. An IR is described as a collection of services that a university or research organization provides to its community members for the administration and distribution of digital content produced by the organization and its constituents.
iii. Greenstone, Fedora, DSpace, Omeka, Eprints etc.

5.9 Further Readings

1. *INSTITUTIONAL REPOSITORIES : A KEY ROLE FORLIBRARIES*. (n.d.). INFLIBNET. https://ir.inflibnet.ac.in/server/api/core/bitstreams/5dce0967-449d-44d9-ac02-32ade8734963/content
2. *METADATA AND DUBLIN CORE – Knowledge Organization and Processing – Cataloguing*. (2025). Inflibnet.ac.in. https://ebooks.inflibnet.ac.in/lisp3/chapter/metadata-and-dublin-core/
3. MUQUEEM, S., & Ambedkar, B. (n.d.). *CHANGING ROLE OF LIBRARY PROFESSIONALS IN THE DIGITAL ENVIRONMENT*. https://drtc.isibang.ac.in/ldl/bitstream/handle/1849/400/p92_shaista_khan.pdf?sequence=1&isAllowed=y
4. *UNIT 7 DIGITAL LIBRARIES Structure*. (n.d.). https://egyankosh.ac.in/bitstream/123456789/25563/1/Unit-7.pdf

5.10 Probable Questions

a) Explain what a digital library is and how it differs from the internet. What essential qualities define a digital collection as a "library"?
b) An Institutional Repository (IR): What is it? Talk about the main goals and advantages of setting up an IR in a university.
c) Describe the Digital Collection Development process. What are the most important factors to consider while choosing materials to digitize?
d) What is Metadata? Explain the significance of each of the three primary categories of metadata descriptive, administrative, and structural for a digital library.

UNIT 6: OPEN ACCESS

UNIT STRUTURE

6.1 Learning Objectives

6.2 Introduction

6.3 Concept of Open Access

6.4 Need and Importance in Digital Library System

6.5 Role of Open Access in Research Visibility and Data Sharing

6.6 Management and Integration of Open Access Resources in Digital Environments

6.6.1 Policy and Collection Development Aspects

6.6.2 Technical Integration and Discovery

6.6.3 User Education and Support

6.7 Let Us Sum Up

6.8 Answers to Check Your Progress

6.9 Further Readings

6.10 Probable Questions

6.1 Learning Objectives

After going through this unit, learners will be able to

- Explain the concept and key definitions of open access.
- Describe why open access is needed in digital libraries and scholarly communication.
- Analyse how open access improves research visibility, impact and data sharing.
- Outline ways to manage and integrate open access resources in digital library systems.
- Discuss key challenges and stakeholder responsibilities in sustaining open access.

6.2 Introduction

Open access is now a central feature of scholarly communication and digital library practice. It describes ways of making scholarly and educational materials available on the internet so

that anyone can read and use them without paying subscription fees or facing unnecessary restrictions. This idea grew from the serial's crisis, when the prices of journals rose much faster than library budgets, and from the rapid growth of digital technologies that made online dissemination easy and inexpensive. In response, scholars, libraries, funding agencies and international bodies such as the Budapest Open Access Initiative and later Berlin and Bethesda meetings proposed new models for sharing research outputs freely, while still respecting author rights and quality control. Digital libraries have benefited greatly from the open access movement. Open access journals, institutional repositories, open monographs, open educational resources and open data services supply a large body of content that can be collected, curated and provided to users through library platforms. This is particularly important for institutions in developing countries that cannot subscribe to all relevant journals yet wish to support high quality teaching and research. At the same time, open access also changes how libraries work by shifting some focus from buying content to managing and preserving the outputs created by their own scholars. In this unit we explore the concept of open access, its need and importance in digital library systems, its contribution to research visibility and data sharing, and the management and integration of open access resources in digital environments. The discussion is based on key academic and professional sources, including Suber's book on open access, Charles Bailey's overview, and modules on digital libraries and open access initiatives.

6.3 Concept of Open Access

The concept of open access has been defined and refined through several key initiatives and scholarly works. The Budapest Open Access Initiative describes open access as the free availability of scholarly literature on the public internet, permitting any user to read, download, copy, distribute, print, search or link to the full text without financial, legal or technical barriers, with the only constraint being the need to protect the integrity of the work and ensure proper acknowledgement and citation of the author. The Bethesda Statement on Open Access Publishing and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities build on this idea by emphasising that authors or copyright holders grant users a free, irrevocable and worldwide licence to use and redistribute the work and that a complete version of the work should be deposited in at least one trusted repository that supports open access, interoperability and long-term preservation. Peter Suber states that open access literature is digital, online, free of charge and free of most copyright and licensing restrictions, indicating that both price barriers and many permission barriers are removed for scholarly works that are typically written without expectation of payment. Charles Bailey refers the core concept open access stating that it removes subscription fees and most copyright and licensing restrictions to make royalty free scholarly literature available with minimal use restrictions, mainly limited to author attribution and protection of work integrity. It stresses that open access is not simply about free reading but about granting explicit reuse rights, often through standard licences such as Creative Commons.

Open access is usually discussed with reference to specific models. In the green route, authors self-archive preprints or post-prints in institutional or disciplinary repositories that are freely accessible on the internet and that often use standard protocols like the Open Archives Initiative

Protocol for Metadata Harvesting. In the gold route, journals publish the final version of record of articles as openly accessible on their websites, sometimes supported by article processing charges paid by authors, institutions or funding bodies, while readers do not pay subscription fees. Some authors also distinguish between gratis open access, where price barriers are removed but users may still face permission restrictions, and libre open access, where both price barriers and at least some permission barriers are removed through licences that allow reuse, adaptation and redistribution. Although early discussions focused mainly on journal articles, the scope of open access now extends to a wide range of scholarly and educational materials. These include theses and dissertations, conference papers, monographs, research reports, technical documents, open educational resources and in many cases research data and software. The open access movement is therefore part of a broader set of practices often described as open science or open scholarship, which aim to make the research process and its outputs more transparent, inclusive and reusable.



Check Your Progress

Q1. According to the BBB (Budapest, Bethesda and Berlin) statements, what key conditions define open access to scholarly literature in terms of access and user rights.

Q2. How do green and gold open access differ with respect to where the scholarly work is made available and who typically bears the cost of publication.

6.4 Need and Importance in Digital Library System

The need for open access in digital library systems arises from both economic and social factors as well as from the possibilities created by digital technologies. For several decades, libraries have struggled with the serial's crisis, which refers to the continuous and often sharp increase in the cost of journal subscriptions that far outpaces the growth of library budgets. This has forced many institutions, including well-funded universities, to cancel subscriptions or restrict access to journals that their researchers need, while poorer institutions have had only limited access to core literature. Open access offers an alternative model that can reduce the dependence on expensive subscriptions by allowing readers to reach research outputs without paying, thus easing the financial burden on digital library systems and widening access. From a social and ethical perspective, open access supports the principle that publicly funded research should be publicly accessible. When research is financed by taxpayers or public institutions, limiting access to those who can afford subscriptions is hard to justify. Open access digital libraries can provide universal access to public funded research publications, national scholarly works, traditional knowledge and documentary heritage collections through institutional repositories and open access journals, which strengthens democracy, supports evidence-based policy and allows communities to use knowledge for local development. For learners and researchers in developing countries, open access helps reduce global information inequalities by making high quality literature accessible irrespective of institutional wealth.

In technological terms, digital libraries are complex information systems that collect, manage and preserve digital content and provide coherent access to organised repositories of information and knowledge. The growth of open access has significantly contributed to digital library collections and technologies, because open materials can be harvested, indexed and integrated from many distributed repositories and journals across the world. Open standards and protocols such as the Open Archives Initiative and OAI-PMH have been designed to support interoperability among repositories and services, allowing digital libraries to provide unified search and access to open content. By using these standards, open access publishing becomes part of a universal digital library system where institutional repositories, subject archives and open journals collectively serve as components of a distributed global collection. Open access is also important for digital libraries because it changes the role of libraries from being only consumers or buyers of information to being producers and stewards of institutional knowledge. Libraries now manage institutional repositories where faculty and researchers deposit their articles, theses, research data and other digital assets. In doing so, they support not only immediate access but also long-term preservation, metadata quality, persistent identifiers and integration with other systems, such as research information management platforms and web ranking services. This gives libraries a stronger role in institutional research strategies and increases the visibility of the institution as a whole.

Open access improves teaching and learning services within digital library systems. Open access books, open courseware and open educational resources can be linked into reading lists, course modules and learning management systems without complicated licence negotiations or restrictions on classroom use. This is very useful for distance and open learning institutions, where students depend heavily on digital resources and often access them from outside the campus network. In all these ways, open access is not an optional extra but a core requirement for building inclusive, sustainable and effective digital library systems.

6.5 Role of Open Access in Research Visibility and Data Sharing

The most important benefits of open access is its positive effect on research visibility and impact. When articles and other outputs are freely available online without subscription barriers, they can be discovered and read by a much wider audience, including scholars in smaller institutions, independent researchers, practitioners, policy makers and members of the public who do not have access to paid databases. Studies and reports referenced by authors such as Mudry and Bailey show that open access articles often receive more downloads and, in many fields, higher citation counts than comparable paywalled articles, largely because they are easier to find and share. For institutions, this increased visibility contributes to a higher web presence, stronger reputational impact, and better measurement of research influence through repository statistics and web-based metrics. Open access also shortens the time between research production and research use. In many subject areas, preprints in open repositories allow new results to be circulated before formal journal publication, which accelerates scientific communication and allows rapid feedback from peers. This is especially valuable in fast moving fields such as medicine, environmental science and information technology, where timely access to findings can influence clinical decisions, policy interventions or technological developments. Open access repositories and journals are indexed by major search engines and

specialised academic search tools, which further enhances discovery and allows researchers to build on each other's work more efficiently.

The relationship between open access and data sharing is increasingly important in the context of open science. Many funders and policy documents now link open access publication of articles with requirements to share underlying research data in trusted repositories, except in cases where ethical or legal constraints prevent open release. The Bethesda and Berlin statements, as well as later open science declarations, emphasise not only free access to texts but also support for interoperability, unrestricted distribution and long-term archiving, which apply equally well to datasets and supplementary materials. Shared data can be reused for new analyses, combined with other datasets, or used to validate published results, thus improving transparency and reproducibility. Digital libraries and institutional repositories play a central role in enabling this visibility and data sharing. They provide the infrastructure for depositing articles and datasets, assigning metadata and persistent identifiers, and linking publications to related data, code and documentation. They also support policy compliance by giving researchers clear workflows for meeting open access and open data mandates. Librarians help by advising on data management plans, recommending appropriate licences for reuse and guiding researchers in selecting suitable repositories, including discipline specific data archives. In this way, open access is not only a matter of publishing articles differently but a broader shift towards more visible, shareable and verifiable research practices.



Check Your Progress

Q3. Why does making research outputs openly accessible usually increase their visibility and potential citation impact.

Q4. How do institutional repositories and digital libraries support data sharing as part of open science practices.

6.6 Management and Integration of Open Access Resources in Digital Environments

Managing open access resources in digital environments requires a planned approach that combines policy, technical workflows and user support. Libraries cannot assume that open access content will automatically reach their users simply because it is on the web. Instead, they need to decide which open access resources are relevant, ensure that these resources are reliably accessible through library systems and help users understand how to use them properly. This section briefly looks at policy aspects, technical integration and user services that support open access management in digital libraries.

6.6.1 Policy and Collection Development Aspects

Many libraries now include open access materials explicitly in their collection development policies. Such policies may state that the library will actively support institutional repositories,

promote local open access journals and integrate selected external open access sources that meet defined quality criteria. When selecting open access resources, librarians look at subject relevance, academic quality, stability of the platform, licensing conditions and technical interoperability, just as they do for licensed electronic resources. For example, a digital library in a science-oriented university might prioritise indexing of high-quality open access repositories in physics, health sciences or engineering along with the institution's own repository. Open access also affects institutional policies about author behaviour. Some universities adopt self-archiving mandates that require faculty to deposit copies of their accepted manuscripts in the institutional repository, subject to publisher policies. Others adopt open access policies that encourage publication in open access journals and provide support for article processing charges when needed. The Charles Bailey text notes that policy registries exist for self-archiving and that different universities have experimented with such mandates, showing that organisational commitment is important for building strong open access collections. Libraries are often involved in drafting, interpreting and implementing these policies, which positions them as key promoters of open access within the institution.

6.6.2 Technical Integration and Discovery

On the technical side, integration of open access resources into digital environments centres on metadata, interoperability and discovery services. Most institutional repositories and many open access journals expose metadata describing their contents using the Open Archives Initiative Protocol for Metadata Harvesting. This allows aggregators and discovery tools to harvest records from many repositories and present them through unified search interfaces. In a digital library, the local discovery layer can therefore index records from the institutional repository, selected subject repositories and directories such as the Directory of Open Access Journals, so that users can search across many open access sources from one place. Electronic resource management systems and link resolvers can also be configured to recognise open access versions of articles. For example, if a user searches for an article and the library does not hold a subscription to the journal, the system may still provide a link to an open access copy found in a repository or an open journal. To support this, librarians ensure that their knowledge bases and link resolver settings reflect available open access collections and that they use identifiers such as DOIs consistently. Open access also needs to be reflected in catalogue records and subject guides, so that users are aware of key open platforms in their fields. Digital libraries further integrate open access resources with learning management systems and research information systems. Links to repository copies of articles, open textbooks and open courseware materials can be embedded directly into online course pages, making it easy for students to access readings without facing paywalls. Research information systems can display repository links for publications when listing staff outputs, helping to meet funder requirements and improving visibility. This type of integration depends on good metadata quality, persistent identifiers and standardised interfaces between systems.

6.6.3 User Education and Support

Successful management of open access resources also depends on user awareness and skills. Librarians train students and researchers to identify trustworthy open access journals and

repositories, to distinguish between legitimate open access and questionable or predatory outlets, and to understand the meaning of various open licences such as Creative Commons. Training often covers how to deposit works in institutional repositories, how to interpret publisher self-archiving policies and how to manage author rights so that open access is possible. In the context of data sharing, librarians and data stewards help researchers to prepare data management plans that include provisions for open deposit where appropriate, selection of suitable repositories and assignment of metadata that support reuse. They also advise on ethical and legal issues such as confidentiality, consent and intellectual property, making clear that open access is not a reason to ignore such responsibilities. Through workshops, one to one consultations and online guides, digital libraries promote informed use and creation of open access resources, strengthening the overall culture of openness and integrity in research.

6.7 Let Us Sum Up

We have learned open access as making scholarly literature digital, online, free of charge and largely free of copyright and licensing restrictions, based on the BBB declarations and authors like Suber and Bailey. It outlined key models such as green self-archiving in repositories and gold open access journals, and noted the distinction between gratis and libre forms according to the extent of reuse rights.

The open access is needed to address the serials crisis and rising subscription costs, to ensure equitable access to publicly funded research, and to strengthen digital library collections through interoperable repositories and journals. It highlighted how open access improves research visibility and impact, and how it links to open data and open science by promoting the sharing and reuse of research data.

Libraries manage and integrate open access resources by developing supportive policies, backing self-archiving, and selecting external open sources. The technical tools such as metadata harvesting, discovery layers and link resolvers connect open content to teaching and research platforms, with user education ensuring that students and researchers use and produce open access materials responsibly.

6.8 Answers to Check Your Progress

Ans to Q1. The BBB statements define open access as the free availability of scholarly literature on the public internet, allowing any user to read, download, copy, distribute, print, search or link to the full texts and use them for any lawful purpose without financial, legal or technical barriers, while copyright is used only to protect the integrity of the work and to ensure proper acknowledgement and citation of the authors.

Ans to Q2. In green open access, authors typically self-archive preprints or postprints in institutional or disciplinary repositories that are freely accessible online, while the journals may remain subscription based, whereas in gold open access, the journals themselves make the final published articles freely available on their websites, often financed through article processing charges or other models in which the costs are not borne by readers.

Ans to Q3. Open access increases visibility and potential citation impact because research outputs are freely available to anyone with internet access, can be indexed widely by search engines and scholarly portals and are not limited by institutional subscriptions, which together expand the readership and the chance that other authors will discover and cite the work.

Ans to Q4. Institutional repositories and digital libraries support data sharing by providing trusted platforms for depositing datasets with appropriate metadata, persistent identifiers and licences, linking these datasets to related publications, and offering guidance on funder policies and ethical requirements, so that data can be discovered, reused and cited as part of open science practices.

6.9 Further Readings

1. Arora, J. (year). Open Access Publishing: An Overview. In G. Chowdhury and S. Foo (eds.), *Digital Libraries and Information Access: Research Perspectives*. London, Facet Publishing.[ppl-ai-file-upload.s3.amazonaws]
2. Bailey, C. W., Jr. (2006). What Is Open Access. In N. Jacobs (ed.), *Open Access: Key Strategic, Technical and Economic Aspects* (pp. 13–26). Oxford, Chandos Publishing.[ppl-ai-file-upload.s3.amazonaws]
3. Budapest Open Access Initiative. (2002). Read the Budapest Open Access Initiative. Available at: http://www.budapestopenaccessinitiative.org.ppl-ai-file-upload.s3.amazonaws+1
4. Chowdhury, G., and Foo, S. (2012). Digital Libraries and Open Access. In G. Chowdhury and S. Foo (eds.), *Digital Libraries and Information Access: Research Perspectives*. London, Facet Publishing.[ppl-ai-file-upload.s3.amazonaws]
5. Mudry, K. (2014). Open Access: The Importance of Open Information. Keystone Paper, LIS2000.[ppl-ai-file-upload.s3.amazonaws]
6. Suber, P. (2012). *Open Access*. Cambridge, MA, MIT Press.ppl-ai-file-upload.s3.amazonaws+1

6.10 Probable Questions

Fill in the blanks

1. According to Peter Suber, open access literature is _____, online, free of charge and free of most copyright and licensing restrictions.
2. In the _____ route to open access, authors self archive preprints or postprints in institutional or disciplinary repositories.
3. The long-term rise in journal subscription prices that outpaces library budgets is often referred to as the _____ crisis.
4. The protocol commonly used for harvesting metadata from open access repositories is called the _____.

Short answer type questions

1. State two reasons why open access is particularly important for digital libraries in developing countries.
2. Briefly distinguish between gratis open access and libre open access with reference to permission barriers.
3. How do institutional repositories increase the visibility of an institution’s research outputs.

Long answer type questions

1. Discuss the concept of open access with reference to the definitions provided by the Budapest Open Access Initiative, the Bethesda Statement, the Berlin Declaration.
2. Examine the need and importance of open access in digital library systems.
3. Describe the main components involved in the management and integration of open access resources in digital environments.

UNIT 7: OPEN ACCESS RESOURCES AND PLATFORMS

UNIT STRUCTURE

7.1 Learning Objectives

7.2 Introduction

7.3 Types of Open Access

7.4 Open Access Resources

7.5 Open Access Platforms

7.6 Major International Repositories

7.7 Major National Repositories

7.8 Let Us Sum Up

7.9 Answers to Check Your Progress

7.10 Further Readings

7.11 Probable Question

7.1 Learning Objectives

After studying this unit, learners will be able to:

- Describe the main types of open access, including gold, green, hybrid, diamond and bronze models.
- Explain what is meant by open access platforms.
- Describe major international and national open access repositories and their roles in research and education.

7.2 Introduction

Open access means that readers can use scholarly works online without paying subscription fees and without facing major legal barriers, while still respecting the basic rights of authors. In this model, literature is usually digital, online, free of charge to read, and free from most copyright and licensing restrictions that would otherwise limit reuse.

Digital libraries and institutional repositories strengthened this trend by offering organised, long-term access to digital collections of articles, theses, data and learning materials. They collect, manage, preserve and provide access to digital content for defined user communities,

often using standard protocols and metadata to make discovery and interoperability easier. As a result, open access resources and platforms now cover journals, books, theses, preprints, learning objects and research data across all disciplines and regions.

The major open access declarations

i. **Budapest Open Access Initiative (BOAI), 2002**

   - First major statement to define "open access" for scholarly literature. It calls for free, unrestricted online access to research articles and supports two main strategies: self-archiving (green OA) and open access journals (gold OA).

ii. **Bethesda Statement on Open Access Publishing, 2003**

   - Focuses on biomedical research and sets conditions for what counts as an open access publication. It emphasised that authors retain copyright and grant users a licence to copy, use and distribute the work with proper attribution.

iii. **Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003**

   - Broadens open access from science to all fields of knowledge, including humanities and cultural heritage. Urges institutions and funders to develop policies and infrastructures that support open access archiving and publishing.

iv. **UNESCO / WSIS documents related to OA, 2003 onwards**

   - World Summit on the Information Society (WSIS) Declaration and Plan of Action briefly endorse open access to scientific information as part of building inclusive knowledge societies. Later UNESCO work links open access with education, development and the right to benefit from scientific progress.

v. **Paris OER Declaration (UNESCO), 2012**

   - Focuses on open educational resources but strongly supports open access principles for teaching and learning materials. Calls on governments and institutions to release publicly funded educational resources under open licences.

vi. **BOAI 10-year and 15-year reaffirmations (2012, 2017)**

   - Reaffirm the original Budapest principles and propose concrete recommendations to make open access the default for publicly funded research. Stress coordinated policies on repositories, licences, and standards across funders and institutions.

7.3 Types of Open Access

Open access is often described using colour and model labels that indicate how articles become freely available and how publishing is financed. Although terminology sometimes varies, a few broad types are widely recognised in policy documents, library guides and publisher instructions.

7.3.1 Gold open access

Gold open access refers to works that are made openly available on the publisher's site at the time of publication, usually in fully open access journals or open access books. In this model, the final version of record is free to read immediately, and the journal often applies a Creative Commons licence to enable reuse under defined conditions. The cost of publication is typically covered by article processing charges (APCs) paid by authors, their institutions or funders, or by other business models such as subsidies, membership or consortia funding.

7.3.2 Hybrid open access

Hybrid open access appears in subscription journals that offer authors an option to make individual articles open upon payment of an APC, while other content in the same issue remains behind a paywall. This model was promoted as a transitional mechanism but has been criticised because publishers sometimes receive both subscription revenue and APCs for the same journal, a practice often described as "double dipping". Hybrid open access can help authors comply with funder mandates but may be less efficient for libraries that still need to pay for overall subscriptions.

7.3.3 Green open access

Green open access, or self-archiving, refers to depositing a version of an article or other work in an institutional or subject repository that is freely accessible online. Authors usually publish in any journal and then make a preprint or the author's accepted manuscript (AAM) available, sometimes after an embargo period specified by the publisher. Green open access is usually APC-free and depends on repository infrastructure, clear publisher policies and, in some contexts, funder or institutional requirements to deposit.

7.3.4 Diamond (or platinum) open access

Diamond open access journals and platforms are free to read and free to publish in, meaning that neither readers nor authors pay fees. Costs are met through other sources such as institutional support, consortia funding, grants or volunteer work, and these journals are often community driven and non-commercial. Recent initiatives by UNESCO and regional bodies have encouraged diamond open access as a fairer model, particularly for researchers from low- and middle-income countries.

7.3.5 Bronze open access and related terms

Bronze open access refers to articles that are temporarily or permanently free to read on a publisher's website but do not carry a clear open licence, so reuse rights remain uncertain. Some guides also use terms such as "platinum" open access to describe APC free journals, or

“black” open access to refer to unauthorised sharing on illegal platforms, which should not be confused with legitimate open access. For practical purposes in policy and library work, the main distinctions are between green, gold, hybrid and diamond routes, along with an understanding of licence conditions.

Check Your Progress

Q1. What is meant by gold open access and how are publication costs usually covered in this model?

Q2. How does green open access differ from hybrid open access in terms of where the article is made available and how?

Q3. Why is diamond open access considered more equitable for many researchers?

7.4 Open Access Resources

Open access resources cover a wide range of scholarly and educational materials that users can access online without subscription or pay per view barriers. For learners, teachers and researchers, these resources include articles, books, theses, course materials, multimedia objects and datasets that can be integrated into teaching, learning and further research.

Open access journals: Open access journals are the most visible resources, delivering peer reviewed articles that are free to read and often free to reuse under Creative Commons licences. Directories such as the Directory of Open Access Journals (DOAJ) index thousands of such titles across disciplines and languages, making it easier to discover reliable open access content. Alongside journals, open access monographs and book chapters are increasingly being published through initiatives like the Directory of Open Access Books (DOAB) and platforms such as OAPEN.

Theses and dissertations: It represents original research that may not yet appear in journals. Many countries have created national or institutional repositories that provide open access to electronic theses and dissertations (ETDs), often using standard software such as DSpace and common interoperability protocols. In India, for example, Shodhganga at INFLIBNET offers a large collection of full text theses and synopses submitted to Indian universities, freely downloadable in multiple languages.

Open educational resources (OER) extend beyond research to include learning and teaching materials such as syllabi, lecture notes, presentations, simulations, question banks and full courses. These materials are usually shared with open licences that allow adaptation and localisation, supporting both formal and non-formal education. National repositories like

NROER in school education and platforms like e Gyankosh or e PG Pathshala in higher education illustrate how OER can support curriculum and teacher development at scale.

7.4.1 Open access journals and books

Open access journals follow peer review processes similar to subscription journals but remove price barriers for readers. Large non-profit and commercial publishers, as well as small scholarly societies and university presses, operate open access titles using combinations of APCs, institutional subsidies, sponsorship and consortia agreements. Many of these journals apply licences like CC BY or CC BY NC, which clearly state how articles may be reused, shared or adapted.

Open access books have grown rapidly, particularly in the humanities and social sciences, where monographs remain a preferred form of scholarly communication. Platforms such as OAPEN and DOAB host peer reviewed open access books from academic publishers, while some universities run their own presses that support open monograph publishing. These resources are especially important in contexts where print books are expensive or hard to obtain, allowing students and researchers to access core readings online without cost.

7.4.2 Open educational resources and courseware

Open educational resources are teaching and learning materials made available under open licences for free use, adaptation and redistribution. They cover curriculum maps, modules, textbooks, lecture notes, assignments, multimedia objects and full courses, and are meant to be reused and remixed rather than just consumed. Internationally, initiatives like MIT OpenCourseWare, Open Yale Courses and large MOOCs have demonstrated how universities can share course content openly with global learners and other example include Open Textbook Library, Lumen/Boundless, JHSPH OpenCourseWare.

In India, e PG Pathshala provides structured postgraduate course content in multiple disciplines, while NPTEL hosts thousands of video lectures and course modules in engineering and related areas. Repositories such as NROER, Sakshat and e Gyankosh support school and higher education by archiving digital textbooks, audio video programmes and interactive resources under Creative Commons licences. Together, these OER initiatives make high quality learning materials available to teachers and learners who may not otherwise have access to expensive textbooks or proprietary e learning packages.

Disciplinary / subject repositories

- Repositories focusing on specific fields, often used for preprints and working papers. Examples: arXiv (physics, maths, CS), PubMed Central (biomedicine), Social Science Open Access Repository (SSOAR), various preprint servers listed in COAR’s directory.

National and specialised open repositories (India and others)

- Shodhganga: Indian national repository of theses and dissertations.

- Shodhgangotri: repository of synopses/research in progress.
- e Gyankosh: national digital repository for IGNOU ODL materials.
- NROER: National Repository of Open Educational Resources (school level).

Open access conference proceedings and preprints

- Free access to conference papers and early versions of research outputs. Examples: many OA proceedings in DOAJ/DOAB, discipline based servers such as arXiv and others indexed by COAR and OpenDOAR.

Open data and government/open science data portals

- Research datasets and government data made freely available for reuse. Examples: Data.gov (US open government data) and other national open data portals.

Aggregators, search engines and directories for OA content

- Make it easier to discover OA resources from many repositories.

  Examples:

  - DOAJ (journals), DOAB (books).
  - OpenDOAR and ROAR (directories of repositories).
  - BASE and OAIster (search engines harvesting OAI PMH repositories).
  - Open Access Directory (OAD) lists tools, repositories, policies and more.

Lists of open educational resources and e resources

- Curated lists from libraries and institutions combining many OA/OER tools. Examples: institutional "Open Access e Resources" pages (e.g., university library guides in India that list DOAJ, DOAB, PubMed Central, Shodhganga, e PG Pathshala, etc.).

Check Your Progress

Q4. What are open educational resources and how do they differ from open access journal articles?

Q5. Name any two Indian initiatives that provide open educational resources or open courseware and briefly describe their focus.

7.5 Open Access Platforms

Open access platforms are the technical and organisational systems that host, manage and deliver open access content to users. They may take the form of journal publishing platforms, institutional or disciplinary repositories, aggregators, preprint servers or specialised search engines that harvest metadata from multiple sources.

Another group of platforms includes journal and book hosting systems that support open access publishing workflows. Open-source software like Open Journal Systems (OJS) and various commercial hosting services allow institutions or societies to manage peer review, production and dissemination entirely in an online environment. At a higher level, directory and discovery platforms such as DOAJ, DOAB, OpenDOAR and aggregators like OAIster act as gateways, enabling users to search across thousands of repositories and journals from a single interface.

Major global open access journal initiatives

- Directory of Open Access Journals (DOAJ)
- Public Library of Science (PLOS)
- BioMed Central (BMC)
- eLife
- Bioline International
- SpringerOpen and Nature's OA journals.
- Wiley and Taylor & Francis OA journals.
- MDPI, Frontiers and other born OA publishers

Indian open access journal initiatives (publishers / platforms)

- CSIR–NIScPR (formerly NISCAIR) journals
- Indian Academy of Sciences journals
- Indian National Science Academy (INSA) journals
- Indian Council of Medical Research (ICMR) and Indian Medlars Centre (MedIND)
- Medknow (now part of Wolters Kluwer)
- Bioline International

An institutional repository is a web-based archive that collects, organises, preserves and provides open access to the scholarly and academic outputs of a specific institution, such as a university or research centre. It typically stores research articles, preprints and postprints, electronic theses and dissertations, conference papers, reports, datasets, course notes and other learning objects in digital form. By serving as an institution based digital asset management and preservation system, it offers a central, searchable platform where this material can be discovered and reused. Institutional repositories increase the visibility and impact of an institution's research, provide a cost effective and immediate route to share scholarly work, and help institutions and authors comply with open access and funder mandates. These

repositories provide open access through web interfaces and standard protocols such as the Open Archives Initiative Protocol for Metadata Harvesting (OAI PMH).

7.6 Major International Repositories

International open access repositories and services are important because they provide global access to disciplinary literature and enable cross institutional discovery. Many of them are disciplinary repositories that serve particular research communities, while others act as large scale harvesting services that build union catalogues of open access content.

arXiv is one of the earliest and most influential disciplinary repositories, covering physics, mathematics, computer science and related fields. It operates as a preprint server where authors can deposit manuscripts prior to formal journal publication, allowing rapid dissemination and community feedback while still enabling later peer reviewed versions. In the life sciences, PubMed Central functions as a full text archive of biomedical and health science journal literature, including articles arising from funder mandates such as those of the NIH.

Other major international repositories and services include Social Science Research Network (SSRN) for social sciences, Europe PubMed Central for European funded biomedical research, and institutional systems like the University of California's eScholarship repository. Global directories and harvesting services, such as OpenDOAR, ROAR, OAIster and BASE, index thousands of institutional and disciplinary repositories and make their contents searchable through unified interfaces. These infrastructures, built on open protocols like OAI PMH, greatly increase the visibility and impact of open access resources.

Major international repositories

i. arXiv
ii. Europe PubMed Central (Europe PMC)
iii. SSRN (Social Science Research Network)
iv. RePEc (Research Papers in Economics).
v. CiteSeerX
vi. CERN Document Server
vii. NASA ADS (Astrophysics Data System)
viii. HAL Institute National de Recherche en Informatique et en Automatique Archive Ouverte
ix. Virginia Tech University Digital Library and Archives
x. University of California eScholarship Repository

7.7 Major National Repositories

Many countries have developed national level repositories or networks of institutional repositories to ensure preservation and visibility of their scholarly output. These systems often focus on theses and dissertations, publicly funded research, or educational resources in local languages, and are usually supported by government agencies or consortia

In India, the INFLIBNET Centre under the University Grants Commission (UGC) plays a leading role in building national open access infrastructure. Shodhganga is a flagship repository of Indian electronic theses and dissertations, aggregating ETDs from hundreds of universities and making them freely accessible online in multiple Indian languages as well as English. Its companion service Shodhgangotri hosts synopses and research in progress, revealing research trends and reducing duplication of topics.

Other Indian repositories include institutional archives at leading universities and research institutes, such as the e prints repository at the Indian Institute of Science, repositories at national laboratories, and thematic services like OpenMed for biomedical literature. In the educational domain, e Gyankosh, NROER and various state level initiatives provide national level platforms for open textbooks, courseware and multimedia resources. Together, these repositories support open access to Indian research and educational content, enhance visibility of local scholarship, and facilitate reuse in teaching and further research.

Major National institutional repositories

1. Indian Institute of Science (IISc) ePrints Repository
2. IIM Kozhikode (IIMK) Institutional Repository
3. Indira Gandhi Institute of Development Research (IGIDR) Repository
4. Central Marine Fisheries Research Institute Institutional Repository
5. Indian Institute of Astrophysics Repository
6. Indira Gandhi National Open University (IGNOU) Institutional Repository
7. INFLIBNET Institutional Repository (IR @ INFLIBNET)
8. NIScPR / NISCAIR Online Periodicals Repository
9. Open Access Repository Publications of Fellows of the Indian Academy of Sciences
10. National Institute of Oceanography India Digital Repository
11. OpenMed National Informatics Centre India
12. National Aerospace Laboratories Institutional Repository

7.8 Let Us Sum Up

Open access makes it possible to read scholarly work online without paying subscription fees or facing heavy legal restrictions, while still respecting what authors are entitled to. It rests on important policy statements such as the Budapest, Bethesda and Berlin declarations, later

reinforced by UNESCO and BOAI, which call for free online access, self-archiving, open licences and supportive infrastructure. In practice, several models exist: gold and hybrid options on publisher sites, green self-archiving in repositories, diamond journals that charge no fees at all, and bronze articles that are free to read but have unclear reuse rights. Today, open access covers journals, books, theses, course materials, preprints and datasets, helped by services like DOAJ, DOAB, OAPEN and large OER programmes. Institutional repositories act as digital asset hubs that store, preserve and share an institution's outputs, using common technical standards to boost visibility and satisfy funder requirements. Together, international and national repositories form a worldwide network that broadens access to research and education.

7.9 Answers to Check Your Progress

Ans to Q1: Gold open access means that the final published version of an article or book is made freely available on the publisher's site immediately upon publication, usually under an open licence. Publication costs are generally covered through article processing charges paid by authors, their institutions or funders, or through alternative funding such as institutional subsidies or transformative agreements.

Ans to Q2: Green open access involves authors depositing a version of their work, such as a preprint or accepted manuscript, in an open repository, sometimes after an embargo, while the journal itself may remain subscription based. Hybrid open access, in contrast, makes specific articles open on the publisher's platform in otherwise subscription journals when authors pay an APC for that option.

Ans to Q3: Diamond open access is considered more equitable because neither readers nor authors pay fees, reducing financial barriers for researchers from underfunded institutions or countries. Its funding relies on shared institutional or community support rather than per article charges, which helps address inequalities created by high APCs in some gold and hybrid journals.

Ans to Q4: Open educational resources are teaching and learning materials such as textbooks, lecture notes, videos and full courses that are shared under open licences allowing free use, adaptation and redistribution. Open access journal articles, while often free to read and sometimes reusable, are primarily research outputs and are usually not designed as modular, remixable teaching resources in the way OER.

Ans to Q5: e PG Pathshala is an Indian initiative that offers structured postgraduate course content across disciplines in open online format, supporting higher education teaching and self-learning. NPTEL provides extensive open courseware and video lectures in engineering and related subjects, developed by IITs and other premier institutions, freely accessible to students and teachers.

7.10 Further Readings

1. Arora, J. Open Access Publishing: An overview. In Digital Libraries and Information Access: Research Perspectives, edited by G. Chowdhury and S. Foo. Facet Publishing.[ppl-ai-file-upload.s3.amazonaws]
2. Budapest Open Access Initiative (2002). "Read the Budapest Open Access Initiative." Available at: http://www.budapestopenaccessinitiative.org.[ppl-ai-file-upload.s3.amazonaws]
3. Butcher, N. (2015). A Basic Guide to Open Educational Resources (OER). UNESCO and Commonwealth of Learning.[ppl-ai-file-upload.s3.amazonaws]
4. Open Access Australasia. "What are the different types of open access?" Available at: https://oaaustralasia.org/what-are-the-different-types-of-open-access/.[oaaustralasia]
5. Open Access Types – Open Access Publishing (JHU Welch Library). "Open Access Types." Available at: https://browse.welch.jhmi.edu/writing_publishing/oa_types.[browse.welch.jhmi]
6. Suber, P. (2004). "A Very Brief Introduction to Open Access." Available at: http://www.earlham.edu/~peters/fos/brief.htm.[ppl-ai-file-upload.s3.amazonaws]

7.11 Probable Questions

Fill in the blanks

1. In the _____ open access model, articles are made freely available on the publisher's website immediately upon publication.
2. Depositing a preprint or author's accepted manuscript in an institutional repository is known as _____ open access.
3. Open access journals that are free to read and do not charge authors any APCs are usually called _____ open access journals.
4. Shodhganga is a national repository in India that provides open access to _____ submitted to universities.

Short answer type questions

1. Define open access and explain how it differs from traditional subscription-based publishing.
2. Briefly describe the role of institutional repositories in supporting green open access.
3. What is meant by bronze open access and why can it be problematic from a reuse perspective?

Long answer type questions

1. Discuss the main types of open access, write their key features, advantages and limitations.

2. Describe different categories of open access resources and explain how open access repositories and platforms support teaching, learning and research in a university environment

Unit 8: Organization of Knowledge

Unit Structure

8.1 Learning Objectives
8.2 Introduction
8.3 Concept of Knowledge Organization
- **8.3.1 Definition**
- **8.3.2 Purpose and Importance of knowledge organization**
- **8.3.3 Types of Knowledge**:

8.4 Library classifications: Concept
- **8.4.1 Classification system**
- **8.4.2 Common Classification Schemes**

8.5 Cataloguing
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 completing this unit, learners will:

- Know the concept of knowledge organization and importance of it in libraries.
- Understand how knowledge organization works in a library.
- Get aware of classification and various classification schemes.
- Know about the concept of cataloguing and its various types.

8.2 Introduction

Knowledge has always been fundation to human civilization. From prehistoric cave drawings to modern-day digital databases, humans have continuously created, stored, and shared knowledge in order to better understand the world and improve their lives. One of the most fundamental concepts in Library and Information Science (LIS) is knowledge organization, because libraries exist not just to store information, but also to make knowledge accessible, meaningful, and usable. Libraries store a large amount of information in many formats including books, newspapers, journals and electronic materials. Over time, the amount keeps on increasing, and managing this information becomes very important to make them quickly and easily available for use. In this scenario, arranging the information in a clear and systematic order is crucial to make it useful and easily findable for the users. This process is termed as knowledge organization. Proper organization helps users to find information easily and more effectively. Without organization of knowledge, a well-stocked library would be difficult to use properly.

8.3 Concept of Knowledge Organization

Knowledge organization is the systematic arrangement of recorded information in such a way that anyone can easily locate, retrieve, and utilize it. In simple terms, it is the process of organizing information in order to prevent confusion and loss. Knowledge organization is the intellectual and systematic arranging of recorded information so that it is easy to find, understand, and use. It discusses the ideas and methods used to organize information in libraries, databases, and information systems. The primary goal of knowledge organization is to minimize complexity and bring order to the vast universe of knowledge. Human knowledge expands constantly across fields and formats. Knowledge organization offers a framework for connecting similar ideas, distinguishing unrelated ones, and establishing meaningful connections between concepts.

8.3.4 Definition

Knowledge organization can be defined as the systematic arrangement, classification, and representation of knowledge to ensure efficient storage, retrieval, and use.

This definition focuses on three key elements:

1. Structure: Knowledge must be organized logically.

2. Representation: Knowledge must be communicated through symbols, words, or codes.

3. Accessibility - Knowledge should be organized for easy retrieval and usage.

In library science, this arrangement is achieved using methods such as classification systems (DDC, UDC, CC), subject headings, thesauri, and metadata standards. They serve as bridges between users and information sources. Without them, libraries would become just book warehouses rather than being a dynamic learning place. This definition focuses on three key elements:

8.3.5 Purpose and Importance of knowledge organization

The importance of organizing knowledge cannot be emphasized enough. Libraries serve people from every background, interests, and aspirations. A well-organized library allows:

- To make information easy to find
- To save the time of user
- To bring together related subjects
- To support research and learning
- To preserve intellectual heritage

In the digital era, knowledge organization has expanded beyond physical libraries to include digital repositories, online catalogues, and search engines.

8.3.6 Types of Knowledge:

Knowledge is not uniform in nature. In library and information science, knowledge is typically divided into two types:

1. Explicit Knowledge.
2. Tacit Knowledge.

Explicit knowledge is knowledge that can be simply communicated, recorded, stored, and distributed. It is structured, systematic and often appears in written or recorded form. Examples include books, research papers, manuals, databases, reports, and digital files. The majority of library collections include explicit knowledge. Classification systems, catalog records, and indexing procedures are mainly designed to arrange this type of knowledge. Explicit knowledge may be codified into symbols or language, transmitted through formal channels, stored for extended periods, and retrieved frequently without loss of data. The main advantage of explicit knowledge is that it is easily shareable. It can be easily distributed throughout time and space. However, explicit knowledge often indicates only the surface level of human comprehension.

Tacit knowledge is extremely personal, experience-based, and difficult to put in words. It resides in people's minds and is influenced by intuition, abilities, principles, and personal experiences. Unlike explicit knowledge, tacit knowledge is difficult to write down or properly document.

Examples of tacit knowledge include a librarian's ability to identify user needs without clear instructions. Experienced problem-solving skills, Professional judgment and decision-making, Cultural and social sensitivity etc. all come under tacit knowledge.

In libraries, tacit knowledge is crucial for the quality of services. An experienced librarian may know which sources are most reliable for a specific topic, even if such information is not written down anywhere else. This type of knowledge is acquired by experience, involvement and reflection.

One of the most difficult aspects of knowledge organization is gathering and organizing tacit knowledge. However, modern knowledge management approaches try to organize tacit knowledge through documentation, best-practice standards, and institutional memory.

Relationship between explicit and tacit knowledge

Explicit and tacit knowledge are not contradictory; rather, they are interrelated. Explicit knowledge is often constructed on the basis of tacit knowledge. For example, an author's personal ideas (tacit) turn into a book (explicit). Similarly, explicit knowledge may become tacit through practice and experience. Understanding both types is crucial for effective organization in libraries. While classification methods organize explicit knowledge, human expertise allows its interpretation, integration, and meaningful utilization.

8.4 Library classifications: Concept

Library classification is the process of organizing library materials in a logical and systematic manner based on its subject content. It arranges documents and information that cover similar topics and distinguishes those dealing with different areas. Classification creates an idea-plan, showing how subjects are linked to one another.

Classification is more than just random shelving; it is based on well-thought ideas and systems that represent the structuring of human knowledge. Each document is assigned a class number that identifies its subject and decides where it sits on the shelf.

Definition of Library Classification:

Library classification can be defined as the process of systematically organizing library materials based on their subject content similarity to make it easier to store, find, retrieve and use.

Objective of library classification:

Some of the main objectives of library classification are as follows:

- To bring related books and materials together
- To help users go through a subject easily
- To save time in finding materials
- To support efficient management of the library materials and many more.

8.4.1 Classification system

A library classification system is a way for organizing books and other resources in a logical sequence. It makes it easier for readers to find what they need and helps librarians to manage collections more efficiently.

Over time, various approaches and systems of classification have emerged. Structurally, classification is broadly divided into two categories:

1. Enumerative classification

2. Faceted classification

Enumerative classification:

Enumerative classification systems compile a list of all subjects and sub-subjects and assign them class numbers. here, each topic already has a pre-determined number. If a book matches a specific topic, it is assigned that number. There is not much need to combine or produce new numbers.

The main advantage of this classification is that it is simple to use and comprehend. On the downside, it is not adaptable for new or developing topics.

Example of this classification include Dewey Decimal Classification (DDC).

Faceted Classification:

This classification is also known as analytico-synthetic classification. In this method divides a topic into various facets, which are then combined to create a class number.

Subjects are analyzed into various components, such as topic, place, time, and form. The exact number is then produced by combining these components.

Benefits of this classification is that it is highly flexible and detailed which is ideal for creating new and specific subjects. But there are some limitations also, like it is difficult to learn and proper training is needed for effective application. Example of this system include Colon Classification by S. R. Ranganathan.

8.4.2 Common Classification Schemes

Dewey Decimal Classification (DDC): it is a numerical system that organizes library materials into ten main subject classes, ranging from 000 to 900. Each main class is further divided into more specific topics using decimal numbers. since it is simple and easy to apply, DDC is widely used in academic and public libraries.

Library of Congress Classification (LCC): it uses a combination of letters and numbers to arrange materials by subject. Each letter represents a broad subject area, and additional letters and numbers provide more specific classification. This system is more detailed and flexible, making it suitable for large academic and research libraries.

Universal Decimal Classification (UDC): it is an international classification system based on the Dewey Decimal system. It allows complex subjects to be classified using symbols, punctuation, and numbers. UDC is commonly used in special and international libraries where detailed subject analysis is needed.

Colon Classification: it was developed by S. R. Ranganathan and is based on the principle of facets, such as personality, matter, energy, space, and time. It uses colons and other symbols to combine these facets into a classification number. Although innovative, it is mostly used for academic study rather than everyday library practice.

8.5 Cataloguing

Concept: A library is often viewed as the heart of an institution, and the catalogue can be called the heart of the library. Cataloguing is the process of making a systematic list of library materials that describes and identifies each item within a collection. It contains extensive information on documents, such as the author, title, subject, edition, and publishing facts. As library collections develop, the catalogue becomes an increasingly important tool for accessing the required materials among a large number of collections. It acts as a guide to the library's whole collection and assists patrons in understanding the resources available. The catalogue informs users about the types of materials available at the library and where to find them. While classification indicates where a document is stored, cataloguing describes what it is. The library catalogue serves as a communication tool between the library and its patrons. It informs people about what the library possesses and how to obtain it.

The term catalogue is derived from the Greek words kata and logos. Kata means "according to," and logos means "order." Thus, a catalogue is an arrangement of reading materials in a systematic and logical manner.

In simple words, cataloguing may be defined as:

"The process of creating bibliographic records which describe library materials and provide their location for easier storage, access and retrieval".

Over time, various library science experts have explained the concept of a catalogue in their own unique ways. one of the first clear definitions was found in C. A. Cutter's book Rules for a Dictionary Catalogue, published in 1876. According to him, a catalogue is a list of books organized in a precise order, and it includes items from a specific library or collection.

Types of Cataloguing: Based on various factors, cataloguing can be divided into various categories.

Cataloguing by Arrangement

Cataloguing by arrangement is concerned with the order in which catalogue entries are arranged for use by readers. It includes:

Alphabetical cataloguing: It arranges entries in strict alphabetical order of author, title, or subject headings.

Example: Author catalogue.

Classified cataloguing: It arranges entries according to a classification scheme based on subjects.

Example: Entries arranged according to Dewey Decimal Classification numbers.

Alphabetico-classified cataloguing: It is a combination of alphabetical and classified arrangements, where entries are classified by subject and then arranged alphabetically within each class**.**

Example: Subject classes arranged numerically, with authors listed alphabetically within each class.

Cataloguing by Function / Content

Cataloguing by function or content is concerned with the information provided in the catalogue to identify and represent documents. It includes:

Descriptive cataloguing: It deals with the physical and bibliographic description of documents, such as author, title, edition, and size.

Example: Recording full publication details of a book.

Subject cataloguing: It deals with the intellectual content of documents and assigns subject headings or classification numbers.

Example: Assigning the subject heading “Library Science” to relevant books.

Cataloguing by Organization

Cataloguing by organization is concerned with how cataloguing work is organized and shared among libraries. It includes:

Cooperative cataloguing: It involves libraries sharing cataloguing work to avoid duplication.

Centralized cataloguing: It is done by a central agency for use by many libraries.

Union catalogue: It lists the holdings of several libraries in a single catalogue.

MARC (Machine Readable Cataloguing): It is a standardized digital format for bibliographic data.

OPAC (Online Public Access Catalogue): It is a computerized catalogue that allows users to search a library’s collection electronically. Users can search by keyword, author, title, or subject, making it fast, accurate, and easy to use. OPACs are the most common type of catalogue in modern libraries.

Cataloguing by Physical Form

Cataloguing by physical form is concerned with the physical format in which the catalogue is presented. It includes:

Book catalogue: It is a printed list of all the materials available in a library, arranged in a systematic order. While it is easy to handle and can be used without equipment, it is difficult to update and quickly becomes outdated.

Card catalogue: It consists of individual cards, each representing a library item, usually arranged alphabetically by author, title, or subject. It was the traditional method of cataloguing before computers and allowed users to browse library holdings manually.

Microform catalogue: It stores catalogue records on microfilm or microfiche.

Knowledge organization offers the theoretical framework, whereas classification and cataloguing are practical implementations in libraries. Classification organizes resources by subject, whereas cataloguing explains them in comprehensive manner, providing efficient knowledge access.

These three components collaborate in turning a collection of documents into a useful and relevant information system.



Check Your Progress

A. Fill in the Blanks

a. Knowledge is generally categorized into two types: Explicit Knowledge and _____.

b. Faceted classification is often referred to as _____ classification.

c. The term catalogue is derived from the Greek word *kata* meaning "according to" and logos meaning _____.

d. _____ cataloguing entails libraries collaborating on cataloguing tasks to prevent redundancy.

8.6 Let Us Sum Up

Knowledge organization, library classification, and cataloguing are essential components of library and information science. They ensure that the ever-growing number of recorded information is easily accessible, organized, and relevant. Classification creates order through subject arrangement, cataloguing provides identity and access, and knowledge organization connects them together into an integrated system. In an age of information overload, these processes remain crucial for guiding users toward relevant Information and fostering lifelong learning.

8.7 Answers to Check Your Progress

A. Fill in the Blanks

a. Tacit Knowledge,
b. Analytico-synthetic
c. Order
d. Cooperative

8.8 Further Readings

1. Bliss, H.E. (1933). *Organization of Knowledge in Libraries and the Subject Approach to book*. New York

2. Needham, C. D. (1965). *Organizing knowledge in libraries: An introduction to classification and cataloguing g*. Grafton Book.

3. Wynar, B. S., & Taylor, A. G. (1992). *Introduction to cataloging and classification* (8th ed.). Libraries Unlimited.

4. Krishna, Kumar (1996). *Theory of Classification*. New delhi : Vikash Publishing House. Chapter5-11

5. Probable Questions

A) Short answer type questions

I. What do you understand by knowledge organization in libraries?
II. Define Classification
III. What are the two main types of knowledge?

B) Long answer type questions

I. Explain various classification schemes.
Discuss different types of cataloguing with examples

UNIT 9 BIBLIOGRAPHIC CONTROL AND SUBJECT HEADING

UNIT STRUCTURE

9.1 Learning Objectives
9.2 Introduction
9.3 Bibliographic Control
9.4 Concept of indexing: Type, Application, Computerized Indexing, Indexing of Internet resources
9.5 Subject Headings
9.6 Let Us Sum Up
9.7 Answers to Check Your Progress
9.8 Further Readings
9.9 Probable Questions

9.1 Learning Objectives

After this unit, learners will be able to

- Explain the concept of bibliographic control in both print and electronic environments.
- Describe the concepts of subject indexing and the different types of indexes.
- Explain the difference between subject headings and keywords, and their functions in information retrieval.
- Describe how knowledge is organized in digital libraries using traditional and new tools.

9.2 Introduction

The purpose of libraries is to provide individuals with access to recorded knowledge. In order to do this effectively, they must maintain organized records of what they possess and provide clear, searchable descriptions of each item. This is the fundamental concept of bibliographic control. It refers to the actions that let visitors find out what papers are available, where they are located, and what topics they cover. This task has traditionally revolved around subject access. Rarely do patrons visit the library with a certain edition in mind; instead, they typically arrive with a query or topic in mind. Subject heading lists, catalogues, and classification systems were the mainstays of bibliographic management during the print era. The world of knowledge has greatly expanded beyond local collections in the digital age. Resources from numerous universities are combined in digital libraries, and access is frequently worldwide. Bibliographic control has become more crucial and more difficult as a result. New methods, such as universal bibliographic control (UBC) and metadata-based description, have evolved to allow discovery across systems and borders

At the same time, indexing and subject analysis have evolved. Digital libraries include keywords, thesauri, ontologies, and other knowledge organization systems (KOS) to facilitate search and browsing, although traditional subject headings and classification are still important. The broad idea of bibliographic control will be covered in this lesson, followed by

the useful tools of topic headings and indexing, and finally knowledge organizing techniques in digital libraries.

9.3 Bibliographic Control

The collection of actions that guarantee recorded information resources are methodically characterized, accessible, and usable when needed is known as bibliographic control. In the past, it developed from national bibliographies that sought to document a nation's publishing output and from initiatives like Gesner's Bibliotheca Universalis, which intended to list all printed books of its time. The concept of Universal Bibliographic Control was created in the 20th century by international organizations like IFLA. Under this system, each nation is responsible for describing its own publications in accordance with established criteria, and these records can subsequently be shared globally.

Bibliographic control at the library level consists of:

- Recording titles, authors, and publication details is known as descriptive cataloging.
- Subject classification and cataloging (specifying the topic of the item).
- Authority control (making sure that subjects and names are used consistently).
- Upkeep of user-searchable catalogues and indexes.

The work and its metadata are now more important in the digital ecosystem than the physical object. A shared web of metadata, including standardized titles, author identities, subject terms, and persistent links, connects printed books, e-books, and digital articles that contain the same intellectual content. Quality metadata has become a core paradigm of bibliographic control. It influences a resource's online visibility and the ease with which people may locate and assess it.

National libraries are no longer the only institutions with bibliographic control. Metadata is currently created and shared by publishers, research repositories, and even individual academics. As a result, bibliographic control is now more dynamic and shared, allowing for the reuse, enrichment, and cross-platform linking of material. However, because of their extensive experience with standards and dedication to dependability and customer service, libraries continue to play a significant role.

9.4 Concept of indexing: Type, Application, Subject Heading and Keyword

Indexing is the process of examining a document's content and condensing it (sometimes using phrases or codes) to facilitate retrieval. To put it simply, indexing provides an answer to the query, "What is this document about?" Users can quickly locate pertinent documents without having to read them all thanks to an effective index.

9.4.1 Types of Indexing

Traditional subject heading systems include significant pre-coordinate components, whereas keyword indexing is typically post-coordinate. These days, a lot of systems incorporate elements from both paradigms.

- Assigned vs. Derived Indexing

Index terms used in derived indexing are extracted straight from the material, frequently from headings, abstracts, or titles. One such example is keyword indexing, which uses words from the text or title as entry points. In assigned indexing, the indexer selects terms based on the subject content from a regulated vocabulary, such as a thesaurus or a set of subject headings.

Since derived indexing uses text that is already present in the document, it is typically simpler to automate. Although assigned indexing can result in more accurate and consistent subject access, it necessitates intellectual examination by a qualified indexer.

- Pre-coordinate indexing and post-coordinate

Complex subject strings are created during pre-coordinate indexing, frequently combining many elements (e.g., "Children – Education – India – History"). Such pre-structured headers are produced by classification-based analysis systems such as chain indexing, PRECIS, and POPSI. The catalogue or index then files these headings in a certain order.

Simple terms (uniterms) are assigned during indexing in post-coordinate indexing, and the user combines them during search. For instance, a document may be indexed with distinct terms such as "Children," "Education," "India," and "History," and the user may utilize Boolean operators (AND, OR, NOT) in a search interface to mix these terms. Post-coordinate indexing is a key component of many contemporary online databases and catalogues.

9.4.2 Application of Indexing

Indexing supports numerous vital roles in a library or digital environment:

- It gives subjects access to records in a database or catalog.
- It enables users to obtain documents related to a subject even if they are not familiar with certain authors or titles.
- It makes browsing more meaningful by grouping related products together.

9.4.3 Computerised Indexing

9.4.3.1 Meaning and Features

Computerised indexing is the use of computer technology in the subject indexing process, where computers are used to assist in the analysis of documents, formulation of index terms, storage of index files, and retrieval of information. Computerised indexing is faster and more economical than manual indexing and is necessary in digital libraries.

The major features of computerised indexing are:

- Speed in the formation of indexes and searching
- Handling large databases
- Consistency in the formulation of index terms
- Automatic, semi-automatic, and interactive indexing
- Easy modification and reorganisation of index files

9.4.3.2 Manual Indexing vs. Computerised Indexing

Manual indexing is totally dependent on the human judgment of the indexer. It is slow and time-consuming. It is only suitable for small collections of documents.

Computerised indexing, on the other hand, uses computer software, algorithms, and databases to assist or automate the indexing process. It is faster and more economical in the long run than manual indexing. It is necessary in digital libraries.

In manual indexing, consistency in the formulation of index terms may vary from indexer to indexer. In computerised indexing, consistency in the application of rules is ensured if the system is properly designed.

9.4.3.3 Advantages and Disadvantages of Computerised Indexing

Advantages are:

- Enables fast indexing and retrieval
- Efficiently processes large data sets
- Reduces repetitive human labor
- Enables online searching and remote access
- Aids in the integration of library management systems

Disadvantages are:

- Involves high initial investment in hardware and software
- Requires expert human labor for operation and maintenance
- Computerized automatic indexing may not be semantically precise
- Output quality is dependent on input quality and algorithms used

9.4.3.5 Components of Computerised Indexing System

A computerised indexing system has the following components:

- Input Component – document input, document analysis, and term selection
- Processing Component – indexing rules, algorithms, and software processing
- Database Component – storage of indexed records and terms
- Output Component – display of search results and index entries
- All the components interact to achieve effective information organisation and retrieval.

9.4.3.6 Categories of Computerised Indexing Systems

Computerised indexing systems can be categorised as follows:

- Automatic Indexing – index terms are produced completely by the computer without human intervention
- Semi-automatic Indexing – computer proposes terms, but final decision is made by the indexer
- Interactive Indexing – continuous interaction between indexer and system during indexing

The categories are mainly distinguished by the level of human participation.

9.4.3.7 Methods of Computerised Indexing

The following are the most important methods of computerised indexing:

- Keyword indexing (KWIC, KWOC)
- Automatic term extraction from text
- Assignment indexing using controlled vocabularies
- Hybrid methods that combine human knowledge and computerised indexing

These methods are widely used in bibliographic and full-text databases.

9.4.3.8 Indexing Internet Resources

- Search Engine Indexing

Search engine indexing is the process of using web crawlers or spiders to browse web pages and build searchable indexes. Pages are indexed according to keywords, metadata, hyperlinks, and content structure.

This method allows users to quickly access information from the web, although the quality of results may vary as there is no subject control.

- Subject / Information Gateway

Subject gateways are carefully compiled lists of internet resources, arranged by subject specialists. Unlike search engines, they emphasize quality over quantity.

The most important features are:

- Resources are selected and evaluated
- Subject-based arrangement
- Useful for academic and research work

- Semantic Web

The Semantic Web is an effort to make web information machine-readable. It employs data structure, metadata, and ontologies to enhance the accuracy of information retrieval.

By allowing computers to understand relationships between concepts, the Semantic Web enables intelligent searching and data integration.

- Taxonomies

Taxonomies are hierarchical systems of classification that are used to organize information from general to specific levels. They are systematic approaches to browsing and retrieval.

Taxonomies are used extensively in:

- Digital libraries
- Knowledge management systems
- Institutional repositories and websites

They enhance consistency and navigation in large information systems.

9.5 Subject Heading

A document's subject is represented using standardized linguistic terms called subject headings. They are typically taken from a list of subject headings that specify each heading's permitted form, scope, and connections to other terms.

9.5.1 One major models of subject heading lists is:

Sears List of Subject Headings:

The first list was compiled by Minnie Earl Sears in 1923 as List of Subject Headings for Small Libraries. It was compiled from headings already in use in nine good small libraries, and it introduced “See” and “refer from” references for the first time. Later issues introduced “See also” references and a teaching chapter entitled “Practical Suggestions for the Beginner in Subject Heading Work”, which was later renamed “Principles of the Sears List of Subject Headings”.

Many editors have since revised the list. Dewey Decimal numbers were introduced, the title was changed to Sears List of Subject Headings, and “for Small Libraries” was removed so that medium libraries could also use it. Symbols x and xx replaced long “refer from” phrases. Later issues used computers to verify headings, met the needs of online catalogues, reduced complicated compound headings and inverted headings, and introduced headings for fiction and children’s materials. Recent issues have updated religion headings, changed “Indians” to Native Americans, and reduced “modern” headings such as Modern history

9.5.2 Formation of Subject Headings

SLSH forms subject headings in various ways.

Single word headings: One noun, such as Art, Agriculture, Education, Religion. Homographs (same word, different meaning) are followed by a scope word, such as Seals (Animals), Seals (Law). Abstract concepts and actions are singular (Credit, Beauty); concrete objects are plural (Books, Flowers).

Phrase headings: Two or more words.

- Compound headings: Two nouns connected by "and," such as Religion and science, Joy and sorrow.
- Adjectival headings: Adjective + noun, such as American literature, Tropical fish.
- Prepositional phrase headings: Connected by a preposition, such as Freedom of information, Electricity in agriculture.

Complex subjects with subdivisions: A main heading and subdivisions.

- Topical: Clothing and dress – History; Religion – Philosophy.
- Form: Chemists – Directories; Geology – Maps.
- Chronological: United States – History – 1775–1783, Revolution.
- Geographic:
    - Subject by place: Agriculture – India; Theatre – Paris.
    - Place by topic: United States – Census; Tibet – Climate.

9.5.3 Entry Format

Since the fifteenth edition, the SLSH follows a thesaurus format. The top heading is in bold, while the cross-reference headings are in regular type. All entries and references are placed on the left side of the page, leaving the right-side blank for local headings and references, which turn the book into a local authority file.

9.5.4 Cross References

The former "see" and "see also" references and x/xx were eliminated and replaced with thesaurus symbols: UF/USE, BT, NT, RT, SA.

- UF (Used for): lists non-preferred terms that refer to the selected heading, such as synonyms, inverted headings, or "and" headings.
- USE: indicates to the cataloger which heading to use instead of another term.
- BT, NT, RT, SA indicate broader, narrower, related, and "see also" terms for searching.

9.5.5 Standard Vocabulary

The SLSH is a standard vocabulary for small and medium libraries. It provides one preferred term for each concept and uses references from other forms, which helps to ensure that headings are consistent and easy to understand.

9.5.6 Filing Order

For subdivided headings, Sears uses the suggested standard order:

[Topical] – [Geographic] – [Chronological] – [Form]. For example, Elderly – Housing – United States is preferred over "Elderly – United States – Housing", and Sports – United States – Statistics over "Sports – Statistics – United States". This order also indicates that if a heading already has a form or chronological subdivision, a place is added before these, for example Physicians – Ohio – Directories.



Check Your Progress

a) What is bibliographic control in the context of a library?
b) How is derived indexing different from assigned indexing?
c) What is the primary purpose of applying the Sears List of Subject Headings (SLSH)?

9.5 Let Us Sum Up

Libraries provide access to recorded knowledge, and therefore they depend on bibliographic control to indicate what exists, where it is, and what it is about. This involves descriptive cataloging, subject cataloging, authority control, and maintaining up-to-date searchable catalogs. In the modern digital age, multiple versions of the same work are often linked together through shared metadata, and therefore high-quality metadata is critical. Indexing assists in searching by answering the question "What is this document about?" through derived or assigned terms, which can be done by computers for efficiency. Subject heading aids such as the Sears List offer straightforward headings and a logical filing order.

9.6 Answers to Check Your Progress

a) It's the group of processes that ensures information resources are described, classified, and searchable so you can access and use them when needed. This involves descriptive and subject cataloging, authority control, and maintaining catalogs and indexes that are easily searchable.

b) Derived indexing derives index terms directly from the document itself—such as from the title or abstract. Assigned indexing, on the other hand, involves a trained indexer selecting terms from a controlled vocabulary (such as a thesaurus or subject headings list) based on the document's subject content.

c) The Sears List provides a simple, standardized set of subject headings, particularly useful for smaller libraries, to ensure that each idea has only one preferred term and that subject access remains consistent and understandable.

9.7 Further Readings

1. SUBJECT INDEXING MODELS. (n.d.). In *egyankosh*. Retrieved January 31, 2026, from https://egyankosh.ac.in/bitstream/123456789/33117/1/Unit-16.pdf

2. Akidi, J. O. (2020). BIBLIOGRAPHIC CONTROL PRACTICES IN THE DIGITAL AGE: CONCEPTUALMODEL AND THEORETICAL FRAMEWORK PERSPECTIVE. *Library Philosophy and Practice*. https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=7434&context=libphilprac%2F1000
3. Bergamin, G., & Guerrini, M. (2022). *Bibliographic Control in the Digital Ecosystem*. Firenze University Press.
4. Guerrini, M. (2022). ["Universal bibliographic control in the digital ecosystem: opportunities and challenges", "Universal bibliographic control in the digital ecosystem: opportunities and challenges"]. *DOAJ (DOAJ: Directory of Open Access Journals)*. https://doi.org/10.4403/jlis.it-12742

9.8 Probable Questions

1. Why has high-quality metadata become so important to bibliographic control in the digital age?
2. State any two advantages and any two disadvantages of computerised indexing.
3. Explain pre-coordinate and post-coordinate indexing with suitable examples, and describe how modern systems combine elements of both methods.
4. Describe the key features of the Sears List of Subject Headings as a model of subject heading list, including the construction of headings, entry format, cross-references, and filing order.

UNIT 10: METADATA AND METADATA STANDARDS IN DIGITAL LIBRARIES

UNIT STRUCTURE

10.1Learning Objectives
10.2Introduction
10.3Metadata: Concept, Types and Functionalities
10.4Metadata Standards: MARC, Dublin Core (DC), METS, etc.
10.5Protocol For Metadata Harvesting (PMH)
10.6Open Archives Initiatives (OAI Model)
10.7Let US Sum Up
10.8Answers To Check Your Progress
10.9Further Readings
10.10 Probable Questions

10.1Learning Objectives

After studying this unit, learners will be able to:

- Define metadata and explain its importance in digital libraries.
- Identify and describe different types of metadata.
- Discuss the functionalities of metadata in resource description and retrieval.
- Explain major metadata standards: MARC, Dublin Core (DC), METS, etc.
- Describe the Open Archives Initiative (OAI) and Protocol for Metadata Harvesting (OAI-PMH).
- Understand how metadata interoperability supports digital library services

10.2Introduction

In the digital era, libraries increasingly depend on structured data to organize, retrieve, preserve, and share information resources. Metadata often described as *"data about data"* plays a central role in enabling effective resource discovery, interoperability, and digital preservation.

Digital libraries, institutional repositories, and archival systems use metadata standards such as MARC, Dublin Core, and METS, along with protocols like OAI-PMH, to ensure uniformity and efficient exchange of bibliographic information across platforms. This unit introduces the concept, types, and functions of metadata, examines major metadata standards, and explains the Open Archives Initiative and the Protocol for Metadata Harvesting (OAI-PMH).

10.3Metadata: Concept, Types and Functionalities

10.3.1Metadata: Concept

Metadata refers to the structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.
In simple terms, metadata answers the questions:

- What is the resource?
- Who created it?
- When and how was it created?
- What is its format and subject?

In digital libraries, metadata acts as a bridge between the user and the digital object, enabling efficient search, organization, and long-term preservation.

10.3.2 Types of Metadata

Metadata is generally categorized into three major types:

1. Descriptive Metadata

- Used for resource discovery and identification.
- Includes elements such as title, author, subject, keywords, abstract.
- Examples: Dublin Core, MARC.

2. Structural Metadata

- Indicates how different components of a resource are organized.
- Useful for complex digital objects such as e-books, image collections, or multimedia.
- Example: METS, which defines the structure of digital objects.

3. Administrative Metadata

- Provides technical information required for managing a resource.
- Includes file type, creation date, access rights, preservation metadata, etc.
- Sub-types include:
    - Technical metadata

- Rights metadata
- Preservation metadata (e.g., PREMIS)

10.3.3 Functionalities of Metadata

Metadata performs several essential functions in digital information management. These include:

1. Resource Discovery – Helps users find relevant information through keywords, subjects, and titles.
2. Identification and Authentication – Provides unique identifiers for digital objects (e.g., DOI).
3. Organization and Classification – Allows systematic arrangement of resources.
4. Interoperability – Facilitates sharing of metadata across different systems (e.g., through OAI-PMH).
5. Digital Preservation – Stores information necessary to maintain and preserve digital resources.
6. Rights Management – Specifies copyright, license, and access conditions.
7. Navigation – Helps users navigate multi-part digital documents through structural metadata.

10.4 Metadata Standards: MARC, Dublin Core (DC), METS

Metadata standards ensure consistency, accuracy, and interoperability across information systems. They define the format, structure, and elements to be used.

10.4.1MARC (Machine-Readable Cataloging)

MARC (Machine-Readable Cataloging) is a standardized format used for representing, storing and exchanging bibliographic and related information in machine-readable form. Developed in the 1960s by Henriette Avram at the Library of Congress, MARC made it possible for libraries to automate cataloging processes and share bibliographic data efficiently across library networks. The primary objective of MARC is to establish a common structure for cataloging records so that they can be understood and processed by computers, facilitating resource sharing, cooperative cataloging, and OPAC-based retrieval. MARC 21 is the most widely used version, which includes several formats such as Bibliographic, Authority, Holdings, Classification, and Community Information records.

A MARC record consists of three main components: the Leader, Directory, and Variable Fields. The Leader is a 24-character fixed field that contains essential processing information. The Directory indicates the location of each field within the record. The Variable Fields are divided into Control Fields (00X range), which do not contain indicators

or subfields, and Data Fields (01X–999), which contain indicators and subfields identified by delimiters such as $a, $b, and $c. MARC records use tag numbers to represent specific data elements, such as 020 for ISBN, 100 for the personal author, 245 for the title statement, 260 or 264 for publication details, 300 for physical description, 650 for subject headings, and 700 for additional authors.

MARC plays a crucial role in modern libraries by enabling standardization, reducing duplication of work, supporting automation, and promoting wide-scale resource sharing through networks like OCLC and DELNET. However, the system is complex and requires professional training, and maintenance can be costly. With the development of new technologies, MARC is gradually being complemented and potentially replaced by BIBFRAME, which is based on linked data and semantic web standards. While MARC is a linear text-string format designed for traditional library systems, BIBFRAME enables more dynamic web-based data connections. Despite emerging alternatives, MARC continues to be the most widely used global standard for cataloging and bibliographic data exchange.

- Developed by the Library of Congress.
- Provides a standardized format for representing bibliographic data in machine-readable form.
- Uses fields, subfields, indicators, and tags (e.g., 100 for author, 245 for title).
- Widely used in libraries for cataloging printed and digital resources.
- Variants include MARC 21, UNIMARC, and CNMARC.

Key features:

- Supports detailed cataloging.
- Ensures interoperability across library management systems.
- Backbone of traditional library cataloging.

10.4.2 Dublin Core (DC)

Dublin Core (DC) is a simple and widely used metadata schema designed to describe digital resources such as webpage, documents, images, videos, and other electronic objects. It was developed in 1995 at a workshop held in Dublin, Ohio, USA, hence the name “Dublin Core,” and is maintained by the Dublin Core Metadata Initiative (DCMI). The primary goal of Dublin Core is to provide a standardized and easy-to-use method for resource description that supports interoperability across different information systems. Dublin Core consists of a set of 15 core metadata elements known as the Simple Dublin Core, which include Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, and Rights. Each element is optional and repeatable, allowing flexibility depending on the resource being described. Dublin Core can also be extended for more complex description, known as Qualified Dublin Core, which adds

refinements and encoding schemes to enhance specificity, such as Date.Created, Date.Modified, or spatial coverage data.

- Simple and widely used metadata standard for digital resources.
- Consists of 15 core elements, such as Title, Creator, Subject, Description, Publisher, Date, Format, etc.
- Designed for simplicity and broad applicability.

Advantages:

- Easy to understand and implement.
- Suitable for institutional repositories, digital libraries, and open access archives.
- Extensible through Qualified Dublin Core.

10.4.3 METS (Metadata Encoding and Transmission Standard)

METS (Metadata Encoding and Transmission Standard) is an XML-based framework developed and maintained by the Library of Congress for encoding and packaging descriptive, administrative, and structural metadata related to digital objects. It is widely used in digital libraries, archives, and institutional repositories for managing and preserving complex digital materials such as manuscripts, images, audio, video, and e-documents. A METS file contains several major components, including the METS Header, Descriptive Metadata, Administrative Metadata, File Section, Structural Map, Structural Links, and Behavior Section, which together describe the content, technical details, relationships, and access mechanisms of a digital object. METS supports interoperability and long-term preservation and is often used with standards like MODS and PREMIS. Although flexible and powerful, METS is complex and requires knowledge of XML and metadata standards.

- XML-based standard developed by the Library of Congress.
- Designed to encode descriptive, administrative, and structural metadata for complex digital objects.
- Often used in digital preservation and large-scale digitization projects.

Components of METS include:

1. Descriptive Metadata – Can embed Dublin Core or MARC.
2. Administrative Metadata – File technical, rights, and preservation metadata.
3. Structural Map – Indicates organization of files within the digital object.
4. File Section – Lists all digital files.

METS acts as a wrapper or container for different metadata types.

10.5 Protocol for Metadata Harvesting (PMH)

The OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) is a protocol that allows the harvesting or collecting of metadata from various repositories.

Protocol for Metadata Harvesting (PMH), commonly known as OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting), is a low-barrier protocol developed by the Open Archives Initiative to facilitate the harvesting and sharing of metadata across digital repositories, libraries, archives, and scholarly communication platforms. Introduced in 2001, OAI-PMH enables interoperability between systems by allowing service providers to collect metadata from different data providers (repositories) using standardized HTTP and XML-based requests. The protocol supports metadata formats such as Dublin Core, making it widely compatible and easy to implement. OAI-PMH defines six key harvesting operations called verbs—Identify, List Metadata Formats, List Sets, Get Record, List Identifiers, and List Records—which help retrieve metadata in bulk or individually. It plays a vital role in open access scholarly communication by enabling centralized search services like institutional repositories, digital libraries, and open-access databases to aggregate metadata from multiple sources. OAI-PMH improves resource discoverability and reduces duplication of cataloging efforts, making it an essential protocol for modern digital information environments.

10.6 Open Archives Initiatives (OAI Model)

10.6.1OAI-PMH Model

OAI-PMH is based on two main components:

1. Data Providers

- Repositories that store metadata and expose it through OAI-PMH.
- Example: DSpace, EPrints, institutional repositories.

2. Service Providers

- Harvest metadata from multiple repositories to build services such as search engines or discovery portals.
- Example: OAIster, BASE, OpenAIRE.

10.6.2 OAI-PMH Verbs (Commands)

OAI-PMH uses six verbs:

1. Identify – Information about the repository.
2. ListMetadataFormats – Available metadata formats (e.g., DC).
3. ListSets – Set structure of the repository.
4. ListIdentifiers – Returns identifiers of items.
5. ListRecords – Returns full metadata records.

6. GetRecord – Retrieves a single metadata record.



Check Your Progress

A, Fill in the Blanks

1. Metadata is frequently characterized as "____ about ___".
2. The three primary categories of metadata are _____, Structural, and Administrative.
3. _____ is a standardized schema established by the Library of Congress for encoding bibliographic information in a machine-readable format.
4. The basic metadata format comprising 15 fundamental elements such as Title, Creator, and Subject is referred to as _____.

10.7Let US Sum Up

This unit discussed the concept, types, and functions of metadata in digital libraries. Metadata provides description, structure, and administrative information about resources, enabling efficient discovery, management, and long-term preservation. Major metadata standards such as MARC, Dublin Core, and METS support interoperability and uniformity in library systems. The Open Archives Initiative and OAI-PMH facilitate open access and metadata sharing across repositories, making them fundamental components of modern digital libraries.

10.8 Answers to Check Your Progress

A. Fill in the Blanks

a. data, data
b. Descriptive
c. MARC
d. Dubline Core

10.9 Further Readings

1. Gilliland, A. (2016). *Introduction to Metadata*. Getty Publications.
2. Library of Congress. (2023). *MARC 21 Format for Bibliographic Data*.
3. DCMI. *Dublin Core Metadata Initiative*.

4. Library of Congress. *METS: Metadata Encoding and Transmission Standard*.
5. Open Archives Initiative. (2023). *Protocol for Metadata Harvesting*.

10.10 Probable Questions

a. Define metadata and explain its importance in digital libraries.
b. Distinguish between descriptive, structural, and administrative metadata.
c. Explain the major features of the MARC 21 format.
d. What are the 15 core elements of the Dublin Core metadata standard?
e. Describe the structure and purpose of METS.
f. What is OAI-PMH? Discuss its six verbs.
g. Explain the difference between data providers and service providers in the OAI model.

UNIT 11 DATA IN LIBRARY

UNIT STRUCTURE

11.1 Learning Objectives
11.2 Introduction
11.3 Concept of data, information, Knowledge and wisdom,
11.4 Types of data and data resources
11.5 Concept of Big data in libraries and information centre
11.6 Open data: characteristics and standards,
11.7 Linked data in library
11.8 Let Us Sum Up
11.9 Answers to Check Your Progress
11.10 Further readings
11.11 Probable Questions

11.1 Learning Objectives

After going through this unit, you will be able to

- Define data, information, knowledge and wisdom, DIKW Model
- Explain types of data
- Learn about data resources and its use in libraries
- Define Big Data and its use in libraries and information centers
- Learn about open data, its characteristics and standards
- Learned about linked data in libraries

11.2 Introduction

We live in an era of information explosion. Every second, every minute new information of various forms is getting generated from every hook and corner, all over the world through the internet, social media, research and technology. Data has become an integral part of our lives. On the other hand, libraries are treasure troves of knowledge; they have huge collections of data in different formats for their users. Along with the rapid technical development of the world, libraries are working hard to level up along with the rising demands of the users. For better services and proper dissemination of information to their users, libraries need to adopt new technologies and ways for the execution of their vision of providing better services to its users. It has become very important for libraries, as centers of knowledge, to learn about data for handling the fast-growing world of information. This unit introduces the concept of data, its types, explores new concepts like Big data, Open data, Linked data and how libraries are using these to move beyond traditional concepts.

11.3 DIKW Model: Concept of data, information, Knowledge and wisdom

- **Data:** The word data came from a Latin word ‘datum’ meaning to give. Data simply means raw, unprocessed facts and figures with no meaning. They are raw information like facts, numbers, words, symbols or details that are collected for a purpose. Example: Numbers, symbols, factual information etc.

- **Information:** The word information came from a Latin word ‘informatio’ meaning a formation. Information means processed or organized data that has meaning. It is the organized data that has meaning and helps us understand something. Example: statistical records, a weather report, a phone number or address etc.

- **Knowledge:** The word Knowledge came from the Latin word ‘cognoscere’ meaning to know or to understand. Knowledge means the state of knowing about any particular fact, a form of understanding, familiarity or awareness about any specific fact or situation. Example: Practical knowledge like knowing how to use Microsoft Excel, theoretical knowledge like why water boils at 100 degree Celsius etc.

- **Wisdom:** The word wisdom came from the Latin word ‘sapiential’ meaning being wise or good sense. The meaning of wisdom is simply the ability to use knowledge correctly to make good decisions and judgments. It is not only about knowing facts but the ability to know what is correct or incorrect and how to use or apply the knowledge in a proper way. Example: Saving money instead of spending, learning from mistakes, helping others without expecting anything back.

11.4 Types of data and data resources

There are different types of data based on structure, nature and source. They are as follows:

11.4.1 Based on structure

- **Structured data:** Structured data are those data organized in proper format that is fixed like rows, columns, spreadsheets. Structured data make it easier for computers for reading, storing and searching as it uses a fixed schema. Examples: Spreadsheets, student records, bank transaction details etc.

- **Semi Structured data:** This type of data is partially organized, has flexible structure and doesn’t follow any table format instead containing some tags or labels for organization of information. Examples: Web pages, XML, JSON etc.

- **Unstructured data:** Unstructured data is data that has no fixed format or organization. These data are in free form and are harder to store and analyze and need special tools like text mining or AI to extract meaning. Examples: Photos, videos, PDF files, audio recordings etc.

11.4.2 Based on nature

- **Quantitative data:** Quantitative data or numerical data are those data that can be measured and analyzed statistically. These data are expressed in numbers and are collected through instruments, tests or counting. Example: data about age, height, no of students etc.

- **Qualitative data:** Qualitative data or non-numerical data are those data that are descriptive in nature, expressed only in words. This type of data describes qualities, characteristics or attributes and are often collected through surveys, interviews or observations. Example: Different tastes of food, various types of fruits etc.

11.4.3 Based on source

- **Primary data:** Primary data is that type of fresh data that is originally collected by the collector, researcher or organization directly from the source for a specific purpose. It is collected through methods like questionnaires, interviews, observations, experiments etc. The collection of these types of data is very time consuming and expensive. Example: Conducting a survey, interviewing people for certain data etc.

- **Secondary data**: Secondary data are those data that are already available or have been already collected by someone else for some purposes. These data are very easy to obtain, are cheaper than primary data and can be reused for analysis. Example: Research papers, information from books, government reports etc.

11.4.4 Data resources in libraries

Data resources are simply sources of information from which data can be collected and can be used or stored like books, newspapers, government reports etc. Libraries are the centers for information and they provide access to a wide range of data resources to the users in both printed form and digital form. Libraries make it very easy for users to find and use these data resources for learning and research purposes. Examples of data resources are books, journals, newspapers, magazines, library collections, digital archives, online databases like Google Scholar, JSTOR etc.

11.5 Concept of Big data in libraries and information centre

Big Data are those humongous sets of data that are impossible to handle or analyze by traditional data processing tools like Excel or simple databases. This data gets generated very quickly and continuously in different forms that can only be managed by advanced technologies like Hadoop or cloud storage. Examples of Big data are data from social media platforms like Facebook, Instagram, YouTube, health data from hospital records or fitness apps as well, data from online shopping websites etc.

11.5.1. Characteristics of Big data or the 5 Vs of Big data

- **Volume:** Volume refers to the big amount of data being generated every second. Example: Social media platforms store extreme amounts of photos, videos and messages.

- **Velocity:** Velocity refers to the speed at which data is getting created, collected and processed. Example: The users in social media are posting every second in various forms.

- **Variety:** Variety simply refers to the various ways or forms of data collected from many sources. Example: text, images, videos, audios etc.

- **Veracity:** Veracity refers to the reliability, accuracy of data. A huge amount of data comes from various sources and it might not be correct or accurate so it is important to check if it is trustworthy or not. Example: Information on social media might display fake information, a sensor might record wrong values due to any technical issue etc.

- **Value:** Value simply refers to the usefulness or importance of data or the benefit gained from analyzing the huge amount of data that helps in making better decisions or improvements. Example: Health data used by hospitals to improve patient care, a company analyzes customer data to provide better products and services etc.

11.5.2. Use of Big Data in libraries

With the rapid development of the world, libraries need to level up their services according to the demands of the users. In this era of information explosion, it becomes very important to collect and analyze information in a proper way. The use of Big data helps libraries to upgrade their services, give personalized services to the users. It really helps by collecting and analyzing huge amounts of information about users, books and digital resources to elevate, improve and offer better library services. Big data helps in understanding user behavior, helps in tracking materials often used by users, enhances digital resources and improves better decision making and planning.

11.6 Open data: characteristics and standards

The meaning of Open data is that the data that is available freely for anyone for use, access, modification, reuse and redistribute without any legal or technical restrictions. Open data is

published online in accessible format so that anyone can use it for their works or purposes. Example: Maps and Geographic information (GIS), government census data and statistics etc.

11.6.1. Characteristics of Open Data

- **Free access:** Open data is freely available online without any legal restrictions and can be accessed by anyone. It can be easily downloaded in usable format like CSV, JSON, XML.
- **Reusability:** Open data can be reuse, redistribute, modify, and share etc by anyone for personal, academic or commercial purposes.
- **Free of charge:** Mostly Open data is free of charge for everyone or sometimes a minimal cost of reproduction is required.
- **Machine readable format:** It is available in machine readable format that computers can easily read and process.

11.6.2. Standards of Open Data

- Open Format Standard like Comma Separated Values (CSV), JavaScript Object Notation (JSON), Extensible Markup Language (XML) and Resource Description Framework (RDF)
- Metadata Standard
- Licensing Standard like Creative Commons (CC-BY, CC) and Open Data Commons (ODC-ODbL, PDDL)
- Interoperability Standard
- Quality Standard
- Accessibility Standard
- Security and privacy standard

11.7 Linked data in library

The meaning of Linked Data is simply a method used for publishing structured data on the web in such a way that it can be linked or interconnected with other sources of data leading to the easy access and use by people and computers as well. Linked data follows a specific type of web standards so that various datasets can interact with each other forming a web of data. It connects related information using common standards with different websites and databases so that data becomes more useful and meaningful. Libraries contain a big amount of information like bibliographic and metadata information and linked data helps libraries to make this

information more connected on the web making it discoverable, enhances metadata, supports the semantic web, data integration and interoperability.

11.7.1. Technologies used in Linked Data

- Uniform Resource Identifier (URI)
- Resource Description Framework (RDF)
- Web Ontology Language (OWL)
- SPARQL Protocol and RDF Query Language (SPARQL)
- Linked Open Data (LOD)
- Simple Knowledge Organisation System (SKOS)
- JavaScript Object Notation for Linked Data (JSON-LD)



Check Your Progress

A. Fill in the Blanks

a. Open data can be freely reused and _____.
b. RDF stands for _____.
c. The highest level of the DIKW hierarchy is _____.
d. Unstructured data is data that has no fixed _____.
e. Big data helps in understanding _____.

11.8 Let Us Sum Up

In the contemporary age of "information explosion," where data is essential to existence, libraries are transforming from conventional "treasure troves" into modern knowledge centers that embrace new technology to satisfy increasing user expectations. This transformation depends on comprehending the DIKW paradigm, which advances from raw "Data" to significant "Information," recognized "Knowledge," and utilized "Wisdom," while also managing various data types classified by structure, nature, and source. To manage the vast influx of information, libraries are utilizing "Big Data," defined by the 5 Vs (Volume, Velocity, Variety, Veracity, and Value), to monitor user behaviour and customize services. Moreover, they are adopting "Open Data," which guarantees unrestricted access and reusability, and "Linked Data," which employs web standards to interlink structured material, enabling libraries to transcend conventional paradigms and enhance discoverability for their patrons.

11.9 Answers to Check Your Progress

a. Redistributed, **b.** Resource Description Framework**, c.** Wisdom**, d.** Organization**, e.** User Behaviour

11.10 Further readings

1. Safdar, Muhammad.,Ur Rehman, Shafiq., Arif, Muhammad & Ashiq, Murtaza (2022). Research data services in libraries: a systematic literature review. *Information Discovery and Delivery, 51*(4)
2. Kamupunga, W.,& Chunting, Yang (2019). Application of Big Data in Libraries. *International Journal of Computer Applications, 178*(16).
3. Ayre, Lori Bowen.,& Ganer, Jim(2017). Open Data: What it is and why you should care. *Public Library Quarterly, 36*(2), 173-184
4. Bizer, Christian., Heath, Tom., & Lee, Tim Berners (2009). Linked Data: The Story So Far. *International Journal on Semantic Web and Information Systems, 5*(3), 1-22
5. Baskarada, Sasa & Koronios, Andy (2013). Data, Information, Knowledge, Wisdom (DIKW): A Semiotic Theoretical and Empirical Exploration of the Hierarchy and its Quality Dimension. *Australasian Journal of Information Systems, 18*(1)

11.11 Probable Questions

A. Short Answer Type questions

1. Explain Data. What are the different types of data?
2. What do you mean by Big Data? Explain the characteristics of Big Data.
3. What is the difference between structured, semi-structured and unstructured data?

B. Long Answer Type Questions

1. Explain how Big data is helping libraries.
2. What is the concept of Open Data? Write three characteristics of Open Data.
3. What are the technologies used in Linked Data?
4. Name some Open Data Standards.
5. What is the role of Linked Data in libraries?
6. Write the difference between data, information, knowledge and wisdom with examples.

UNIT 12 DATA, STORAGE, SECURITY AND DATA REPOSITORIES

UNIT STRUCTURE

12.1 Learning Objectives
12.2 Introduction
12.3 Data storage
12.4 Data protection; security, sharing and ethics
12.5 Data repository; types, functions, resources and software for data repositories
12.6 Let Us Sum Up
12.7 Answers to Check Your Progress
12.8 Further Readings
12.9 Probable Questions

12.1 Learning Objectives

After this unit, learners will be able to

- Define the basic concepts of data storage in digital libraries and the best practices associated with different types of data storage.
- Define the crucial role of data protection, security, and ethics in digital information management.
- Define the concept of data repositories and classify them, along with the primary software used for their management.
- Discuss the challenges and future trends in digital library security.

12.2 Introduction

The digital revolution has completely changed the way we handle information. For so long, the library was a place of wonder, a place that was full of physical things: books, manuscripts, journals, all lined up on shelves. But now, the true prize of a modern library may not be visible at all. Data is the true prize of a modern library, and as we continue to move further into the information age, digital information is growing at an exponential rate. The library is no longer just a shelf full of paper; it is a complex digital environment that collects, interprets, stores, and disseminates massive amounts of information.

The transition from physical objects to binary code presents new challenges. While in the physical world, protecting a book meant keeping it dry and away from fire, in the digital world, protecting information means fighting off invisible cyber-attacks, preventing bit rot that can damage files over time, and controlling access. A “digital library” is a complex computer-based information system designed to collect, organize, and deliver digital information to users, whether on a local area network or over the Internet.

This unit examines the technological and moral foundations of this new world. We will begin with the fundamentals of data storage, how information is actually stored on a computer and in

the devices that hold it. Next, we will navigate the crucial landscape of data security, examining the dangers that threaten libraries and the measures that must be taken to safeguard user privacy and intellectual property. Finally, we will examine data repositories, the specialized archives that fuel modern research, facilitating the sharing and preservation of scholarly data for future generations.

12.3 Data storage

Storage of data is the basic building block on which digital libraries are based. It is more than just the process of storing a file; it involves the long-term sustainability of the data as well. In a computer system, all data, whether textual, graphical, auditory, or visual, is finally converted into a universal form that is a series of binary patterns (0s and 1s).

12.3.1 Types of Storage

In the real world, information manifests itself in the form of books, lectures, paintings stuff we can touch and feel. But in the world of computers, all these need to be converted to one common language: binary.

Bits and Bytes: At the most basic level, everything is represented as a series of bits—0s and 1s. These bits are grouped together to form bytes (8 bits) and are the basic units of measurement for data.

Text Representation: Text data is represented by assigning a unique binary code to each character. ASCII (American Standard Code for Information Interchange) uses 7 or 8 bits to represent English letters and symbols, while Unicode uses 16 bits or more to represent characters from almost all languages.

Image Representation: Images are broken down into a matrix of tiny dots called pixels. Each pixel represents a binary number that holds information about its color and intensity. High-resolution images require millions of bits to represent all that data.

Audio and Video: Audio signals are represented as an analog wave, which needs to be sampled at a rate fast enough to convert it into digital form (a process called quantization). Video combines these audio samples with a rapid series of digital images to produce the illusion of motion.

12.3.2 Storage Media and Devices

Digital libraries use a variety of hardware for storing their binary data, and these devices are normally categorized based on their speed, capacity, and the level of permanence.

1) Internal Storage (Primary Memory):

This is the computer's fast work area.

- **RAM (Random Access Memory):** Fast memory chips that store data the CPU is actively using. It's volatile, so power loss means data disappears.

- **ROM (Read-Only Memory):** Permanently set instructions (like the boot sequence) that cannot be altered.

2) Magnetic Storage:

A traditional mainstay of digital storage, which uses magnetic patterns on a surface to encode bits**.**

- **Hard Disk Drives (HDD):** Made up of magnetic platters that spin as a head accesses them. They provide large storage capacities at a low price, making them the norm for library databases and server files.
- **Magnetic Tapes:** Slower than disks but very inexpensive and long-lasting. Often used for backups and long-term storage storing massive amounts of data that don't need to be accessed immediately (Cold Storage).

3) Optical Storage:

Lasers are used to read and write data on plastic discs.

- **CD-ROM (Compact Disc Read-Only Memory):** Holds 650 MB of data the equivalent of hundreds of books. Played a crucial role in the early digital age for distributing static databases and multimedia content.
- **DVD (Digital Versatile Disc):** Holds between 4.7 GB and over 17 GB of data, becoming an absolute necessity for video storage and larger software packages.

4) Solid State and Flash Storage:

- **SSDs (Solid State Drives):** Flash memory-based storage that replaced spinning disks. Much faster and more reliable than HDDs, though often pricier per gigabyte.
- **USB Drives:** Portable flash memory sticks that have replaced floppy disks for the transfer of personal data. Data protection; security, sharing and ethics

With the growing amount of digital data in libraries, they are increasingly vulnerable to cyber threats. Data protection refers to the process of securing data against compromise, loss, or damage**.**

12.3.3 Storage Architectures

Libraries do not depend on a single hard drive; they operate on complex systems to manage data on a large scale.

- **Network-Attached Storage (NAS**): It is a dedicated file server that is connected to the library network, allowing many users to access the same files simultaneously.
- **Cloud Storage:** Many libraries are migrating their data to the cloud, which is a remote server managed by companies such as AWS or Google. This approach provides immense scalability and disaster recovery, but it also raises concerns about data ownership and possible long-term costs.

12.4 Data Protection; Security, Sharing, and Ethics

With the rise of digital libraries, the library is faced with the security challenges that come with the digital world. Information Security is the protection of information against unauthorized access, use, disclosure, disruption, modification, or destruction. This is not just a technical issue; it is a fundamental responsibility to maintain the trust of library users.

12.4.1 Cybersecurity Threats

The library faces a variety of cyber threats, ranging from nuisance attacks to serious criminal attacks.

- **Malware:** This is a generic term for malicious software, including viruses, worms, and spyware. Ransomware is particularly pernicious, as it involves hackers encrypting the library's data and then demanding payment in exchange for the decryption key.
- **Phishing:** These are deceptive emails that masquerade as trusted entities (such as a bank or a boss) in an attempt to trick library staff into divulging their passwords or other sensitive information.
- **Data Breaches:** These occur when unauthorized people access confidential information. One such incident occurred in 2014 at the University of Maryland Libraries, where hackers gained access to a database containing 300,000 records of students and faculty.
- **DDoS Attacks (Distributed Denial of Service):** These involve hackers flooding the library's servers with traffic from multiple sources, causing the website or online catalog to become unavailable and preventing legitimate users from accessing library services.

12.4.2 Security Measures and Standards

To protect against threats, libraries should adopt a Défense in Depth strategy, integrating technical, physical, and administrative security measures.

1) Technical security measures:

Encryption: Encrypt data both at rest (on storage media) and in transit (over networks). In this way, even if data is stolen, it cannot be read without the decryption key.

Access control: Implement strict user authentication, including the use of strong passwords and Multi-Factor Authentication (MFA) – a password and a second code (sent to a phone, for example) to gain access.

Backups: Implement the 3-2-1 rule: three copies of data, stored on two different types of media, with one copy offsite. In this way, data can be restored after an event such as a ransomware attack or a fire.

2) Physical security measures:

Surveillance: Install CCTV cameras in server rooms and public areas to prevent and detect the theft of devices.
RFID: Install RFID tags on books and equipment to prevent unauthorized removal and to track inventory.

3) Administrative security measures (policies and standards):
Libraries should adopt international standards to inform their security policies:
ISO/IEC 27001: The international standard for Information Security Management Systems (ISMS), which describes risk management framework.
ISO/IEC 27002: A code of practice providing guidance on security controls.
ISO 15489: Emphasizes records management, ensuring that digital records are managed and disposed of securely.

12.4.3 Ethics and Data Sharing

Ethics and Data Sharing in libraries occupy a very sensitive point at which the aim is to achieve open access. This is a matter of balance between security and serving the community, and this creates very real questions.

Privacy and analytics: Libraries collect data about how people use the collection to refine what is being offered, such as identifying what books are in trend. However, there is a need to anonymize the data in such a way that no one can track their reading history.

Copyright and IPR: Digital libraries contain copyrighted materials. They use Digital Rights Management to limit the problem of copying, while still allowing fair use of the material for researchers.

Informed consent: When research data is shared, it is important for libraries to ensure that any human subjects have given their consent to share their data.

12.5 Data Repository; Types, Functions, Resources and Software

Data Repository is basically a digital library carefully designed to store, manage, preserve, and share datasets. Unlike conventional digital libraries that showcase completed works like books and research papers, data repositories focus on the unprocessed data that supports the findings of research papers.

12.5.1 Types of Data Repositories

Data repositories are commonly classified depending on their focus and target audience:

Institutional Repositories (IR):

These are controlled by an institution (like a university) to store the intellectual output of its own community. A university can use an IR to store theses, dissertations, and research data

files created by faculty and students, thus capturing the intellectual output of the institution and increasing its reputation.

Disciplinary (Subject) Repositories:

These are subject-based and not institution-based. Researchers from across the world contribute to these. Examples include GenBank (genetic data) and Dryad (general scientific data). Researchers like these because they make data visible to the community of people most likely to reuse the data.

Project-Specific Repositories:

Large-scale research projects (like the Human Genome Project) often establish project-specific repositories to handle the complex data flows that are characteristic of their projects.

12.5.2 Functions of a Data Repository

A good data repository is more than just a hard drive. It goes through a cycle of manual management steps:

Ingest: Data intake. This step checks file types, looks for viruses, and ensures that the person submitting the data has the legal right to submit it.

Metadata Management: Data is useless if it can't be located. Depositors enter Metadata—information about the data itself—such as title, author, date, and subject terms, often using standards such as Dublin Core.

Preservation (Archiving): The essential long-term task. This involves techniques such as migration (converting, for instance, older formats like Word 95 files to PDF/A) and replication (copying data to different locations) to ensure data readability for many decades.

Access Control: Not all data should be accessible to all. Data repositories allow researchers to restrict access to authorized users only, or to set embargos (private for a year while papers are written).

Discovery and Citation: Data repositories assign a unique identifier, such as a DOI (Digital Object Identifier), so that data can be cited in academic papers just like a book, giving proper credit to the original author.

12.5.3 Software for Data Repositories

These data repositories are usually hosted on software platforms by libraries. Open-source software is particularly in vogue because of its flexibility and economy.

DSpace: This software was born out of collaboration between MIT and HP. It is the most widely used software for institutional repositories. It is particularly useful for text-heavy

repositories like theses, papers, and so on. Its organizational structure (Community -> Collection -> Item) mirrors that of a university.

CKAN (Comprehensive Knowledge Archive Network): This is a full-featured data management system used by many governments worldwide for Open Data portals (like data.gov). It is designed with datasets in mind and has in-browser data visualization capabilities (charts, maps). For instance, the IIM Ahmedabad library used CKAN to build its research data repository.

Dataverse: This software was developed at Harvard and is specifically designed for research data. It is particularly useful for tabular data (spreadsheets) and has superior citation management capabilities. It has “federation” capabilities that allow multiple university Dataverses to be linked together.

Fedora Commons: This software is designed for complex digital archives that have unusual or complex digital objects. It is often used as the storage layer for other software systems.

EPrints: This software has been around since the days of the University of Southampton and is widely used for open-access publication repositories.



Check Your Progress

a) What is data repository?
b) What are some common cyber security threats?
c) What are optical storage devices?

12.3 Let Us Sum Up

In this unit, we have traversed the technological backbone of the modern digital library. We began with Data Storage, the foundation of the digital edifice. We learned that all digital information is ultimately a sequence of binary bits, stored on a hierarchy of media ranging from high-speed RAM and SSDs to high-capacity magnetic tapes for archival preservation. We discussed the importance of storage architectures like NAS and Cloud storage in providing scalable and reliable access to these resources. We then moved to the critical area of Data Security. As libraries become digital custodians, they face sophisticated threats like ransomware, phishing, and data breaches. We explored the "Defense in Depth" approach, which combines technical tools (encryption, firewalls), physical security (CCTV, RFID), and administrative policies (ISO standards) to safeguard user privacy and institutional assets. Finally, we examined Data Repositories, the specialized engines of the open science movement. We distinguished between institutional and disciplinary repositories and outlined

their core functions: ingest metadata management, preservation, and access control. We also surveyed the landscape of open-source software like DSpace, CKAN, and Dataverse that enables libraries to build these vital infrastructures.

12.4Answers to Check Your Progress

a) Data Repository is basically a digital library carefully designed to store, manage, preserve, and share datasets. Unlike conventional digital libraries that showcase completed works like books and research papers, data repositories focus on the unprocessed data that supports the findings of research papers.

b) Malware, Phishing, Data Breaches, DDoS Attacks (Distributed Denial of Service)

c) Optical storage devices are a form of secondary storage that uses the laser technology to read and write data on a plastic disc coated with a special layer. The data is represented in the form of microscopic pits and lands on the disc surface, and the laser beam reads it without touching the disc. Examples of optical storage devices include the CD-ROM (Compact Disc Read-Only Memory), which can store about 650 MB of data, and the DVD (Digital Versatile Disc), which can store data ranging from 4.7 GB to over 17 GB of data.

12.5 Further Readings

1. *COAR*. (n.d.). COAR. https://coar-repositories.org/
2. *Data Storage*. (n.d.). Retrieved January 30, 2026, from https://web.ntnu.edu.tw/~tcchiang/IntroCS/2_Data%20Storage%20(I).pdf
3. Geng, L. (2015). The Research of Digital Library Mass Information Storage System Architectur. *International Symposium on Computers & Informatics* .
4. *Library Guides: Research Data Management: Data Storage*. (2026). Du.edu. https://libguides.du.edu/Research_Data_Management/Data_Storage?utm_source=chatgpt.com
5. Magnus Osahon Igbinovia, & Bolaji David Oladokun. (2024). Information Security in Libraries, Librarianship, and Information Science. *Elsevier EBooks*. https://doi.org/10.1016/b978-0-323-95689-5.00162-0
6. *Research Guides: Data Management and Sharing Plans: Data Security & Ethics*. (2025). Wm.edu. https://guides.libraries.wm.edu/dmp/data-security?utm_source=chatgpt.com
7. Roeschley, A. (n.d.). *Defining Data Ethics in Library and Information Science*.

8. Roeschley, A., & Khader, M. (n.d.). *Functional Requirements for Research Data Repositories*.
9. Saha, R. (2024). Data Privacy and Cyber Security in Digital Library Perspective: Safe Guarding User Information Rudrani Saha. *INTERANTIONAL JOURNAL of SCIENTIFIC RESEARCH in ENGINEERING and MANAGEMENT*, *08*(04), 1–5. https://doi.org/10.55041/ijsrem30761

12.6 Probable Questions

a) Discuss the "Defense in Depth" strategy for protecting digital libraries. What specific technical, physical, and administrative controls are involved?
b) Describe two open-source software systems that are frequently used to create powerful data repositories and outline their main features.
c) Distinguish "Institutional Repositories" from "Disciplinary Repositories."
d) Discuss Cyberattacks andits various types.

UNIT 13 RESEARCH DATA MANAGEMENT

UNIT STRUCTURE

13.1 Learning Objectives
13.2 Introduction
13.3 Data Management
13.4 Concept of research data management
13.5 RDM software and service: Free and Open source
13.6 Let Us Sum Up
13.7 Answers to Check Your Progress
13.8 Further Readings
13.9 Probable Questions

13.1 Learning Objectives

After this unit, learner will be able to:

- Learn about the concept and significance of research data management (RDM) in academic institutions.
- Identify the stages of the research data lifecycle and the activities involved in managing data effectively.
- Discuss the role of libraries and librarians in supporting RDM services.

13.2 Introduction

Innovation and societal advancement are fueled by research, which is the foundation of academic and scientific advancement. Data, the raw material from which knowledge is created, is at the core of any research project. In the past, research data was frequently kept on external hard drives, personal laptops, or even physical notebooks, which made it difficult to exchange and susceptible to loss. However, the amount and complexity of research data have skyrocketed with the introduction of Big Data and digital technology. The area of research data management (RDM) emerged as a result of this change, which required a more organized method of managing data.

Traditionally the guardians of knowledge, academic libraries are now playing a crucial role in overseeing this new type of intellectual output. In order to make sure that research data is arranged, stored, and made available for future verification and reuse, they are going beyond books and journals. The principles of RDM, the lifecycle of research data, and the instruments and tactics that organizations can use to efficiently manage their data assets are all covered in this subject.

13.3 Data Management; data collection, data collection tools

The processes of gathering, storing, and utilizing data in a safe, effective, and economical manner are all included in the large field of data management. It is applicable to all types of organizational data, not only research. Effective data management guarantees that data is securely stored, readily retrieved, and accessible for replication or confirmation of results.

Data management in the context of an academic library entails the ongoing statistical documentation of several library departments, including resource acquisition, circulation, and user interactions. But the scope has grown considerably. It now contains the policy and technical frameworks needed to manage digital assets across the course of their lifecycle. This covers every step of the process, from the original planning of data creation to its long-term storage or disposal. Data loss is avoided, data integrity is guaranteed, and adherence to legal and ethical requirements is made easier with effective data management.

13.4 Concept of research data management

Data management for Research, on the other hand, is a niche of data management dedicated to data generated during scholarly research. Data management has been defined as the active management and appraisal of data over the lifecycle of scholarly and scientific interest. RDM is not only about archiving data at the end of a project; it is an ongoing process beginning before the research actually starts.

Research data lifecycle

To better understand Research Data Management (RDM), it is helpful to look at the research data lifecycle, which describes the stages in which data progress. A common way of describing this is to break the process down into the following stages:

- Creating Data: This stage involves planning (through Data Management Plans), designing data collection tools, and collecting data. At this stage, decisions need to be made regarding file formats, file naming conventions, and organizational Schemas.
- Processing Data: After data collection, data entry, digitization, translation, or transcription takes place. Data are then validated and cleaned to remove errors.
- Analyzing Data: The cleaned data are analyzed to reveal meaningful research results. This stage often results in the creation of research outputs, such as publications.
- Preserving Data: Data are migrated to formats that are suitable for long-term preservation, backed up, and archived, ensuring that they remain accessible in the future.
- Giving Access to Data: This stage involves data sharing, managing access through licensing or copyright, and enabling data reuse.
- Reusing Data: Data are reused by other researchers or by the original researchers for further research or educational activities.

Importance of RDM:

RDM is very important for a number of reasons:

- **Verification:** It allows other researchers to verify their findings, hence increasing the validity of the research.
- **Reuse:** Data can be reused for other research, hence saving time that would have been used to collect the data.
- **Visibility:** The data can increase the visibility of the research.
- **Compliance:** Many funding bodies require Data Management Plans (DMPs) and make it mandatory to make the data publicly available.

13.5 RDM software and service: Free and Open source

For effective Research Data Management (RDM), institutions rely on software solutions that enable the storage, organization, and sharing of research data. While there are commercial solutions available, institutions are increasingly turning to Free and Open-Source Software (FOSS) solutions that are cost-effective and can be customized.

13.5.1 CKAN (Comprehensive Knowledge Archive Network)

CKAN is a powerful open-source data portal solution used by governments and research organizations around the world. It is intended to facilitate the publication, dissemination, discovery, and use of data.

Features: The solution allows dataset uploading with metadata enrichment and data visualization using graphs and maps. It also provides a comprehensive API that makes it easier to integrate with other systems.

Use Case: The Vikram Sarabhai Library at IIM Ahmedabad has successfully used a research data repository solution based on CKAN. The repository serves as a single point of access for research data at the institute, facilitating preservation and reuse.

13.5.2 DSpace

While DSpace is more commonly known as an institutional repository solution for publications, it is also widely used as a research data solution.

Features: It supports a wide variety of file types and has strong preservation tools. The solution also has strong support from the community, ensuring that it is regularly updated with a variety of plugins that can be used to improve data management functionality.

Suitability: It is suitable for institutions that are already using DSpace for publications and would like to manage data within the same environment.

13.5.3 Dataverse

Dataverse is an open-source web application developed by Harvard University for sharing, preserving, citing, exploring, and analyzing research data.

Features: It provides professional citation for each dataset, thus encouraging researchers to share their data. It also provides fine-grained control over data access permissions.

Community: It has a strong global community and is used by many top-class universities.

13.5.4 EPrints

Similar to DSpace, EPrints is a traditional player in the repository arena. Although initially developed for text-based research outputs, it has been extended to handle complex data objects.

Features: It is very flexible and can be customized to handle specific data types and metadata formats as demanded by different disciplines.

13.5.5 Zenodo

Zenodo is a cloud-based service and not a self-hosted software package, although it is open-source (based on Invenio software). It is free to use by researchers and is developed by CERN.

Features: It allows research outputs in all scientific fields and assigns a Digital Object Identifier (DOI) to all uploaded files, making them citable.



Check Your Progress

a) Give an example of an open-source program for building data repositories.
b) What is the Dataverse platform's salient feature?

13.7 Let Us Sum Up

In this unit, the scope of Research Data Management (RDM) has been explored. It started with the assumption that data management involves the secure and efficient management of data to ensure its long-term value. The conversation then progressed to the particular concept of RDM, which deals with the data life cycle in research workfrom the point of initial creation and processing to analysis, preservation, and finally, reuse.It is claimed that RDM is more than a technical challenge; it is a strategic imperative for institutions of higher learning. RDM improves research quality, collaboration, and meets the ever-tightening requirements of funding agencies. The libraries are seen at the forefront of this effort, transforming from passive book storage facilities to active participants in the research process.The discussion ends with a review of the available technology that supports RDM. Open-source solutions such as CKAN, DSpace, and Dataverse offer effective and affordable solutions that allow institutions to build their own data repositories. By embracing these approaches and technologies, institutions of higher learning can ensure that their research data is a valuable resource for future generations.

13.8 Answers to Check Your Progress

a) DSpace

b) Dataverse is an open-source web tool for exchanging, storing, referencing, examining, and evaluating research. It encourages academics to publish their data by giving each dataset a professional citation. Additionally, it offers fine-grained control over permissions for data access.

13.9 Further Readings

1. Dora, M., & Kumar, A. (2015). Managing Research Data in Academic Institutions: Role of Libraries. *CALIBER*. https://www.researchgate.net/publication/278021825_Managing_Research_Data_in_Academic_Institutions_Role_of_Libraries
2. Flores, J., Brodeur, J., Daniels, M., Nicholls, N., & Turnator, E. (n.d.). *Libraries and the Research Data Management Landscape*. Retrieved January 30, 2026, from https://www.clir.org/wp-content/uploads/sites/9/RDM.pdf
3. Md. Ashikuzzaman. (2024, March 15). *Data Management for Libraries : The Role of Libraries in Data Management*. Library & Information Science Education Network; Library & Information Science Education Network. https://www.lisedunetwork.com/data-management-for-libraries-the-role-of-libraries-in-data-management/#google_vignette
4. Nongo, Celina & Udofot, Columbus & Alabi, Comfort. (2022). DATA MANAGEMENT IN ACADEMIC LIBRARIES: EMBRACING INNOVATIVE QUALITY ASSURANCE TEMPLATES IN AN ERA OF CHANGE IN INFORMATION SERVICES.
5. *Research Data Management for librarians*. (n.d.). Retrieved January 30, 2026, from https://www.dcc.ac.uk/sites/default/files/documents/events/RDM-for-librarians/RDM-for-librarians-booklet.pdf
6. Yoon, A., & Schultz, T. (2017). Research Data Management Services in Academic Libraries in the US: A Content Analysis of Libraries' Websites. *College & Research Libraries*, *78*(7). https://doi.org/10.5860/crl.78.7.920
7. Zotoo, I. K., & Liu, G. (2019). Research Data Management (RDM) Strategy for Academic Libraries in Ghana: Setting a National Development Agenda. *OALib*, *06*(04), 1–24. https://doi.org/10.4236/oalib.1105383

13.10 Probable Questions

a) Explain Research Data Management (RDM) and why it is considered a strategic focus for today's academic institution.

b) Outline the Research Data Lifecycle. List the key activities that are performed at each phase of the research data lifecycle.

c) Compare and contrast two open-source software tools that can be used for managing research data. Which software tool would you recommend for a university library, and why?

UNIT 14: EMERGING TRENDS IN LIBRARY DATA MANAGEMENT

UNIT STRUCTURE

14.1 Learning Objective

14.2 Introduction

14.3 Artificial Intelligence

14.4 Machine Learning

14.5 Blockchain

14.6 Cloud Solution in Data Management

14.7 Data Privacy

14.8 Protection and Cyber Security Trends

14.9 New Data Governance Strategies for Libraries

14.7 Data privacy

14.8 Protection and cyber security trends

14.9 New Data Governance Strategies for Libraries

14.10 Let Us Sum Up

14.11 Answers to Check Your Progress

14.12 Further Readings

14.13 Probable Questions

14.1 Learning Objectives

After studying this unit, learners will be able to:

- explain the basic ideas of artificial intelligence, machine learning, blockchain, and cloud solutions in the context of library data management
- describe current trends in data privacy, protection, and cybersecurity for digital libraries and electronic resources.
- discuss emerging data governance strategies that define roles, policies, and standards for library data.

14.2 Introduction

Libraries are now working with very large volumes of digital data that come from catalogues, institutional repositories, subscription databases, learning management systems, and user interaction logs. This data is valuable because it supports teaching, research, and evidence-based decision making, but it also creates new responsibilities for librarians as data managers and data stewards. In recent years, technologies such as artificial intelligence, machine learning, blockchain, and cloud computing have begun to reshape how libraries organize, store, secure, and govern this data. At the same time, cyber-attacks on major libraries and the rapid growth of electronic resources have made data privacy, protection, and security core issues in library management.

Emerging trends in library data management therefore cut across technology and policy. They involve intelligent tools for discovery, analytics, and automation; secure and scalable infrastructures in the cloud and stronger frameworks for privacy, cybersecurity, and data governance. Libraries are moving from seeing data only as something to store, to treating it as a strategic asset that must be curated, protected, and used ethically throughout its life cycle. In this unit, learners will explore these developments and consider how library professionals can respond to technical, ethical, and organisational challenges in a practical way.

14.3 Artificial Intelligence

Patrick Henry Winston in 1999 had stated artificial intelligence is the subfield of computer science concerned with understanding the nature of intelligence and constructing computer systems capable of intelligent action.

Artificial intelligence is a branch of computer science that aims to build systems capable of tasks that normally require human intelligence, such as learning, reasoning, pattern recognition, and language understanding. In libraries, AI systems rely heavily on large datasets, including bibliographic records, usage logs, full text content, and user profiles, to automate decisions and improve services. AI in this environment is usually implemented through components like natural language processing, recommender systems, expert systems, and intelligent search interfaces.

The use of AI in libraries can improve information retrieval. Traditional catalogues and databases mainly match keywords, but AI based tools can support semantic search, where the system tries to understand the meaning of a query and its relationship with documents. This allows users to retrieve more relevant items even when they do not know the exact controlled vocabulary or subject headings. AI can also power question answering systems that respond in natural language, which is especially useful for novice users who feel uncomfortable with complex search strategies.

Another important application is personalization. AI algorithms can analyse patterns in a user's previous searches, downloads, and borrowing behaviour, and then suggest books, articles, or datasets that may be of interest. In current awareness and SDI services, AI can automatically track new publications that match a researcher's profile and send timely alerts, making the

service more precise and efficient. This kind of personalization can increase usage of resources, but it also raises questions about how much user data should be collected and how it should be protected, themes that will appear again in later sections.

Libraries are also using AI to create chatbots and virtual assistants on library websites. These chatbots can answer basic reference queries, provide guidance on how to use the OPAC or digital library, and offer information about rules, timings, and services, at any time of the day. For example, a chatbot can help a remote learner locate e books on a topic, explain how to access a database off campus, and remind them of due dates or fines. Studies in higher education show that AI chatbots can improve responsiveness and reduce routine workload for staff, provided that complex queries are still referred to human librarians.

AI also supports back-end data management tasks. Tools can assist in automatic subject indexing, name authority control, duplication of records, and detection of anomalies in usage statistics. Visual analytics tools driven by AI can reveal patterns in collection use, collaboration networks among researchers, or emerging research themes, which can guide collection development and data driven planning. At the same time, library professionals must be aware of biases in training data and need to ensure that AI tools do not reinforce inequalities or exclude minority languages and user groups.



Check Your Progress

Q1. Mention any two ways in which AI powered chatbots can support library users.

Q2. Why is user data protection an important concern when libraries apply AI for personalized services?

14.4 Machine Learning

Machine learning is a subset of AI that focuses on algorithms which learn patterns from data and make predictions or decisions without being explicitly programmed for each case. In library data management, ML techniques are especially useful where there are large, complex datasets, such as usage logs from e resource platforms, circulation data, web analytics, and research data repositories. These algorithms can uncover trends and relationships that are difficult to detect manually, and they can adapt as new data arrives.

One major area is usage analytics. By applying clustering and classification methods, ML can segment users by behaviour, discipline, or level of engagement, and identify under used collections or services. Predictive models can forecast demand for certain databases or journals, informing subscription decisions and helping libraries to manage budgets more effectively. In

institutional repositories, ML can support automatic metadata extraction from full text, such as identifying author names, affiliations, keywords, and funding agencies.

ML also plays a growing role in cybersecurity for academic libraries. ML based systems can monitor network traffic patterns, detect anomalies, and flag potential intrusions, malware, or suspicious login activity much faster than manual monitoring. For instance, if there is a sudden spike in failed logins from unfamiliar locations, or unusual download patterns from a single user account, ML algorithms can recognise these as possible threats and trigger alerts or automated responses. This is valuable when protecting digital collections and subscription resources from abuse and data leakage.

In user services, ML supports recommender systems similar to those used in e-commerce, where the system learns from the behaviour of many users to suggest documents that others with similar interests have used. ML models are also being explored for automatic classification of documents into subject categories, sentiment analysis of user feedback, and even identification of research fronts from citation networks. These tools can improve the visibility and discoverability of library resources, and provide richer analytics to administrators and faculty.

However, ML requires high quality, well curated datasets, and the models must be transparent and explainable when they influence access to information or assessment of users. There are risks of bias if training data does not represent all user groups or subject areas fairly, which can lead to unequal recommendations or misclassification. Libraries need to combine technical rigour with ethical oversight, including clear policies on what data is collected, how long it is retained, and how it is anonymised before being used for model training.

14.5 Blockchain

Blockchain is a distributed ledger technology where records are grouped into blocks that are cryptographically linked and stored across many nodes in a network. Once information is added to a block and validated through consensus mechanisms, it is extremely difficult to alter, which provides strong guarantees of integrity and provenance. Early uses of blockchain centred on cryptocurrencies, but later generations of the technology support smart contracts and many other applications beyond finance.

Blockchain is a distributed ledger technology where records are grouped into blocks that are cryptographically linked and stored across many nodes in a network. Once information is added to a block and validated through consensus mechanisms, it is extremely difficult to alter, which provides strong guarantees of integrity and provenance. Early uses of blockchain centred on cryptocurrencies, but later generations of the technology support smart contracts and many other applications beyond finance

In scholarly communication, blockchain has been proposed as a way to create verifiable, time stamped versions of research outputs, support transparent peer review, and record contributions to collaborative work. For example, a journal article or dataset could be registered on a blockchain with its metadata and hash, making it easy to check that a given copy is authentic

and has not been tampered with. Some experiments show that blockchain can help to audit scientific studies and preserve evidence of research processes, which may be of interest to academic libraries concerned with research integrity.

There are also library specific prototypes such as LibChain, a blockchain based distributed library management system. In such systems, lending transactions, including e lending and inter library loans, are recorded on a shared ledger, allowing libraries to cooperate in resource sharing without depending on a single central server. Blockchain based digital rights management models have been suggested to handle permissions for e books and other digital content, potentially offering more flexible loan models while maintaining clear rights information.

At the same time, blockchain has important limitations. Public blockchains can raise privacy issues, because all nodes hold copies of the ledger, and strong encryption and access controls are needed to protect sensitive data. The technology can also be resource intensive and may be complex to integrate with existing library systems. Therefore, many current discussions focus on permissioned blockchains with restricted participation, and on specific use cases where the benefits of immutability and decentralisation outweigh the costs.

14.6 Cloud Solution in Data Management

Cloud computing provides on demand access to shared computing resources such as servers, storage, databases, and software over the internet. In library data management, cloud solutions have become central for hosting integrated library systems, discovery platforms, digital repositories, and storage for large digital collections. Instead of running all services on local servers, libraries can use infrastructure as a service, platform as a service, or software as a service models offered by commercial or consortium providers.

Cloud based library systems can reduce capital expenditure on hardware and allow more flexible scaling as usage grows, which is particularly helpful when digital collections and user numbers increase. They also support collaborative services, such as shared catalogues and regional repositories, by allowing multiple institutions to access the same platform. For example, cloud-based repository services can host institutional research data with tools for submission, metadata management, and long-term preservation, while relieving individual libraries from some technical burdens.

However, moving data and services to the cloud raises questions about data ownership, jurisdiction, privacy, and security. Professional guidelines stress that libraries should retain ownership of their data, ensure that vendors comply with relevant privacy and data protection laws, and specify how data is stored, processed, and shared. Contracts should cover issues such as data portability at the end of a contract, backup and disaster recovery, and incident response in case of security breaches. For sensitive datasets, libraries may adopt hybrid models where some components remain on premises while others move to the cloud.

Securing cloud-based library data requires encryption of data at rest and in transit, careful configuration of access controls, and continuous monitoring for vulnerabilities. Libraries need

to work closely with IT units and cloud providers to implement standards like ISO 27001 and related information security frameworks, which define best practices for managing risks. Staff training is also vital, because misconfigurations and weak account management are frequent causes of cloud incidents.



Check Your Progress

Q3. State two advantages of cloud-based library systems for data management.
Q4. Mention any two security measures that should be applied to library data in the cloud.

14.7 Data Privacy

IBM states “data privacy is the principle that a person should have control over their personal data, including deciding how organizations collect, store and use it.”

Data privacy in libraries refers to the protection of personal information about users, such as identity details, borrowing histories, search logs, and usage patterns of electronic resources. Historically, libraries have defended the right of users to read and seek information without unjustified monitoring, and professional codes of ethics emphasise confidentiality of records. With the shift to digital services and detailed transaction logging, the amount of personal data collected has increased, making privacy protection more complex and more important.

Modern privacy challenges include tracking by third party vendors, cross platform analytics, and integration of library systems with campus wide identity management and learning analytics platforms. Libraries must decide what data is collected by default, how long it is kept, and for which purposes it can be used. International standards and regulations, such as ISO 27701 and data protection laws like the General Data Protection Regulation of the European Union (GDPR), encourage practices such as data minimisation, purpose limitation, and user consent.

Library users should be informed clearly about privacy policies, that they have options to control how much personal data is collected, and that they can review or request deletion of their data where possible. Libraries are also encouraged to de identify or aggregate usage data for reporting and research, so that individuals cannot be easily re identified. As AI and analytics tools are introduced, privacy preserving techniques such as differential privacy and federated learning can help analyse behaviour while keeping raw data local or protected.

14.8 Protection and Cyber Security Trends

Cybersecurity deals with protecting information systems, networks, and data from unauthorised access, damage, or disruption. Academic and research libraries have become attractive targets for cyber-attacks because they host valuable digital collections, subscription resources, and sensitive user and research data. Recent incidents, including attacks on major

national and academic libraries, show that disruptions can affect online catalogues, repositories, reading room services, and staff operations for long periods.

Main cyber threats faced by libraries

- Data breaches: unauthorised access to personal data, research data, or licensed content.
- Malware and ransomware: malicious software that damages, steals, or encrypts data and demands payment.
- Phishing and social engineering: deceptive emails or messages that trick staff or users into revealing passwords or clicking harmful links.
- Denial of Service (DoS/DDoS): attacks that overload servers and make catalogues or repositories unavailable.
- Insider threats: misuse of access by staff, contractors, or students, whether intentional or accidental.

Use of structured security frameworks

- Many libraries follow standards such as ISO/IEC 27001 for systematic risk assessment, control selection, and periodic review of security measures.
- Frameworks like the NIST Cybersecurity Framework help organise work into identify, protect, detect, respond, and recover, ensuring both prevention and incident handling.

Technical protection measures

- Network security: deployment of firewalls, secure configuration of routers/switches, and intrusion detection or prevention systems.
- Patch and vulnerability management: regular updates of operating systems, library applications, and middleware to close known security holes.
- Backup and recovery: frequent, tested backups, including offline or off-site copies, to restore systems after ransomware or hardware failure.

Access control and identity management

- Strong authentication: wider use of multi factor authentication for staff and administrators, especially for remote and privileged access.
- Role based access control: defining roles so each user or system only sees the data needed for their work (principle of least privilege).
- Network segmentation: separating public, staff, and critical server networks so that compromise in one segment does not spread easily.

Intelligent, data driven defence

- Log and traffic analysis: collecting system logs and network data for continuous monitoring and forensic investigation.
- Anomaly detection: using analytical or machine learning tools to spot unusual behaviour, such as abnormal download patterns or login locations.
- Automated responses: predefined actions (for example, temporarily locking an account or isolating a server) when suspicious patterns are detected.

Human factors and security culture

- Awareness training: regular short programmes to teach staff and users how to recognise phishing, use strong passwords, and handle data safely.
- Clear procedures: simple, well-known steps for reporting suspicious emails, strange system behaviour, or lost devices.
- Emphasis in developing countries: where budgets are tight, low-cost education and basic hygiene (passwords, updates, backups) can greatly improve protection.

Legal and ethical dimensions

- Compliance with data protection laws: meeting obligations under regulations such as the GDPR, including appropriate security and, where required, breach notification.
- Documentation: policies, risk assessments, and records of controls to demonstrate due care and support audits or investigations.
- Ethical balance: protecting personal data and confidentiality while avoiding excessive monitoring or creating unfair barriers to information access.

14.9 New Data Governance Strategies for Libraries

Data governance is the set of policies, roles, processes, and standards that guide how data is managed, shared, and protected across an institution. In libraries, data governance now goes beyond bibliographic control to include research data management, analytics data, user data, and records about electronic resource usage and licences. Effective governance aims to ensure that data is accurate, secure, usable, and aligned with legal and ethical requirements.

Recent work stresses that library data professionals should be key partners in institutional data governance structures. Their experience with metadata standards, preservation, access, and user rights positions them to advocate for transparency and equity in data policies. Governance models increasingly include roles such as data owners, data stewards, and data custodians, with clear responsibilities and lines of accountability. In a university setting, this may mean library representation on data governance committees that oversee research data, learning analytics, and administrative data.

Best practice guidelines for data governance recommend clear data policies and standards, definition of data quality rules, and documentation of data flows between systems. Libraries can contribute by defining standards for metadata, retention schedules for different categories

of data, and access rules for sensitive datasets such as thesis drafts, embargoed research, or usage logs. Governance strategies should also address equity and inclusion, ensuring that marginalised communities are represented in decision making about data and that data practices do not reinforce existing biases.

In practical terms, new governance strategies may include establishing institutional research data policies, setting up data repositories with clear licensing options, and creating guidelines for sharing data with external partners, including commercial vendors and consortia. Libraries often take the lead in research data management training, helping researchers design data management plans and understand obligations related to funder policies and data sharing mandates. An integrated approach to data governance will help libraries balance openness, reuse, and innovation with the need to protect privacy and maintain trust.



Check Your Progress

Q5. _____is the set of policies, roles, processes, and standards that guide how data is managed, shared, and protected across an institution.

Q6. What is data privacy?

14.10 Let Us Sum Up

AI and machine learning are increasingly used to improve information retrieval, personalization, usage analytics, and cybersecurity. Blockchain adds possibilities for tamper evident records, stronger rights management, and collaborative systems such as distributed library management.

Cloud solutions underpin library systems and repositories by offering scalable, cost effective platforms, with careful attention to contracts, standards, and security arrangements so that data ownership and confidentiality are maintained. Across all these technologies, libraries need to handle data privacy responsibly, especially for user records and electronic resource usage, in line with ethical codes and legal regulations, while new cybersecurity approaches emphasise coherent frameworks, robust technical controls, and ongoing education to protect digital collections and services from evolving threats. New data governance strategies place libraries at the heart of institutional debates about how data is collected, documented, shared, and preserved, and active participation in governance structures, policy development, and research data management enables library professionals to keep technological change aligned with enduring values of intellectual freedom, equity, and trust.

14.11 Answers to Check Your Progress

Ans to Q1: First, AI powered chatbots can answer routine reference questions and guide users through the library website, OPAC, and digital library at any time, helping especially remote learners. Second, they can provide information about rules, services, and account status, such as due dates and fines, which reduces routine workload for library staff.

Ans to Q2: User data protection is important because personalized AI services depend on collecting and analysing detailed information about individuals' reading, searching, and learning behaviours, which can reveal sensitive aspects of their interests and identities. If this data is mishandled, shared without consent, or breached, it can undermine user trust and conflict with library ethical commitments and data protection laws.

Ans to Q3: Cloud based library systems reduce the need for local hardware and allow services to scale easily as collections and users grow, which is useful for institutions with limited infrastructure. They also support shared platforms for catalogues and repositories, enabling collaboration and cost sharing among libraries and improving access for users.

Ans to Q4: Two necessary security measures are encryption of data at rest and in transit, and strong access controls such as multi factor authentication and role-based permissions for staff and systems. Regular patching, vulnerability management, backup, and monitoring of cloud configurations are also essential to reduce the risk of breaches.

Ans to Q5: Data governance

Ans to Q6: Data privacy is the principle that a person should have control over their personal data, including deciding how organizations collect, store and use it.

14.12 Further Readings

1. Ajo Geevarghese Thomas, & Neethu Mohanan. "A Descriptive Study on Digital Innovations and Technologies in Libraries." Rajagiri College of Social Sciences, Kalamassery.[ppl-ai-file-upload.s3.amazonaws]
2. Ayooluwa Aregbesola, & Ekene Lawrence Nwaolise. "Securing Digital Collections: Cyber Security Best Practices for Academic Libraries in Developing Countries." Landmark University, Omu Aran.[ppl-ai-file-upload.s3.amazonaws]
3. Persadha, P. D., Judijanto, L., Susanti, M., & Reza, H. K. "Data Privacy and Security Protection Strategies in Library Electronic Resources Management." Holistik Analisis Nexus, Vol. 1, No. 7, 2024.ppl-ai-file-upload.s3.amazonaws+1
4. American Library Association. "Library Privacy Guidelines for Library Management Systems." 2016.[ala]
5. Gaikwad, M., & Bilawar, P. "Effects of Emerging Technologies on Library Services: Artificial Intelligence, Machine Learning, Cloud Computing, and Blockchain."[journals.lcu.edu]

6. “The Role of Library Data Professionals in Data Governance.” ScholarWorks, Indiana University–Purdue University Indianapolis, 2022.[scholarworks.indianapolis.iu]

14.13 Probable Questions

Fill in the blanks

1. _____ search uses the meaning of a query rather than only exact keywords to retrieve relevant documents.
2. In library data management, _____ learning algorithms analyse large datasets to identify patterns and make predictions without being explicitly programmed.
3. A distributed, tamper evident ledger of transactions maintained by a network of nodes is known as _____ technology.
4. In cloud-based services, libraries should ensure that they retain _____ of all data stored and processed by vendors.
5. Protecting users’ personal information such as borrowing records and search histories is an important aspect of data _____ in libraries.

Short answer type questions

1. Briefly describe two applications of artificial intelligence in academic libraries.
2. What are the main privacy challenges created by the use of detailed usage data in digital libraries?
3. State two key elements of a good data governance framework for library data.

Long answer type questions

1. Discuss the role of artificial intelligence and machine learning in transforming library data management.
2. Explain the potential applications of blockchain in libraries for acquisitions, rights management, and preservation. What are the main advantages and limitations of using blockchain.
3. Critically examine the challenges of ensuring data privacy, protection, and cybersecurity in digital libraries.