RESEARCH DATA MANAGEMENT: AN OVERVIEW

RDM Outreach Team
INFLIBNET Centre
Gandhinagar

INFLIBNET

DataCite

RDM OUTREACH

Coverage

01 What is Research Data?

02 Types of Research Data

03 Classification of Research Data

04 Open Data and Open Science

05 Plan S & Ten Principles of Plan S

06 What is Open Research Data?

07 The benefits of Open Data

08 What is Research Data Management (RDM)?

09 RDM Lifecycle

10 Importance of RDM

11 **FAIR**ification of Research Data

12 Research Data Repository (RDR)

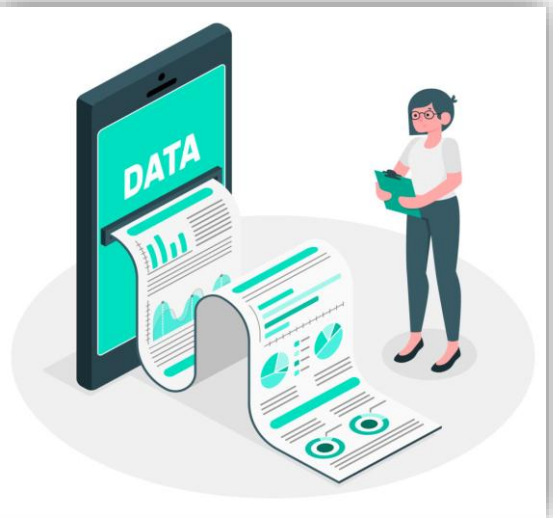
13 Desirable Characteristics of Research Data Repository

14 Institutional Repository vs Data Repository

15 Registry of Research Data Repository

What is Research Data?

01



Research Data is collected, observed, or created, for purposes of analysis to produce original research results.

Research Data can be generated for different purposes and through different processes, and that can be divided into different categories. Each category may require a different type of “treatment”.

Types of Research Data

02

Observational

1 Data captured in real-time, usually irreplaceable. For example, sensor data, survey data, sample data, neurological images

Experimental

2 Data from lab equipment, often reproducible, but can be expensive. For example, gene sequences, chromatograms, toroid magnetic field data

Simulation

3 Data generated from test models where model and metadata are more important than output data. For example, climate models, economic models

Derived or Compiled

4 Data is reproducible but expensive. For example, text and data mining, compiled database, 3D models

Reference or Canonical

5 A collection of smaller (peer-reviewed) datasets, most probably published and curated. For example, gene sequence databanks, chemical structures, or spatial data portals

Classification of Research Data

03

Method

Experimental, observational, simulation, derived or compiled...

Format

Spreadsheets, databases, images, maps, audio files, un(structured) text...

Content

Numerical, textual, audiovisual, multimedia...

Nature

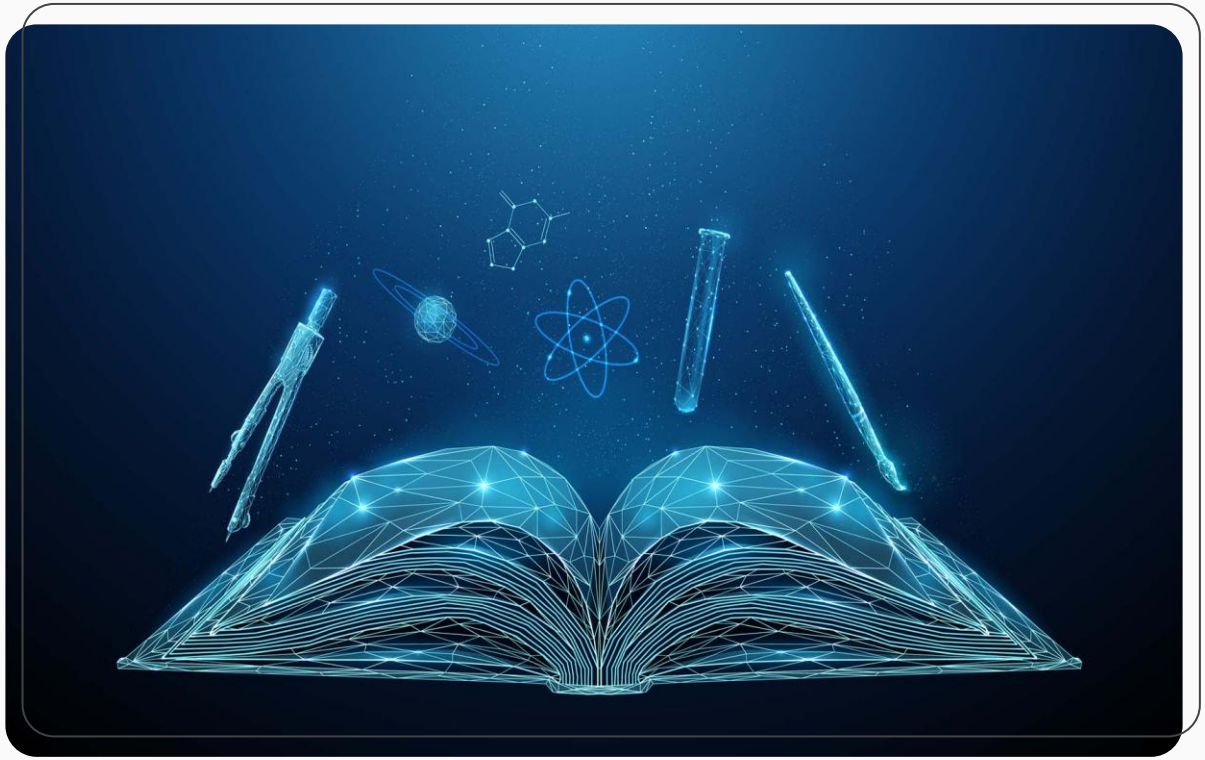
Digital (born digital or digitized) or non-digital (e.g., paper surveys)

Source

Primary or secondary

Processing

Raw or processed



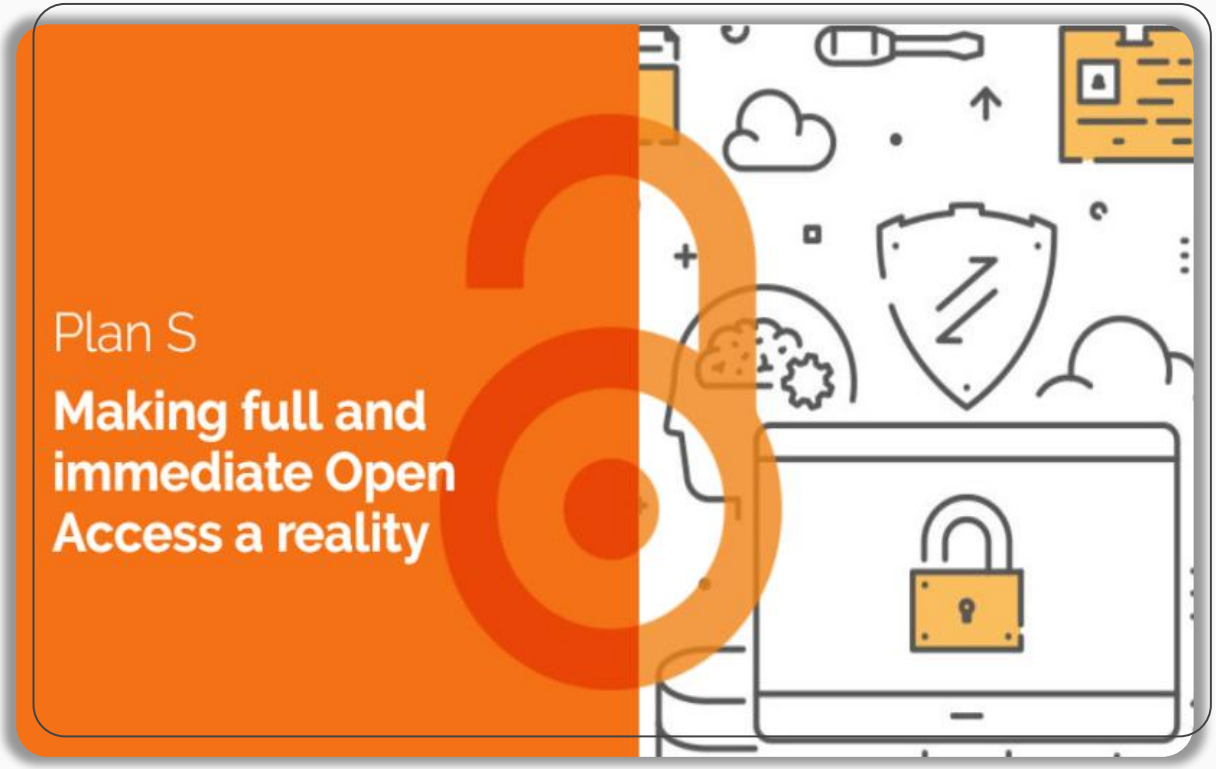
Open Data and Open Science

04

01 A data is said to be open if anyone is free to readable, and redistribute that without any legal, technological or social restriction – **Open Data** shall be machine-readable and it should also be easily accessible.

02 Open data benefit the open science movement in a number of ways e.g., prevent duplicating the collection of data across organizations

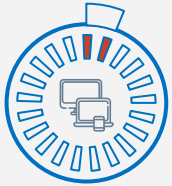
03 **Open Science** refers to the practice of making scientific research, data and findings accessible to everyone through data sharing and collaboration without any restrictions



Plan S
&
Ten Principles of Plan S

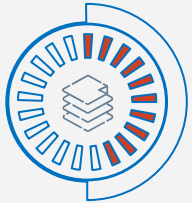
05

Plan S

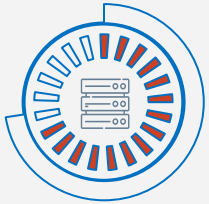


Plan S is an initiative for open-access science publishing launched in 2018 by "cOAlition S", a consortium of national research agencies and funders from twelve European countries.

The plan requires scientists and researchers who benefit from state-funded research organisations and institutions to publish their work in open repositories or in journals that are available to all by 2021.



The key principle states that "With effect from 2021, all scholarly publications on the results from research funded by public or private grants provided by national, regional and international research councils and funding bodies, must be published in Open Access Journals, on Open Access Platforms, or made immediately available through Open Access Repositories without embargo."



The plan was structured around ten principles.

Ten Principles of Plan S

1

Authors should retain copyright on their publications, which must be published under an open license such as Creative Commons;

2

The members of the coalition/The Funders should establish robust criteria and requirements for compliant open access journals and platforms;

3

The Funders should also provide incentives for the creation of compliant open access journals and platforms if they do not yet exist;

4

Publication fees should be covered by the funders or universities, not individual researchers;

5

Such publication fees should be standardized and capped;

Ten Principles of Plan S

6

Universities, research organizations, and libraries should align their policies and strategies;

7

For books and monographs, the timeline may be extended beyond 2021;

8

Hybrid open-access journals are not compliant with the key principle;

9

Members of the coalition should monitor compliance and sanction non-compliant beneficiaries.

10

Open archives and repositories are acknowledged for their importance;

IN INDIA...

In the year 2012, India initiated the “National Data Sharing and Accessibility Policy (NDSAP)” to share non-sensitive data available either in digital or analog forms that is generated using public funds by various Ministries/Departments/Subordinate offices/Organizations/Agencies of Government of India as well as States. The NDSAP policy is designed to promote data sharing and enable access to Government of India owned data for national planning, development and awareness.

What is Open Research Data?

06

Open Scientific Data OR Open Research Data is a type of open data focused on publishing observations and results of scientific activities available for anyone to analyze and reuse.

It is:

- to allow the verification of scientific claims, by allowing others to look at the reproducibility of results.
- to allow data from many sources to be integrated to give new knowledge.

The benefits of Open Data

07

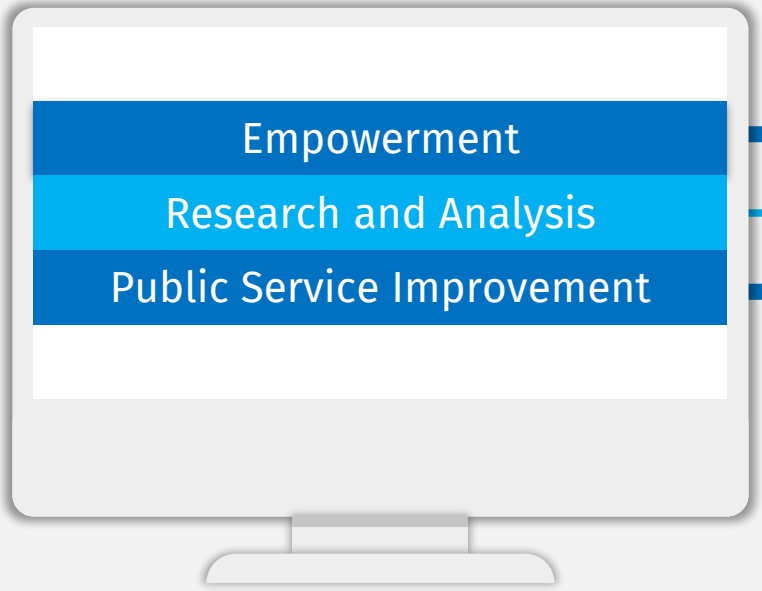
Open data promotes transparency by making information accessible to the public, fostering trust in government/institutions.

Accessible data encourages innovation by enabling individuals/businesses/researchers to develop new products, services & solutions

Open data can help in economic growth by creating opportunities for entrepreneurship and driving efficiency in public and private sectors.

By providing access to information, open data holds organizations and governments accountable for their actions and decisions.

Open data encourages collaboration between governments, organizations, and individuals, leading to the sharing of best practices and the co-creation of solutions to complex problems.



Open data empowers citizens by giving them the tools and information they need to participate in civic processes and make informed decisions

Researchers and analysts can use open data to conduct studies, analyze trends, and generate insights across various fields, like healthcare, education, and the environment

Accessible data can help identify areas for improvement in public services, leading to better resource allocation and more effective policymaking

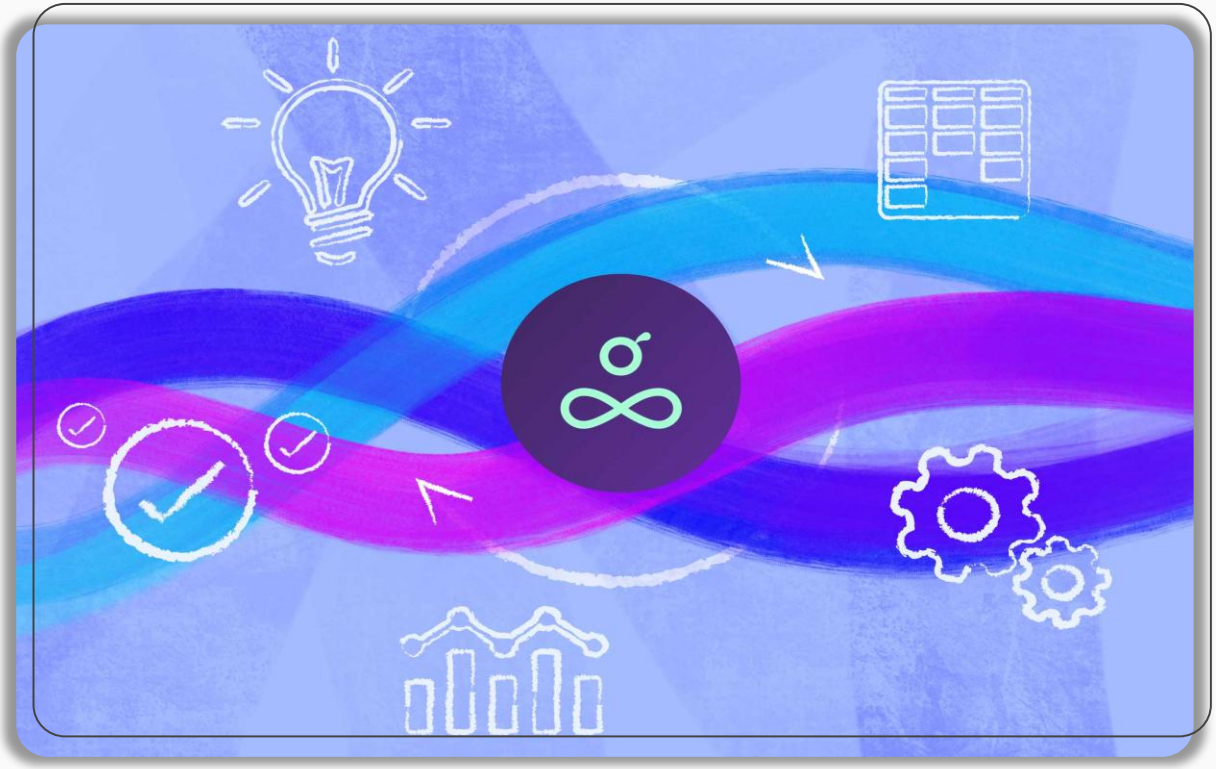
What is Research Data Management (RDM)?

08

Research data management is the organization, documentation, storage, and preservation of the data resulting from the research process, *where data can be broadly defined as the outcome of experiments or observations that validate research findings.*

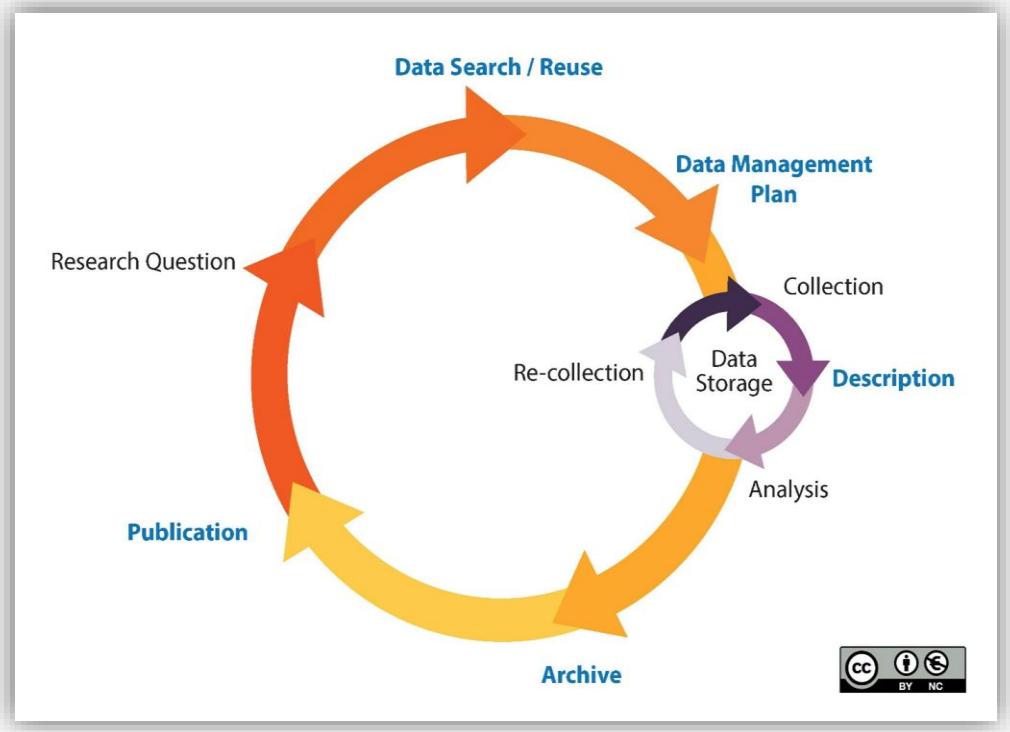
It can take a variety of forms including numerical output (quantitative data), qualitative data, documentation, images, audio, and video

Source: https://www.nnlm.gov/guides/data-glossary/research-data-management



RDM Lifecycle

9

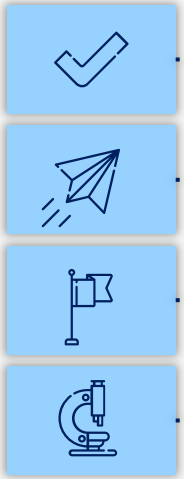


- Choosing file formats
- File organization & naming conventions
- Version control
- Document all project/file details
- Access control & security
- Backup & storage
- File format conversions
- Sharing and preservation

source: http://guides.library.ucsc.edu/datamanagement

Importance of RDM

10



Quality Assurance	Ensures the accuracy, reliability, and completeness of research data for drawing valid conclusions and making informed decisions.
Reproducibility and Transparency	Allows other researchers to replicate the study and verify its findings, enhancing the credibility and trustworthiness
Compliance and Ethical Considerations	Adhering to data management standards and regulations ensures that research data are handled ethically, respecting the rights and privacy of participants, and complying with legal and institutional requirements
Efficient Data Retrieval and Analysis	Organized and properly documented data are easier to find, access, and analyze which saves time and resources by enabling researchers to focus on data interpretation and knowledge generation

Data Security and Protection

Effective data management strategies safeguard sensitive information from unauthorized access, loss, or corruption, reducing the risk of data breaches and ensuring confidentiality and integrity

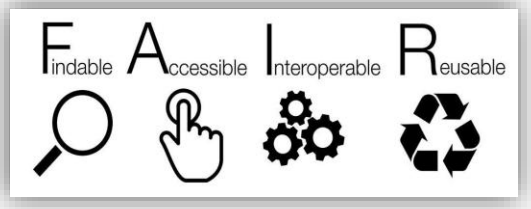
Facilitating Collaboration

Properly managed data can be easily shared and collaborated on among research teams, fostering interdisciplinary collaboration and accelerating the pace of scientific discovery

Long-Term Preservation and Access

Planning for the long-term preservation of research data ensures that valuable assets are archived and accessible for future studies, preventing data loss and promoting knowledge continuity

FAIRification of research data



Findable

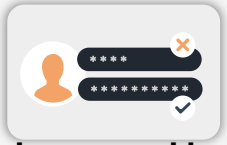
Data should be easy to find for both humans and computers.

This involves assigning globally unique and persistent identifiers (such as DOIs), providing metadata that describe the data in a standardized way, and ensuring that data are indexed by search engines

Accessible

Data should be easily accessible once found.

This means providing clear and open access permissions, ensuring that data are stored in repositories with appropriate access controls, and providing clear instructions on how to access the data



Interoperable

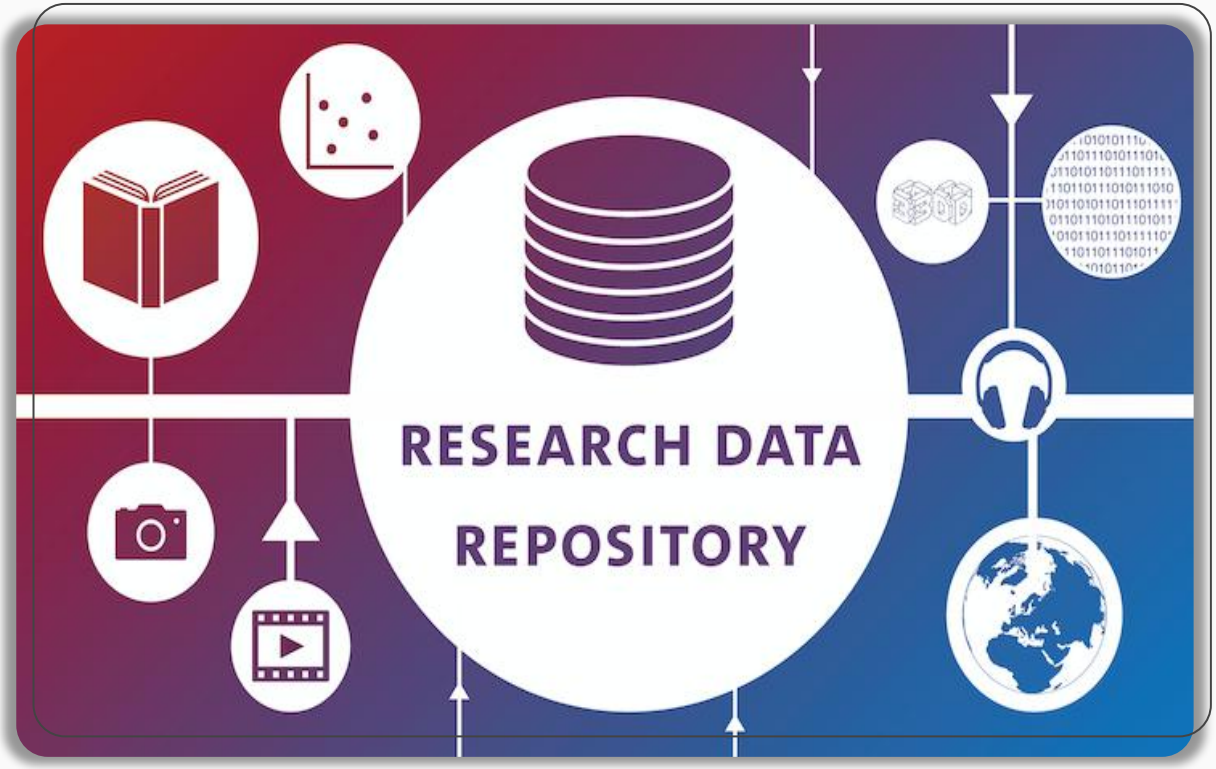
Data should be structured and formatted which can be integrated with other data sources and analyzed.

This involves using standard data formats, vocabularies, and ontologies, as well as providing clear documentation on the structure and meaning of the data

Reusable

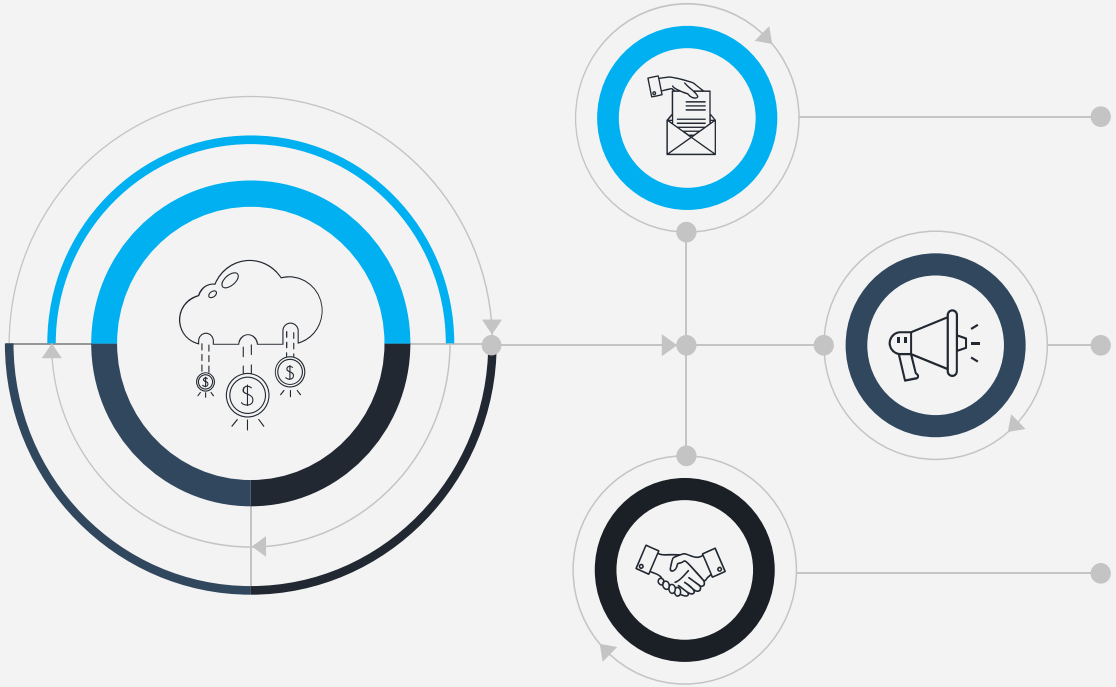
Data should be easy to reuse for future research purposes.

This involves providing clear licensing information, data should be well-documented, annotated and stored in a sustainable and long-term repository, providing sufficient contextual information to understand how the data were collected and processed



Research Data Repository (RDR)

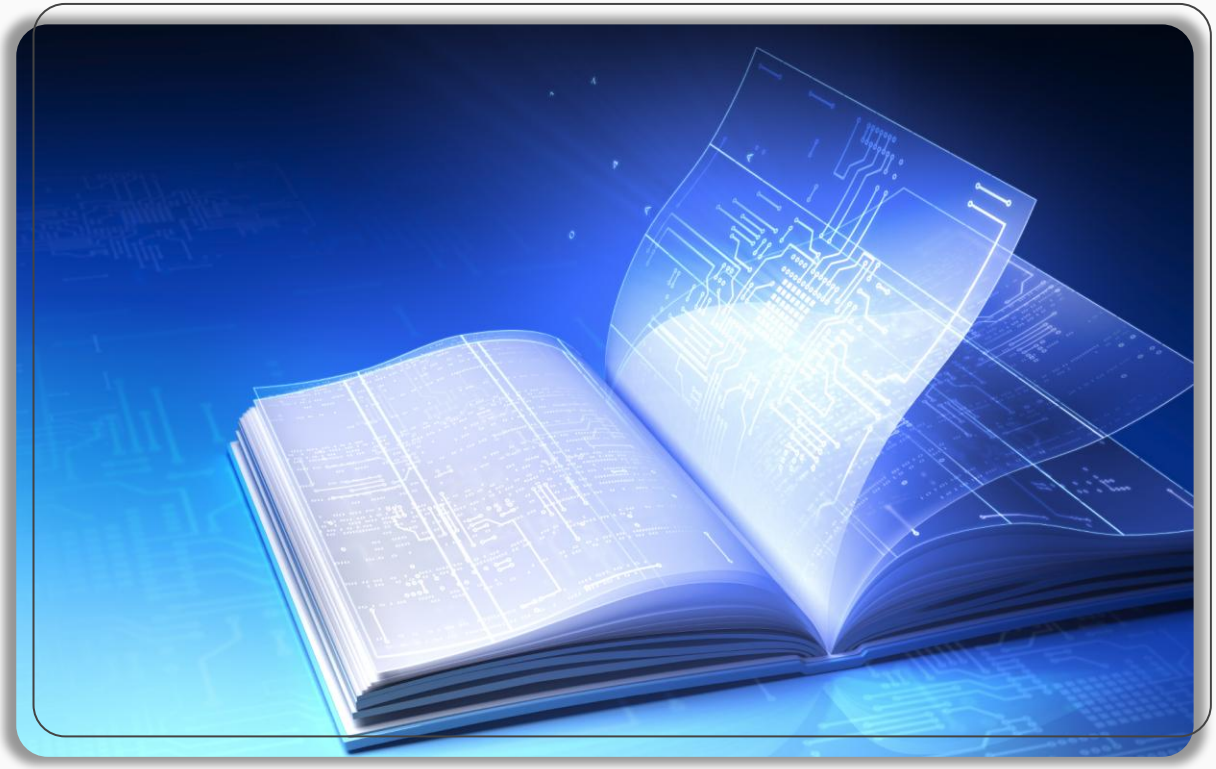
I2



It is Data library or Data archive

A unified data management platform; used to store and manage datasets from disparate sources.

The datasets can be used for sharing, analysis and reporting purpose.



Desirable Characteristics of Research Data Repository

13

Organizational Infrastructure

Free and Easy Access
The repository provides broad, equitable, and maximally open access to datasets and their metadata free of charge in a timely manner after submission, consistent with legal and policy requirements related to maintaining privacy and confidentiality, national data sovereignty, and protection of sensitive data.

Clear Use Guidance
The repository ensures datasets are accompanied by documentation describing terms of dataset access and use (e.g., reuse licenses and need for approval by a data use committee).

Risk Management
The repository has documented capabilities for ensuring that administrative, technical, and physical safeguards are employed to comply with applicable confidentiality, risk management, and continuous monitoring requirements for sensitive data.

Retention Policy
The repository provides documentation on policies for data retention.

Long-term Organizational Sustainability
The repository has a plan for long-term management of data, including maintaining integrity, authenticity, and availability of datasets; has contingency plans to ensure data are available and maintained during and after unforeseen events.

Digital Object Management

Unique Persistent Identifiers
The repository assigns a dataset a citable, unique persistent identifier (PID or DPI), such as a digital object identifier (DOI), to support data discovery, reporting (e.g., of research progress), and research assessment (e.g., identifying the outputs of Federally funded research).

Metadata
The repository ensures datasets are accompanied by metadata to enable discovery, reuse, and citation of datasets, using schema that are appropriate to, and ideally widely used across, the communities that the repository serves.

Curation and Quality Assurance
The repository provides or facilitates expert curation and quality assurance to improve the accuracy and integrity of datasets and metadata.

Broad and Measured Reuse
The repository ensures datasets are accompanied by metadata that describe terms of reuse

Common Format
The repository allows datasets and metadata to be accessed, downloaded, or exported from the repository in widely used, preferably non-proprietary, formats consistent with standards used in the disciplines the repository serves.

Provenance
The repository has mechanisms in place to record the origin, chain of custody, version control, and any other modifications to submitted datasets and metadata.

Technology

Authentication
The repository supports authentication of data submitters. The repository has technical capabilities that facilitate associating submitter PIDs with those assigned to their deposited digital objects, such as datasets.

Long-term Technical Sustainability
The repository has a plan for long-term management of data, building on a stable technical infrastructure and funding plans.

Security and Integrity
The repository has documented measures in place to meet well established cybersecurity criteria for preventing unauthorized access to, modification of, or release of data, with levels of security that are appropriate to the sensitivity of data (e.g., the NIST Cybersecurity Framework).

Institutional Repository vs Data Repository

14

Institutional Repository

- It focuses on collecting, preserving and disseminating scholarly outputs produced by a institute such as research articles, theses, dissertations, conference papers and other academic materials.
- It serves as a digital archive for intellectual output of the institution's faculty, student and staff.
- Softwares used to create IR: DSpace, Eprints, Fedora, Hydra, Drupal

Data Repository

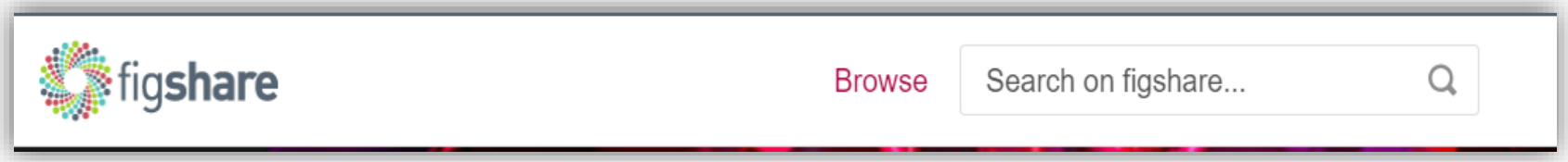
- It primarily focuses on storing and sharing research datasets
- It provides a platform for researchers to deposit, discover and access datasets related to various disciplines.
- It plays crucial role in promoting data sharing, reproducibility and transparency in research.
- Softwares used to create RDRs: CKAN, NADA, DataVerse

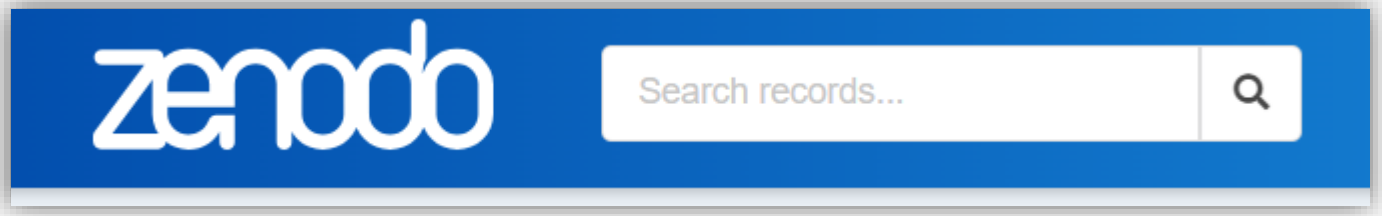
In nutshell…

Research Data Repository is essential for:

- Data Preservation
- Data Sharing
- Compliance with Funding Agency Policies
- Citation and Recognition
- Data Management and Discovery
- Data Security and Privacy

General Purpose Repository (examples)





Data Specific Repository (examples)

Registry of Research Data Repository

I5

It is an open science tool that offers researchers, funding organizations, libraries, and publishers an overview of existing international repositories for research data.

It is a global registry of research data repositories from all academic disciplines.

Operating since 10+ years and provides a curated index of over 3,000 research data repositories around the world from all disciplines.

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

THANKS!

rdm.outreach@inflibnet.ac.in