MLIS S1-02

ONE YEAR MASTER'S DEGREE COURSE

(CBCS in Accordance with NEP 2020)

SEMESTER - I

E RESOURCE MANAGEMENT AND RDM IN LIBRARIES

LIBRARY AND INFORMATION SCIENCES

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

Subject Experts

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

Dr. Dipen Deka, DLIS, Gauhati University, Guwahati

Dr. Deepjyoti Kalita, DLIS, Cotton University, Guwahati

Course Co-ordinator: Komal Kumari, KKHSOU

SLM Preparation Team

UNITS	CONTRIBUTORS
1, 2,3,4, 5,6 ,7 ,8, 9 & 14	Mr. Pranjit Kalita, KKHSOU
10,11,12,13	Dr. Nishant Kashyap Ghatowar, Librarian, Bikali College

Editorial Team

Content & Language:	Dr. Deepjyoti Kalita, Assistant Professor, DLIS, Cotton University
	Mr. Dibyajyoti Patgiri, Assistant Librarian, KKHSOU
	Ms Jayashree Konwar, KKHSOU
Structure, Format & Graphics :	Pranjit Kalita, KKHSOU

August, 2026 ***ISBN: 978-93-7839-013-5***

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

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

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

Headquarter : Patgaon, Rani Gate, Guwahati - 781017

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

CONTENTS

UNIT 1: E RESOURCES IN LIBRARIES **7-16**

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

UNIT 2: DIGITAL RIGHTS MANAGEMENT

Digital information resources (copyright, DRM, other IPR issues, licensing issues); Commercial Publishers and Digital Resource products. **17-27**

UNIT 3: ERM AND LIBRARY CONSORTIA

Collection Development of Electronic Resources: Evaluation, Subscription models, Pricing models, Policy, Renewal issues; Consortia model for electronic resources: Need, Importance, Benefits **28-41**

UNIT 4: DIGITAL REPOSITORIES

Digital Repositories: Concept, Need, Significance, Institutional Repositories, Proprietary and open-source software. National and international Digital Repositories **42-50**

UNIT 5 WEB TECHNOLOGIES IN LIBRARY

Web technologies: Meaning and significance, Concept of HTML, UIRLs, WEB browsers, search engines, websites, directory, blogs and portals, Internet protocols and Internet security, AI in Library development **51-60**

UNIT 6: DATA MANAGEMENT PLANNING

Creating DMPs: key components, practical examples, essential tools for drafting; FAIR Principles: framework for Findable, Accessible, Interoperable, Reusable data. **61-68**

UNIT 7: DATA ETHICS AND TOOLS

Privacy and Ethics: considerations in data handling, privacy regulations, sensitive data protection; Data Management Tools: types overview, popular software comparison, selection for research needs. **69-77**

UNIT 8: FUNDAMENTALS OF RESEARCH DATA MANAGEMENT

Introduction to various aspects of Data; Research Data Life-cycle; Research Data Management: Meaning, Definition, Scope, Purpose, Role of librarian **78-85**

UNIT 9: RDM SOFTWARE AND TOOLS **86-94**
RDM software & services: Commercial, Free and Open-Source examples, features and benefits

UNIT 10: RDM IN ACADEMIC LIBRARIES **95-103**
Roles of Academic Libraries; Skills Requirements for Research Data Management (RDM), best practices and case studies

UNIT 11: RDM TRAINING **104-112**
Data Literacy: program design, workshops; Librarian roles, evaluation methods; Promoting RDM culture.

UNIT 12: RDM POLICIES **113-124**
Institutional Policies: development, funder mandates Compliance, legal guidelines; Auditing, improvement.

UNIT 13: RESEARCH DATA REPOSITORIES **125-133**
Data Repositories: concept, types, features; Metadata standards, access policies; Preservation strategies.

UNIT 14: EMERGING TRENDS IN LIBRARY DATA MANAGEMENT **134-145**
AI, machine learning, blockchain, and cloud solutions in data management, Data privacy, protection Cyber-security trends, new data governance strategies for libraries

INTRODUCTION

The course "E-Resource Management and Research Data Management in Libraries" addresses the paradigm shift in information generation, storage, and distribution in the twenty-first century. Libraries, 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 rapid 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. This course introduces the concept of "electronic resources," or "e-resources," which are now the foundation of contemporary research and education, including everything from sophisticated databases and multimedia archives to online journals and e-books.

The course explores the fundamental transition from "ownership" of hard copies to "access" of digital content, which brings many advantages like round-the-clock availability, multi-user access, and robust search capabilities. Learners will examine how libraries have evolved from print-based establishments to hybrid information hubs, understanding the definition and types of e-resources, the importance of aggregators in organizing massive volumes of data, and how libraries are adjusting to satisfy the evolving information demands of patrons in the digital age. The course traces the evolution from traditional manual cataloguing systems and physical volumes to digitized library services, Online Public Access Catalogs (OPACs), and the widespread availability of electronic journals and databases through the World Wide Web. This transformation has also changed the role of librarians from custodians of physical collections to "digital librarians" or "information managers" who require expertise in web design, server architecture, and digital rights management.

A critical component of the course examines Digital Rights Management (DRM), copyright law, intellectual property rights (IPR), and licensing frameworks, which have become increasingly crucial for library and information professionals. Information creation, storage, distribution, and access have all fundamentally changed as a result of digital technology development. Different obstacles beyond those encountered in traditional print environments have emerged because digital content is typically not owned by libraries in the same manner as print materials. Rather, libraries grant access to material under license, frequently for a predetermined amount of time and subject to particular

requirements established by the publisher or aggregator. The course comprehensively covers the Indian Copyright Act of 1957 and its multiple amendments, the protection of original works including literary, dramatic, musical, artistic, cinematographic films, and sound recordings, along with the duration of copyright protection for different work types. Learners will understand DRM as a systematic approach to copyright protection for digital works in cyberspace, consisting of technology methods intended to regulate and control how digital content is used, duplicated, and distributed.

The course also addresses commercial publishers of digital resources and the complex issues surrounding intellectual property rights, including resource piracy, cultural piracy, economic piracy, internet piracy, and copyright infringement challenges. Learners will learn about the WTO's TRIPS Agreement and international frameworks governing IPR protection, the different types of intellectual property including patents, trademarks, industrial design rights, and trade secrets, as well as practical licensing issues faced by libraries. Upon completion, learners will possess comprehensive understanding of both the opportunities presented by e-resources such as searchability, multi-user access, publication speed, multimedia integration, and space-saving benefits and the challenges including infrastructure costs, the digital divide, preservation and archiving difficulties, health concerns from screen reading, and the complexities of managing licenses and Digital Rights Management software that libraries must carefully navigate.

UNIT 1: E RESOURCES IN LIBRARIES

UNIT STRUCTURE

1.1 Learning Objectives
1.2 Introduction
1.3 Definition
1.4 Need of E Resources
1.5 Evolution from Traditional Print to Electronic Collections
1.6 Types of E Resources
1.7 Advantage and Challenges of E Resources
1.8 Let Us Sum Up
1.9 Answers to Check Your Progress
1.10 Further Reading
1.11 Model Questions

1.1 LEARNING OBJECTIVES

After going through this unit, you 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 abbreviated 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. Through different perspective, 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.

1.4 NEED FOR 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.

1.5 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 High Wire 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: It is 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 yous

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

1.11 MODEL 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: DIGITAL RIGHTS MANAGEMENT

UNIT STRUCTURE

2.1 Learning Objectives

2.2 Introduction

2.3 Digital Information Resources

2.3.1 Copyrights, DRM, Other IPR Issues, Licensing Issues

2.4 Commercial Publishers of Digital Resources

2.5 Let Us Sum Up

2.6 Answers to check Your Progress

2.7 Further Readings

2.8 Model Questions

2.1 LEARNING OBJECTIVES

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

- Explain about the concept of Digital Rights, its kinds, importance.
- Discuss about concept of copyright, DRM, IPR and Licensing issues.
- Know about commercial publishers of digital resources.

2.2 INTRODUCTION

Information creation, storage, distribution, and access have all fundamentally changed as a result of the development of the internet and digital technology. Libraries are rapidly purchasing and offering access to electronic resources including e-journals, e-books, online databases, and digital archives. Traditionally, libraries have been custodians of physical books, journals, and manuscripts.

Different obstacles beyond those encountered in a traditional print environment have been brought about by this transition from print to digital. When a library bought a printed book in the past, it was legally the owner of the physical object and could lend, save, or preserve it for future users without constantly requesting permission. Digital resources, however, operate according to a completely new paradigm. Digital content is typically not owned by libraries in the same manner as print materials. Rather, they grant access to material under license, frequently for a predetermined amount of time and subject to particular requirements established by the publisher or aggregator.

For this reason, it is increasingly regarded as crucial for library and information professionals to comprehend Digital Rights Management (DRM), copyright law, intellectual

property rights (IPR), and licensing frameworks. DRM has a direct impact on what a library can and cannot do with the resources it offers to its patrons. DRM refers to the technologies, rules, and techniques used to regulate how digital content is accessed and used. At the same time, a new set of problems with cost, access equality, and long-term information availability have been brought about by the expansion of huge commercial publishers that create and sell digital resources at scale.

In order to give students a comprehensive grasp of the digital information environment in which contemporary libraries function, this course addresses all of these related topics.

2.3 DIGITAL INFORMATION RESOURCES

Digital resources are broadly classified into two main categories:

- Online digital sources and
- Other electronic resources.

Online digital sources include materials that are accessed through the internet, such as e-journals, e-books, and online databases. On the other hand, other electronic resources refer to offline or physical digital formats like CD-ROMs and offline databases, which do not require continuous internet access.

An e-journal is a scholarly journal available in electronic form. It may either have a printed counterpart or exist solely in digital format, known as a born-digital journal. These journals often replicate the print version but provide additional features such as:

- Hyperlinks to related content
- Interactive visuals and multimedia
- Easy navigation and search options

The concept of the e-book dates back to the 1960s, when Van Dam of Brown University first used the term. An e-book is essentially a digital publication that can be accessed through electronic devices or online platforms. These resources are available in different forms:

- Open access e-books i.e. freely downloadable
- Subscription-based e-books available through institutional licensing

Online databases are another important category of digital resources and can be divided based on their content and purpose. These include:

- Indexing and Abstracting Databases (e.g., SCOPUS, LISA, LIST) provide bibliographic information and summaries
- Full-text Databases (e.g., ScienceDirect) provide complete articles and documents

- Reference Databases include digital dictionaries, encyclopedias, and almanacs
- Statistical Databases (e.g., IndiaStat) provide numerical and statistical data

Digital resources play a crucial role in modern libraries by transforming how user access and utilize information. They offer several advantages, such as:

- Quick and efficient information retrieval
- Simultaneous access for multiple users
- Remote accessibility from different locations

Beyond convenience, digital resources significantly support academic research. Studies have shown that tools like databases, institutional repositories, and library websites are essential for scholars, with university library websites often being considered even more valuable than general search engines like Google.

2.3.1. COPYRIGHT, DRM, IPR ISSUES & OTHER LICENSING ISSUES

COPYRIGHT

The original works of creators, including literary, artistic, musical, and digital creations, are protected by copyright, a sort of intellectual property protection. It guarantees authors control over the usage and dissemination of their work. However, copyright only applies when the work is expressed in a tangible form, such as writing, recording, or digital format; it does not protect ideas, concepts, discoveries, or titles. The World Intellectual Property Organization defines copyright as the legal term for the rights that authors have over their creative works, sometimes known as author's right. Books, music, paintings, sculptures, movies, computer programs, databases, ads, maps, and technical drawings are all protected by copyright.

It is crucial to academic and professional life since it demands that users properly credit their sources and refrain from misusing the work of others.

A broad variety of unique creative works are protected by copyright law, yet laws don't always offer a comprehensive list. Generally speaking, it protects the various ways that ideas are expressed in the literary, artistic, digital, and economic spheres.

The following groups are safeguarded under copyright law.

- Literary works include plays, novels, poetry, reference books, and newspaper articles.
- Computer-related works, such as databases and software
- Performing and audio-visual arts, such as movies, music, and dance

- Artistic creations like sculptures, paintings, sketches, and photos
- Architectural creations, such as building plans and constructions
- Commercial and applied works such as technical drawings, maps, and adverts.

Indian Copyright Act, 1957

The Copyright Act, 1957, which governs copyright in India, has undergone multiple amendments (1983, 1984, 1985, 1991, 1992, 1994, 1999, and 2012) to stay up with both domestic and international regulations.

Historical Background

The Indian Copyright Act, 1847, which was passed to enforce English copyright laws in India, is the country's first copyright law. It was created by the British during the East India Company era. This rule was later repealed, replaced, and extended to all British territories, including India, by the Copyright Act of 1911. Additionally, it was altered once again in 1914 by the Indian Copyright Act, 1914, which continued to be in effect in India until the parliament of sovereign India replaced it with the Copyright Act, 1957.Certain types of intellectual property, commonly referred to as works of authorship, are given property rights under copyright laws. By prohibiting people from replicating an original work in any other way, copyright rules safeguard the creator's legal rights.

Works Protected Under Copyright (Section 13)

The following categories of **original works** receive copyright protection in India:

- **Literary Works**: books, articles, journals, computer programs, databases, translations, adaptations
- **Dramatic Works**: plays, choreography, recitations, dance
- **Musical Works**: compositions with graphical notation (lyrics and music have *separate* rights)
- **Artistic Works**: paintings, sculptures, drawings, photographs, maps, architectural works
- **Cinematographic Films**: video recordings, short clips, soundtracks of movies
- **Sound Recordings**: CDs, phonograms, standalone audio recordings

Type of Work	Duration
Literary, Dramatic, Musical, Artistic works	Life of author + 60 years
Posthumous, anonymous, pseudonymous works	60 years from publication
Cinematographic films & Sound recordings	60 years from publication
Government works	60 years from publication
Broadcasting organisation	25 years from broadcast
Performer's right	50 years from performance

Digital Rights Management

A methodical approach to copyright protection for digital works in cyberspace is known as digital rights management, or DRM. It consists of a number of technology methods intended to regulate and control how digital content is used, duplicated, and distributed. Copyright holders have lost a great deal of control over their works due to the expansion of the Internet, file sharing, and easy access to digital content including music, movies, software, and texts. By limiting unwanted access and eliminating internet piracy, DRM has evolved as a technological answer to this problem. Additionally, it seeks to guarantee that users abide by the terms and conditions established by rights holders and automate licensing procedures.

The objective of DRM is to regulate digital content usage and stop illicit online dissemination. It guarantees that content is only accessible by authorized users and that usage adheres to predetermined guidelines, including terms of use, duration, and payment. DRM also aids in tracking usage of protected works and determining who owns them.

Key Features of DRM are:

- limits the spread and duplication of digital content.
- enforces use conditions and automates licensing.
- limits unwanted access and duplication.
- aids in identifying responsible users and infringement.

IPR and Other Licensing Issues

Original works of human creativity that have not been copied from anyplace or by anybody are referred to as intellectual property (IP). Inventions, literary and artistic creations, names, symbols, pictures, and designs utilized in trade are a few examples. The legal rights that creators and innovators have over their original works of art and commercial work are known as intellectual property rights. Similar to how a person owns physical property, an intellectual property right (IPR) grants the creator exclusive ownership; unauthorized use is subject to legal action, and others must get prior permission before using the work.

The Primary Objectives of IPR are a follows:

- Protecting the ownership and originality of the creator’s work.
- Providing recognition to the creator or concerned authority.
- Allowing owners to gain financial benefits from their creations.
- Providing financial incentive for investment in creative work.
- Motivating creativity and contribute to economic growth.

- Offering economic monopoly advantage to the holder of the right.
- Enhancing the financial status of individuals and the national economy.

Types of IPR are as follows:

- **Copyright:** a privilege granted to the proprietor of a work of literature or art, such as books, music, software, paintings, articles, etc. Publication, dissemination, and adaption are all under the owner's sole authority. Duration: 50 to 100 years after death plus the creator's lifespan; for anonymous works, 95 years after publication or 120 years after creation.
- **Trademark:** A term, symbol, logo, phrase, or combination that sets one organization's products or services apart from another. For instance, the logos of UNICEF, WHO, LG, and Reebok. The symbol ® is used for registered trademarks, while ² is used for unregistered ones. Misuse may result in legal action, and registration may be done domestically or internationally.
- **Patents:** New inventions, such as new machinery, goods, or processes, are granted rights. The invention ought to be original, creative, and useful in the workplace. The exclusive right to stop anyone from producing, utilizing, or marketing the innovation without authorization belongs to the patent holder. 20 years after the application date.
- **Industrial Design Rights**: Preserve an object's visual design, including its shape, colour, pattern, or any combination of these. Novelty, uniqueness, and aesthetic appeal are taken into consideration while granting the right. Duration: 10 to 25 years; national or EU-wide registration is required.
- **Trade Secrets:** rights to proprietary formulas, procedures, techniques, designs, or concepts that provide a company with a financial edge over rivals. There is no official registration, in contrast to other IPRs; instead, the owner is responsible for safeguarding the trade secret. Anyone can utilize it once it is made public.

The WTO's TRIPS Agreement (Trade-Related Aspects of Intellectual Property Rights) was ratified by India and went into effect on January 1st, 1995. TRIPS establishes basic requirements for IPR protection in each of its member nations. Copyrights, trademarks, patents, geographical indications, industrial designs, trade secrets, and integrated circuit layout designs are among the main topics covered under TRIPS. India was granted a five-year transition period (until January 2000) as a developing nation, plus an extra five years for product patents in agricultural chemicals and pharmaceuticals.

Despite its significance, IPR still faces a number of issues that cast doubt on the idea and procedure of intellectual property protection. Some of them are:

- **Resource Piracy**: It occurs when natural and biological resources from one nation are taken and even trademarked by another. A prime example is the unfettered flow of neem seeds to companies like W. R. Grace for export, and the transfer of Basmati rice types from India to strengthen the US rice market. Because there are no protective regulations in place, no legal action is feasible.
- **Intellectual and cultural piracy:** When communities' distinctive cultural history, customs, and intellectual contributions are freely appropriated without acknowledgment or consent, it is known as intellectual and cultural piracy. For example, US companies adopting the trade name "Basmati" for their fragrant rice or Pepsi using the trade name "Bikaneri Bhujia" are both improper and legally dubious practices.
- **Economic piracy**: When trade names are used to take both home and foreign markets, local economies where the original innovation occurred are destroyed. This is referred to as economic piracy. Millions of small farmers are negatively impacted by this as well. For instance, Grace took over the US market from small-scale Indian manufacturers of neem-based biopesticides, while US rice dealers usurped European markets.
- **Internet Piracy:** In today's digital age, internet piracy is becoming a bigger problem. Because the internet and intranet provide easy ways for private and unique information to leak, it is becoming more and more difficult for countries and organizations to protect their uniqueness and trade secrets.
- **Copyright infringement:** It is still a common issue where rivals utilize literary or creative works for commercial purposes without getting consent from the relevant rights holder, which is a clear legal infraction.

Since agreements, not ownership, regulate the licensing of electronic resources, it is a complicated procedure. Each license that a library signs or accepts is a formal agreement that outlines the vendor's and the institution's rights, obligations, and restrictions. Libraries frequently pay for access rather than ownership since electronic resources, in contrast to print resources, are typically accessed through subscriptions.

- Ownership versus Access: The difference between ownership and access is one of the fundamental problems with licensing. While most e-resources are rented through yearly subscriptions, which restricts long-term management, purchasing content grants ownership. Libraries frequently have to pay maintenance costs, and if a subscription expires, access may be terminated, making collection development unclear.
- Access Control and Authorized Users: Licenses specify exactly who is permitted

to use the resource, including alumni, walk-in users, and institutional members. Providing remote access via proxy servers, controlling off-site access, and guaranteeing license compliance all present challenges. Equal access to information may be hampered by limitations on user groups.

- Permitted Uses and Limitations: The use of resources is outlined in licensing agreements. Large-scale copying or redistribution is frequently prohibited, even if users are typically permitted to browse, download, print, and store individual items for study or instruction. Even sharing articles is typically restricted to authorized individuals, and abuse may result in the institution's access being suspended.
- Issues with Interlibrary Loans: One major issue with e-resource licensing is interlibrary lending (ILL). Many vendors only let printed copies to be provided by fax or postal mail and prohibit digital sharing. While some enforce stringent regulations like the CONTU "rule of five," others forbid even this. These restrictions make it more difficult for libraries to share resources.
- Conditions & Terms: Maintenance fees, perpetual access, archive rights, and definitions of core journals are just a few of the conditions found in licensing agreements. These phrases' ambiguity can lead to misunderstandings, particularly when it comes to subscribed content preservation and long-term access.
- Issues of Law and Jurisdiction: The agreement's controlling law is another crucial consideration. Many licenses are initially based on foreign laws, which may later need to be modified to conform to local legal systems. Enforcement and conflict settlement become more difficult as a result.
- Modifications and Vendor Management: Certain agreements let vendors to modify services or license terms at any time. Libraries face uncertainty as a result of the possibility of unforeseen changes in access and usage conditions that could impact service delivery.
- Abuse and Systematic Downloading: Penalties may result from unauthorized actions like excessive use or systematic downloading. Monitoring and awareness are crucial because a single user's activities can frequently result in the entire institution losing access.
- Non-Disclosure Provisions: Confidentiality clauses in licenses may limit the sharing of terms or prices. However, these provisions might be at odds with the need for transparency, particularly in public organizations.
- Preservation and Archiving: The problem of archival rights is crucial. Although licenses may not always permit downloading or storing material for long-term usage,

libraries anticipate permanent access to paid content. The sustainability of digital collections is impacted by this.

- Warranty and Liability Concerns: Disclaimers that materials are supplied “as is” and that providers are not responsible for mistakes or access issues are frequently included in licenses. Users may be adversely affected by problems like service outages or a lack of reimbursement for downtime.
- Usage Data and Assessment: Although usage statistics are crucial for assessing e-resources, problems might occur because of inconsistent data, a lack of standardization, and incorrect reporting formats. Standards like SUSHI and COUNTER are not always consistently adhered to.

2.4 COMMERCIAL PUBLISHERS OF DIGITAL RESOURCES

The supply of digital materials to university and research libraries is mostly and frequently dominated by the commercial publishing sector. Libraries can access databases, e-books, e-journals, and other digital content via vast platforms created by major publishers and aggregators. Librarians in charge of resource management and collection growth can benefit from knowing these publisher’s identities, services, and business practices.

Elsevier (via its ScienceDirect platform), Springer Nature, Wiley-Blackwell, SAGE Publications, Taylor & Francis, Oxford University Press (Oxford Journals Collection), and Cambridge University Press (Cambridge Journals Online) are some of the most well-known commercial publishers. Each of these publishers has sizable collections of e-books and journal titles in the fields of science, technology, social sciences, humanities, and medicine. Through institutional subscriptions, their platforms provide library patrons with access to thousands of full-text books and papers.

Aggregators combine content from several publishers onto a single platform; unlike individual publishers, they may not always own the content they offer. Examples of aggregated collections in the humanities and biological sciences include JSTOR and BioOne, respectively. Another aggregated portal that focuses on social science and humanities journals is ProjectMUSE. Smaller libraries that cannot afford individual subscriptions to dozens of publishers will find these services very helpful.

Academic libraries can also benefit from subscription-based services like Web of Science and Scopus. These are databases that collect, indexes and manages the bibliographic details of contents published in various subject domains. These databases acts as single point search point by author, institution, country, subject etc for resources.

These are very useful for detecting high-impact research, bibliometric studies, and citation analysis. These databases are subscribed to by several academic institutions in order to facilitate the assessment and discovery of research.

A significant substitute for the conventional subscription model is open access (OA) publishing. Digital content is made freely accessible to all users without the need for a subscription under open access. Certain publishers provide hybrid journals, in which certain articles are available without charge while others are protected by a paywall. Others have made the complete switch to “gold” or “diamond” open access models. In order to increase access to research output, libraries have actively supported preprint platforms, institutional repositories, and open access efforts.

Additionally, commercial publishers have adjusted to the development of huge language models and artificial intelligence. Publishers and libraries are reconsidering how their resources are incorporated into these platforms and how academic integrity can be upheld when AI-generated content melds with published research as more students and researchers use tools like ChatGPT and Gemini for information discovery. In addition to licensed digital materials, libraries are being asked to offer advice on the moral application of AI-generated content.

Libraries continue to be concerned about the financial aspect of commercial publishing. The “serials crisis,” a phenomenon where major publishers frequently raise subscription costs dramatically each year, has prompted many libraries to cut their collections and terminate memberships. Growing support for open access models, government-funded digital resources, and cooperative licensing through national and international library consortia has resulted from this conflict between commercial interests and the public good of information access. With time, libraries have also explored consortia based models to access information resources.

CHECK YOUR PROGRESS

Fill in the Blanks

i. ___ usage and usage are managed through the use of Digital Rights Management (DRM).

ii. Libraries typically pay for ____ instead of ownership when it comes to electronic resources.

iii. Only works that are expressed in a ____ form is protected by copyright.

iv. The ___ loan of electronic resources is associated with the CONTU “rule of five.”

2.5 LET US SUM UP

The units describe how libraries are changing from print to digital settings, where ownership is replaced by access. It emphasizes how crucial digital information resources like databases, e-books, and e-journals are to contemporary research. Important ideas that govern the usage of digital content are covered, including copyright, intellectual property rights (IPR), digital rights management (DRM), and licensing concerns. While licensing establishes user rights and limitations, DRM restricts access and guards against abuse. Additionally, the course highlights the importance of open access models and commercial publishers. All things considered, it offers a thorough grasp of how to control, access, and manage digital resources in libraries.

2.6 ANSWERS TO CHECK YOUR PROGRESS

i. Digital Content, ii. Access, iii. tangible, iv. interlibrary loan

2.7 FURTHER READINGS

1. Ashraf, T. (n.d.). *E-Resource Management & Licensing Issues*. Retrieved March 28, 2026, from https://library.iitd.ac.in/arpit_2019-2020/Week%203-%20Module%202-%20E-Resource%20Management%20and%20Licensing%20Issues.pdf

2. Shimray, S. R., Subaveerapandiyan, A., & Ahmad, N. (2025). Digital transformation in academic libraries: e-resources, OPACs and AI in information discovery. *Reference Services Review*, 1–18. https://doi.org/10.1108/rsr-12-2024-0078

3. Warwick, C., Terras, M., Galina, I., Huntington, P., & Pappa, N. (2008). Library and information resources and users of digital resources in the humanities. *Program*, *42*(1), 5–27. https://doi.org/10.1108/00330330810851555

4. Zia, Y., Tehseen, S., & Rajesh Kathane. (2020). Information (ICT) security and Libraries. *IP Indian Journal of Library Science and Information Technology*, *5*(1), 43–45. https://doi.org/10.18231/j.ijlsit.2020.009

2.8 MODEL QUESTIONS

a) Define DRM.

b) What is meant by licensing of e resources? What are the various issues related to licensing of e resources?

c) Define digital information resources.

d) Briefly discuss about copyright and IPR.

*** ***** ***

UNIT 3: ERM AND LIBRARY CONSORTIA

UNIT STRUCTURE

3.1. Learning Objectives

3.2. Introduction

3.3. Collection development of Electronic Resource

- 3.3.1 Evaluation
- 3.3.2 Subscription Models
- 3.3.3 Pricing Models,
- 3.3.4 Policy
- 3.3.5 Renewal Issues

3.4. Consortia Model for Electronic Resources

- 3.4.1 Need
- 3.4.2 Importance
- 3.4.3 Benefits

3.5. Let Us Sum Up

3.6. Answers to Check your Progress

3.7. Further Readings

3.8. Model Questions

3.1 LEARNING OBJECTIVES

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

- Learn about collection development policies for e resources
- Learn about importance of consortia for e resources
- Key issues relating to e resources

3.2 INTRODUCTION

One of the most challenging and significant tasks in contemporary library administration is the management of electronic resources. The processes of selection, procurement, and upkeep were comparatively simple when libraries depended nearly exclusively on print items. A book was chosen, placed on a shelf, ordered, received, and cataloged. Electronic resources, however, do not take this straightforward route. They have several access platforms, fluctuating price models, licensing agreements, technical requirements, and renewal deadlines that need to be closely monitored and handled year after year. Electronic resource management, or ERM, encompasses all of the systems and

procedures that a library employs to manage its electronic resources, starting with initial assessment and continuing through subscription maintenance and final cancellation or renewal. ERM includes not only the choice and purchase of electronic resources, but also their licensing, access control, usage tracking, and the choices made when to renew or cancel subscriptions due to financial restrictions.

The emergence of library consortia has been a significant advance in the field of library resource management. A library consortium is a collection of libraries that band together to collaborate and exchange resources, especially in order to bargain with publishers and aggregators for cheaper prices on electronic resources. Collective action through consortia has made it possible to obtain a significantly greater range of academic materials than any one library could purchase on its own because individual libraries, particularly smaller ones, have limited budgets and bargaining strength.

Both of these closely related topics are covered in this unit; the model of library consortia, including its necessity, significance, and advantages in the context of managing electronic resources, and the collection development of electronic resources, including evaluation, subscription, pricing, policy, and renewal.

3.3 COLLECTION DEVELOPMENT OF ELECTRONIC RESOURCE

In the context of electronic resources, collection development is the deliberate process by which a library finds, assesses, chooses, purchases, and maintains electronic materials to satisfy its patrons' information needs. The management of electronic resources adds layers of complexity that print materials never needed, even though the fundamental concepts of collection development such as relevance to the subject matter, content quality, and appropriateness for the target users remain the same.

A library's collection may contain a wide variety of electronic materials. An electronic resource is defined as material (data and/or programs) encoded for manipulation by a computerized device" in the AACR2, 2005 Update. This content can necessitate the usage of a computer network connection or a peripheral that is physically attached to a computerized device. In simple words e resources includes e-books, e-journals, full-text aggregation databases, databases for indexing and abstracting, reference databases like directories and encyclopedias, databases for statistics and numbers, e-images, and audio-visual materials.

Today's academic libraries work in what is known as a "dual print and electronic

environment," which means they have to manage both print and digital collections concurrently, balance their budgets between the two, and decide whether to cancel print in favor of electronic access or subscribe to the electronic version of a title that already exists in print. This is not a straightforward financial computation. It entails evaluating the additional features of electronic access, the possibility of losing archival rights upon cancellation, the technological prerequisites for using the resource, and whether digital or print format best meets the needs of the library's patrons.

The task of collection development has become more challenging for libraries, especially academic ones, due to a number of economic challenges. Library budgets are simply unable to keep up with the growth of published literature due to the combined effects of inflation, currency fluctuations, and increased publication activity in newer multidisciplinary fields. Scholarly journals raise their subscription prices by approximately 9 to 10 percent each year. As a result, libraries are forced to consider collection expansion more strategically, concentrating on how to maximize the worth of every rupee spent in addition to what to purchase.

3.3.1 EVALUATION

The evaluation of electronic resources is the first and most crucial step in the collection development process. Unlike print materials, it requires coordination among multiple departments such as technical, acquisitions, cataloguing, and licensing, because e-resources involve complexities beyond content, including access, licensing, and pricing.

- Involves multiple stakeholders in decision-making
- Addresses issues like compatibility, licensing, access, and cost

Content evaluation focuses on the relevance and quality of the resource. It must support institutional research goals, enhance the existing collection, and maintain academic standards. Additionally, electronic resources require checking for extra features beyond print versions.

- Aligns with institutional objectives: If the selected electronic resource is actually required for the parent institution. For example an institution dealing with social sciences subject may not require resources dealing exclusively with topic of sciences.
- Ensures quality: Quality of the resources can be identified by the quality of production, for examples journals should follow essential norms of peer review.
- Adds value over print versions: One has to check if the already existing print collection is already satisfying the particular information need that the e reesources might

add.

- Includes backfiles/archives: Before purchasing library has to check if the backfiles, i.e.e older volumes are accessible or not along with the current volumes.
- Updated regularly: Frequency of content updation in the e-resource package is an important factor to be considered.

Technical evaluation is equally important to ensure smooth access and usability. Libraries must verify how the resource can be accessed and whether it fits existing systems.

- Access type: Remote and standalone both access should be available.
- Remote access supports multiple users and off-campus use: Should have the capacity ot handle different type of access requests at single time.
- Checks authentication systems: Only authenticated users should be able to access.
- Ensures compatibility with hardware/software
- Verifies formats (HTML, PDF, XML, EPUB)

Functionality and reliability determine how effectively users can interact with the resource. A user-friendly interface and strong search features are essential.

- Easy navigation and interface
- Advanced search options (Boolean, keyword, full-text)
- Export features (print, email, download)
- 24/7 availability with minimal downtime

Vendor support is another critical but often overlooked aspect. Efficient vendor services improve user satisfaction and reduce operational issues.

- Timely technical support
- Training and product demonstrations
- Trial access before purchase
- Usage statistics (COUNTER-compliant)
- Provision of MARC records

Evaluation is a continuous process and must also be done before renewal decisions. It ensures that the library provides high-quality resources within budget constraints. It is an ongoing assessment, not one-time, essential before renewal, focus on quality over quantity.

3.3.2 SUBSCRIPTION MODEL

Libraries can access electronic resources through a range of subscription

arrangements, and making wise purchasing decisions requires an understanding of these models. Direct subscription from the publisher or vendor by the library and subscription via a library consortium are the two most common subscription methods.

There are various types of direct subscriptions. A library may choose only the journals or databases that are most pertinent to its patrons by subscribing to specific titles from a publisher's catalog. As an alternative, a library can decide to sign up for a subject-based bundle that organizes books according to academic fields. The most contentious form is the "big deal," where a publisher offers a library access to all of its publications for a single fee, even if the library hasn't specifically chosen those titles. The massive agreement can provide access to a lot of content at a cheaper price per item, but it also gives libraries less control over their collection and can force them into difficult-to-break long-term financial obligations.

Another option is subscriptions based on aggregators. Content from several publications is combined and made accessible through a single platform by aggregators like JSTOR, ProQuest, and EBSCO. Smaller libraries that can't afford to bargain directly with several publishers may find this helpful. Aggregated collections do have some disadvantages, though, such as duplicate titles in several databases and the combination of desired and less helpful titles.

Patron-Driven Acquisition (PDA), also known as demand-driven acquisition, is a relatively recent concept in which libraries initiate a purchase when a user actually consumes the content rather than committing to a resource subscription up front. Libraries can avoid paying for content that is never utilized by using this model, which matches spending with proven user needs. Multi-year agreements, in which a library agrees to a subscription for two or three years in return for fixed pricing, are also typical and provide budget stability in return for less flexibility.

3.3.3 PRICING MODEL

For electronic resources, there isn't a single standard price strategy. The pricing offered to a library can vary greatly based on a number of factors, including the size of the institution, the number of simultaneous users allowed, the type of access required, and whether the library is subscribing as an individual institution or as part of a consortium. Different publishers take different approaches.

The Full-Time Equivalent (FTE) model, which bases the membership cost on the number of full-time employees or students at the subscribing institution, is one of the most popular price bases. For the same resource, a larger university with more registered

users will pay more than a smaller one. Particularly when the resource is highly specialized, the FTE computation should be based on the actual target user group for the resource rather than the entire institutional population.

Institutions are divided into tiers depending on size or category under tier-based pricing, and each tier has a unique fee. Concurrent user pricing, on the other hand, bases a cost on the number of users who can access the resource simultaneously. Regardless of how often the resource is used, flat-fee pricing provides access to all users for a single annual payment. In contrast, pay-per-view or pay-per-article pricing models don't require a subscription; instead, customers pay each time they access a piece of content. Libraries with low or erratic demand for a specific resource can benefit from this arrangement.

Because the price may rise significantly after the initial term, introductory pricing—where a publisher gives a lower fee in the first year to entice subscribers needs careful consideration. Before agreeing to any membership, libraries should always find out about anticipated yearly price rises. Publishers typically raise prices by three to six percent annually, though this is frequently greater. The compounding effect over a number of years can put a heavy burden on library budgets.

By combining their purchasing power, institutions can receive cheaper prices through consortia pricing, which will be covered in more detail in the following section. Another model that libraries should be aware of is separate pricing for content and access. In this model, the subscription fee is billed separately from the hosting or access fee, which increases transparency regarding the continuing expenses of preserving access to content that the library has already paid for.

3.3.4 POLICY

A collection development policy for electronic resources serves as a guide for library employees in making consistent, transparent, and well-thought-out decisions about the purchase, management, and selection of electronic resources. A collection development policy, according to the American Library Association, "defines the scope of a library's existing collections, plans for the continuing development of resources, identifies collection strengths, and outlines the relationship between selection philosophy and the institution's goals."

An electronic resources policy must go further and handle a number of concerns unique to the digital medium, whereas standard collection development strategies cover characteristics including topic area, level of material, and target audience. Technical viability (whether the library has the infrastructure to provide access), format preference (when a

title is available in both print and electronic versions), duplication (whether to subscribe to a title that is already held in another format), and access and preservation issues (whether perpetual access rights are secured) are some examples.

The expectations of the library with relation to technological standards, vendor support, licensing conditions, and pricing models are communicated to acquisition librarians and subject selectors by a sound policy. It also lays out the procedure for incorporating users in the creation of collections, including ways to notify users of purchases and get input on possible new resources.

Format preference should be explicitly addressed in the policy. For instance, it should specify that in terms of content currency, an electronic resource having a print equivalent shouldn't lag behind its print counterpart. It should stipulate that the electronic version must offer enough value over its print counterpart, whether through improved functionality, increased accessibility, or the capacity to support many users at once. The policy should outline the conditions under which having a title in several formats is acceptable when duplication is allowed.

Policies for collection creation must be evaluated and revised frequently as the e-resource ecosystem changes; they are not one-time documents. The policy must change to accommodate new publishers, new platforms, new licensing methods, and changes in user behavior. The difficulties presented by open access publication, patron-driven acquisition, and the incorporation of artificial intelligence tools into the research process might not be resolved by a policy that was sufficient five years ago.

3.3.5 RENEWAL ISSUES

One of the hardest things in Electronic Resource Management (ERM) is renewing electronic resources. It necessitates reassessing current subscriptions based on performance, value, and relevance, in contrast to new acquisitions. This entails analyzing both qualitative input and quantitative data.

- Focuses on whether to keep using or stop using current resources.
- Analyzing utilization, cost, feedback, and alternatives is necessary.

When making judgments about renewals, use statistics are crucial. To accurately evaluate real consumption and compare resources, libraries rely on standardized data, particularly COUNTER-compliant reports.

- Standardized measurement is ensured by counter statistics.
- Downloads are used to measure full-text databases.

- Database indexing is determined by searches.
- aids in determining cost per use
- Beneficial for evaluating the worth of various materials

However, consumption data alone shouldn't be the sole basis for renewal decisions. Qualitative factors offer a more profound understanding of a resource's actual worth.

- Low utilization could be the result of ignorance.
- User and faculty input is crucial.
- Verify the content's originality
- Assess usability and accessibility
- Take into account the responsiveness of the vendor

Decisions about renewals are more complicated due to financial limitations. While membership fees continue to rise, e-resources frequently take up a significant amount of the library's budget.

- Exorbitant and growing subscription fees
- Budgets for libraries are either fixed or limited.
- Risk of unreviewed automatic renewals
- A proper renewal review procedure is required.

Another crucial element in the regeneration process is timing. Better decision-making and negotiation are ensured by careful planning.

- Vendors must give at least two months' notice.
- Time required for consultation and assessment
- Negotiation or cancellation scope
- Each member library in a consortium must give their consent.

Before choosing not to renew, libraries must take post-cancellation access rights into account. Continued access to previously paid content is determined by these rights.

- Verify your rights to eternal access.
- After cancelation, access to previously purchased content
- Certain publishers might impose maintenance costs.
- These terms must be stated explicitly in the license agreement.

3.4 CONSORTIA MODEL FOR ELECTRONIC RESOURCES

Some of the major consortia for e resources are:

- e-shodhsindhu initiated by INFLIBNET in 2015 for universities and colleges under UGC 12(B) and 2(f)

- National Knowledge Resource Consortium initiated by CSIR and Department of Science and Technology in 2009
- INDEST- AICTE Consortium initiated in 2003 by Ministry of Human Resource Development, Govt. of India
- N-LIST (National Library and Information Services Infrastructure for Scholarly Content) started in 2010 by INFLIBNET for general Degree colleges under Sections 12(B) and 2(f) of UGC
- CeRA (Consortium for e-Resources in Agriculture) initiated in 2009 by Indian Council of Agricultural Research for National Agricultural Research and Education System (NARES) Libraries
- DRDO E-Consortium: Initiated in 2009, managed by DESIDOC, serves research labs under the Defence Research and Development Organization.
- ONOS (One Nation One Subscription): The ONOS started its operations on January 1st, 2025. For the years 2025, 2026, and 2027, ONOS phase-I has been authorized. The Information and Library Network (INFLIBNET) Center, an independent inter-university center of the University Grants Commission (UGC), Gandhinagar, Gujarat, is in charge of organizing nationwide subscriptions for journal access. The government of India's Department of Higher Education will provide the universities with access to the journals via a single portal called "One Nation One Subscription" (https://www.onos.gov.in/). The One Nation One Subscription's usage and the works of Indian authors published by these organizations will be reviewed on a regular basis by the ANRF.

3.4.1 NEED

A formal partnership between libraries to acquire, manage, and share resources for mutual benefit is known as a library consortium. Although resource sharing has long occurred through union catalogs and interlibrary loans, the emergence of expensive electronic resources made organized collaboration imperative.

- A formal pact for mutual benefit between libraries
- Prioritize resource sharing, management, and acquisition.
- developed significantly as e-resources grew.

Due to expensive prices and technical constraints, few elite institutions in India had access to electronic resources at first. Prior to the introduction of significant consortium initiatives, the majority of colleges and universities lacked access to high-quality intellectual content.

- Access was previously restricted to IITs, IIMs, and IISc
- Access was enhanced by the INDEST Consortium (2003) and UGC-INFONET (2004), although smaller institutions lacked infrastructure and funding.
- Made databases and e-journals widely accessible.

Inequalities in the publishing market are also addressed by the consortium approach. It is challenging for individual libraries to successfully negotiate because large publishers have significant price power.

- Pricing is dominated by major publishers.
- Libraries don't have much negotiating leverage.
- Collectively, consortiums bargain for better terms.
- Stronger bargaining position equals larger group

Effective resource sharing is another crucial requirement. Shared subscriptions are very successful since electronic resources enable several users to access content at once.

- Allows for multiple users to access it at once.
- Minimizes cost duplication
- Optimizes the use of constrained funds
- Particularly advantageous for emerging areas

3.4.2 IMPORTANCE

In addition to lowering expenses, library consortiums are crucial for guaranteeing equitable and long-term involvement in the system of intellectual communication. They give libraries an organized framework for effectively managing and gaining access to electronic resources. Library consortiums boost the information infrastructure generally, improve access, increase negotiating power, and standardize procedures.

- Go beyond the financial advantages
- Encourage equitable and sustainable access
- Increase involvement in academic communication

Consortiums provide a single-window access solution that satisfies various academic and research requirements from the standpoint of individual libraries. Additionally, they offer professional negotiation assistance, which is essential for managing intricate licensing contracts.

- Using a single platform to access several resources
- Availability of professional negotiating and licensing expertise

- aids in comprehending words such as:
- Users with authorization
- Interlibrary loan
- Unending access
- Rights to archives
- Terms of cancellation

The development of India's national information infrastructure has been greatly aided by consortiums. The establishment of e-Shodh Sindhu in 2015 was a significant step in bringing several efforts together.

- e-Shodh Sindhu combined:
- The INDEST-AICTE UGC-INFONET N-LIST program
- developed a cohesive and effective system
- Better access to colleges, universities, and technical schools

Creating best practices and standards for electronic resource management is another important function of consortiums. They aid in the development of structures that guarantee uniformity and justice.

- Encourage model license contracts
- Adhere to COUNTER use statistics guidelines
- Promote equitable pricing practices
- Examples consist of:
- CDL (California Digital Library) agreements and SERU (NISO initiative) licensing models.

3.4.3 BENEFITS

Libraries, users, academics, and even publishers can profit greatly from the consortium approach to electronic resource management. It establishes a cooperative framework that increases access, lowers expenses, and improves the process of scholarly communication in general.

- Benefits a variety of parties, including users, publishers, scholars, and libraries.
- Encourages cooperation and increased access to information

Better license conditions and cost savings are the biggest benefits for libraries. Libraries can bargain with publishers more successfully if they band together.

- Lower subscription fees as a result of collective bargaining
- Improved license conditions (perpetual access, ILL, more users)

- Clear pricing and better contracts
- Long-term and stable resource access

Consortiums are especially beneficial to smaller institutions because they provide them with access to resources that would otherwise be out of reach.

- Thousands of articles and databases are accessible
- Overcomes financial constraints
- lessens institutional inequality
- encourages intellectual development in smaller universities

Consortiums greatly improve scholarly work's visibility and accessibility for users, researchers, and writers. Better research results and knowledge sharing follow from this.

- Greater availability of scholarly materials
- A rise in citations and reading
- promotes the productivity of research
- permits the distribution of documents between institutions

The consortia approach benefits publishers as well, especially in terms of efficiency and stable revenue.

- One agreement as opposed to several discussions
- Consistent and steady income
- Greater institutional reach
- Enhanced exposure of publications

CHECK YOUR PROGRESS

a. _____ is the standardised usage reporting system used by libraries to measure electronic resource usage across platforms.

b. _____ is the subscription model in which a publisher makes its entire catalogue of journals available to a subscribing library for a single package price, rather than allowing the library to select only the titles it needs.

c. _____ is a merged result of INDEST-AICTE, UGC-INFONET, and N-LIST happened in December 2015.

d. _____ is the pricing model in which the subscription cost of an e-resource is calculated based on the number of full-time students or staff at the subscribing institution.

3.5 LET US SUM UP

The selection, assessment, procurement, licensing, access, and renewal of digital

materials in libraries are all part of the intricate and ongoing process known as electronic resource management, or ERM. E-resources, in contrast to print materials, necessitate careful consideration of technological, economical, and legal considerations. While subscription and price structures dictate how resources are accessed and paid for, evaluation guarantees quality, relevance, usefulness, and cost-effectiveness. Consistent decision-making is guided by a clearly defined collection development policy, which also adjusts to evolving digital contexts. To guarantee the best possible use of resources, renewal decisions are based on usage data, cost analysis, and user input.Consortiums of libraries have become an essential way to deal with growing expenses and tight budgets. They facilitate better licensing conditions, increased access to scholarly content, and collective bargaining. India's national information infrastructure has been reinforced by programs like e-Shodh Sindhu. By encouraging cooperation, cutting expenses, and improving the distribution of knowledge, consortiums help libraries, users, academics, and publishers. They continue to be crucial for effective and fair access to technological resources despite obstacles.

✓ 3.6 ANSWERS TO CHECK YOUR PROGRESS

a) COUNTER, b) Big Deal, c) E-Shodh Sindhu, d) FTE (Full Time Equivalent)

3.7 FURTHER READINGS

1. Ashraf, T. (n.d.). *E-Resource Management & Licensing Issues*. https://library.iitd.ac.in/arpit_2019-2020/Week%203-%20Module%202-%20E-Resource%20Management%20and%20Licensing%20Issues.pdf
2. 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
3. Baruah, S., & Chanda, D. A. (2022). *COLLECTION DEVELOPMENT POLICY FOR E-RESOURCES*. SJSU ScholarWorks. https://digitalcommons.unl.edu/libphilprac/7004?utm_source=digitalcommons.unl.edu%2Flibphilprac%2F7004&utm_medium=PDF&utm_campaign=PDFCoverPages
4. Bentil, W. (2025). Analysis of Collection Development Policies for Electronic Resources at Selected Academic Libraries in Ghana. *Collection Management*, 1–15. https://doi.org/10.1080/01462679.2025.2497833
5. Hiremath, U. (2001). Electronic consortia: resource sharing in the digital age. *Collection Building*, *20*(2), 80–88. https://doi.org/10.1108/01604950110388716

6. Johnson, S., Evensen, O., Gelfand, J., Lammers, G., Sipe, L., Zilper, N., Fronty, J., Hafner, J., Mansfield, J., Okerson, A., & 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.pdf
7. Kaur, N. (2007). *E-resources and Collection Development: Emerging Issues for The Academic Libraries*. http://ir.inflibnet.ac.in:8080/ir/bitstream/1944/1435/1/599-607.pdf
8. Sankar, P., & Kavitha, E. S. (2016). Collection Development of Electronic Resources: A Guideline for Librarians from Selection to Renewal. *Indian Journal of Information Sources and Services*, *6*(1), 33–39. https://doi.org/10.51983/ijiss.2016.6.1.438

3.8 MODEL QUESTIONS

1. What are the main topics that should be covered in an electronic resource collection development policy?
2. Describe the steps involved in renewing an electronic resource membership and the things a library should take into account before doing so.
3. What are the primary obstacles to managing subscriptions to electronic resources that library consortia encounter?
4. Explain the electronic resource management consortium paradigm. Talk about its necessity, significance, and advantages using the experiences of Indian library consortiums including INDEST, UGC-INFONET, and e-Shodh Sindhu.

*** ***** ***

UNIT 4: DIGITAL REPOSITORIES

UNIT STRUCTURE

4.1 Learning Objectives

4.2 Introduction

4.3 Digital repositories

- 4.3.1 Concept and Need
- 4.3.2 Significance
- 4.3.3 Institutional Repositories
- 4.3.4 Proprietary and Open-source Software

4.4 National and International Digital Repositories

4.5 Let Us Sum Up

4.6 Answers to Check Your Progress

4.7 Further Readings

4.8 Model Questions

4.1 LEARNING OBJECTIVES

After going through this unit, you will be able to

- learn about digital repositories; its need significance.
- know about institutional repositories (IRs) and different software for creating IRs.
- know about national and international digital repositories.

4.2 INTRODUCTION

Libraries and academic institutions all around the world have transcended their conventional position as the physical caretakers of print collections in recent decades. It is now both feasible and essential for institutions to save, manage, and distribute their scholarly and research output in digital form due to the development of digital technology, the internet, and open publication norms. The technology that enables this is a digital repository. It offers an organized, searchable, and user-friendly environment for gathering, storing, and making digital items from academic papers and theses to datasets, pictures, and multimedia files available.

Two significant advancements in academia have coincided with an increase in the demand for digital repositories. The first is the exponential increase in research output, which has made it more challenging for researchers to find, access, and expand upon past work in the absence of a structured digital infrastructure. The second is the growing

expense of subscribing to commercial journals, which has forced academics to adopt open publishing models, in which research findings are made publicly available instead of being restricted by costly subscription restrictions. Digital repositories especially institutional repositories have emerged as a crucial tool for accomplishing this.

India offers a crucial background for comprehending digital repositories. India, a nation with hundreds of universities and research facilities many of which struggle financially has witnessed substantial government investment in the creation of national-level digital repositories. To guarantee that the intellectual output of Indian scholars is not only maintained but also available to everyone, regardless of institutional affiliation or geographic location, initiatives such as Shodhganga, the National Digital Library of India, and Vidyanidhi have been established.

This unit presents the idea of digital repositories, explains their importance, talks about the unique class of institutional repositories, looks at the software options available for creating them, and gives a summary of the main national and worldwide programs for digital repositories.

4.3 DIGITAL REPOSITORIES

4.3.1 CONCEPT AND NEEDS

A digital repository is a system for the systematic administration, access, and storage of digital objects. Publicly available papers, pre-print manuscripts, theses, dissertations, datasets, technical reports, images, audio recordings, video files, and software are all examples of digital objects. A digital repository differs from a basic file storage system in that it is structured, searchable, and based on a set of metadata standards that make it easy for users to locate and retrieve particular objects.

The functions of various digital repository types vary. A university or research organization's scholarly work is the focus of an institutional repository. A discipline or subject repository, commonly referred to as a thematic repository, gathers materials from several institutions within a certain field, like the social sciences, physics, or medicine. A national repository unifies content from all of the nation's institutions on a single platform.

Digital repositories are necessary because of a number of interrelated issues that both libraries and scholars deal with. An institution's research output was dispersed throughout departmental servers, individual hard drives, printed volumes in locked archives, and commercial publications that required costly subscriptions to access prior

to the existence of repositories. This resulted in duplication of effort, gaps in the literature, and a limited return on the public investment that sponsored the original research because a thesis submitted at one university can be entirely undetectable to academics at another institution pursuing closely related work.

By compiling the dispersed intellectual output of a community or institution into a single, well-organized, and publicly accessible platform, digital repositories directly solve this issue. For a university, this implies that everyone with an internet connection can access the theses of previous students, faculty articles, research project reports, and materials created by different departments in one location. It makes it possible for a national organization to compile research results from hundreds of institutions in a way that would not be feasible otherwise.

Digital repositories have a compelling preservation case as well. Digital files are susceptible. Formats get outdated, hard drives malfunction, and websites go down. Stable file formats, backup mechanisms, and rules for controlling the technology refresh required to maintain digital information accessibility as platforms and devices change over time are all part of a well-managed digital repository, which is constructed with long-term preservation in mind. The digital record of intellectual heritage could easily vanish in the absence of such infrastructure.

4.3.2 SIGNIFICANCE

Digital repositories are essential to the transformation of scholarly communication; they are not merely cost-saving or convenient tools. They assist libraries adapt to the digital world and alter how research is disseminated, accessible, and assessed. Digital repositories-

- Act as an essential component of contemporary academic communication
- Boost research impact, visibility, and accessibility
- Support the changing roles of libraries

Promoting open access, particularly via the green route, where authors submit their work for free access to an online platform even if it is published in commercial publications, is one of the main purposes of digital repositories.

- Turn on unrestricted access without paywalls
- Encourage repository-based green open access
- Open access (OA) journals that complement gold

Repositories greatly increase the effect and visibility of research. Scholarly production

has a wider audience since open access papers are more frequently read and cited.

- Research accessibility on a global scale
- Increased rates of citations
- Increased influence on academia

By storing and disseminating institutional research, repositories help libraries transition from being passive content purchasers to active knowledge managers.

- Organize and distribute the results of local research
- Maintain institutional expertise
- bolster the library's primary purpose

By organizing intellectual output, repositories also aid in institutional design and research evaluation.

- Monitor the output of your research
- Determine your gaps and strengths.
- Help with audits and evaluations

4.3.3 INSTITUTIONAL REPOSITORIES

An institutional repository (IR) is a particular kind of digital repository created by and for a specific institution, usually a university or research organization, in order to gather, store, and A digital archive used to gather, store, preserve, and disseminate an institution's intellectual output, particularly research findings, is called an institutional repository (IR).

- an IR stores and oversees an organization's digital research
- It aids academics in collaborating and sharing their work
- It increases the prominence of scholarly work

A university's service for managing and disseminating digital content produced by its community is commonly referred to as an IR.

An IR comprises items such as:

- Preprints and post-prints of research articles
- Dissertations and theses (ETDs)
- Monographs and books
- Reports and datasets
- Posters, training materials, and conference papers
- Working papers and technical reports, which are not officially published, are examples of grey literature that can be found in repositories.

Primary objectives of an IR include,

- Boosting research's visibility internationally
- Maintaining institutional digital content
- Making it accessible by self-archiving
- Controlling the intellectual production of the institution make publicly accessible the digital scholarly output created by its members.

4.3.4 PROPRIETARY AND OPEN-SOURCE SOFTWARE FOR BUILDING IR

Commercial suppliers create proprietary software for digital repositories, which usually includes a number of services like installation assistance, frequent upgrades, and technical helpdesks. Because the vendor takes care of the majority of the technical management, proprietary systems have the advantage of frequently requiring less internal technical capability. The drawback is that universities become reliant on the commercial vendor for continuous support and development, and licensing fees might be high. The institution might be in a challenging situation if the vendor discontinues the product or modifies its pricing.

In contrast, institutions can use, modify, and expand the software without having to pay license costs thanks to the open-source method. It is important to have a thorough understanding of both DSpace and EPrints, the two most popular open-source repository technologies in libraries and academic institutions. Details of these software are discussed in the paper titled MLIS-104 "Digital Libraries" .

4.4 NATIONAL AND INTERNATIONAL DIGITAL REPOSITORIES

In addition to specific institutional repositories, a number of significant digital repositories function on a national and international scale, combining content from numerous institutions and making it available to a worldwide audience. These repositories have been crucial in establishing the infrastructure for open access research and guaranteeing that scholarly work, especially that generated in nations with limited resources, gets the attention it merits.

Indian electronic theses and dissertations (ETDs) are stored in Shodhganga, the country's most significant digital repository. It is kept up to date by the University Grants Commission (UGC) of India's independent inter-university INFLIBNET Centre. The repository was created in response to a June 2009 UGC mandate that required all universities recognized by the UGC Act to submit soft copies of their MPhil and PhD

theses for hosting in the INFLIBNET repository. The Sanskrit words “Shodha,” which means research, and “Ganga,” which refers to the Ganga River, are the sources of the name “Shodhganga,” which implies a constant flow of research information. Currently, more than 500,000 theses and more than 13,000 synopses from Indian universities and colleges are stored in the repository. The international academic community can freely view Indian doctoral research because all full-text materials are accessible for reading and downloading in open access format. The multilingual nature of Indian research is reflected in the repository’s support for several Indian languages, including Hindi, Tamil, Gujarati, Sanskrit, Urdu, and Malayalam, in addition to English.

Another significant digital archive in India is Vidyanidhi, which specializes in doctoral theses submitted by Indian universities. With assistance from the National Information System for Science and Technology (NISSAT) and subsequently from the Ministry of Culture and MHRD, it was founded at the University of Mysore. One of the first initiatives in India to establish open access research infrastructure is Vidyanidhi, which serves as a digital library and archive for Indian doctorate theses.

Developed at IIT Kharagpur as part of the National Mission on Education through Information and Communication Technology (NME-ICT), the National Digital Library of India (NDL) is another larger project started by the Ministry of Human Resource Development (now the Ministry of Education). Under a single metadata harvesting interface, NDL seeks to incorporate content from a variety of digital libraries, institutional repositories, and educational platforms throughout India. This was Initiated in April 2015, with a pilot launched in May 2016, and officially dedicated to the nation on June 19, 2018. Regardless of their institution or location, students, researchers, and educators throughout India will be able to access a wide variety of educational and research resources through a single window. Instead of acting as a primary repository, NDL acts as an aggregator, gathering metadata from many sources, such as Shodhganga, and making it searchable via a single platform.

A significant international organization devoted to encouraging the production, use, and accessibility of electronic theses and dissertations is the Networked Digital Library of Theses and Dissertations (NDLTD). NDLTD was founded in 1996 and offers a global platform for the accessibility of ETDs from institutions worldwide. It maintains a union catalogue that enables searching across ETD collections from its member institutions worldwide, encourages the implementation of open access policies for theses and dissertations, and provides training and assistance to institutions establishing their own

ETD programs. India's Shodhganga repository connects Indian PhD research with the global academic community by gathering metadata for NDLTD.

One of the world's oldest and most popular disciplinary repositories is arXiv. Established in 1991 at Los Alamos National Laboratory and now hosted at Cornell University Library, arXiv is a repository for pre-print manuscripts in physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, and related fields. In order to make their findings accessible to the public months or even years before official publication, researchers in these fields usually submit their articles to arXiv either before or concurrently with submission to peer-reviewed journals. Over two million papers are currently stored on arXiv, which receives hundreds of thousands of new contributions annually. It has significantly altered how computer science and physics research is disseminated, and its paradigm has influenced the larger open access movement.

The National Center for Biotechnology Information (NCBI), a division of the US National Library of Medicine (NLM), is responsible for maintaining PubMed Central (PMC), a free online database. Full-text academic publications published in biomedical and life sciences journals are archived by PMC and made publicly accessible. It was established in 2000 and has grown into one of the largest repositories of its kind, currently archiving over 11.9 million articles. PMC is more than a simple document store; submissions are indexed and formatted for enhanced metadata, medical ontology, and unique identifiers, which allows the content to be linked with other NCBI databases and searched through the Entrez system. Funding agencies such as the US National Institutes of Health (NIH) mandate that research funded by their grants must be deposited in PMC, which has been a powerful driver of its growth.

The broader landscape of national and international repositories also includes the Directory of Open Access Repositories (OpenDOAR), which is not itself a repository but a directory that lists and describes open access repositories worldwide, allowing researchers and librarians to identify appropriate repositories for specific subject areas or geographic regions. Similarly, the Registry of Open Access Repositories (ROAR) tracks the growth of open access repositories globally and provides statistics on their content and development.

These national and international platforms together constitute the open access repository ecosystem. Each operates at a different scale and with a different scope, but all share the common goal of making research more visible, more accessible, and more durable. For Indian researchers, the connection between national repositories like Shodhganga and NDL, on one hand, and international networks like NDLTD and

OpenDOAR, on the other, represents an important channel through which Indian scholarship can reach and engage with the global research community.

CHECK YOUR PROGRESS

a. Shodhganga, is the national digital repository of Indian electronic theses and dissertations, maintained by I_____ under the UGC.

b. _____one of the most widely used open-source repository platforms, was originally developed through a collaboration between the MIT Libraries and Hewlett-Packard Labs.

c. ____is the standard protocol that allows other systems to automatically harvest metadata from digital repositories and make the content discoverable through aggregators.

d. ____ is the practice by which researchers deposit their own published or pre-publication work directly into a digital repository

4.5 LET US SUM UP

- Digital repositories are structured systems that gather, store, maintain, and make digital scholarly content including theses, papers, databases, and multimedia accessible. They arose as a result of the demand for free access and the quick expansion of research output, which helped address the problems of dispersed information and restricted accessibility. Repositories facilitate long-term digital preservation while increasing visibility, accessibility, and research effect through the centralization of intellectual output.
- In particular, institutional repositories (IRs) oversee and distribute an organization's research, fostering cooperation and international recognition. Additionally, they assist libraries in becoming active knowledge managers. Repositories employ either open-source software, such as DSpace and EPrints, which give flexibility without licensing fees, or proprietary software, which provides technical support but comes at a cost.
- Repositories like Shodhganga, NDL, arXiv, and PubMed Central encourage open access and transnational information exchange on a national and international scale. All things considered, digital repositories are essential to contemporary scholarly communication and research distribution.

4.6 ANSWERS TO CHECK YOUR PROGRESS

a) INFLIBNET, b) DSPACE, c) OAI-PMH, d) Self-archiving

4.7 FURTHER READINGS

1. Kiryakova, G. (2013). *Digital repositories – a tool for effective management of institutional assets*. Academia.edu; Trakia Journal of Sciences, Volume 11, Supplement 1, ISSN 1313-7050, pp 377-383. https://www.academia.edu/5343299/Digital_repositories_a_tool_for_effective_management_of_institutional_assets
2. MPUNDU MFULA, C. (n.d.). *CREATING DIGITAL REPOSITORIES*. https://www.ifla.org/wp-content/uploads/2019/05/assets/services-for-parliaments/preconference/2017/chama-mpundu-mfula-creating-digital-repositories.pdf
3. Rajput, P. (n.d.). *DL Software: Greenstone and Eprints*. Retrieved March 31, 2026, from https://mlsu.ac.in/econtents/5699_DL-MLISc%20Unit%20III%20Software%20EPrints%20&%20Greenstone.pdf
4. *Trusted Digital Repositories: Attributes and Responsibilities An RLG-OCLC Report RLG Mountain View, CA*. (2002). https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf

4.8 MODEL QUESTIONS

a. Define institutional repositories, and what challenges do libraries face in establishing and maintaining them?
b. Discuss the difference between proprietary and open-source software for digital repositories, and state two advantages and one disadvantage of each.
c. Discuss in brief on PubMed Central and arXiv as examples of international digital repositories.
d. Discuss about different national and international digital repository initiatives.
e. Elaborate the concept and necessity of digital repositories. In what ways have repositories transformed the storage, sharing, and preservation of research, and what significance do they hold in the open access movement?

*** ***** ***

UNIT 5: WEB TECHNOLOGIES IN LIBRARY

UNIT STRUCTURE

5.1 Learning Objectives

5.2 Introduction

5.3 Web Technologies

5.3.1 Meaning and Significance

5.3.2 Concept of HTML, URLs, Web Browsers, Search Engines, Websites, Directory, Blogs and Portals

5.4 Internet Protocol and Internet Security

5.5 Let Us Sum Up

5.6 Answers to Check Your Progress

5.7 Further Readings

5.8 Model Questions

5.1 LEARNING OBJECTIVES

After going through this unit, you will be able to

- learn about common web technologies,
- learn about parts of web technologies like HTML, URLs, Web Browsers.

5.2 INTRODUCTION

The term Web Technologies denotes a standard enabling developers to construct web applications utilizing predefined classes, objects, methods, and properties found in markup languages, style sheet languages, or programming languages. online technology is comprehensively defined as the utilization of diverse tools and standards employed to not only publish content on the online but also to organize, structure, and facilitate its accessibility. was designed by **Tim Berners-Lee in 1989** and has since grown to become an indispensable tool for education, commerce, governance, and communication.

The effective operation of web technologies relies on the collaboration of various integrated technologies, including Client-Server Technologies, Web Browsers, and document markup languages such as HTML (HyperText Markup Language), XML (Extensible Markup Language), and RDF (Resource Description Framework).

The internet has progressed through four distinct generations:

- **Web 1.0** (early 1990s): Read-only; introduction of the World Wide Web and first search engines

- **Web 2.0** (mid-2000s): It is the 2nd generation of web development and design facilitating features like interactive sharing of information, interoperability, user cantered design and platforms such as YouTube, X, and Facebook emerged.
- **Web 3.0** (late 2000s–2010s): Semantic markup, web services, and machine-readable data
- **Web 4.0** (emerging): Symbiotic Web, where systems intelligently communicate with each other

Web technology is crucial in all sectors such as education, government, business, healthcare, and library services revolutionizing global access to and dissemination of knowledge.

5.3.1 MEANING AND SIGNIFICANCE

Web technologies refer to the online technology encompasses the array of tools, languages, protocols, and systems employed to develop, administer, and operate websites and online applications on the internet. It constitutes the cornerstone of the contemporary digital landscape by facilitating communication, interaction, and information dissemination via web browsers. These technologies facilitate the global information network that links users around the globe.

Web technology fundamentally comprises client-side and server-side components. Client-side technologies, including HTML, CSS, and JavaScript, facilitate the design of web page structure, aesthetics or interface, and interactivity, enabling direct user engagement with content. On the other hand, server-side technologies, encompassing programming languages such as PHP and Python, in conjunction with databases and protocols like HTTP, oversee data processing, storage, and application logic in the background.

Web technology encompasses several tools and frameworks that improve performance, responsiveness, and user experience. It functions via a client-server architecture, facilitating uninterrupted communication between browsers and servers. Web technology facilitates fast information distribution and real-time interaction, ranging from basic static pages to intricate dynamic applications.

In recent years, the domain of web technology has broadened considerably, integrating sophisticated elements such as real-time communication, artificial intelligence, and decentralized systems. Web technology is essential in shaping contemporary online experiences in the filed of library and information science along with other domains by

facilitating easy access to information and services while enhancing interactivity.

The significance of web technologies in library includes:

- Gives an interactive platform for loading all types of web elements easily
- Provides cross-platform and device-independent access to content from anywhere in the world 24X7.
- Supports multimedia, dynamic content, and database-driven web applications
- Facilitates services like e-mail, digital libraries, OPAC, and real-time communication
- Enables organizations to manage and disseminate information efficiently at low cost
- Supports the evolution from static websites to smart, intelligent web applications

5.3.2 CONCEPT OF HTML, URLS, WEB BROWSERS, SEARCH ENGINES, WEBSITES, DIRECTORY, BLOGS AND PORTALS

HTML:

HTML (Hypertext Markup Language) is the standard markup language employed for the creation and design of web pages. It establishes the fundamental framework of a webpage, facilitating the appropriate presentation of content on web browsers. HTML is a markup language, not a programming language, that employs tags to delineate items inside a page.

HTML comprises a collection of elements that instruct the browser on content presentation. These elements serve to structure and designate many components of a webpage, including headings, paragraphs, links, images, and more elements. HTML can denote “this is a heading,” “this is a paragraph,” or “this is a link,” so assisting browsers in accurately interpreting and rendering the content.

Each HTML element is typically encapsulated by tags that delineate its commencement and conclusion. These tags organize the webpage and guarantee that the content is shown in a coherent and comprehensible manner.

HTML serves as the foundational framework for all web pages, imparting structure and significance to the content presented on the internet. A sample of HTML is given below:

```
<!DOCTYPE html>
<html>
<head>
<title>Page Title</title>
```

```
</head>
<body>
<h1>My First Heading</h1>
<p>My first paragraph.</p>
</body>
</html>
```

Here,

- <!DOCTYPE html> The declaration specifies that this document is an HTML5 document.
- <html> It is the root element of an HTML page
- <head> this element contains meta information about the HTML page
- <title> element gives a title for the HTML page that is shown in the browser s title bar or in the page's tab)
- <body> this element defines the document's body, and contains all the headings, paragraphs, images, hyperlinks, tables, lists, etc.
- <h1> element defines a large heading and <p> element defines a paragraph.

URLs:

A URL or Uniform Resource Locator is a concise sequence of characters utilized by computers to identify and retrieve resources on a network, particularly on the internet. It is frequently referred to as a web address, facilitating users' access to web pages, documents, photos, videos, and many online resources hosted on servers. For example https:// https://kkhsou.ac.in/ is a URL.

The standardization of URLs occurred in 1994 by the Internet Engineering Task Force (IETF), with input from Tim Berners-Lee. Originally restricted to ASCII characters, URLs now accommodate several languages via International Resource Identifiers i.e. the IRIs.

A URL consists of many parts, each having a specific function in locating and accessing a resource on the internet:

- **Scheme (Protocol):** It defines the method used to access the resource, such as HTTP, HTTPS, or FTP.
- **Subdomain:** It Represents a subdivision of the main domain, commonly seen as *www*.
- **Domain:** The main name of the website that identifies it uniquely (e.g., example).
- **Top-Level Domain (TLD):** The extension of the domain, such as *.com*, *.org*, or *.in*, often indicating purpose or country.

- **Port Number:** Specifies the communication endpoint on the server (default: 80 for HTTP and 443 for HTTPS).
- **Path:** Indicates the exact location of a resource or page within the website.
- **Query String Separator:** The symbol **(?)** that begins query parameters in a URL.
- **Query String / Parameters:** Data passed in key–value pairs (e.g., key=value) separated by **&**.
- **Fragment:** Denoted by **#**, it directs the browser to a specific section within a webpage.

For example, here in

https://www.geeksforgeeks.org/computer-networks/components-of-a-url/

This URL can be divided into the following parts:

1. Scheme / Protocol

- https://
- Indicates how the browser communicates with the server.
- HTTPS (HyperText Transfer Protocol Secure) ensures secure data transfer.

2. Subdomain

- www
- An optional prefix that represents a specific section of the website, often the main web service.

3. Domain Name

- kkhsou.ac.in
- The unique address of the website.
- It consists of:
- Name: kkhsou
- Top-Level Domain (TLD): .ac.in (commonly used for academic organizations in India)

4. Path

- /computer-networks/components-of-a-url/
- Specifies the exact location of the resource within the website.
- Broken further into directories:
- computer-networks ’! main section
- components-of-a-url ’! specific page/article

WEB BROWSERS

It is a software application used for retrieving, presenting, and navigating information

on the World Wide Web. The concept of the web browser originated with World Wide Web, the first browser created by Tim Berners-Lee in 1990; the first publicly available browser, Mosaic, was released in 1993. Widely used browsers include:

- Internet Explorer (1995, Microsoft)
- Netscape Navigator (1994)
- Mozilla Firefox (2002)
- Google Chrome (2008)
- Opera (1995)

SEARCH ENGINES

A search engine is a software application intended to retrieve information from a dataset, predominantly on the World Wide Web. It assists users in obtaining pertinent results by analyzing queries inputted into a search bar. Upon entering a query, the search engine examines its indexed database and generates a list of web pages (documents or files) that include or pertain to the specified terms. These results are typically presented in order of significance.

Search engines enable users to enhance their queries with operators like AND, OR, and NOT, facilitating the acquisition of accurate results. They offer sophisticated search functionalities, allowing users to seek specific content kinds, like images, videos, news, phrases, or individual websites.

A search engine serves as a mechanism that enables users to efficiently find and get information from the extensive resources available on the internet.

Components of a Search Engine are:

- Web Crawler (Spider/Bot): This automatically traverses the internet to discover, fetch, and scan web pages, following links to update the index.
- Indexer: It analyzes, tokenizes, and processes text from crawled pages, generating an inverted index that associate's terms with their positions in the document corpus.
- Database (Index Repository): A comprehensive, structured database that houses all indexed information, metadata, and keywords to facilitate swift retrieval.
- Ranking Algorithm/Search Mechanism: Assesses the pertinence of indexed documents in relation to queries by employing diverse signals (e.g., keywords, user intent) to prioritize results.
- Search Interface: The user-oriented component, such as a browser search bar, where users input queries and obtain search engine results pages.
- Caching and Optimization: Retains frequently accessed or recent search results

to improve speed and alleviate strain on the primary search infrastructure.

WEBSITES

The inaugural website was developed by Tim Berners-Lee on August 6, 1991, at CERN, and it elucidated the concept of the World Wide Web. A website comprises a series of interconnected web pages (HTML documents) hosted on a server, typically commencing with a Home Page that offers content and connects to various sections of the site. Websites may be categorized as:

- Static - immutable HTML text rendered immediately (origin: 1991)
- Dynamic: It content that is database-driven and generated in real-time by server-side technologies like as Servlets, JSP, or PHP (origin: late 1990s)

DIRECTORY

A Web Directory is a systematically arranged, human-curated compilation of websites classified by subject or category. The Yahoo! Directory, established in 1994, was among the earliest and most prominent web directories, preceding the prevalence of automated search engines. Directories functioned as the principal means of locating websites prior to the invention of search engines capable of indexing and ranking billions of pages.

BLOGS AND PORTALS

A blog, an abbreviation for “web log,” is a website typically managed by an individual, featuring regular postings of commentary, event descriptions, or various media such as graphics or video. The blog concept emerged about 1994 when Justin Hall released one of the earliest personal blogs; the name “weblog” was introduced by Jorn Barger in 1997 and subsequently abbreviated to “blog” by Peter Merholz in 1999. Blogs are progressively utilized by libraries, enterprises, and educational institutions as promotional, notification, and marketing instruments.

A Web Portal is a singular access point website that offers a wide array of integrated services and resources from a consolidated interface. The notion originated in the mid-1990s with platforms such as Yahoo! (1994) and AOL providing integrated services that combined news, search, email, and directories on a single interface. Each library portal is designated as the “Mirror of the Library,” encompassing all digital services such as e-journals, online databases, and OPAC. Portals are constructed on Client-Server Architecture, utilizing either two-tier or three-tier (multi-tier) frameworks to improve usability, scalability, and performance.

5.4 INTERNET PROTOCOL AND INTERNET SECURITY

The Internet Protocol (IP) is the essential communication protocol that delineates the transmission of data packets across the internet. IP was officially delineated in 1974 by Vint Cerf and Bob Kahn in their seminal article “A Protocol for Packet Network Intercommunication,” establishing them as the progenitors of the internet. Each host on the internet is allocated a distinct IP address:

- IPv4: It was deployed in 1981 and is a 32-bit numeric dot decimal IP address. Example 192.168.5.18. It supports Manual Configuration.
- IPv6 (introduced in 1998): 128-bit alphanumeric hexadecimal addresses designed to address IPv4 limitations. Example: 50b2:6400::6c3a: b17d: 0:10a9

Benefits of IPv6 over IPv4 are

- **Larger Address Space:** IPv6 uses 128-bit addresses whereas IPv4 uses. 32-bit allowing many more devices to connect.
- **Improved Security:** IPv6 has built-in features like data authentication and encryption, making connections safer than IPv4.
- **Simpler Header:** IPv6 has a streamlined header that enhances the speed and reducing processing cost.
- **Better Support:** IPv6 provides improved Quality of Service, enhancing video, audio, and website performance.
- **Mobile-Friendly:** IPv6 offers better, faster, and more secure support for mobile connections.

Internet security is a subset of cybersecurity aimed at safeguarding the users, devices, and networks from cyber threats during internet connectivity. It employs techniques such as firewalls, encryption, and antivirus software to safeguard personal information, prevent unwanted access, and reduce hazards including malware, phishing, and identity theft.

Cybercrimes in Libraries

Cybercrime refers to the unlawful conduct in which computers or networks serve as instruments or objectives. The growing reliance on ICT in libraries has rendered systems vulnerable to dangers including data theft, hacking, and illegal access. These offenses impact information security, intellectual property, and user privacy.

Types of Cybercrime in Libraries:

- **Software-related crimes:** This includes unauthorized access, viruses, Trojan attacks, software piracy
- **Data-related crimes:** This includes data leakage, data manipulation (data

diddling), data spying

- **Physical crimes:** Includes crimes like theft of hardware, destruction of storage media
- **Internet-based crimes:** Covers email spoofing, hacking, password attacks, cyber stalking and common incidents include website defacement, insider misuse, and breach of confidential data

Followings are some of the Prevention Measures used in libraries

- Use of strong passwords and multi-layer authentication
- Installation and regularly update antivirus and firewall systems
- Restricting unauthorized downloads and access
- Continuous data backup and encryption of sensitive files
- Avoiding suspicious emails and unknown links
- Securing Wi-Fi networks and avoid public networks for transactions
- Conducting user awareness and staff training programs

CHECK YOUR PROGRESS

a) HTML stands for_____.
b) The World Wide Web was designed by___ in the year____
c) IPv4 uses____ -bit addresses, while IPv6 uses____ -bit addresses
d) Institutions, including libraries, are one of the targeted areas _____of attacks

5.5 LET US SUM UP

Web technologies are instruments and standards that facilitate the development and administration of websites and online applications. HTML organizes web pages through tags; URLs distinctly identify resources on the internet; web browsers access and present content; search engines catalog and evaluate web pages; websites consist of a series of HTML documents hosted on a server; directories serve as curated collections of websites; blogs function as personal commentary platforms; and portals act as integrated service gateways. Internet Protocol (IP) regulates the transmission of data packets across the internet. Libraries are increasingly confronted with cybersecurity concerns such as hacking, malware, data theft, and ransomware. The implementation of firewalls, robust passwords, antivirus software, encryption, and staff training effectively safeguards library resources and patron data.

5.6 ANSWERS TO CHECK YOUR PROGRESS

a) Hyper Text Markup Language

b) Tim Berners-Lee; 1989

c) Answer: 32-bit; 128-bit

d) Ransomware

5.7 FURTHER READINGS

1. Md. Ashikuzzaman. (2017, April 3). *What is the Internet Protocol (IP)?* Library & Information Science Education Network; Library & Information Science Education Network. https://www.lisedunetwork.com/internet-protocol/
2. Pooja, & Pai, R. D. (2025). Best Practices for Cyber Security in Academic Libraries. *Information Security Education Journal (ISEJ)*, *14*(1), 1–9. https://doi.org/10.6025/isej/2025/12/1/1-9
3. Prakash Rastogi. (2025, January 14). *What Is Web Technology? Definition & Examples 2025*. VirtuBox Blog; VirtuBox. https://www.virtubox.io/blog/what-is-web-technology
4. Thomas, R. (2017). IMPACT OF WEB TECHNOLOGIES ONLIBRARY AND DOCUMENTATION CENTRES. *International Journal of Library & Information Science (IJLIS)*, *6*(5). https://iaeme.com/MasterAdmin/Journal_uploads/IJLIS/VOLUME_6_ISSUE_5/IJLIS_06_05_004.pdf
5. W3Schools. (2024). *Introduction to HTML*. W3schools.com. https://www.w3schools.com/html/html_intro.asp

5.8 MODEL QUESTIONS

1. What is HTML? Name any two HTML tags and their functions.
2. What is a URL? List any four components of a URL.
3. What are the different types of cybercrime that affect academic libraries and the measures used to prevent them.
4. Critically examine the role of Web Technologies including HTML, URLs, Web Browsers, Search Engines, Websites, Directories, Blogs, and Portals in transforming library and information services. Illustrate with examples.

*** ***** ***

UNIT 6: DATA MANAGEMENT PLANNING

UNIT STRUCTURE

6.1 Learning Objectives

6.2 Introduction

6.3 Creating DMPs

6.3.1 Key Components

6.3.2 Practical Example, Essential tools for drafting

6.4 FAIR Principles: Framework for findable, Accessible, Interoperable, Reusable data

6.5 Let Us Sum Up

6.6 Answers to Check Your Progress

6.7 Further Readings

6.8 Model Questions

6.1 LEARNING OBJECTRIVES

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

- learn about the concept of data management planning
- learn about key components and tools for creating DMPs
- learn about FAIR principle

6.2 INTRODUCTION

A data management plan delineates the categories of data, code, and other research outputs that a project will produce, as well as the methods for organizing, formatting, documenting, storing, and preserving these resources. It also discusses the information that will be shared with other researchers and the methodologies by which such data will be distributed. In recent years, funding organizations and academic publications have increasingly underscored the need of disseminating research data and associated products. This expectation is based on the necessity for transparency, reproducibility, and validation of research findings, which are essential for the progress of academic work. A data management plan must encompass both the handling of data throughout the research process and its management and dissemination post-project completion.

In contemporary libraries, the formulation of a Data Management Plan (DMP) has become imperative for the efficient management of research and institutional data. Key points suggesting the importance or need of creating DMP are as follows:

- Guarantees data integrity by appropriate collection, storage, and documentation

methodologies.

- Facilitates reproducibility through careful documentation of research methods and methodologies.
- Improves transparency by explaining data sources, methodologies, and constraints.
- Enhances collaboration by outlining roles, duties, and communication methods.
- Facilitates compliance with funding agency conditions for grant applications and research initiatives.
- Guarantees adherence to ethical, legal, and regulatory standards, particularly concerning sensitive data.
- Facilitates the enduring preservation and ongoing accessibility of research data
- Enhances research effect by rendering data discoverable, accessible, and reusable.

6.3 CREATING DMPS

A data management plan is a formal document that specifies the strategies, methods, and guidelines to be followed for managing, describing, storing, analyzing, preserving, and sharing the data collected and produced during the course of a research project.

Data Management Planning (DMP) is an important technique in modern libraries to guarantee research data are managed systematically throughout their life cycle. With data-intensive research, libraries have broadened their conventional functions to encompass data curation, preservation, and access. DMPs are increasingly required by funding sources and are an important part of research planning and evaluation.

Role of Libraries

- Institutional repositories offer a safe environment for gathering and storing research data.
- Libraries ensure the long-term preservation of data, frequently beyond the usual timescales of technology.
- Librarians provide consultancy services to assist researchers arrange data and apply proper metadata standards.
- Libraries promote reproducibility by archiving raw datasets for verification and validation of research.
- Libraries help researchers deposit their data in the right repositories to enable open access.

- Librarians are “data concierges” guiding scholars through data management and storage alternatives

Steps for creating Data Management Plan involves the following

- The process begins with determining types, formats and sources of data that will be created by the investigation.
- Researchers should clearly describe the techniques of data gathering including the tools, instruments and procedures used.
- The file formats and standards should be chosen to ensure long-term usage and interoperability.
- Create a plan for data organization using standardized file naming conventions and metadata standards.
- The data storage requires the identification of secure storage solutions such as institutional repositories or cloud platforms.
- Backup solutions should be designed to avoid data loss and guaranteed data security.
- Data cleaning and quality control procedures shall be described to preserve correctness and reliability.
- Ethical, legal and privacy issues should be considered, especially in the case of sensitive data.
- Policies for data sharing, including access, licensing and embargo periods, must be explicitly established.
- Plan for the long-term preservation of information, including retention periods and sustainable formats.
- The roles and responsibilities of researchers, librarians and technical workers should be clearly defined.
- Draft and develop the plan using appropriate tools such as DMPonline or DMPTool
- The DMP should be evaluated and updated continuously over the study lifecycle.

6.3.1 KEY COMPONENTS

The Key components that need to be considered while creating DMPs are

- **Data Description and Types:** Describes the data types to be collected (e.g. surveys, photos, software), formats, volumes, and methods.
- **Documentation and Metadata Standards:** Defining how data will be documented, including metadata schemas for example Dublin Core or others in order to ensure it is

discoverable and interpretable.

- **Backup and Storage Strategy:** Details of immediate storage solutions, security measures (encryption, restricted access), and backup routines during the project.
- **Ethics, Legal, and Intellectual Property (IP) right consideration:** Addressing privacy, confidentiality, and data anonymization, along with ownership, copyright, and licensing for reuse.
- **Data Sharing and Reuse Plan:** Outlines how, when, and where data will be shared, including any embargo periods, restrictions, or data repositories used.
- **Archiving and Preservation**: Details of long-term plans to preserve data in repositories, including costs and file formats to ensure future usability.
- **Roles and Responsibilities:** Identifying different staff members responsible for data management, preparation, and curation throughout the project lifecycle.
- **Budgetary Considerations**: Covering the costs related to data management, such as repository fees, storage space, or data preparation services.

6.3.2 PRACTICAL EXAMPLE, ESSENTIAL TOOLS FOR DRAFTING

Some practical examples are discussed below:

- **Institutional Repository Administration:** University library receives, maintains, makes available theses, dissertations, research datasets, faculty publications through one unified system. The data are organized according to metadata standards, stored in structured forms and provided with open access policies. The library preserves materials long term through regular backups and format migration. This shows that libraries can systematically manage academic research outputs.
- **Cultural Heritage Digital Archive:** A library launches an initiative to digitize manuscripts, pictures, oral histories and local cultural records. Data collection is performed utilizing scanning and field recordings and is packaged with metadata for simple retrieval. General materials are shared openly. Sensitive or community-based data is limited.
- **Research Data Support Service:** The library assists sponsored project researchers in developing Data Management Plans It helps to organize, clean, store and share data using proper repositories. The library supports compliance with funding agency regulations and encourages data reuse.

Tools for Libraries to Manage Data

- **DMP Creation Software**

DMPonline provides researchers with the tools to develop organized data management plans that meet funder requirements. DMPTool offers templates and collaborative features for data management planning. (https://dmponline.dcc.ac.uk/)

- Storage and Repository Tools

Libraries have widely adopted DSpace for storing and managing digital resources.

- Data Cleaning & Organization Tools

OpenRefine reduces inconsistencies and improves data quality. Microsoft Excel is used to organize, sort and do basic analysis of data collections.

- Metadata & Documentation Tools

Dublin Core is used to describe digital materials in libraries. Zotero assists in managing documentation and research references.

- Tools for Backup & Conservation

LOCKSS stores the content in the long term by distributing the storage. Google Drive offers backup and simple access to your files.

6.4 FAIR PRINCIPLES: FRAMEWORK FOR FINDABLE, ACCESSIBLE, INTEROPERABLE, REUSABLE DATA

The FAIR Principles are a set of principles to make sure research data is Findable, Accessible, Interoperable and Reusable, therefore improving its use and long-term worth. The principle is interlinked yet independent, so that they can be applied in a phased manner and in varied combinations, depending on the institutional capacity and resources. The design aims to make it easier for academics and institutions to get involved, encouraging a slow adoption of better data management methods. The FAIR Principles offer a comprehensive framework for the appropriate management of research data in the digital age.

Libraries are crucial to the implementation of FAIR practices through researcher consultation, data curation, and sustainable access.

Applying FAIR principles makes data more usable, accessible and impactful, and strengthens the overall research ecosystem

Elements of the FAIR Principles

- Findable

Data needs to be easily found by humans and robots, which means it needs accurate meta data, indexing and unique IDs. This enables datasets to be effectively discovered in repositories and databases.

- Accessible

Data should be retrievable through standard protocols and access methods, with well-defined conditions for access when limitations are applied. Even if data is restricted, there should be metadata to let consumers know that the data exists.

- Interoperability

Data must be formatted in accordance with standard formats, vocabulary and metadata so that it may be integrated with other datasets and systems. This allows for interoperability and cross-platform and cross-discipline data sharing.

- Reusable

Data should be adequately recorded, licensed and kept for usage for future study and other reasons. Clear usage rights and good documentation increase the long-term worth of data.

Role of librarians in implementing the FAIR principles

- Librarians advocate for FAIR principles throughout their universities and research groups.
- Librarians use FAIR concepts into Data Management Plans (DMPs), preservation policies and processes.
- Librarians organize and enrich research data with metadata, standards and documentation.
- Librarians offer training to researchers and staff on metadata standards, vocabularies, and data technologies.
- Librarians aid researchers in depositing data in FAIR-compliant repositories.
- Librarians assess institutional data procedures and collections for FAIR compliance.

CHECK YOUR PROGRESS

a) What is a Data Management Plan (DMP)?

b) What does FAIR stand for?

c) Name any one tool used for creating a DMP.

6.5 LET US SUM UP

This subject examines the concept and importance of Data Management Planning (DMP) in libraries, and its role in organising, storing, conserving and distributing research data effectively. It requires that DMP maintains data integrity, transparency, repeatability,

and compliance with funding constraints. This lesson covers the principal stages of developing a DMP, including data gathering, documentation, storage, backup, ethical issues and long-term preservation. It also covers core components such as metadata, data sharing and stakeholder responsibilities. Practical applications and tools such as DMPonline and DSpace are presented. Finally, the FAIR principles are proposed as a framework for making data findable, accessible, interoperable and reusable.

6.6 ANSWERS TO CHECK YOUR PROGRESS

a) A Data Management Plan (DMP) is a formal document that describes how research data will be collected, organized, stored, protected and shared throughout a research project.

b) FAIR stands for **Findable, Accessible, Interoperable, and Reusable**.

c) One tool used for creating a DMP is **DMPonline**.

6.7 FURTHER READINGS

1. Angelo, A., & Lund, P. (2026). *Data management plans: what are they and how libraries can help.* Handle.net; University of Canterbury. Library. http://hdl.handle.net/10092/7894
2. Bhardwaj, R. K. (2018). Data Management: A Practical Guide for Librarians. *Journal of Scientometric Research*, *7*(2), 142–144. https://doi.org/10.5530/jscires.7.2.24
3. *Data Management Planning | University Libraries*. (2026). University of Maryland Libraries. https://www.lib.umd.edu/research/data/data-management
4. Natal, G. (2017). Data Management: A Practical Guide for Librarians. *Journal of the Medical Library Association*, *105*(4), 421. https://doi.org/10.5195/jmla.2017.336
5. PETRUNOVSKA, S. V. (2025). Successful Practices and Challenges in Implementing Fair Principles for Academic Libraries. *University Library at a New Stage of Social Communications Development. Conference Proceedings*, *10*, 115–121. https://doi.org/10.15802/unilib/2025_343173
6. Riley, A. C. (2017). Data management and curation: Professional development for librarians needed. *College & Research Libraries News*, *76*(9), 504–506. https://crln.acrl.org/index.php/crlnews/article/view/9385/10557

6.8 MODEL QUESTIONS

a) Explain the role of libraries in Data Management Planning.

b) Write the key components of a Data Management Plan.

c) Describe the importance of Data Management Planning in libraries.

d) Explain the concept of Data Management Planning (DMP) and discuss its steps, key components, tools, and the role of FAIR principles in effective data management.

*** ***** ***

UNIT 7: DATA ETHICS AND TOOLS

UNIT STRUCTURE

7.1 Learning Objectives

7.2 Introduction

7.3 Privacy and Ethics: Consideration in Data Handling, Privacy Regulations, sensitive data protection

7.4 Data Management Tools: Types overview, Popular software comparison, Selection for research needs

7.5 Let Us Sum Up

7.6 Answers to Check Your Progress

7.7 Further readings

7.8 Model Questions

7.1 LEARNING OBJECTIVES

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

- learn about data ethics in the context of library and information science.
- different tools used in libraries in order to handle library and research data.

7.2 INTRODUCTION

Data are basic facts, statistics or observations that are collected, analyzed and structured to provide information that has meaning. Data is critical for libraries to manage resources, monitor user behavior and improve services. Data at the library can be classed broadly into bibliographic data (catalog records, metadata), user data (membership information, borrowing history), transactional data (circulation records, access logs) and digital usage data (e-resource access, search queries).

Ethics is very crucial while dealing with library data since it contains sensitive personal and intellectual information of users. Libraries are built on the values of privacy, confidentiality, and intellectual freedom, which guarantee that users may access material without fear of surveillance or improper use of their data. Ethical issues: Improper use of data, unlawful access, excessive gathering or sharing of user data may violate user trust and professional ethics. Research also suggests that poor data procedures in libraries could lead to privacy hazards and ethical difficulties, particularly in the case of analytics-based investigations.

There are various technologies used for managing and organizing library data such

as Integrated Library Systems (ILS) like Koha, Library Management Software such as SOUL 3.0, database management systems, digital repositories and data analytics tools. These tools help in effective storage, retrieval, organizing and safe handling of library data.

7.3 PRIVACY AND ETHICS: CONSIDERATION IN DATA HANDLING, PRIVACY REGULATIONS, SENSITIVE DATA PROTECTION

Privacy and ethics are foundational principles of ethical use of library data, as they guarantee users intellectual freedom, confidentiality, and trust. With the increase of big data and analytics, libraries are expected to have rigorous ethical standards and regulatory compliance to ensure that user data is processed securely and ethically as digital systems.

7.3.1 CONSIDERATION IN DATA HANDLING

Ethical data handling is a general term that refers to the responsible acquisition, storage, use and dissemination of data, respecting the rights of users.

The Core Principles of Ethics are:

Confidentiality:

- To protect users' personal and reading records from unwanted access.
- Safeguarding privacy of what people read, search or borrow.

Privacy:

- Users should have the right to access information freely and without surveillance or monitoring.
- Libraries should not track user behavior unless necessary.

Intellectual Freedom:

- Users should not be restricted or evaluated in their choice of information.
- Good stewardship is not censorship but open inquiry.

Permission:

- Users should be aware of what data is gathered and how it is utilized.
- Consent has to be specific, not implied.

Transparency:

- Libraries need to be transparent about their data practices.
- Users should know how long their data will be stored and who can access it.

Data Minimisation:

- Collect only the data essential for library services.
- Don't gather too much or irrelevant info.

Accountability:

- Library personnel are required to conform to codes of ethics (e.g., ALA Code of Ethics).
- Institutions must be held accountable for data misuse or breaches.

7.3.2 PRIVACY REGULATIONS AND SENSITIVE DATA PROTECTION

Privacy is a core value of libraries, and is crucial for ensuring user confidence and access to It matters as

- It encourages open access to information without fear of being judged or monitored.
- Builds trust between the library and its users.
- Prevents the misuse of personal and intellectual data.
- Supports democratic values including freedom of expression and inquiry.

Lack of privacy may prevent users from accessing some sites and so restricting learning chances.

Some of the common privacy risks

- Access to user records without authorization.
- Hacking and data security breaches.
- Tracking users without their knowledge.
- Sharing data with third parties (vendors, institutions)

Some Major Privacy Regulations are

General Data Protection Regulation (GDPR) - European Union:

- User permission, data protection and the right to be forgotten are stressed.
- Demands transparency of data gathering and processing

Information Technology Act, 2000 (India)

- Recognizes electronic data legally and protects against cybercrime.

Digital Personal Data Protection Act (India 2023):

- Prioritizes consent-based data processing.
- Guarantees the right of people on their personal data.

FERPA (U.S.)

- Protects student education records, such as library usage data.

ALA/IFLA Guidelines;

- Fight for less data gathering and more privacy protection.
- Oppose surveillance and over-tracking of users.

IFLA Recommendation for Libraries

- Library and information services should respect and advance privacy both at the level of practices and as a principle.
- Library and information services should promote advocacy activities at national, regional and worldwide levels (e.g. by human rights and digital rights organisations) to defend the privacy of the individual and his/her digital rights and motivate library workers to consider these concerns.
- Library and information services shall not engage in electronic surveillance and any sort of unlawful monitoring or collecting of users personal data or information behaviour which would jeopardize their privacy and undermine their rights to seek, receive and impart information. They should take steps to minimize the collection of personal information about their users and the services they use.
- While the government's intrusion into the information or communications of users can't be entirely prevented, library and information services should ensure that any such intrusion is underpinned by legitimate principles for such practices and is necessary and proportionate to legitimate aims (e.g. "International Principles on the Application of Human Rights to Communications Surveillance").
- When library and information services provide access to resources, services or technology that may jeopardize users' privacy, libraries should urge users to be aware of the implications and offer guidance in data protection and privacy protection.
- Library and information services should enable users to make educated decisions, to take legitimate actions and to weigh up risks and advantages in their communications and use of services on the Internet.
- Data protection and privacy protection should be included in media and information literacy training for library and information service users. This should include teaching on the technologies to employ to keep their privacy.
- The education of library and information professionals should encompass ideas and practices of data and privacy protection in a networked world.

Sensitive data is personal or confidential information that is at a higher risk of being misused and so requires greater protection.

Types of Sensitive Data in Libraries

- Name, address, contact information or other individual identifiable information.

- History of borrowing and reading tastes.
- Academic records and research data
- Digital traces (search queries, login data).
- Demographic information (age, gender, socioeconomic status)

Research shows that library analytics can often contain sensitive data such as academic performance, demographics, and usage habits, which might pose privacy problems if not managed effectively.

Measures to Protect Sensitive Data

Technical Measures:

- Encryption: Protects data in storage and in transit.
- Access Control: Ensures that only authorized users can access the data.
- Security Systems & Firewalls Keep the unwelcome out.
- Anonymization/De-identification: Remove identifiers of individuals.

Administrative Rules:

- Data Governance Policies: Clear regulations on how data is used and managed.
- Regular Audits: Monitor adherence and find weaknesses.
- personnel Training Train library personnel on data ethics and security.
- Data Retention Policies: Remove data when it is no longer necessary.

User-Oriented Measures:

- Provide transparent privacy notices
- Let users select if they want to opt in to data collecting.
- Allow users to decide who can access their personal data, and give them the right to have it deleted.

Ethical Issues in Current Library Data Usage

Learning Analytics:

- Using user data for academic performance prediction could infringe privacy.
- This often lacks adequate permission and openness.

Big Data and AI:

- Data collecting on a large scale increases the potential of misuse.
- Privacy and service improvement - ethical problems.

Third-Party Vendors:

- Data shared with external platforms can be confidential.

Re-identification Risks:

- Sometimes anonymised data can be traced back to individuals.

7.4 DATA MANAGEMENT TOOLS: TYPES OVERVIEW, POPULAR SOFTWARE COMPARISON, SELECTION FOR RESEARCH NEEDS

Data management technologies are useful in organizing, storing, processing and conserving library and research data. These technologies enhance research activities and decision-making, by ensuring effective data processing, accessibility and long-term usage. The tools can be broadly classified into the following catagories:

- **Integrated Library Management Systems:** This are used for managing library operations such as cataloguing, circulation, and acquisitions. Examples:

Koha, SOUL 3.0

- **Database Management System (DBMS):** Used to store, retrieve, and manage structured data efficiently. Examples: MySQL, PostgreSQL.
- **Institutional Repositories and Digital Libraries:** Preserve and provide access to scholarly outputs. Examples: DSpace, EPrints. Details of this tools are discussed in the paper titled Digital Libraries.
- **Data analysis and data Visualisation Tools:** They are used for analyzing datasets and generating insights. Examples: Microsoft Excel and R; a free and open-source computer language for statistical analysis and data visualization in research. It enables complex techniques like regression, data mining, text. It is commonly used for bibliometric analysis, research data analysis and trend investigations. It does require some basic programming expertise though that could be a disadvantage for novices.
- **Reference Management Tools:** They are Help organize citations and research references. Examples: Zotero, a free, open-source reference management application that helps you collect, organize, and cite your research. It allows users to save articles, books and web resources straight from browsers and organize them in folders. Zotero automatically generates citations and bibliographies in many styles (APA, MLA, Chicago, etc.). It also allows for cloud syncing and collaboration with other scholars. It's easy to use, making it a great pick for students and researchers who don't have a lot of technical expertise.

Mendeley**,** a reference manager software founded in 2007 by PhD students Paul Foeckler, Victor Henning and Jan Reichelt and bought by the Dutch academic publishing giant Elsevier in 2013. It is used to distribute research papers and create bibliographies for academic topics.

- **Data Management Planning Tools**: Assist in creating Data Management Plans (DMPs). Examples: DMPTool, DMPonline

Software	Type	Key Features	Advantages	Limitations
Koha	ILS	Cataloguing, OPAC, circulation	Open-source, customizable	Needs technical support
SOUL 3.0	ILS	Designed for Indian libraries	User-friendly, supported by INFLIBNET	Less flexible
DSpace	Repository	Digital archiving, metadata support	Widely used, scalable	Complex setup
EPrints	Repository	Self-archiving, simple interface	Easy to use	Less customizable
MySQL	DBMS	Structured data storage	Fast, reliable	Limited analytics
R	Analysis	Statistical modelling, visualization	Powerful, open-source	Requires coding
Microsoft Excel	Analysis	Data entry, charts	Easy to use	Not for big data
Zotero	Reference	Citation management	Free, simple	Limited storage
Mendeley	Reference	PDF management, collaboration	Good for researchers	Privacy concerns

The selection of relevant tools is dependent upon the nature, scope and aims of the research.

Some of the important factors for selection are

- **Data Types**
- For Structured data: DBMS (such as MySQL)
- For Unstructured data: Repository systems (e.g. DSpace)
- **Purpose of research:**
- Data analysis: Excel, R
- Literature review: Zotero, Mendeley.
- **Level of data:**
- Small Data Sets: Excel
- Massive data collections: DBMS or sophisticated analytics tools
- **Technical Know How**
- Beginner: Excel, Zotero, Advanced: R, PostgreSQL
- **Budget Limitations:**
- Preference for open-source tools (Koha, DSpace, R) Affordable
- **Collaboration Requirements:**
- Tools with cloud support and sharing (Mendeley, Zotero)

- **Institutional Needs:**
- Existing system compatibility
- Data policy and standards compliance

CHECK YOUR PROGRESS

Fill in the blanks

a) _____ is a popular Open-Source Integrated Library System for library automation.

b) _____ is a digital repository software for storing scholarly information.

c) _____ is a statistical computing tool for complex data analysis.

d) _____ is a reference management tool for citation generating and research organization

7.5 LET US SUM UP

- This unit introduces the concept of data and the importance of data in libraries to manage resources and improve services. It describes several sorts of library data such as bibliographic data, user data, transactional data and digital data. The unit stresses the importance of ethics, privacy and secrecy in the handling of sensitive user information, hence safeguarding intellectual freedom and trust. It also talks about privacy legislation such as GDPR and India's Data Protection Act, and threats such as data breaches and misuse. The paper introduces several data management tools (ILS, DBMS, repositories, reference managers). Proper selection of tools based on research needs to achieve efficient, secure and successful data management.

7.6 ANSWERS TO CHECK YOUR PROGRESS

a) KOHA, b) DSpace, c) R, d) Zotero

7.7 FURTHER READINGS

1. Asher, A. (2017). *Risk, Benefits, and User Privacy: Evaluating the Ethics of Library Data*. Iu.edu; Rowman & Littlefield. https://scholarworks.iu.edu/dspace/items/8d885eea-019a-4170-9d06-d85a6b6d617b
2. Floridi, L., & Taddeo, M. (2016). What Is Data ethics? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2083). https://doi.org/10.1098/rsta.2016.0360
3. *IFLA Statement on Privacy in the Library Environment – IFLA*. (n.d.). https://

www.ifla.org/publications/ifla-statement-on-privacy-in-the-library-environment/

4. Lund, B., & Wang, T. (2022). What Does Information Science Offer for Data Science Research? A Review of Data and Information Ethics Literature. *Journal of Data and Information Science*, *0*(0). https://doi.org/10.2478/jdis-2022-0018
5. Salo, D. (2024). Data Ethics in Library Learning Analytics. *Journal of Librarianship and Scholarly Communication*, *12*(1). https://doi.org/10.31274/jlsc.16245
6. Tony Robert Fish. (2021). Data ethics. *Routledge EBooks*, 131–137. https://doi.org/10.4324/9781003201182-27

7.8 MODEL QUESTIONS

a) What is data ethics? What are core principles of data ethics?

b) Discuss the IFLA recommendations for data privacy in LIS.

c) Briefly discuss about different data management tools used in libraries and research.

d) What are the important factors that influence the selection of data tools.

e) What are the general types of data related to libraries?

*** ***** ***

UNIT 8: FUNDAMENTALS OF RESEARCH DATA MANAGEMENT

UNIT STRUCTURE

8.1 Learning Objectives

8.2 Introduction

8.3 Introduction to various aspects of Data

8.4 Research Data Life-style; Research Data Management; Meaning, Definition, Scope, Purpose, Role of Librarian

8.5 Let Us Sum Up

8.6 Answers To Check Your Progress

8.7 Further Readings

8.8 Model Questions

8.1 LEARNING OBJECTIVES

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

- learn about various aspects of data in research
- learn various concept of RDM
- role of librarian in research data management

8.2 INTRODUCTION

All academic work is built on research data. Data; whether it's statistics from an experiment, words from an interview, photos from a survey or entries from a database; gives the evidence that leads to new understanding. In today's digital age, the volume of research data has grown and careful collection, organisation, preservation and exchange of research data is crucial. Research data management, or RDM, is a structured method to data handling throughout its lifecycle and is the answer to this difficulty.

If not properly managed, data can be lost, misconstrued or rendered unusable. Kothari states that the research is a systematic search for information and so there is a need to collect, handle and analyse data in a reliable way. Research designs affect the gathering and understanding of data, and Creswell stresses the necessity for clear management methods from the outset. Researchers and libraries need to go down to basics as more and more funding organizations demand data management strategies and open sharing.

This unit provides an introduction to key ideas in research data management. It starts with the nature of data, then proceeds to the research data lifecycle, definition, scope and goal of RDM, and the important role of librarians in RDM.

8.3 INTRODUCTION TO VARIOUS ASPECTS OF DATA

A research data set is any recorded factual material that is generally recognized by the scientific community as being essential for the validation of research findings. It encompasses a variety of materials that are produced or gathered during the course of a research project, including physical files, digital datasets, documentation, artifacts, recordings, and more. Kothari characterizes it as information that is essential for deriving conclusions and is obtained through observation or experimentation.

Data Types

There are two primary categories of data:

- ***Quantitative data:*** Statistically measurable and numerical data, such as test scores or survey ratings. It enables the utilization of instruments such as correlations and averages for analysis.
- ***Qualitative data:*** Descriptive, such as interview transcripts, and involving text, images, or observations. It concentrates on the contexts and meanings.

Additional distinctions include:

Type	Detailed Description	Examples
Quantitative	Numerical values representing amounts or counts; allows mathematical operations like addition or statistical tests (e.g., t-tests, ANOVA). Discrete: whole numbers; continuous: any value on a scale.	Test scores (discrete), height measurements (continuous), survey ratings (1-5 scale).
Qualitative	Non-numerical descriptions capturing qualities, opinions, or contexts; rich in meaning but requires thematic coding.	Interview transcripts, field notes, photographs, thematic narratives.
Primary	Original data generated directly by the researcher for the specific study; highest control over quality but resource-intensive.	Custom surveys, lab experiments, direct observations.
Secondary	Pre-existing data from external sources; cost-effective but may lack specificity or require cleaning.	Census records, published datasets, archival documents.

Data Formats

Long-term accessibility and interoperability are influenced by the format selection. In

order to prevent obsolescence, it is recommended to adhere to non-proprietary, unadulterated, open standards. Incompatibility is a potential issue with proprietary formats such as Excel (.xlsx); instead, consider using CSV.

Category	Common Formats
Text/Documents	TXT, PDF/A, XML, DOCX, RTF
Tabular/Spreadsheets	CSV, TSV, XLSX, SAV (SPSS), DTA (Stata)
Images	JPEG, TIFF, PNG, BMP
Audio/Video	WAV, MP3, MP4, AVI, FLAC

Variables

The word variable comes from the word vary, meaning to alter or be different. Variables are those features that are not fixed but rather vary from circumstance to scenario. A variable is a number or characteristic that takes on different values from one observation, individual, place, or time to another. The possible values of a variable are termed variates.

For example, quantity is a variable, because it varies from commodity to commodity. Also price, age, weight, height, income, productivity etc. are variables as their values are not fixed. These variables are measured in different units. For example, quantity can be measured in units or tons, height can be measured in centimetres or inches, and age can be measured in years.

Variable Type	Detailed Explanation	Examples
Independent (IV)	Manipulated or selected by researcher to observe effects; cause in causal studies.	Dosage level, teaching method etc.
Dependent (DV)	Outcome affected by IV; measured to assess change.	Weight loss, test performance etc.
Control	Held constant to isolate IV effects; reduces noise.	Participant age range, lab conditions etc.
Moderator	Influences IV-DV strength/direction; interacts with IV.	Gender moderating therapy effect etc.
Mediator	Explains how/why IV affects DV; intermediate mechanism.	Attitude mediating ad exposure on purchase
Confounding	Uncontrolled variable correlating with IV/DV; biases results.	Exercise confounding diet-weight link etc.

8.4 RESEARCH DATA LIFE-STYLE; RESEARCH DATA MANAGEMENT; MEANING, DEFINITION, SCOPE, PURPOSE, ROLE OF LIBRARIAN

Research data is not something that comes and goes throughout a project but flows through a series of stages from initial planning to eventual re-use by others. This trend is commonly known as the research data life cycle. Knowing this lifespan helps researchers and librarians prepare for good research data management, sometimes referred to as RDM. RDM offers a way to manage data so that it is useable, trustworthy and safe over time.

Research Data Lifecycle

Data lifecycle runs from planning to data collection, preparation and analysis, archiving, publication, and finally reuse.

Planning: Researchers plan the study and develop a Data Management Plan (DMP) that describes data kinds, formats, storage, access, and future usage; frequently a requirement of funding agencies.

Data Collection (Create): Data is collected using techniques such as experiments, questionnaires, or observations. Data collection is done with ethical and legal standards.

Preparation and Analysis (Explore): Data is cleaned, sorted and analysed. Proper documentation is done for clarity and reproducibility.

Archiving (Preserve): Store valuable material in repositories with appropriate formats, metadata and access requirements for long term preservation.

Publication (Share): Data are given with identifiers (e.g. DOIs) and permissions to allow transparency and validation.

Reuse: Archived data is reused for study, education or verification. Promoted by projects such as European Open Science Cloud.

Research Data Management: Meaning and Definition

Research data management (RDM) is the set of actions that support data across all of these stages in the data lifecycle. According to Crystal Lewis, RDM is the arrangement, storage, preservation and dissemination of data generated or collected during a research endeavour. This definition points out that RDM is much more than merely storing datasets on a computer; it begins at the planning stage and extends long beyond the completion of data collecting. National Library of medicine defines RDM as the organisation, documentation, storage and preservation of data resulting from the research process, where data can be broadly defined as the outcome of experiments or observations that validate research findings and can take a variety of forms including numerical output

(quantitative data) and qualitative data, documentation, image, audio and video.

In other words, RDM is the systematic planning, organizing, documenting, protection and stewardship of research data It is about how raw data is processed to make clean, analysable datasets, and how that data is subsequently protected so that future users may appropriately interpret and utilize it. Good RDM: makes data usable, interpretable, trustworthy and reproducible.

In practice RDM is usually made up of:

- Development and writing of data management plans.
- Choosing formats and constructing folder structures
- Applying standards to file names, variables and metadata;
- Reproducible cleaning and transformation of data.
- Documenting all processing steps and decision making
- Storing data safely and backing it up.
- Preparation of data and documents for archiving and dissemination.
- Access rights and permissions, managing ethical restrictions

Scope

The RDM scope includes:

- **Type of data:** Data may be quantitative (numbers), qualitative (text or images), primary (collected by the researcher), secondary (re-used), or derived (produced by calculations or transformations).
- **Data formats and structures:** Tabular files (e.g. CSV), text files, relational databases, GIS files, or specific formats such as DICOM for medical pictures. The management strategy should outline acceptable formats and conversion methods for preservation.
- **Metadata and documentation:** Description of content, structure, way of data collection, units of measurement, coding schemes, significance of each variable and processing history. This also includes the contextual information such as study topics, instruments used and consent requirements.
- **Storage and backup:** What to do with data during and after the project: institutional servers, cloud services, repositories. Policies should stipulate frequency of backup, accountability and methods of recovery.
- **Security and access control:** Measures such as encryption, secure transfer protocols, and controlled access lists to safeguard sensitive data. This is particularly critical when it involves personal or sensitive data, including compliance with rules such as GDPR.

- **Preservation & reuse:** Long-term preservation approaches to ensure data is accessible and understandable in the future, including format migration, checksums for integrity, and ongoing repository support. It also includes preparing for licences that enable reuse, while respecting ethical and legal constraints

Purpose

The primary purpose of RDM is to

- **Save time:** With good data management you can find, interpret and reuse your own data without having to trawl through disorganised folders or recreate deleted files. Good organization, logical naming of files, and current documentation reduce the time spent on mundane chores and increase time for analysis and writing.
- **Improve workflow and standardise research practice:** When a research group decides on uniform protocols for formats, file names, documentation and backup, the whole workflow becomes smoother and more predictable. Having the same data for multiple team members makes it easier to work consistently, reducing errors and helping to sustain projects for a long time.
- **Meet funding or publishing requirements:** Many funding organizations and publications now require researchers to submit a data management plan, to keep data securely, and to make data available, where possible, in an appropriate repository. By following RDM best practices researchers can comply with these standards without last minute complications, improving the odds of successful grants and publications.
- **Promote responsible research:** Accurate, safe, and ethically and legally managed data, including privacy and confidentiality for human participants, are essential for responsible research. RDM supports these obligations by encouraging rigorous documentation, explicit consent procedures, proper access controls, and long-term retention of data required to verify results.
- **Enable new discoveries and public accessibility:** When data is carefully documented and shared through reliable repositories other researchers can mix it with their own work, test new questions, avoid repeating efforts. This improves the effect of the original project, fosters open science and makes publicly financed research more available to the wider community, including policy makers and the public.
- **Responding to retraction allegation:** Research Data Management (RDM) plays a critical role in **responding to retraction allegations** by providing verifiable evidence that supports the integrity and transparency of the research process. When a publication is questioned, well-managed research data can help determine whether the concerns stem from an honest error, methodological limitations, or research misconduct.

Role of Librarian

Librarians have a role in the research data lifetime (typically based on models like the USGS(U S Geological Survey) Data lifetime, which covers steps like planning, collection, quality assurance, preservation, discovery, and reuse).

- **Help with Data Management Plans (DMPs):** Many funding agencies (e.g. NSF, NIH or equivalents in other countries) require DMPs as part of grant applications. Librarians are there to help researchers write these plans, guiding them through data formats, storage, sharing regulations, intellectual property and ethical issues.
- **Data Curation and Repository Support:** Advising researchers on placing data in institutional or discipline repositories. This contains advise on file formats for long-term preservation, versioning and licensing (e.g. Creative Commons).
- **Metadata & Documentation Services:** Development or consultation on metadata schemas, data dictionaries and readme files to guarantee data is comprehensible and reusable by others.
- **Training and Education:** Creating and delivering workshops, webinars, online tutorials or embedded training on topics like data organization, version control (e.g. Git), safe storage, data cleaning, or technologies like DMPTool, Figshare or Zenodo.
- **Consultation and Referral:** Individual consultations on best practices in RDM. Often, they do not need to be subject experts in statistics or programming but function as skilled brokers that link researchers with specialized campus resources (IT, legal, ethical boards).
- **Policy Development and Advocacy:** Involvement in the development of institutional RDM policies, advocacy for open science, and consideration of concerns including as data sharing mandates, privacy (e.g., GDPR or IRB compliance), and equity in data access.
- **Collaboration and Team Science:** Working as part of research support teams, sometimes as “data librarians” or “informationists” engaged in projects.

CHECK YOUR PROGRESS

Fill in the Blanks

a) RDM involves the organisation, storage__ and dissemination of research study data.

b) In the research data lifecycle, ___is the stage in which data is prepared for long term preservation in a trusted repository.

c) Data lifecycle runs from planning to data collection, preparation and analysis, archiving, finally ___.

8.5 LET US SUM UP

- The unit introduces Research Data Management (RDM) which is a crucial component in managing data through its lifetime from planning and collecting to analysis, preservation, sharing and reuse. It includes data types (quantitative/qualitative, primary/secondary), formats (such as CSV, PDF/A), variables (independent, dependent, etc.). RDM helps keep data structured, recorded, safe and replicable, saving time, meeting funder requirements and enabling discoveries. Librarians help with data management plans, curation, training, metadata and policy lobbying. Good RDM helps keep research trustworthy and ethical in the digital age.

8.6 ANSWERS TO CHECK YOUR PROGRESS

a) Preservation, b) Archiving, c) publication and reuse.

8.7 FURTHER READINGS

1. Creswell, J. W. (2009). *Research Design : Qualitative, Quantitative, and Mixed Methods Approaches* (3rd ed.). Sage.
2. Kothari, C. R. (2004). *Research Methodology : Methods & Techniques* (2nd ed.). New Age International (P) Ltd., Publishers, Cop.
3. Lewis, C. (2024). *Data Management in Large-Scale Education Research*. https://doi.org/10.1201/9781032622835
4. *The Research Data Lifecycle*. (2024). Univie.ac.at. https://RDM.univie.ac.at/what-is-research-data-management/the-research-data-lifecycle/

8.8 MODEL QUESTIONS

a) What is research data management?
b) When does research data management begin and when does it end?
c) What is the Role of Librarians in RDM?
d) Discuss various aspects of data.
e) Discuss various stages of Research data lifecycle.
f) Discuss the scope and purpose of RDM.

*** ***** ***

UNIT 9: RDM SOFTWARE AND TOOLS

UNIT STRUCTURE

9.1 Learning Objectives

9.2 Introduction

9.3 RDM Software and Services; Commercial, Free and Open-Source Examples, Features and Benefits

9.4 Let Us Sum Up

9.5 Answers To Check Your Progress

9.6 Further Readings

9.7 Model Questions

9.1 LEARNING OBJECTIVES

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

- learn about popular RDM software
- commercial and free and open source RDM software and their features.

9.2 INTRODUCTION

Research Data Management (RDM) is now part of the research lifecycle ensuring that data are correctly organised, stored, conserved and shared according to ethical, regulatory and institutional criteria. With the growing emphasis on open science, data transparency and reproducibility, scientists are seeking specialized software and tools to handle their data efficiently**.** The unit provides yous with an overview of a selection of RDM software and services to support different stages of the data lifecycle, including planning, collecting, storage, sharing and preservation. This includes common tools such as DMPTool and DMPonline to create data management plans, REDCap for secure data collection, Open Science Framework (OSF) for collaborative research and data sharing, Registry of Research Data Repositories (re3data) to find suitable repositories, and figshare to share research outputs. These tools highlight how digital platforms help researchers fulfill funding demands, increase the visibility of their study, and facilitate data reuse. The unit also classifies RDM tools as commercial, free and open source and discusses their important features, advantages and disadvantages. Understanding these differences will enable yous to choose appropriate tools based on their research needs, institutional backing and resource availability.

9.3 RDM SOFTWARE AND SERVICES

Planning, organizing, storing, characterizing, sharing, and maintaining research data throughout its lifecycle are all made easier with the help of Research Data Management (RDM) tools and services. These tools facilitate the reuse and reproducibility of research outputs, assist researchers adhere to funder and institutional mandates, and preserve data quality.

RDM Tools

Applications and platforms that support particular phases of the research data lifecycle; planning, collection, storage, documenting, sharing, and preservation; make up RDM software. They can be displayed in subcategories:

- **Data Management Planning Tools:** These are the tools that help researchers prepare Data Management Plans (DMPs) aligned with funder and institutional requirements.

DMP TOOL (https://dmptool.org): It is a free, open-source program called the DMP Tool assists researchers in developing data management plans (DMPs). Many funding organizations demand these plans in order to submit grant proposals. A click-through method for constructing a DMP that satisfies funder requirements is offered by the DMP Tool. Additionally, it has tools for data management best practices, help text for questions, and direct access to funder websites.

The initial DMP Tool was a grassroots initiative that started in 2011 when eight universities collaborated to offer in-kind staff and development contributions. Funding organizations like the National Science Foundation and the National Institutes of Health demanded that researchers make plans for handling their research data, which prompted the initiative. As a result, by working together, the participating institutions were able to cut expenses and combine expertise in meeting data management objectives. The University of California Curation Center (UC3) at the California Digital Library, DataONE, Digital Curation Center (DCC-UK), Smithsonian Institution, University of California, Los Angeles Library, University of California, San Diego Libraries, University of Illinois, Urbana-Champaign Library, and the University of Virginia Library were the initial contributing institutions. The Alfred P. Sloan Foundation provided funds to the founding partners to develop a second version of the DMP Tool, which was published in 2014, due to the tool's first success.

The DMP Tool and its UK-based equivalent, DMPonline, have seen a surge in demand due to the global spread of open data regulations. In order to codevelop and maintain a single open-source platform, UC3 and DCC decided to formalize our relationship in 2016. We might expand our reach and advance best practices by offering a fundamental infrastructure for DMPs, enabling us to take part in a truly global open scientific ecosystem. The DMP Tool is still available for community assistance. The Editorial Board of the tool offers guidance and experience to keep grant requirements and related best practices up to date.

Key Features includes

- Step-by-step wizard for DMP creation
- Compliance with funder requirements
- Links to funder websites
- Help text and best practices guidance
- Institutional customization options
- Machine-Actionable DMPs: The DMP Tool now supports **machine-**actionable DMPs (maDMPs), which transform static plans into dynamic, interoperable systems. These enable metadata integration, real-time updates, and automated workflows.

How Does DMP works

The DMP Tool functions as a guided, interactive platform that helps institutions and academics create efficient Data Management Plans (DMPs). Because of its customization features, institutions can modify templates, guidelines, and resources to meet their unique needs.

The click-through wizard, which walks users through the process of establishing a DMP that satisfies funder requirements, is a crucial element. In order to help in appropriately answering questions, the tool also provides resources for Research Data Management (RDM) best practices, help text, and direct links to funding agency websites.

The application provides best practice assistance to guarantee quality and standardization, assisting researchers in efficiently organizing and optimizing their plans. In order to facilitate monitoring and integration, it also includes a system for registering a DMP-ID, which gives each data management plan a unique identification.

The technology also facilitates a collaborative environment in which participating firms can evaluate plans, offer suggestions for enhancements, and offer feedback. This ongoing feedback guarantees that the tool is current, pertinent, and in line with changing funder and research requirements.

DMPOnline (https://dmponline.dcc.ac.uk/)

DMPonline is an intuitive, web-based application intended to assist academics in formulating Data Management Plans (DMPs) mandated by funding agencies during proposal development and subsequent to funding authorization. It offers funder-specific templates and best practice recommendations, allowing researchers to create high-quality and compliant Data Management Plans (DMPs).

The tool provides comprehensive tailored recommendations and illustrative responses, given by entities including the Digital Curation Centre (DCC), research funding bodies, and academic institutions. It enables users to examine publicly accessible DMPs, which act as benchmarks for formulating their own plans.

DMPonline operates via a template-driven approach, wherein users respond to preliminary inquiries to ascertain the suitable template according to funder or institutional stipulations. Assistance is offered at every stage to aid users in understanding and responding appropriately.

The initiative is underpinned by a robust community-driven framework, including complimentary access to researchers, a responsive helpdesk service, and proactive user involvement to enhance the tool perpetually.

DMPonline offers customisation and subscription functionalities for institutions and funders. These encompass the creation and modification of templates, provision of customized instruction, branding of the tool, monitoring of user statistics, utilization of APIs for data extraction, and allocation of user permissions. Various subscription tiers are offered according to institutional requirements. DMPonline is crucial for enabling effective research data management planning and ensuring adherence to funder stipulations.

Data Collection and Capture Tools: These are the applications for structured, secure data collection, especially for surveys, clinical and social science data.

The Open Science Framework (OSF) (https://osf.io/): It is a free, open-source online platform developed by the Center for Open Science. It enables researchers to efficiently organize, collaborate, store, and share different components of their research projects in one place.

OSF provides personal dashboards that allow users to manage multiple projects, whether private or public. It supports collaboration within and across institutions, making it easier to share data and work with research teams. The platform includes version control features to track changes and maintain records, along with access control options at both project and file levels.

Additionally, OSF assigns persistent identifiers (DOIs) to research outputs and offers a system to register and share work at any stage whether ongoing or completed. It also integrates with external storage services such as GitHub, Google Drive, and Dropbox, allowing users to bring together files, data, and code from different sources.

Overall, OSF is designed to fit seamlessly into researchers' workflows and supports flexible sharing either privately among collaborators or publicly with proper citation and identification.



Source: (https://osf.io/)

RedCap: REDCap is a secure web-based platform designed for constructing surveys and is widely recognized for gathering clinical trial data. REDCap was established in 2004 at Vanderbilt University and is currently endorsed by a worldwide partnership of non-profit organizations. REDCap has a longstanding reputation for gathering clinical trial data, however it is applicable for all types of surveys. It is offered in two variants: a licensed edition with IT support and a complimentary institutional version backed by user groups.

Access and Configuration of RedCap: Researchers must obtain access through ethical permission and necessary documentation. Upon approval, they are designated as project managers to develop and oversee research data.

Ket Features include

- Digital and physical instruments (Online Designer & Data Dictionary) for survey development

- Data validation (automated error verification)
- Various instruments for systematic data collection
- Generation of reports and fundamental statistical analysis
- Export data for sophisticated analysis (e.g., STATA)
- Data Protection and Confidentiality

REDCap guarantees safe data storage, access via login credentials, and permissions depending on user roles. It facilitates anonymization, pseudonymization, and electronic permission, while preserving logs of all activities for transparency. How it has limitations including

- Inappropriate for prolonged data retention
- Potential for data loss if not adequately backed up
- Researchers are required to export and securely store anonymized data post-utilization. Data Management Accountability.

Researchers must ensure secure storage, ethical management, and permanent deletion of data at project conclusion, adhering to institutional and legal protocols.

is a comprehensive tool that helps not only in data capture, but also provides solution to store and collab with existing research data.

- **Data repositories and sharing platforms:** These are software and hosted services used for publishing and preserving datasets and related research outputs.

re3data.org: It is an international registry of research data repositories encompassing all academic fields. It assists researchers in locating appropriate repositories for the storage, sharing, and discovery of datasets, including experimental data, surveys, multimedia files, software, and scientific data. Researchers utilize re3data to identify suitable repositories for data deposition, whereas funders and publishers employ it to verify adherence to data sharing mandates. It is beneficial for librarians, educators, and students to facilitate research, learning, and data literacy.

The platform enables users to search and filter repositories according to criteria such as certification, metadata standards, open access, and persistent identifiers. Results are presented with icons for enhanced comprehension.

re3data fosters transparency by offering metadata as open data (CC0) and facilitates integration via APIs and standardized formats such as XML. It is administered by entities including the Karlsruhe Institute of Technology and Purdue University Libraries, under the direction of a working group and an editorial board.



Source: https://www.re3data.org/

Figshare: It is an open research repository platform that provides infrastructure for storing, managing, and sharing research outputs. It enables researchers and institutions to make their work discoverable, citable, and transparent.

The platform supports a wide range of research materials, allowing users to showcase and disseminate outputs effectively. It also helps in tracking and reporting research impact, promoting open science and accessibility.

Institutional and open source repository software: These are the software that institutions deploy for local repositories and data services. Examples includes DSpace, Dataverse, EPrints, Invenio, etc., often customized by university libraries for datasets and other research outputs.

Key features include:

- Flexible metadata schemas,
- OAI PMH support, integration with identity systems,
- configurable workflows, and support for long term preservation policies.

Some other tools include

Dataverse is an open-source repository used by academic institutions for dataset sharing, citation, and long-term preservation.

CKAN is an open-source system for publishing and discovering large datasets, widely used in open data initiatives.

LabArchives provides electronic lab notebooks with data management features, especially useful in laboratory-based research.

Dryad is a curated repository for datasets linked to scholarly publications, ensuring

quality and long-term preservation.

Synapse is designed for collaborative biomedical research, offering strong data governance and controlled access features.

Zenodo is an open-access repository that allows free sharing of datasets, software, and publications with DOI assignment.

ICPSR specializes in social science data, providing curated datasets and secure access for long-term use.

CHECK YOUR PROGRESS

FILL IN THE BLANKS

a. _____ ensures that research data are correctly, stored, conserved, and shared.

b. ____is an international registry that helps researchers identify suitable research data.

c. The ___ supports project organization, collaboration, storage, and of research components.

d. LabArchives is _____ software

9.4 LET US SUM UP

- Research Data Management (RDM) is an essential part of the research lifecycle, ensuring that data are systematically planned, organized, stored, shared, and preserved in line with ethical and institutional requirements. This unit introduces major RDM tools and services that support key stages such as planning, collection, collaboration, storage, and dissemination of data. Yous are exposed to widely used platforms including DMPTool, DMPonline, REDCap, the Open Science Framework (OSF), re3data, figshare, and institutional/open-source repositories like DSpace and Dataverse. The unit also distinguishes between commercial, free, and open source tools, highlighting their features, strengths, and limitations to help researchers select appropriate solutions based on their needs and available support

9.5 ANSWERS TO CHECK YOUR PROGRESS

a) RDM, **b)** re3data.org, **c)** Open Science Framework, **d)** Proprietary

9.6 FURTHER READINGS

1 DFREE. (2018, April 17). *Keeping Up With… Research Data Management*.

Association of College & Research Libraries (ACRL). https://www.ala.org/acrl/publications/keeping_up_with/RDM

2 Dube, T. V. (2025). Research data management in academic libraries: institutional repositories as a reservoir for research data. *Library Management*. https://doi.org/10.1108/lm-06-2024-0070

3 *Library Guides: Research Data Management: REDCap for data management*. (2025). Rcsi.ie. https://libguides.rcsi.ie/RDM/redcap-for-datamanagement

4 *Research Data Management*. (2026). OCLC. https://www.oclc.org/research/areas/research-collections/RDM.html

9.7 MODEL QUESTIONS

1. What is the main purpose of using DMPTool or DMPonline in the research data lifecycle?
2. How does the Open Science Framework (OSF) support collaboration and reproducibility in research projects?
3. Why is it important to classify RDM tools as commercial, free, and open source for researchers and institutions?
4. Explain the role of RDM software and tools in supporting different stages of the research data lifecycle, using examples such as DMPTool, REDCap, OSF, re3data, and figshare.

*** ***** ***

UNIT 10: RDM IN ACADEMIC LIBRARIES

UNIT STRUCTURE

10.1 Learning Objectives
10.2 Introduction
10.3 Roles of Academic Libraries
10.4 Skills Requirements for Research Data Management
10.5 Best Practices and Case Studies
10.6 Let Us Sum Up
10.7 Answers To Check Your Progress
10.8 Further Readings
10.9 Model Questions

10.1 LEARNING OBJECTIVES

After studying this unit, you will be able to:

- understand the concept and importance of research data management (RDM) in academic libraries, including the roles and responsibilities of libraries in managing research data.
- identify and explain the essential skills required for RDM, along with best practices for organizing, storing, sharing, and preserving research data effectively.
- analyse real-life case studies of RDM implementation in academic libraries and apply this knowledge to develop practical approaches for managing research data in institutional settings.

10.2 INTRODUCTION

Research Data Management (RDM) refers to the systematic organization, storage, preservation, and sharing of data generated throughout the research process. In the modern academic environment, research data has emerged as a vital scholarly output alongside traditional publications such as journal articles and books. With the growing emphasis on open science, transparency, and reproducibility, RDM has become an essential component of research practices in academic institutions.

RDM involves a structured approach to handling data from its creation to its long-term preservation and reuse. It includes activities such as data planning, documentation, storage, backup, preservation, and sharing. Proper data management ensures that research data remains accurate, secure, and easily accessible, thereby supporting

validation and replication of research findings.

Academic libraries, which have traditionally focused on managing books and digital information resources, are now expanding their roles to support research data across its entire lifecycle. Libraries are well-positioned to provide RDM services due to their expertise in metadata standards, information organization, digital preservation, and access management. They assist researchers in organizing data, applying appropriate metadata, and selecting suitable repositories for data storage and sharing.

Effective RDM enhances the visibility and impact of research by enabling data reuse and promoting collaboration among researchers. It also helps institutions comply with funding agency requirements and ethical standards related to data handling. Overall, RDM plays a crucial role in improving research quality, ensuring long-term data accessibility, and advancing knowledge in the academic community

10.3 ROLES OF ACADEMIC LIBRARIES

Academic libraries play an important role in managing research data. Earlier, libraries mainly managed books and journals, but now they also help researchers manage their research data. Libraries support researchers at every stage of the research process, from planning to sharing data. Their roles can be understood as follows:

a) Policy Development and Advocacy: Libraries help institutions create rules and policies for managing research data. They ensure that researchers follow the requirements of funding agencies, especially related to data sharing and preservation. Libraries also promote open data, which means making research data available to others. They guide researchers about copyright, licensing, and ethical use of data through workshops and awareness programs.

b) Data Management Planning (DMP) Support: Libraries assist researchers in preparing Data Management Plans (DMPs), which are often required for research projects. They help in deciding what type of data will be collected, how it will be stored, and how it will be shared. This ensures that data is properly managed from the beginning.

c) Metadata Creation and Standardization: Libraries help in describing data using metadata (data about data). Proper metadata makes data easy to find, understand, and reuse. Librarians guide researchers in using standard formats like Dublin Core.

d) Data Storage and Repository Management: Libraries provide safe storage for research data through institutional repositories. They also help researchers choose suitable repositories. Libraries assign identifiers like DOIs, which help in citing and accessing

data easily.

e) Data Curation and Preservation: Libraries check the quality of data, improve documentation, and ensure that data remains usable over time. They also help in preserving data by storing it in reliable formats and systems.

f) Training and Capacity Building: Libraries organize training programs, workshops, and seminars to teach researchers about RDM practices, tools, and data sharing methods.

g) Research Support Services: Libraries provide guidance on organizing data, handling sensitive information, and using tools for data analysis and visualization.

h) Collaboration with Stakeholders: Libraries work together with IT staff, research offices, and faculty members to develop a proper system for managing research data.

In simple terms, academic libraries act as partners in research by helping researchers manage, store, and share their data effectively.

10.4 SKILLS REQUIREMENTS FOR RESEARCH DATA MANAGEMENT

Research Data Management (RDM) in academic libraries requires librarians to learn new skills in addition to their traditional library knowledge. As libraries now support research activities, librarians need technical, analytical, and communication skills to manage research data effectively.

a) Technical skills: Librarians should understand different types of data formats such as CSV, XML, and JSON. They need basic knowledge of databases and data repositories used in academic institutions. Familiarity with simple programming languages like Python or R and data management tools helps in organizing and handling research data more efficiently.

b) Metadata and cataloguing skills: Librarians must know how to describe data properly using metadata (data about data). This helps in making datasets easy to find and use. Knowledge of standard systems like controlled vocabularies and metadata standards ensures better organization and retrieval of data.

c) Data curation skills: This includes cleaning data, checking for errors, and maintaining quality. Librarians also help in documenting data clearly using files like README. They must understand how to preserve data so that it can be used in the future.

d) Legal and ethical knowledge: Librarians should be aware of copyright laws and licenses such as Creative Commons. They must also understand data privacy and ensure

that sensitive or confidential data is handled carefully and ethically.

e) Communication and training skills: Librarians need to teach researchers and students about RDM practices. They conduct workshops, training sessions, and provide guidance on how to manage and share research data properly.

f) Project management skills: Managing RDM services requires planning and coordination. Librarians must work with researchers, IT staff, and institutions while managing time, resources, and tasks effectively.

g) Analytical and critical thinking skills: Librarians should be able to evaluate data quality, check repository standards, and help in developing data-related policies.

h) Domain knowledge: Understanding subject-specific research practices helps librarians provide better support to researchers in different fields.

Overall, the development of these skills enables academic librarians to effectively support research data management and contribute to the advancement of research quality and scholarly communication.

10.5 BEST PRACTICES AND CASE STUDIES

Implementing best practices in Research Data Management (RDM) is essential for ensuring efficient, reliable, and sustainable handling of research data in academic libraries. These practices support the entire research lifecycle and enhance data quality, accessibility, and reuse.

a) Data Management Planning: Academic libraries guide researchers in preparing a Data Management Plan (DMP) at the beginning of a research project. Librarians also help in updating the plan throughout the research lifecycle, ensuring proper data handling, storage, and sharing strategies.

b) Data Organization and Documentation: Libraries promote the use of consistent file naming conventions and proper folder structures to organize data systematically. They encourage version control to track changes and support the creation of detailed documentation, such as README files, to improve data understanding and reuse.

c) Metadata Standards: Librarians assist researchers in applying standardized metadata schemas to ensure datasets are easily discoverable and interoperable. Proper metadata enhances visibility and facilitates data sharing across platforms.

d) Storage and Backup: Academic libraries provide or recommend secure storage systems, including institutional repositories and cloud services. They emphasize the importance of regular backups, both locally and remotely, to prevent data loss.

e) Data Security and Privacy: Libraries ensure that sensitive data is protected through encryption, access controls, and compliance with legal and ethical requirements. They guide researchers in handling confidential information responsibly.

f) Data Sharing and Accessibility: Libraries encourage researchers to share data in open repositories whenever possible. They assist in selecting appropriate licenses and defining access conditions to promote transparency and collaboration.

g) Long-term Preservation: Libraries support the use of sustainable file formats and ensure regular migration of data to updated formats. They also maintain trusted repositories for long-term data preservation.

h) Use of Persistent Identifiers: Assigning persistent identifiers such as DOIs helps in proper citation and long-term access to datasets, a service often facilitated by academic libraries.

i) Documentation and Transparency: Libraries emphasize recording research methodologies and processes to ensure reproducibility and credibility of research outcomes.

In simple terms, these best practices help academic libraries ensure that research data is well-organized, secure, accessible, and reusable, thereby improving the quality, transparency, and impact of research.

Case Studies of RDM in Academic Libraries

a) University of Edinburgh (United Kingdom)

Context: A global pioneer that established one of the first formal institutional RDM policies in 2011.

Strategy: The Roadmap Approach. The library focused on a top-down strategy, ensuring that university policy, technical tools, and human support services (like help desks) were all aligned and launched together.

Key Innovation: The creation of Edinburgh DataShare, a multidisciplinary repository, and its integration with DMPonline. This allowed researchers to create Data Management Plans (DMPs) that connected directly to the storage system.

Outcome: High researcher compliance with international funder mandates. It successfully shifted the academic culture so that data is now viewed as a primary scholarly output, similar to a published journal article.

b) Monash University (Australia)

Context: Known for the Monash Model, which balances technical IT power with the specialized expertise of librarians.

Strategy: A Federated Model. Instead of one department doing everything, the Library

manages soft infrastructure (metadata standards, training, and ethics), while the e-Research Centre manages hard infrastructure (high-speed servers and storage).

Key Innovation: Store.Monash Portal and Bridges (General Data Repository). This system allows researchers to move their active data to a public repository with a single click, automatically generating the necessary descriptions (metadata).

Outcome: It significantly reduced the administrative burden on researchers. By making data easy to publish, Monash improved the global findability and citation rates of its institutional data.

c) Indian Institute of Technology (IIT) Delhi (India)

Context: A premier technical institute that links RDM to its goals for Open Science and global research rankings.

Strategy: Outcome-Based Integration. The Central Library acts as a central hub, connecting RDM services with the institution's Intellectual Property Rights (IPR) cell and research impact analysis.

Key Innovation: Using DSpace (an open-source platform) to expand the Institutional Repository. They also conduct mandatory RDM literacy workshops for PhD scholars to teach the FAIR principles (Findable, Accessible, Interoperable, Reusable).

Outcome: Increased visibility of Indian engineering research and better preparation of scholars for international collaborations where strict data sharing is required.

d) Jawaharlal Nehru University (JNU) (India)

Context: A leading model for RDM in social sciences and humanities, where data is often non-numerical (like interviews or field notes).

Strategy: The Data Curation Focus. The library emphasizes the curation of diverse data types, including qualitative interviews, GIS maps, and government archives.

Key Innovation: The JNU Data Portal. This is a single-window interface that gives researchers access to both subscribed datasets and open-source national data, such as Census or NSSO reports.

Outcome: Established the library as a specialized data hub for national policy-related research and improved data discovery for social science scholars across India.

d) University of Oxford (United Kingdom)

Context: Oxford emphasizes RDM as a core part of research integrity and institutional prestige.

Strategy: The Cross-Departmental Support Model. RDM is not just a library task but a shared responsibility between the Bodleian Libraries, IT Services, and the Research Support Office.

Key Innovation: ORA-Data (Oxford Research Archive for Data). This is a dedicated archive for researchers to deposit datasets that underpin their publications, ensuring long-term preservation and permanent links (DOIs).

Outcome: Established a unified portal called “Research Data Oxford” which provides a single point of entry for training, ethics guidance, and data storage tools.

e) Harvard University (United States of America)

Context: Harvard is the home of the Dataverse Project, which has set the global standard for open-source data repositories.

Strategy: Global Open Access. The library supports a model that encourages researchers from any discipline to share their data publicly to increase the impact of their work.

Key Innovation: Harvard Dataverse. A globally recognized open data repository that allows researchers to create their own “Dataverse” (a mini-repository) to manage, share, and cite their datasets with professional metadata standards.

Outcome: Promoted “Open Science” by making millions of datasets findable and reusable by the global scientific community.

f): Massachusetts Institute of Technology (MIT) (United States)

Context: MIT focuses on the technical lifecycle of digital objects and long-term digital preservation.

Strategy: The Lifecycle Management Model. MIT Libraries collaborate deeply with technical departments to ensure that research data remains readable even as technology changes over decades.

Key Innovation: DSpace@MIT. MIT was one of the co-developers of DSpace, the software now used by most Indian libraries. They use it to store everything from technical reports and datasets to video and audio files.

Outcome: Created a “perpetual” archive where research outputs are never lost, serving as a model for how libraries can act as digital preservation centres.

g) Indian Institute of Science (IISc) (India)

Context: As India’s top research university, IISc has modernized its library into a high-tech data hub.

Strategy: The “Single Window” Data Governance Model. The library recently initiated the “Office of Data,” which treats institutional data as a strategic asset for ranking and research excellence.

Key Innovation: ePrints@IISc and etd@IISc. These are specialized repositories that capture every publication and thesis produced by the institute. They also lead in “Faculty

Profiling," linking a researcher's data directly to their professional profile and citations.

Outcome: IISc became the first university in India to have a dedicated Office of Data, placing the library at the centre of institutional decision-making and global ranking strategies (like NIRF and QS).

CHECK YOUR PROGRESS

A. Fill in the blanks

a) Research Data Management ensures that research data is _____, accessible, and reusable. *(Answer:)*

b) A Data Management Plan (DMP) is usually required by _____ agencies. *(Answer: funding)*

c) Metadata helps in making research data _____ and easy to retrieve. *(Answer: discoverable)*

10.6 LET US SUM UP

- Research data management is a critical component of modern academic research. Academic libraries are evolving from traditional information providers to active partners in the research process. By offering RDM services, libraries enhance research quality, promote transparency, and support compliance with funding requirements.
- The successful implementation of RDM requires a combination of technical expertise, institutional support, and cultural change. Librarians must develop new skills and collaborate with researchers and its professionals to build sustainable RDM infrastructures.
- Adopting best practices and learning from global case studies can help academic libraries effectively manage research data and contribute to the advancement of knowledge

10.7 ANSWERS TO CHCEK YOUR PROGRESS

a) Organized, B) FUNDING, C) DISCOVERABLE

10.8 FURTHER READINGS

1. Cox, a. M., & verbaan, e. (2018). Exploring research data management. Facet publishing.

2. Pryor, g., jones, s., & whyte, a. (eds.). (2013). Delivering research data management services: fundamentals of good practice. Facet publishing.
3. Borgman, c. L. (2015). Big data, little data, no data: scholarship in the networked world. Mit press.
4. Koltay, t. (2021). Research data management and data literacies. Elsevier.

10.9 MODEL QUESTIONS

a) What is Research Data Management (RDM)?

b) What is a Data Management Plan (DMP)?

c) Mention any two roles of academic libraries in RDM.

d) Explain the roles of academic libraries in Research Data Management (RDM).

e) Discuss the skills required for effective RDM services in academic libraries.

f) Describe the best practices in Research Data Management with suitable examples.

*** ***** ***

UNIT 11: RDM TRAINING

UNIT STRUCTURE

11.1 Learning Objectives
11.2 Introduction
11.3 Data Literacy: program design, workshops
11.4 Librarian roles, evaluation methods; Promoting RDM culture.
11.5 Let Us Sum Up
11.6 Answers to Check Your Progress
11.7 Further Readings
11.8 Model Questions

11.1 LEARNING OBJECTIVES

After studying this unit, you will be able to

- explain the concept of data literacy and its key components in the context of Research Data Management (RDM).
- develop and design RDM training programs, including workshops and instructional strategies suited to different user groups.
- describe and critically examine the roles of librarians in delivering RDM training and supporting research data services.
- apply appropriate methods to evaluate RDM training programs and suggest strategies for promoting an effective RDM culture in academic institutions.

11.2 INTRODUCTION

RDM training focuses on equipping researchers, students, and faculty members with the knowledge and skills required to manage research data effectively throughout its lifecycle. In the present digital research environment, there is an increasing need for structured training that helps users handle data in a systematic, organized, and responsible manner. Such training enables individuals to develop competencies in data organization, documentation, storage, sharing, and preservation.

RDM training also supports researchers in meeting institutional and funding requirements while improving the quality, transparency, and reproducibility of research outcomes. Libraries play a central role in delivering RDM training due to their expertise in information organization, metadata standards, digital preservation, and user education. Through workshops, seminars, and guidance services, libraries help build data literacy

and promote better data management practices within academic and research communities.

11.3 DATA LITERACY: PROGRAM DESIGN, WORKSHOPS

Concept of Data Literacy

Data literacy refers to the ability to understand, interpret, and evaluate data in a meaningful way. It also includes the skills required to manage data effectively, use it ethically and responsibly, and communicate insights derived from data clearly. In the modern research environment, data literacy has become an essential competency for students, researchers, and professionals.

In the context of Research Data Management (RDM), data literacy plays a crucial role in ensuring that research data is handled properly throughout its lifecycle. It enables researchers to organize and document their datasets systematically, which improves accessibility and usability. Data literacy also helps in applying appropriate metadata standards, ensuring that datasets are well-described and easily discoverable. Furthermore, it contributes to maintaining data quality and supports reproducibility of research results. Another important aspect is the ability to share data through repositories in a structured and compliant manner.

Components of Data Literacy

Data literacy consists of several key competencies that support effective data management.

i. Data discovery involves the ability to locate relevant datasets from various sources such as repositories, databases, and archives.

ii. Data organization refers to structuring files properly using consistent naming conventions and folder systems.

iii. Metadata creation involves describing data accurately using standard formats to enhance discoverability.

iv. Data analysis awareness provides a basic understanding of how data is processed and interpreted.

v. Data ethics ensures responsible handling of data, including issues of privacy and intellectual property.

vi. Data sharing focuses on understanding policies related to open access and licensing, and

vii. Data preservation involves knowledge of long-term storage and backup strategies.

Program Design for Data Literacy

Designing a data literacy program requires a systematic, well-planned, and user-focused approach. The main aim of such a program is to develop the ability of yous to understand, manage, and use data effectively in academic and research activities. The program should be simple, practical, and adaptable so that it can meet the needs of different types of yous with varying levels of knowledge and experience.

The first step in designing the program is identifying the target users. Different groups of users have different requirements and expectations. Undergraduate students generally need a basic understanding of data concepts, such as what data is and how it is used. Researchers, on the other hand, require more advanced knowledge related to data handling, analysis, and documentation. Faculty members are more concerned with data management practices, data sharing, and compliance with institutional and funding requirements. Therefore, it is important to design the program according to the needs of each group.

The next important step is conducting a needs assessment. This helps in understanding the existing skills and identifying gaps in knowledge related to data literacy. It also involves analyzing discipline-specific data needs, as different subject areas deal with different types of data. In addition, institutional requirements such as policies, infrastructure, and available resources must be considered. Methods such as surveys, interviews, and focus group discussions are commonly used to collect this information and ensure that the program is relevant and effective.

Setting clear learning objectives is another essential part of program design. These objectives guide the entire training process and help in measuring its success. The objectives may include understanding the importance of data, learning how to organize and manage data properly, applying metadata standards, and sharing data responsibly. These objectives should follow the SMART principle, meaning they should be specific, measurable, achievable, relevant, and time-bound.

Curriculum design involves organizing the content in a logical and structured manner. The program should cover key topics such as data discovery, data organization, metadata creation, data ethics, data sharing, and data preservation. The content should progress from basic to advanced levels and include real-life examples to make learning more meaningful.

Effective teaching methods are also important for successful learning. Lectures can be used to explain basic concepts, while group discussions encourage interaction among participants. Case studies help in understanding real-life situations, and hands-on

exercises allow yous to practice and develop their skills.

Finally, the mode of delivery should be flexible. Data literacy programs can be conducted through face-to-face classroom sessions, online courses using learning platforms, or blended learning that combines both methods. This flexibility helps in reaching a wider audience and accommodating different learning preferences.

Workshops for Data Literacy

Workshops are one of the most effective methods for teaching data literacy because they provide a practical and interactive learning environment. They allow participants to actively engage with data concepts and develop hands-on skills that are essential for effective data management.

Types of Workshops

Workshops can be organized at different levels depending on the knowledge and needs of participants.

- Introductory workshops focus on basic concepts of data, helping beginners understand what data is and how it is used.
- Intermediate workshops cover topics such as metadata creation and data documentation, which require a deeper understanding.
- Advanced workshops are designed for experienced yous and include training on data analysis, visualization, and the use of specialized tools.

Workshop Structure

A well-designed workshop follows a clear and organized structure.

- It usually begins with an introduction that provides an overview of the topic and its importance.
- This is followed by an explanation of basic concepts in a simple and understandable manner.
- Demonstration sessions are included to show how tools or techniques work in practice.
- Hands-on practice is a key part of workshops, where participants perform activities themselves.
- Finally, a discussion session allows participants to ask questions and clarify their doubts.

Key Features

Effective workshops are practical and activity-based, allowing yous to apply their knowledge. They use real-life examples to make concepts easier to understand. Interactive

participation is encouraged so that yous remain engaged throughout the session. Step-by-step guidance is also provided to ensure that participants can follow and learn effectively.

Workshop Activities

During workshops, participants are involved in various activities such as organizing datasets using proper file naming conventions, creating metadata for datasets, preparing simple data management plans, and uploading data to repositories. These activities help in developing practical skills.

Use of Technology

Technology plays an important role in enhancing workshops. Tools such as PowerPoint presentations, data management software, online repositories, and video tutorials are commonly used to support learning.

Benefits

Workshops help in improving practical skills, encouraging active learning, providing immediate feedback, and building confidence in handling data. Overall, they are essential for developing strong data literacy and supporting effective research practices.

11.4 LIBRARIAN ROLES, EVALUATION METHODS; PROMOTING RDM CULTURE

Librarians play a vital role in Research Data Management (RDM) training and support within academic and research institutions. Their responsibilities have expanded beyond traditional library services to include data-related guidance and training.

As educators and trainers, librarians conduct workshops and seminars to teach users about data management practices. They also develop training materials and provide one-on-one consultations to help researchers understand specific data-related issues.

Librarians also act as data management consultants. In this role, they assist researchers in preparing Data Management Plans (DMPs), selecting suitable repositories for data storage, and applying appropriate metadata standards to ensure proper documentation and accessibility.

Another important role of librarians is that of advocates of open science. They promote open access publishing, encourage data sharing practices, and support the implementation of FAIR principles, which ensure that data is findable, accessible, interoperable, and reusable.

As metadata specialists, librarians design metadata schemas, ensure that datasets are easily discoverable, and maintain the quality and consistency of data. They also serve as repository managers by managing institutional repositories, ensuring long-term preservation of data, and providing access to datasets for users.

Additionally, librarians contribute as policy developers by helping to formulate institutional RDM policies, data sharing guidelines, and ensuring compliance with funder requirements.

Evaluation Methods of RDM Training

Evaluation is an essential part of Research Data Management (RDM) training, as it helps measure the effectiveness and impact of the program. It ensures that the training meets its objectives and provides useful outcomes for participants.

Importance of Evaluation

Evaluation helps in determining whether the training program is effective or not. It identifies the strengths and weaknesses of the training and provides insights for improvement. It also helps in enhancing the quality of future programs by addressing gaps and user needs.

Types of Evaluation

There are two main types of evaluation used in RDM training.

- Formative evaluation is conducted during the training process to improve teaching methods and delivery. It allows trainers to make immediate adjustments.
- Summative evaluation is conducted after the training to assess overall outcomes and measure learning achievements.

Evaluation Techniques

Various techniques are used to evaluate RDM training.

- Feedback forms are commonly used to collect participants' opinions about content, delivery, and relevance.
- Pre- and post-tests help measure knowledge gained during the training.
- Observation is used to assess participant engagement and participation.
- Assignments and practical tasks evaluate the development of skills.
- Interviews and focus group discussions provide detailed insights into participants' experiences.

Key Performance Indicators

Key Performance Indicators (KPIs) are important tools used to measure the success and effectiveness of Research Data Management (RDM) training programs. They provide measurable evidence of how well the training has achieved its objectives.

Common KPIs include the number of participants trained, which reflects the reach of the program. Improvement in data management skills can be assessed through tests, assignments, or practical activities completed by participants. Another key indicator is the increased use of institutional or subject repositories, which shows that participants

are applying what they have learned. The number of Data Management Plans (DMPs) created after training is also a useful measure of success. In addition, overall user satisfaction, gathered through feedback forms, helps in understanding participants' experiences and perceptions of the training.

Continuous Improvement

Continuous improvement is an essential part of RDM training programs. The results obtained from evaluation and KPIs should be carefully analyzed to identify areas that need enhancement. Training content can be updated to include new tools, technologies, and best practices in data management. Teaching methods can be improved by incorporating more interactive and practical approaches.

Additionally, feedback from participants helps in addressing their specific needs and expectations. Continuous improvement ensures that RDM training remains relevant, effective, and aligned with the evolving requirements of the research community.

Promoting RDM Culture

Promoting a strong Research Data Management (RDM) culture is essential for ensuring sustainable, transparent, and effective research practices in academic and research institutions. A well-developed RDM culture improves the quality of research data, enhances accessibility, and ensures long-term usability and preservation. It also supports open science and increases the visibility and impact of research outputs.

Meaning of RDM Culture

RDM culture refers to the awareness, understanding, and acceptance of good data management practices among researchers, students, and institutions. It involves proper planning, organization, documentation, sharing, and preservation of research data. It also reflects an institutional commitment to managing data responsibly and integrating RDM practices into everyday research workflows. When a strong RDM culture exists, researchers follow standard practices naturally, without external pressure, and recognize the value of managing data effectively.

Strategies to Promote RDM Culture

- **Awareness programs**: They play a crucial role in promoting RDM culture. Activities such as seminars, workshops, conferences, and training sessions help researchers understand the importance of data management. Campaigns, posters, and outreach initiatives further increase awareness and encourage participation.
- **Policy Development:** Institutions can strengthen RDM culture by developing clear policies and guidelines. Institutional RDM policies provide a framework for managing

and sharing data. Mandates from funding agencies also encourage researchers to adopt proper data management practices.

- **Integration into Curriculum:** Introducing RDM concepts into academic curricula ensures that students develop data literacy skills from an early stage. Including RDM training in undergraduate and postgraduate programs helps build long-term awareness and prepares future researchers.
- **Incentives for researchers:** Providing incentives such as recognition, awards, and citation of datasets motivates researchers to adopt good data practices. Acknowledging data as a valuable research output encourages sharing and proper management.
- **Collaboration**: This among libraries, IT departments, and research offices is essential for effective RDM implementation. These stakeholders work together to provide infrastructure, technical support, and training services.

CHECK YOUR PROGRESS

Fill in the Blanks

a) Data literacy is the ability to understand, interpret, and _____ data effectively.

b) A Data Management Plan is commonly known as a_____.

c) The FAIR principle stands for Findable, Accessible, Interoperable, and _____.

d) Metadata helps in improving the _____ of data.

11.5 LET US SUM UP

- RDM training is a vital component of modern library and information science. It ensures that researchers can manage their data effectively, comply with institutional and funder requirements, and contribute to the advancement of knowledge.
- Data literacy forms the foundation of RDM training, while well-designed programs and workshops ensure effective knowledge transfer. Librarians play a central role as educators, consultants, and advocates. Evaluation methods help improve training outcomes, and promoting an RDM culture ensures long-term sustainability.

  In the digital age, libraries are no longer just repositories of books but are evolving into dynamic centres for data management and research support. By actively engaging in RDM training, libraries can significantly contribute to the research ecosystem and knowledge society.

11.6 ANSWERS TO CHECK YOUR PROGRESS

a) Analyze. b) DMP, c) Reusable, d) discoverability

11.7 FURTHER READINGS

1. Cox, A. M., & Verbaan, E. (2018). *Exploring research data management*. Facet Publishing.
2. Pryor, G., Jones, S., & Whyte, A. (Eds.). (2013). *Delivering research data management services: Fundamentals of good practice*. Facet Publishing.
3. Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. MIT Press.
4. Koltay, T. (2021). *Research data management and data literacies*. Elsevier.

11.8 MODEL QUESTIONS

a) What is data literacy? Explain its importance in RDM.
b) What are the key components of data literacy?
c) What is the role of librarians in RDM training?
d) What are the main strategies for promoting RDM culture? Discuss the design and development of a data literacy program in detail.
e) Explain the role of workshops in promoting data literacy with suitable examples.
f) Describe the evaluation methods of RDM training programs.

*** ***** ***

UNIT 12: RDM POLICIES

UNIT STRUCTURE

12.1 Learning Objectives
12.2 Introduction
12.3 Institutional Policies, development, funder mandates
12.4 Compliance, legal guidelines, Auditing, improvement
12.5 Let Us Sum Up
12.6 Answers to Check Your Progress
12.7 Further Readings
12.8 Model Questions

12.1 LEARNING OBJECTIVES

After studying this unit, you will be able to:

- understand and explain the concept of Research Data Management (RDM) policies, including institutional policies, funder mandates, and their significance in libraries.
- analyse compliance requirements, legal guidelines, and auditing mechanisms involved in managing research data effectively and ethically.
- evaluate strategies for improving RDM practices and recognize the role of libraries in supporting policy implementation, data curation, and open science initiatives.

12.2 INTRODUCTION

Research Data Management (RDM) policies are formal frameworks that guide how research data is created, organized, documented, stored, shared, preserved, and reused within an institution or research environment. These policies are essential for ensuring that data remains accessible, reliable, secure, and usable over the long term, thereby supporting transparency and reproducibility in research.

In the digital era, the rapid growth in the volume, variety, and complexity of research data has made systematic data management increasingly important. As a result, academic and research libraries have emerged as key players in supporting RDM through policy development, infrastructure provision, training, and advisory services.

RDM policies operate at multiple levels, including institutional (universities and research organizations), national (government regulations), funder (granting agency requirements), and international frameworks such as open science initiatives and FAIR (Findable,

Accessible, Interoperable, Reusable) principles, ensuring consistency and standardization in data practices

12.3 INSTITUTIONAL POLICIES, DEVELOPMENT, FUNDER MANDATES

Definition

Institutional Research Data Management (RDM) policies are formal, authoritative documents developed by universities, colleges, and research organizations to regulate how research data is handled throughout its entire lifecycle. This lifecycle includes stages such as data creation, processing, documentation, storage, sharing, preservation, and eventual reuse or disposal. These policies act as a guiding framework that ensures consistency and standardization in managing research data across different departments and disciplines.

Such policies are essential in modern research environments where data plays a central role in knowledge creation. They help researchers follow best practices while also aligning institutional activities with national and international standards. Moreover, institutional RDM policies clarify expectations and responsibilities, enabling researchers, librarians, IT staff, and administrators to work collaboratively. By providing clear guidelines, these policies reduce risks associated with data loss, misuse, or ethical violations, and promote responsible data stewardship.

Objectives of Institutional RDM Policies

The primary objectives of institutional RDM policies are multifaceted and reflect the growing importance of data in research.

- One key objective is to ensure data integrity and quality, meaning that data remains accurate, consistent, and trustworthy throughout its lifecycle. High-quality data enhances the credibility and reproducibility of research findings.
- Another major objective is to promote data sharing and reuse. By encouraging researchers to make their data accessible, institutions support collaboration, innovation, and the advancement of knowledge.
- Closely related to this is the goal of supporting open access and open science, which aims to make research outputs freely available to the public whenever possible.
- Institutional policies also emphasize the need to protect sensitive and confidential data, especially in disciplines involving personal, medical, or classified information. This ensures compliance with ethical standards and legal requirements.

- Furthermore, these policies help institutions meet funder and legal requirements, as many funding agencies now mandate proper data management and sharing practices.
- Finally, they aim to facilitate long-term preservation, ensuring that valuable research data remains accessible and usable for future generations.

Key Components of Institutional Policies

A comprehensive institutional RDM policy consists of several essential components that together provide a complete framework for data management.

i. Scope and Applicability

This section defines the extent of the policy by identifying who must follow it, such as faculty members, researchers, students, and staff. It also specifies the types of research data covered, including digital datasets, experimental data, survey data, and more.

ii. Roles and Responsibilities

Clear allocation of responsibilities is crucial for effective implementation. Researchers are primarily responsible for creating and managing data, while librarians provide guidance on metadata, documentation, and preservation. IT staff handle technical infrastructure such as storage systems and security, and institutional authorities oversee policy enforcement and compliance.

iii. Data Management Planning

Institutional policies typically require researchers to prepare a Data Management Plan (DMP) at the beginning of a research project. A DMP outlines how data will be collected, organized, stored, and shared, ensuring systematic planning from the outset.

iv. Data Storage and Security

This component specifies appropriate storage solutions, such as institutional servers, cloud platforms, or data repositories. It also includes guidelines on backup strategies and security measures to protect data from loss, unauthorized access, or cyber threats.

v. Data Sharing and Access

Policies encourage open access to research data whenever possible, promoting transparency and collaboration. However, they also provide mechanisms for controlled or restricted access in cases involving sensitive or confidential data.

vi. Data Preservation

Long-term preservation is a key concern. Policies define retention periods, archiving procedures, and the use of institutional or subject-specific repositories to ensure that data remains accessible over time.

vii. Ethical and Legal Compliance

This section ensures adherence to ethical standards, intellectual property rights, and privacy laws. It provides guidelines for handling sensitive data and maintaining confidentiality, particularly in research involving human participants.

Development of Institutional RDM Policies

The development of institutional RDM policies is a collaborative and iterative process that involves multiple stages and stakeholders.

i. Needs Assessment

The first step is to identify institutional requirements and evaluate existing data management practices. This helps in understanding gaps and challenges.

ii. Stakeholder Consultation

Policy development requires input from various stakeholders, including faculty members, librarians, IT professionals, and legal experts. Their perspectives ensure that the policy is practical and comprehensive.

iii. Drafting the Policy

The policy is then drafted by aligning it with national regulations and international best practices, such as the FAIR principles (Findable, Accessible, Interoperable, Reusable).

iv. Review and Approval

The draft policy undergoes review by institutional committees or governing bodies to ensure its validity and feasibility.

v. Implementation

Once approved, the policy is implemented through awareness campaigns, training programs, and workshops that educate researchers and staff about its provisions.

vi. Monitoring and Revision

RDM policies are not static; they require regular monitoring and updates based on technological advancements, feedback, and changing research needs.

Role of Libraries in Policy Development

Libraries play a pivotal role in the development and implementation of institutional RDM policies. They possess expertise in organizing, managing, and preserving information, which is directly applicable to research data. Librarians contribute by developing metadata standards that enhance data discoverability and interoperability.

They also support data curation and long-term preservation through institutional repositories and digital archives. Additionally, libraries conduct training sessions and workshops to raise awareness about RDM practices among researchers and students. By offering consultation on data management planning and compliance, libraries act as key facilitators in the successful implementation of RDM policies.

In summary, institutional RDM policies provide a comprehensive framework for managing research data effectively, and libraries serve as central partners in ensuring their success.

Funder Mandates in RDM

Definition

Funder mandates in Research Data Management (RDM) refer to the formal requirements set by funding agencies that researchers must follow while managing, storing, and sharing their research data. These mandates are increasingly important in ensuring that data generated through publicly funded research is treated as a valuable resource. They emphasize that such data should be accessible, transparent, and reusable, thereby promoting accountability and maximizing the impact of research outcomes. By enforcing these requirements, funding bodies ensure that research data contributes to broader scientific progress and societal benefit.

Key Features of Funder Mandates

i. Data Management Plans (DMPs)

One of the most common requirements is the submission of a Data Management Plan at the proposal stage. This document outlines how data will be collected, organized, stored, preserved, and shared throughout the research lifecycle. It ensures that researchers plan systematically for responsible data handling.

ii. Data Sharing Requirements

Funder mandates strongly encourage or require open access to research data. Researchers are expected to deposit their data in recognized repositories, making it available for reuse unless there are valid restrictions such as confidentiality or privacy concerns.

iii. Data Preservation

Researchers must ensure that their data is preserved for a specified period, typically ranging from 5 to 10 years or more. This ensures long-term availability and supports future verification or reuse of research findings.

iv. Compliance Monitoring

Funding agencies often monitor compliance by reviewing data management practices during and after the research project. Non-compliance may affect future funding opportunities.

Examples of Funder Mandates

Research Data Management (RDM) Funder Mandates ensure that publicly funded

research is transparent and reusable. Major funders now require specific compliance steps:

- The National Institutes of Health (NIH), USA: Requires a Data Management and Sharing (DMS) Plan for all research generating scientific data, emphasizing timely sharing usually no later than the time of publication.
- Horizon Europe (European Union): Mandates that data be FAIR (Findable, Accessible, Interoperable, Reusable). Open access is the default, though "as open as possible, as closed as necessary" allows for privacy exceptions.
- UK Research and Innovation (UKRI), (United Kingdom): Expects data to be made openly available with as few restrictions as possible, requiring long-term preservation in recognized repositories.
- The Department of Biotechnology (DBT) and the Department of Science and Technology (DST), Ministry of Science & Technology Government of India

Importance of Funder Mandates

Funder mandates are crucial for strengthening the overall research ecosystem.

- They promote transparency by ensuring that research data is openly available for verification and validation.
- By encouraging data sharing, they help prevent duplication of research efforts, saving time and resources.
- These mandates also foster collaboration and innovation, as researchers can build upon existing datasets to generate new knowledge.
- Furthermore, they ensure that publicly funded research delivers maximum value and impact, benefiting society at large.

Overall, funder mandates enhance accountability, improve research quality, and support the global movement toward open science.

12.4 COMPLIANCE, LEGAL GUIDELINES, AUDITING, IMPROVEMENT

Definition

Compliance in Research Data Management (RDM) policies refers to the extent to which researchers and institutions adhere to established rules, policies, and standards governing the management of research data. This includes following institutional policies, meeting funder mandates, and complying with legal and ethical requirements. Effective compliance ensures that research data is handled responsibly, securely, and transparently

throughout its lifecycle, thereby maintaining the integrity and credibility of research. As data-driven research becomes the global norm, compliance has shifted from a “best practice” to a mandatory requirement for securing and maintaining research funding.

Types of Compliance

i. Institutional Compliance

Institutional compliance requires researchers, students, and staff to follow the RDM policies set by their organization. This includes proper data storage, documentation, and sharing practices as defined by the institution.

ii. Funder Compliance

Funder compliance involves adhering to the requirements specified by funding agencies, such as submitting Data Management Plans (DMPs), sharing data in repositories, and maintaining data for a specified period.

iii. Legal Compliance

Legal compliance ensures that research data management aligns with relevant laws and regulations. These include data protection laws, intellectual property rights, and confidentiality agreements, especially when dealing with sensitive or personal data.

iv. Ethical Compliance

Ethical compliance focuses on responsible conduct in research. It involves obtaining informed consent from participants, ensuring the protection of human subjects, and using data in a fair and responsible manner.

Mechanisms for Ensuring Compliance

Institutions ensure RDM compliance through a combination of administrative, technical, and professional mechanisms:

- Administrative: Mandatory Data Management Plans (DMPs) act as blueprints for data handling, often tied to grant approvals and ethics committee reviews.
- Technical: Automated tools like DMPOnline track progress, while repositories enforce metadata standards (e.g., Dublin Core) and assign DOIs for traceability.
- Professional: Data librarians perform curational audits; ensuring datasets align with FAIR principles and long-term preservation requirements.

These mechanisms bridge the gap between policy and practice, transforming raw data into citable, compliant research assets.

Challenges in Compliance

Despite its importance, achieving compliance in RDM can be challenging. Many researchers lack awareness or training in proper data management practices. Limited

infrastructure, such as inadequate storage facilities, can hinder implementation. Additionally, some researchers may resist data sharing due to concerns about misuse or loss of ownership. The complexity of legal and regulatory requirements further adds to the difficulty of ensuring full compliance.

Legal Guidelines in RDM

Importance of Legal Guidelines

Legal guidelines are essential in Research Data Management (RDM) as they provide a structured basis for handling research data responsibly and lawfully. These guidelines ensure that research data management respects ownership rights, protects privacy and confidentiality, and complies with both national and international laws. By following legal frameworks, institutions and researchers can avoid legal disputes, maintain ethical standards, and build trust among stakeholders. Proper legal compliance also enhances the credibility and reliability of research outputs.

Key Legal Aspects

i. Intellectual Property Rights (IPR)

Intellectual Property Rights play a crucial role in defining the ownership and control of research data. They determine who has the authority to use, distribute, or modify the data. IPR covers areas such as copyrights, patents, and licensing agreements, ensuring that the rights of creators and institutions are protected.

ii. Data Protection and Privacy

Data protection laws are designed to safeguard personal and sensitive information. Researchers must ensure that data involving individuals is handled securely and, where necessary, anonymized to prevent identification. This is particularly important in fields such as medical and social research.

iii. Confidentiality

Confidentiality is critical when dealing with sensitive data, including medical records, personal information, or classified research. Researchers must take appropriate measures to restrict access and prevent unauthorized disclosure.

iv. Licensing

Licensing specifies how research data can be reused by others. Open licenses are commonly used to promote data sharing while still protecting the rights of data creators. Clear licensing terms help avoid misuse and encourage responsible reuse.

Ethical Considerations

Legal guidelines in RDM are closely linked with ethical principles. Researchers must obtain informed consent from participants before collecting data and ensure ethical approval

for studies involving human subjects. Responsible data sharing practices must also be followed to avoid harm or misuse of information.

Auditing in RDM

Definition

RDM auditing refers to the systematic evaluation of research data management practices within an institution or research project to determine whether they comply with established policies, standards, and guidelines. It involves reviewing how research data is created, documented, stored, secured, shared, and preserved. Auditing helps institutions ensure that researchers follow proper data management procedures and that data resources are handled responsibly and efficiently throughout the research lifecycle.

Objectives of Auditing

The main objective of RDM auditing is to ensure compliance with institutional policies, funder requirements, and legal regulations. Auditing also helps identify gaps or weaknesses in existing data management practices. Through regular evaluation, institutions can improve data quality, strengthen data security measures, and promote better documentation and preservation practices. Another important objective is to enhance accountability by ensuring that researchers and departments take responsibility for proper data handling and management.

Types of Audits

i. Internal Audits

Internal audits are conducted within the institution by designated committees, library professionals, or research management offices. These audits help institutions review their own practices and identify areas that need improvement.

ii. External Audits

External audits are carried out by funding agencies, regulatory bodies, or independent organizations. These audits ensure that institutions and researchers are complying with funder mandates and recognized standards in research data management.

Key Areas of Audit

RDM audits typically focus on several important aspects of data management.

i. These include reviewing data management plans prepared by researchers, evaluating the effectiveness of data storage and security systems, and assessing the quality of metadata used to describe datasets.

ii. Auditors also examine how institutional repositories are used and whether researchers are properly sharing their data according to policy requirements.

Audit Tools and Standards

Various tools and frameworks are used to support RDM auditing. These include data repository certification standards such as CoreTrustSeal and ISO 16363, which evaluate the reliability and trustworthiness of repositories. In addition, best practice guidelines and tools like the DCC Checklist support effective data curation and preservation. In India, organizations such as Council of Scientific and Industrial Research follow these standards to improve data management and repository practices. Such frameworks help institutions maintain consistent, transparent, and high-quality research data management practices.

Role of Libraries in Auditing

Libraries play an important role in RDM auditing. They maintain institutional data repositories, monitor data deposits made by researchers, and ensure that datasets follow proper metadata standards. Libraries also assist in preparing documentation and provide guidance to researchers, helping institutions improve overall compliance with RDM policies.

Improvement of RDM Practices

Need for Continuous Improvement

Research Data Management (RDM) is a continuously evolving field that requires regular improvement to remain effective and relevant.

Rapid technological advancements, such as cloud computing, big data analytics, and artificial intelligence, are transforming how research data is generated and managed.

- At the same time, research practices are changing, with increased emphasis on collaboration, interdisciplinary studies, and open science.
- Additionally, legal and regulatory frameworks related to data protection, privacy, and intellectual property are constantly evolving.

These factors make it essential for institutions to regularly review and enhance their RDM practices to ensure efficiency, compliance, and sustainability.

Strategies for Improvement

i. Policy Revision

Regular revision of RDM policies is necessary to incorporate new technologies, updated legal requirements, and emerging best practices. This ensures that policies remain relevant and effective.

ii. Training and Awareness

Conducting workshops, seminars, and training programs helps researchers and staff understand proper data management practices. Capacity-building initiatives improve skills related to data handling, documentation, and sharing.

iii. Infrastructure Development

Improving technical infrastructure is crucial for effective RDM. This includes developing secure storage systems, high-quality repositories, and reliable backup solutions to support data preservation and access.

iv. Adoption of Standards

Implementing recognized standards, such as the FAIR principles (Findable, Accessible, Interoperable, Reusable), enhances the quality, accessibility, and interoperability of research data.

v. Collaboration

Collaboration between institutions promotes the sharing of knowledge, tools, and best practices. Inter-institutional cooperation strengthens RDM frameworks and fosters innovation.

Role of Libraries in Improvement

Libraries play a proactive and central role in improving RDM practices.

- They provide specialized RDM services, assist researchers in preparing Data Management Plans (DMPs), and manage digital repositories for data storage and access.
- Libraries also promote open science initiatives and organize training programs to raise awareness.
- By acting as facilitators and knowledge hubs, libraries significantly contribute to the continuous enhancement of RDM practices within institutions.

CHECK YOUR PROGRESS

FILL IN THE BLANKS

a) Research Data Management policies ensure that data remains _____, reliable, and reusable over time.

b) A _____ is a document that outlines how research data will be managed during a project.

c) _____ Compliance refers to following laws related to data protection, intellectual property, and confidentiality.

12.5 LET US SUM UP

- This unit examines Research Data Management (RDM) strategies as organized frameworks for overseeing research data during its lifecycle, guaranteeing accessibility, dependability, and reusability. It elucidates institutional policies, their

aims, elements, and formulation processes, as well as the pivotal function of libraries. The unit emphasizes funder obligations that necessitate Data Management Plans, data exchange, and long-term preservation. It analyses compliance categories like institutional, legal, ethical, and funder and the mechanisms that underpin them. Legal considerations like intellectual property, privacy, and secrecy are examined. The section encompasses auditing processes and underscores continual enhancement via training, infrastructure, standards, and collaboration.

✓ 12.6 ANSWERS TO CHECK YOUR PROGRESS

a) Accessible, b) Data Management Plan / DMP, c) Legal

12.7 FURTHER READINGS

1. Donnelly, M. (2014). *Five Steps to Developing a Research Data Policy*. Digital Curation Centre.
2. Briney, K. (2015). *Data Management for Researchers: Organize, Maintain and Share Your Data for Research Success*. Pelagic Publishing.
3. Pryor, G. (Ed.). (2012). *Managing Research Data*. Facet Publishing.
4. Cox, A. M., & Pinfield, S. (2018). *Exploring Research Data Management*. Facet Publishing.

12.8 MODEL QUESTIONS

a) Define Research Data Management (RDM) policies.
b) What are funder mandates in RDM?
c) Mention any two objectives of institutional RDM policies.
d) What is a Data Management Plan (DMP)?
e) State the role of libraries in RDM.
f) Discuss the key components of institutional RDM policies in detail.
g) Explain funder mandates and their importance in research data management.
h) Describe the legal and ethical guidelines involved in RDM.
i) Examine the role of auditing and compliance in ensuring effective RDM practices.

*** ***** ***

UNIT 13: RESEARCH DATA REPOSITORIES

UNIT STRUCTURE

13.1 Learning Objectives
13.2 Introduction
13.3 Data Repositories: concept, types, features
13.4 Metadata standards, access policies; Preservation strategies
13.5 Let Us Sum Up
13.6 Answers to Check Your Progress
13.7 Further Readings
13.8 Model Questions

13.1 LEARNING OBJECTIVES

After studying this unit, you will be able to:

- understand the concept and importance of research data repositories in modern research.
- identify the key features of research data repositories, such as metadata, access policies, and preservation methods.
- explain the role of academic libraries in managing and supporting research data repositories

13.2 INTRODUCTION

Research Data Repositories are digital platforms designed for the storage, organization, preservation, and sharing of research data. In the present digital research environment, large volumes of data are generated through experiments, surveys, simulations, and observations. Managing this data effectively has become essential, and repositories provide a structured solution for this purpose.

A research data repository is not just a storage system but a specialized platform that ensures data is properly documented, preserved, and made accessible for future use. These repositories allow researchers to deposit their datasets, which can then be accessed by other researchers for verification, reuse, or further study. This supports transparency, reproducibility, and collaboration in research.

Unlike basic storage systems such as personal computers or cloud storage, research data repositories include advanced features such as metadata support, persistent identifiers, and preservation strategies. Academic libraries play a key role in managing

and promoting the use of such repositories within institutions.

13.3 DATA REPOSITORIES: CONCEPT, TYPES, FEATURES

Concept of Data Repositories

The concept of data repositories is closely connected with Research Data Management (RDM). A repository acts as a centralized system where research data is collected, organized, and stored in a systematic way. It ensures that data is not lost and can be accessed in the future.

Repositories serve multiple important functions.

- They store research data securely, preserve it for long-term use, and make it available to others.
- By doing so, they promote the sharing of knowledge and support the validation of research findings.
- They also help in maintaining research integrity by allowing others to verify results.

Another important concept associated with repositories is the FAIR principles. These principles ensure that data is findable, accessible, interoperable, and reusable. Repositories follow these principles to maximize the usefulness of research data.

Types of Research Data Repositories

Research data repositories are specialized digital archives designed to store, manage, and share research outputs. They are classified based on their purpose, scope, and the community they serve.

1. Institutional Repositories:

Institutional repositories are managed by universities or research institutions to store the research output of their members, including datasets, theses, and publications. These platforms help in preserving institutional knowledge and promoting open access.

For example, many universities use DSpace software to build their repositories, such as DSpace@MIT or the University of Cambridge's Apollo repository. These systems ensure that the intellectual legacy of an organization remains accessible to the public and future scholars.

2. Subject-Specific (Disciplinary) Repositories:

Subject-specific repositories are designed for particular disciplines, storing data related to specific fields such as biology, social sciences, or physics. These repositories follow discipline-specific standards, making data more meaningful and useful for researchers in that field.

A prominent example is GenBank, which serves the genetic research community by providing an annotated collection of all publicly available DNA sequences. Similarly, the ICPSR (Inter-university Consortium for Political and Social Research) acts as a specialized hub for social science data, while arXiv serves the physics and mathematics communities.

3. General-Purpose Repositories:

General-purpose repositories accept data from all disciplines and are particularly useful when no subject-specific repository is available. These platforms support interdisciplinary research and provide flexible options for data sharing.

Well-known examples include Zenodo, which was developed under the European OpenAIRE program and is operated by CERN, and Figshare, a commercial platform that allows researchers to make their outputs discoverable and citable. Another example is Dryad, which focuses on data underlying scientific publications across various life sciences.

4. Project-Specific Repositories:

Project-specific repositories are created for particular research projects to store data generated during the study's lifecycle. They may be used for a limited or long-term period depending on the project's funding and goals.

For instance, the Open Science Framework (OSF) is frequently used by research teams to host project-specific files, pre-registrations, and code during the active phase of a study. Another example is the Time-sharing Experiments for the Social Sciences (TESS), which provides a dedicated platform for experimental social science data, or the ELIXIR TeSS portal, which serves as a specialized repository for life science training materials and project outputs.

5. National or Government Repositories:

National or government repositories are developed by government agencies to store large datasets for public use. These repositories support public access to data and promote transparency in governance and taxpayer-funded research.

Data.gov in the United States is a primary example, providing a massive catalog of federal data. In the United Kingdom, the UK Data Service provides access to a wide range of social and economic data, while the NASA Exoplanet Archive serves as a national resource for astronomical data collected through government-funded space missions.

Features of Research Data Repositories

Research data repositories have several important features that make them different from ordinary storage systems and ensure effective management of research data**.**

1. Metadata Support:

Every dataset is accompanied by structured descriptive information, including titles,

contributors, and methodologies. This documentation ensures that data is not only discoverable via search engines but also interpretable by future researchers.

2. Persistent Identifiers (PIDs):

Repositories assign unique, permanent links such as Digital Object Identifiers (DOIs) or Handles. Unlike standard URLs, PIDs remain stable even if the physical location of the file changes, ensuring reliable long-term citation.

3. Curation and Quality Control:

Dedicated repositories perform data validation and cleaning. This curation process ensures that datasets are logically organized, properly formatted, and sufficiently documented for academic reuse.

4. Granular Access Control:

Platforms provide tiered access levels, ranging from fully open-access to restricted or embargoed data. This allows researchers to protect sensitive information or comply with intellectual property requirements while still enabling controlled sharing.

5. Long-term Preservation:

Repositories are designed for digital longevity, employing strategies like format migration and bit-level preservation to protect data against technological obsolescence.

6. Interoperability and Standards:

By supporting standardized protocols and formats, repositories allow data to be seamlessly exchanged between different systems. This aligns with the FAIR principles (Findable, Accessible, Interoperable, and Reusable), which are the global benchmark for modern data stewardship.

7. Discovery Interfaces:

Sophisticated search features and user-friendly interfaces allow scholars to filter datasets by discipline, date, or license, significantly enhancing research efficiency.

13.4 METADATA STANDARDS, ACCESS POLICIES; PRESERVATION STRATEGIES

Metadata Standards in Data Repositories

Metadata standards provide the foundational language and rules that allow research data to be structured, discovered, and shared across global digital platforms. By using uniform schemas, repositories prevent “dark data”—information that exists but cannot be found—allowing researchers to share and cite datasets as formal scholarly outputs.

The following standards are essential for managing modern digital repositories:

1. Dublin Core (DC):

A universal, 15-element set (e.g., Title, Creator, Date) used for basic cross-disciplinary discovery. Its simplicity makes it the "lingua franca" of the metadata world.

2. DataCite Metadata Schema:

The industry standard for research datasets. It is specifically optimized for assigning Digital Object Identifiers (DOIs), ensuring that data is permanently citable and linked to traditional publications.

3. MODS (Metadata Object Description Schema):

An XML-based derivative of MARC. It offers a richer, more flexible bibliographic description than Dublin Core, making it ideal for library-managed digital collections.

4. METS (Metadata Encoding and Transmission Standard):

A hierarchical "wrapper" or digital container. It bundles descriptive, administrative, and structural metadata together, ensuring that complex objects (like a book consisting of 100 separate images) stay organized during transmission.

5. PREMIS (Preservation Metadata: Implementation Strategies):

The gold standard for digital preservation. It acts as a "digital health record," tracking the provenance, authenticity, and technical integrity of files to ensure they remain usable as technology evolves.

6. Disciplinary Standards:

Specialized schemas that capture domain-specific nuances, such as Data Documentation Initiative -DDI (Social Sciences), Darwin Core (Biology), and The Text Encoding Initiative; TEI (Humanities). for structural text encoding.

Adhering to these schemas ensures that data follows the FAIR principles- making it Findable, Accessible, Interoperable, and Reusable. While descriptive standards like Dublin Core facilitate identification, administrative standards like PREMIS guarantee long-term survival. Ultimately, these standards transform raw data into valuable, interoperable assets that support the global research ecosystem.

Access Policies in Data Repositories

Access policies define the rules and conditions under which research data can be viewed, downloaded, or reused. These policies are essential for balancing the goal of open science with the need to protect sensitive information.

1. Open Access:

The most common policy, where data is freely available to the public without restrictions. This promotes transparency and allows for maximum data reuse across the global research community.

2. Restricted Access:

Data is available only to users who meet basic criteria, such as having a verified institutional account. Users may be required to sign a Data Use Agreement (DUA) to ensure the data is handled ethically and not redistributed.

3. Controlled Access:

A high-security "gatekeeper" model used for extremely sensitive information, such as human genomic data. Access requires a formal application to a Data Access Committee (DAC), ethical approval from an Institutional Review Board (IRB), and often requires analysis to take place within a secure virtual environment rather than being downloaded.

4. Embargoed Access:

Data is kept private for a specific period, usually until a related study is published. This allows researchers to have exclusive use of their findings for a set time before the data is automatically released to the public.

5. Closed Access:

Only the metadata (the description of the dataset) is visible to the public, while the actual data files remain hidden. This is typically used for highly sensitive, confidential, or proprietary information that cannot be shared legally but must still be archived for the record.

By implementing these tiered policies, repositories ensure that data sharing complies with ethical standards, legal requirements, and participant privacy. Proper access policies ensure ethical use of data, protect sensitive information, and support compliance with legal requirements. Academic libraries guide researchers in selecting appropriate access levels and licenses.

Preservation Strategies in Data Repositories

Preservation strategies are the systematic processes used by repositories to ensure that research data remains accessible, readable, and usable over long periods, regardless of changes in hardware or software. These strategies are essential for combating digital obsolescence and ensuring the long-term integrity of the scholarly record. The key preservation strategies are:

- **Bit-level Preservation:** This is the most basic level of preservation. It involves making exact copies of the original data and using checksums digital fingerprints to monitor files for corruption. If a checksum changes, it indicates "bit rot," and the file must be restored from a clean backup.
- **Migration:** This is the most common active strategy. It involves periodically moving files from an aging or proprietary format to a modern, open-standard format (e.g., converting

an old Excel file to a CSV). This ensures the data can still be opened by current software.

- **Emulation:** Instead of changing the file itself, emulation uses modern software to "mimic" an obsolete computer system. This allows a researcher to run original, legacy software to view data exactly as it appeared decades ago.
- **Replication:** This involves storing multiple copies of the data in different physical locations, known as geographic redundancy. This protects the collection against localized disasters, such as fires or server failures.
- **Normalization:** Upon deposit, the repository automatically converts files into a few "preferred" or "standard" formats that are known to be stable and easy to preserve. This reduces the variety of file types the system must manage.

Supporting Frameworks for Preservation

- Persistent Identifiers (PIDs): Mechanisms such as DOIs (Digital Object Identifiers) are essential for long-term access. While preservation strategies keep the *file* readable, PIDs ensure the *link* to that file never breaks, even if the repository moves to a new server or domain.
- Trusted Digital Repositories (TDRs): Reliability is guaranteed when repositories follow international standards, such as ISO 16363. Being a "Trusted" repository means the organization has the financial, legal, and technical sustainability to protect data for decades, not just years.
- FAIR Principles: The FAIR guidelines (Findable, Accessible, Interoperable, Reusable) serve as the strategic goal for preservation. A preserved file is only valuable if it remains "Interoperable" with future software and "Findable" through standardized metadata.
- Institutional Role of Academic Libraries: Libraries provide the essential human and physical infrastructure for preservation. Beyond just hosting servers, they manage the curation lifecycle, offer expertise in metadata, and provide training to ensure researchers follow best practices from the start of their projects.

Effective preservation typically adheres to the OAIS (Open Archival Information System) reference model, which provides a functional framework for moving data from the producer to the long-term archive. By combining these strategies, repositories ensure that the data remains a functional asset for future generations rather than becoming unreadable legacy code.

CHECK YOUR PROGRESS

Fill in the Blanks:

a) Research data repositories help in making data _____ and reusable.

b) Metadata is often described as “data about _____.”

c) Academic libraries play a key role in managing _____ data.

d) Persistent identifiers like DOI help in _____ datasets easily.

13.5 LET US SUM UP

- Research data repositories are very important in today’s research world. They help turn raw data into organized, long-lasting, and easy-to-find information. By offering a proper system for storing and sharing data, these repositories make research more transparent, reliable, and easy to verify. They also follow international standards. What makes them different from simple storage systems is the use of proper metadata (data about data), clear access rules, and strong methods to preserve data for the future.
- Academic libraries play a key role in managing these repositories. They connect researchers with the technical systems needed to store and share data. Libraries also guide researchers on issues like copyright and help protect data from becoming outdated due to changing technology.
- In today’s digital age, research data repositories are no longer optional they are essential. A well-managed repository ensures that today’s research data remains useful, accessible, and valuable for future researchers around the world.

13.6 ANSWERS TO CHECK YOUR PROGRESS

a) Accessible, b) data. c) research, d) identifying

13.7 FURTHER READINGS

1. Cox, A. and Pinfield, S. (2013) Research Data Management and Libraries. Journal of Librarianship and Information Science.
2. Borgman, C. L. (2015). Big Data, Little Data, No Data: Scholarship in the Networked World.
3. Pryor, G. (Ed.). (2012). Managing Research Data. Facet Publishing.

4. Rice, Robin & Southall, John. (2016). The Data Librarian's Handbook. London: Facet Publishing, Chartered Institute of Library and Information Professionals (CILIP).

13.8 MODEL QUESTIONS

a) What is a data repository?

b) Mention any two types of data repositories.

c) What do you mean by metadata standards?

d) Why are access policies important in data repositories?

e) Explain the concept and different types of data repositories in detail.

f) Discuss the key features of data repositories with suitable examples.

g) Describe metadata standards and their importance in data management.

h) Explain the different types of access policies in data repositories and discuss their importance in ensuring ethical use and protection of research data.

i) Explain various preservation strategies used in research data repositories.

*** ***** ***

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

14.10 Let Us Sum Up

14.11 Answers to Check Your Progress

14.12 Further Readings

14.13 Model Questions

14.1 LEARNING OBJECTIVES

After studying this unit, you 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, yous 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 you 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 yous. 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.
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.
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 MODEL QUESTIONS

a) Briefly describe two applications of artificial intelligence in academic libraries.

b) What are the main privacy challenges created by the use of detailed usage data in digital libraries?

c) State two key elements of a good data governance framework for library data.

d) Discuss the role of artificial intelligence and machine learning in transforming library data management.

e) Explain the potential applications of blockchain in libraries for acquisitions, rights management, and preservation. What are the main advantages and limitations of using blockchain.

*** ***** ***