

Development of Interoperable RDRs System in India: A Conceptual Proposal

Manu T R¹ and Bhakti Gala²

¹Central Library, Indian Institute of Technology Delhi, Delhi

²School of Library and Information Science (SLIS), Central University of Gujarat,
Gandhinagar

Abstract

RDM services in India is at the nascent level. Most of the libraries are still exploring opportunities and gaining knowledge in this emerging area to implement RDM services in the future fully. RDR is one of the essential RDM services offered by libraries. In India, there are over 50 RDRs are covered in re3data.org. But the maximum of them are general data repositories that cover factual data sources. Also, these repositories are separate among institutional, ministries/funding agencies and national data repositories and have duplicate data elements. In India, there is a need to develop a national RDR that is interoperable with ministry/ funding agencies level and institutional level RDRs and need to establish the national level research data services like UK Data Service, Research Data Canada, Research Data Management - Science Europe, National Research Data Infrastructure Germany, Swedish National Data Service (SND), Landelijk Coördinatiepunt Research Data Management (LCRDM) - Netherlands etc. to take the initiative of developing the national level interoperable RDR system in India. Therefore, in this research study, researchers have proposed developing an interoperable RDR system in India. It also identified the pillars of RDM interoperability and RDM stakeholders in India. It proposed the RDR system in India, which comprises institutional, disciplined or ministries and higher education departments, and national level central RDR.

Keywords: Repositories, Discipline Data Repositories, OAI-PMH, RDR, Research data stakeholders

1. Introduction

Research Data Repositories (here after called RDRs) provides platform academic researchers to deposit, share, and access the research support data generated throughout research lifecycle. The repositories play crucial role in ensuring the accessibility, preservation and long-term availability for future access via linking between data and publications, regularly requirements of the scholarly publishers and funding organization (Manu & Gala, 2021). It covers the collection of various types of data including datasets, texts, spreadsheets, diaries, questionnaires, transcripts, images, pictures, and slides of research data. Most of the RDRs are

available freely and researchers can deposit the data. RDRs can be divided as institutional, national, international, disciplinary, multidisciplinary, and project-specific repositories. According to the Registry of RDRs (re3data.org) there are 3119 (as on May 31, 2023) repositories available worldwide. These include institutional access restricted repositories and open data repositories which host research data openly. In India, are 51 repositories found in re3data.org, including open government data (factual repositories), institutional repositories, and multi-purposed science data repositories. Though enormous repositories are available, a researcher may need help selecting suitable open data repositories by repository reputation, data protection guidelines, types of data deposits, support in tracking usage etc. Additionally, they also required the central platform where they can access all kinds of research data at a single point.

2. Literature Review

The literature provides the synthesis and evaluation of existing literature and published documents on data repositories, software and application tools. Many case studies and research studies on the comparison, critical assessment, features, advantages, and downsides of using various RDRs to store, archive, and share the research data management with other researchers have been published by significant scholars. The majority of these research focused on the evaluation of data repositories and open source software like “DataVerse, CKAN, Digital Commons, DSpace, EPrints, EUDAT, Fedora, Figshare, Greenstone, Invenio, Omeka, SciFLOW, WEKO, Zenodo linkub.ch, GeRDI, Invenio, RENKU, Blue Brain Nexus, Leibniz Data Manager, Google Dataset, BePress, Digitool, IQSS etc”. (Cudre-Mauroux, 2020) are available for making RDM repositories and data deposits and most them used by the ARL academic institutions (Johnston & Coburn, 2020). These benefits to data sharing practices include reuse, evidence-based claims should be made transparently. RDRs policies are needed to govern submission, selection, usage, and service levels (Witt, 2008). The continued efforts of the data ecosystem, liaison work, and continuing design to make the research data available via repositories are essential in research data management. A data ecosystem is defined as the individuals and technologies responsible for gathering, managing, and utilizing data as well as their connections. Consultation, negotiation, advocacy, integration, synthesis, translation, and mutual learning are all components of liaison work. Ongoing design outlines an adaptable data repository method. (Baker & Duerr, 2017). Data repositories that were publicly accessible and allowed to data deposit.

A data repository also known as a data archive or RDR is a structured collection of data, datasets, digital objects that are stored, managed, curate and provides the searchable and query-able interfacing entity. It manages the location where research data is registered, permanently stored, made accessible and retrievable, and curated (Johnston, 2017). RDRs are an opportunity for librarians to leverage their expertise in curation, outreach, and preservation while strengthening their long-standing relationships with academic departments to implement robust repositories that satisfy the needs of their communities (Gerwig, 2017). Treloar & Wilkinson, (2008) argue that RDRs should be supported for easy access to data and other information in reliable and consistent forms. Gradually data repositories are increasingly replacing institutional repositories of universities (Gowen & Meier, 2020) because institutional repositories platforms (EPrints, DSpace, Digital

Commons and OJS) also support data access, sharing, search, DOI management, accountability and impact. (Alsaad, O'Hara, & Carr, 2019; Macgregor, 2020).

The re3data.org indexes metadata of the RDRs established and maintained round the world. It offers services to the researchers, funding organizations, libraries, and publishers. It launched on May 28th, 2013. It indexed over 2300+ RDRs from around the world, and it also presented in typologies categories like institutional, disciplinary, and multidisciplinary project-specific repositories (Pampel, H. et al., 2013). FAIRsharing registry is also a collection of public RDRs that provides the standards, databases, policies and collection/ recommendations (Suhr et al., 2020).

According to Gramstadt (2012), a RDR is a digital platform where researchers can share the results of their work with others in the community. For universities and colleges to effectively capture, manage, and utilize institutional assets, repositories are crucial (Jacobs, Thomas, & McGregor, 2008). It can assist with administrative, research, and learning activities. Repositories are typically created with research data needs, interdisciplinary research projects, long-term preservation & access, file-based data storage, database-based metadata storage, and an intuitive web interface in mind (Curdt & Hoffmeister, 2015). The infrastructure and rules for data citation, deposit, sharing, and reuse are provided by a qualitative data repository (QDR). Furthermore, repositories also support intellectual property & copyright and privacy concerns and interconnection with other citation and reference management software i.e. Zotero (Karcher, Kirilova, & Weber, 2016). The RDRs establishment requires sustainable planning, identifying partners to reach the target audience, identifying the researcher's incentive and benefits, developing learning opportunities, considering the data repository as the one-stop solution and getting feedback for continuous improvement (Hiom et al., 2015; Gerwig, 2017).

The evaluation of the RDRs and catalogue platform describes the platforms' capabilities which comply with funder mandates, publisher requirements, researcher perception and user & organization expectations (Arguillas, Heslop, & Whyte, 2015). The capabilities majorly policy control & reporting and content organization, publishing & preservation capabilities such as Information security; domain and content scope; support for domain-specific workflows; discoverability; domain and context metadata management; version control; persistent identification; preservation metadata; file format checking, data integrity/fixity checking, and continuity support; open interface; clear user license and user community support (Ali, 2019). When selecting software to serve as the foundation of a repository for research data, several criteria must be taken into account, and the software should support both long-term preservation and access (Barsky et al., 2015). Gordon et al. (2015) found that successful data repositories needed protocols for technology and metadata development. Staff members also needed to learn new skills for using the repository software and for interacting with IT departments. The open access, documentation, accountability, openness, and crucial data security standards of research data management may be absent from the cost-effective method. Technology, data and metadata, rules, operation support & management, security, network infrastructure, shared scientific platform instrument, online storage, and are the nine influencing aspects for

scientific data repositories (Ishak et al., 2019; Macgregor, 2020). RDM stakeholders will be in charge of determining a data repository's primary needs for managing research outputs and identifying well-known repositories (Amorim et al., 2017).

The construction of a RDR may be hampered due to factors such as cost-effectiveness, a lack of open access principles, documentation, accountability, openness, and crucial data security components (Müller, Heiss, & Oberhoffer, 2017). Müller, Heiss, & Oberhoffer, (2017) used three configurable elements to develop the RDMS in the OAS-based OpenCampus system. Those elements are forms (stored as nodes), study trees and input/output plugins to use a different format for information inputs/outputs actions. NFDI4 BioDiversity Consortium has been established to support the biodiversity, ecology & environmental community with data management in German (Glöckner & Diepenbroek, 2019). RESTful is the data service platform that provides real-time data storage and a standard interface.

The meta-repositories of data are created to support national missions to ensure quality data and accessibility. Such repositories are Australian Research Data Commons (ANDS), Bielefeld Academic Search Engine (BASE) and Connecting Repositories (CORE), Data.gov, Data Archiving and Networked Services (DANS), DataBridgE, DataCite, EDAT, ICS/World Data System (WDS), OpenAIRE, OpenDOAR, OneRepo and ShARE (Vitale, 2017). Neurofibromatosis (NF) data portal allows data to be accessible to data scientists and help researchers and funders identify gaps (Allaway et al., 2019). NTUData, a prototype researcher-centered system built by National Taiwan University, helps in data management and curation and encourages data sharing (Chiang, Lee, & Jeng, 2020). As key metadata providers, researchers and library professionals are responsible for describing the data.

The most important aspects of RDRs to promote research data deposit and sharing are statistics, international interest, good visibility in scholarly search engine results, strong community support for the repository, and making the repository more aesthetically pleasing. Encouragement and best practices for managing research data, own IPR, and removing third-party rights are provided (Gramstadt, 2012). JISC has been encouraged in the creation of several repositories like "ETHOS (<http://www.ethos.ac.uk/>), JorumOpen (<http://www.jorum.ac.uk/>), Depot (<http://depot.edina.ac.uk/>)" etc. in the U.K. and providing the user services, preservation services & shared infrastructure services (Jacobs, Thomas, & McGregor, 2008).

3. RDRs in India

Platforms known as RDRs host research data in the open (Patel, 2016). As of August 23, 2023, re3data.org listed 51 Indian RDRs or centers on a variety of topics, including life science, medicine, natural science, agriculture, forestry, horticulture, and veterinary medicine, geosciences (including geography), humanities, and social sciences, biology, social and behavioral sciences, economics, and basic biological and medical research. The major ones are "ESSO-INCOIS, Hyderabad, ICRISAT Dataverse Network, ICSSR Data Service (INFLIBNET), Krishi (Knowledge-based Resource Information Systems Hub for Innovations in Agriculture) (Gutam, 2017), India Meteorological Department, Indian Oceanographic Data Centre, Indian Space Science

Data Center (ISSDC), Kodaikanal Solar Observatory (KSO), National Informatics Centre, World Data Centre for Geomagnetism, Mumbai” etc. (Anilkumar, 2018; Bhardwaj, 2019; Gunjal & Gaitanou, 2016). These repositories support multiple subjects, content types, formats, persistent identifiers, data access, restriction, metadata standards, APIs, and other provision features (Bhardwaj, 2019). Indian open data resources are available online, but most do not comply with the FAIR (Findable, Accessible, Interoperable and Reusable) Data and not registered with any international data repositories such as Re3Data.org (Francis & Das, 2019).

Research data software is a platform to build, deposit and made access to research data openly without or minimum intellectual property rights restrictions. Harvard Dataverse & Dryad data repositories were discussed by their features (Patel, 2016). Mendeley Data is an open RDR allowing the researcher to upload their research data and make it available openly through the platform (Bhoi, 2018). The comparative study of open-source data repository software found that Dataverse focuses mainly on social science data, large datasets of geospatial data handled, and CKAN supports visualization features and 223 extensions for making it a more powerful Data Repository tool (Mahato & Gajbe, 2012).

4. National Level Initiatives

With a number of efforts, the Indian government has prioritized the degree of research quality (Pal & Singh, 2019). In order to support national planning and development, they want to promote data exchange (Payal, Awasthi, & Tripathi, 2019). A portal called Open Government Data (OGD) Platform India was created to facilitate the Government of India’s open data project (Open Government Data (OGD, 2020)). In India, the INFLIBNET Centre has set up open access repositories, such as the Shodhganga repository, which offers a place for researchers to deposit their Ph.D. theses and make them accessible to the public, and the Shodhgangotri: Repository of Indian Research In Progress, which offers access to approved synopses or research in-process deposited by the universities for registering for doctoral programs (Tripathi et al., 2017). They have also provided the scholarly network known as VIDWAN and the research information management system known as IRINS, which distribute the scholarly profiles of researchers as well as their research content at a larger level. National Repository of Open Research Data (NRORD), which Patel (2016) proposed, is a national repository of open research data for managing research data. The hosting of NRORD can be handled by a central national agency; as an example, INFLIBNET and nodal institutes have been chosen based on their particular research specialties. INFLIBNET may create the Indian Academic RDR (IARDR) for universities to deposit their research data with the help of libraries, according to Pal & Singh’s (2019) suggestion.

5. Institutional Level Initiatives

Tripathi et al., (2017) assess the researcher’s perception of research data sharing & library support at Jawaharlal Nehru University (JNU), India. Most of the researchers used the computer storage options such as CDs, DVDs, portable hard disks & pen drives for further use, there is no culture of data sharing in their fields, and they faced data storage problems during data creation. Most research scholars needed support

from the libraries to organize, maintain and preserve their raw research data. The laboratory notes, experimental measurements, statistical data, clinical measurements, sample data, survey responses, questionnaires, photographs, films, and observations are produced by the researcher at Aligarh Muslim University (AMU). They are willing to share their research data (Saeed & Ali, 2019) through a personal website, open data repositories, academic, social networks, supplementary files, and make data available within a research group. Doctoral research scholars of Bharathiar University are unaware of the RDM services offered (Fazal & Chakravarty, 2019). Libraries of major academic and research institutes in India offer research support services, institutional repositories for archiving scholarly publications and thesis & dissertation, institutional archives and conducts training programs, seminars, and workshops to create awareness on research data and impact in research data sharing (Gunjal & Gaitanou, 2016). Now, all institutions and universities must think about managing research datasets effectively (Tripathi & Pandey, 2018).

The conceptual framework presents the workflow of the data lifecycle & outline of data management in its various phases (Patel, 2016). Singh, Monu, & Dhingra (2018) has attempted to provide an outline for research and academic institutes for managing research data at the institutional level. The outline includes framing data sharing policies, motivating researchers, accessibility of infrastructure, orientation program, costing, and well-trained staff. Tripathi & Pandey, (2018) presented the framework that includes the data management phase, data storage & hosting phase and data usage phases to provide guidelines to implement the RDM in higher education institutes. Surveying the top twenty university libraries in the world for RDM services, Tripathi, Shukla, and Sonker (2017) proposed a model for RDM services that university libraries might use to manage RDMS. Agadi (2017) investigated how to integrate RDM services by taking into account academic scholarships, data discovery services, knowledge development, discovery application, and data sharing in the university system.

6. Objective of the Study

This conceptual proposal aims to develop an interoperable RDRs system in India. The focused objectives are as follows:

- ❖ To identify the pillars and their role in the development of interoperable RDR systems in India.
- ❖ To propose a workflow and process of RDRs system in India.
- ❖ To identify the significant stakeholder of RDR in India.

7. Methodology

A Systematic Literature Review (SLR) has been undertaken to find existing centralized and interoperable RDR systems worldwide. Authors used known citations and bibliographical databases such as “Scopus, Web of Science, Dimensions AI, EBSCO, Library, Information Science & Technology Abstracts (LISTA), Library & Information Science Source (LISS), Indian citation index” and other scholarly engines including “Google Scholar, Microsoft Academic Search, BASE (Bielefeld Academic Search Engine), CORE” to find

case studies, working papers and research reports. The Library and Information Technology Association (LITA) data management guide, Yang Hu Ran three-level model for developing research data management and repositories and other similar conceptual frameworks were studied. With obtained literature, authors identified the major pillars and their role in the development of the RDRs system in India. They developed the workflow and process for the same.

8. Results

8.1. Pillars of RDM Interoperability in India

RDM Services at the national level required the involvement of the many RDM stakeholders in India including data creator, facilitator, Government at large and secondary data users. Figure 1. presents the pillars of RDM interoperability in India. The each of these pillars need to work collaboratively and should have interoperability to establish the sustained RDM services in India. Data creator and facilitator at the institutional level would have more responsibility to manage, store, archive and share / submission of research datasets, along with thesis, publications and/or project reports under data governance and as per the institutional policy guidelines / compliance by the Government at large which force to undertake the RDM practices.

The Government funding agencies including Ministry of Education and Higher Education Departments in India are still at the nascent level in identifying the value of research data created from funded research and implementing policies related to data management, sharing and accessibility. Since these funding agencies have been providing the monetary grant and facilitating the required advanced infrastructure to undertake the scientific research in various disciplines in an institution. There are several scheme(s), scholarships, research grants provided to excellence in research and promote scientific knowledge. Therefore, they need to be built the ethical guidelines, policies and regulations which researchers are required to adhere throughout their research projects. Such documents will be focused on DMP requirement, scope of data, data license, responsibilities of data storage, security & access/sharing, data quality, awareness, and training, expected data types and formats, supported metadata standards, information on data embargo, data centre/repository, publication provision in the open access journals, ethical guidelines, and regulatory framework etc. The known international funding agencies like National Science Foundation (NSF), National Institutes of Health (NIH), Wellcome Trust, federal funding and other research funding agencies have been playing major role in development RDM practices with their compliance and requirements. These funding agencies does encourage researchers through incentive of additional funding to create data management plans, manage, store, and publish.

National research data services centre by the Government of India required to support the RDM initiatives by the institute and ministries and higher education departments. National research infrastructure of RDM and facilities required to develop a coherent research environment that enables researchers to find, access, contribute and effectively use research data services to produce quality research and impact. It will play the

significant role in the advocating and co-ordinate (as national coordination point) the data management practices for researchers, institutions, and research funders; brings together experts in the field including research data librarians/data stewardship, policy makers, ICT specialists, managers of diverse research institutions and funding agencies; creation of working, task group and discussion forum; establish the RDM networking and international collaborations; provides the trainings and services on RDM skill and competencies development; enabling data discovery through central supernatural RDRs; development of policies and standards and provides the shared additional services which benefits to the institution, to the researcher and to research funder. National research data services centre will also be responsible for developing policies and imposed them to adopt and mandate to develop RDM services in the institutes in India.

National research data services of other countries like ‘Australian Research Data Commons (ARDC), UK Data Service, Research Data Canada, Research Data Management - Science Europe, Research Data Netherlands, National Coordination Point Research Data Management (LCRDM) - Netherlands, National Research Data Infrastructure, Germany, Swedish National Data Service (SND)’ etc. are national level government initiative aimed at addressing the issues confronting researchers in the management of their research data. They have been working on developing the road map towards implementing a national RDM services at their respective national level and working the RDM stakeholders to ensure research data is openly available to support innovation that benefits to all research community.

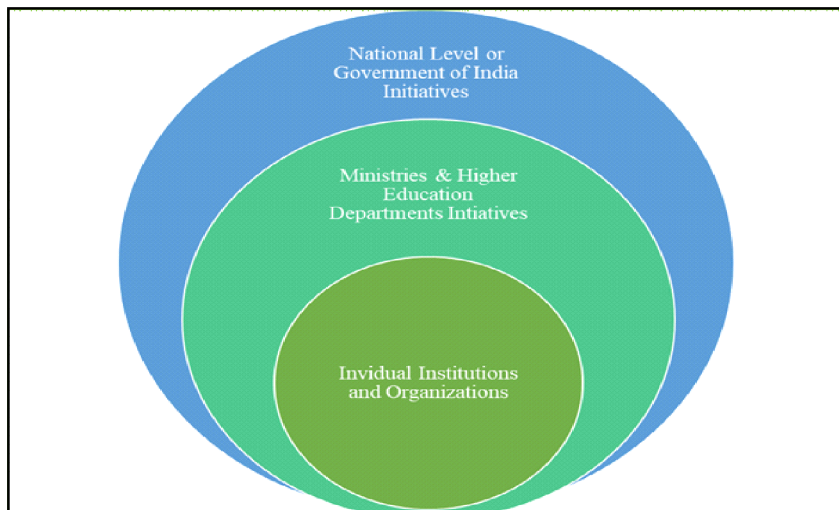


Figure 1: Pillars of RDM interoperability in India

8.2. RDM Stakeholders in India

The major stakeholders of RDM involved the Who created research data? Who helps to create research data? Who facilitates management of research data? And who uses the research data? RDM is a collaborative effort, and it requires the involvement of all support units to establish the RDM practices at the institute

level to the national level. Figure 3 provides the major RDM stakeholders in India who are pillars in carrying out RDM practices in any institute. It has identified majorly four categories of stakeholders i.e., data creator, facilitator, government at large and users. Individual stakeholders have their roles and responsibilities at large in developing RDM best practices.

8.2.1. Data Creator

Data creators or generators are the primary stakeholders because they will hold legal and ethical rights to the data they have generated or gathered. The researchers, faculty members, and scientists who have been involved in the primary research can be called first and foremost RDM stakeholders. The research supervisor or principal investigator who provide guidance to researchers are also major RDM stakeholders. The institution which facilitates infrastructure to undertake research and funding agencies that are granted the fund on research projects are stakeholders of RDM. All the researchers, faculty and scientific members, research supervisors, principal investigators, institutions, and funding agencies are playing a vital role in primary data creation or generation.

8.2.2. Facilitator

The facilitators are the actual working group for carrying out RDM practices including data acquisition, classification, quality control, setting internal data policies, providing training infrastructure for data management, curating, preserving, and providing access to data in any institutes. It is being required the collaborative work among institute leadership, academic departments/schools, R&D unit, library, library committee, research support unit, IT support unit, administrative offices, scholarly publishers and professional associations and networking bodies to establish best practices in the RDM. Therefore, they are a major RDM stakeholder in India.

8.2.3. Government at large

Government support and cooperation are required to build any academic and research institutes system. The S&T institutes in India are hierarchically developed by various ministries, higher education departments and organizations of the Government of India. Therefore, these units can play a significant role in developing RDM practices in their autonomous, parent and associated institutions. The major education ministries and departments are in India: Ministries of education, ministry of science and technology, department of higher education and other ministries and departments. If these ministries / departments can develop institutional common policies on research integrity, ethics, RDM, and sharing data as a public good and make them mandatory to adopt such policies in every institution that will encourage in institutional leadership including libraries work towards the implementation of RDM in their campuses.

8.2.4. Users

The end-users who will reuse the research data and get the benefits from RDM practices are another category of RDM stakeholders in India. Such end users may be the students, other researchers, project

members, faculty / scientific members, government policymakers, journalists, government officials and others. These users can verify the published research, conduct secondary analysis, and use data for teaching purposes etc.

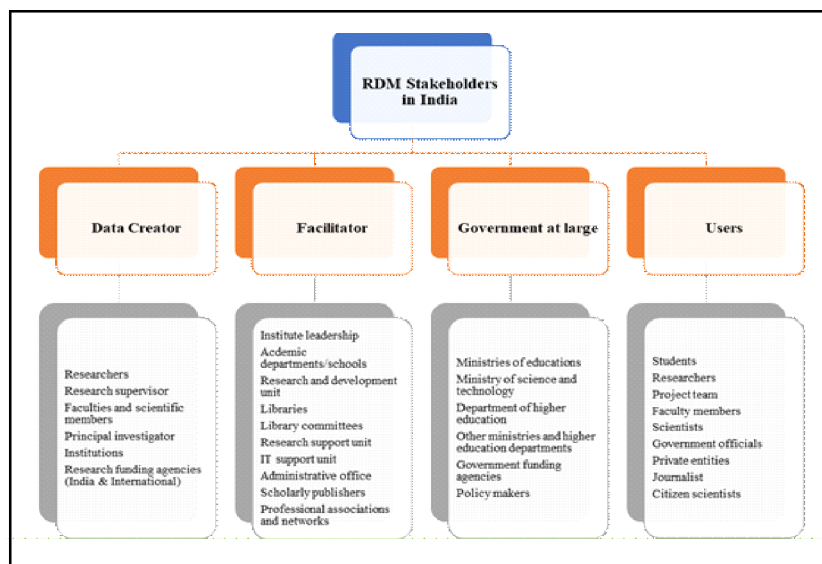


Figure 3: RDM stakeholders in India

8.3. RDRs in India

RDR can be divided into institutional, national, international, disciplinary, multidisciplinary and project specific repositories. According to Registry of RDRs (re3data.org) there are 3879 (as on 05th March 2023) RDR are available around the world. These include the institutional access restricted repositories and open data repositories which host research data openly. In India are 50 RDR found in re3data.org which are open government data (factual data repositories), institutional repositories, and multi-purposed science data repositories. RDRs are few in numbers. Since enormous repositories are existed, researcher may not be able to select the suitable open data repositories by repository reputation, safely, repository can take the data you want to deposit, repository sustain the data value and support analysis and tracking data usage.

Therefore, the academic and research institutions need to take the initiatives for establishing the Open Archive Initiative – Protocol for Metadata Harvesting (OAI-PMH) compliant RDR using open-source software that allow ministries / higher education departments / disciplined RDR and central national level agency to expose and harvest their metadata of research data. The figure 5.3 the presents the workflow of RDR System in India. The primary responsibility of building the data repository includes policy compliance, guidelines, and quality control of data to be deposited into institute level RDR. Discipline / ministries / agency/ higher education department wise and central national repository can only harvest the research data from their autonomous and parent institutions and organizations. The cumulative research data will be accessed from

the disciplined repositories as well as central national level RDR. It enables research data more discoverable and enhance data usage. The users will benefit to retrieve research data across these three layers of repositories. However, research data may find duplicates among these repositories.

8.3.1. Institutional Level RDR

Institutions has sole responsibilities in building the OAI-PMH compliant RDM in their respective institutions. Therefore, institutions stakeholders including researchers, research supervisor/faculty members, library team, IT support team and institute leaderships etc. will be required to work jointly to establish repositories. The Institutional based on RDM stakeholders will play essential role in developing policy compliance, guidelines, data quality control and research data curation activities including the research dataset collection/receiving, cleaning, classification, metadata creation, assign license conditions, and build trusted data repositories to publish research data at the institutional level. The institutional based RDR will provides the registered and restricted (limited) access within campus and only processed data, published data, and analyzed data available for access with other researchers and confidential data and data which has the personal identifiable information will not be made available. Such data also not allowed for harvesting through OAI-PMH compliant. By the large, each institution will host their own institutional RDRs and act as data providers administrative system (Lagoze & Sompel, 2015) that supports the OAI-PMH to harvest metadata to service provider i.e., disciplined or ministries and higher educational department level and national level RDR. Institutions will also ensure the data available for access with all obligations around the data to maintain research data ethics.

8.3.2. Disciplined or Ministries and Higher Education Department Level RDR

The second level RDRs are hosted by the group of disciplined institutions or institutions funded by the Ministries of Education and Higher Education Departments, Government of India. These second level RDRs will host the research data generated by their all-respective institutions and researchers. Research data will be harvested through the individual OAI-PMH compliant institutional RDRs regularly. The coordinating institution of these second level RDR need to be central server setup, install the open metadata harvest system including configuration of OAI base URL and then issues OAI-PMH requests for collecting metadata from their respective institutional RDR (Lagoze & Sompel, 2015). These are value added repositories along with institutional RDRs that enables the research communities to view these repositories to search and retrieve the research data by discipline wise and data generated (research results) by funded ministries and higher education departments at once search without any restrictions. However, users will have to follow source URL of institutional RDR to download full datasets. Ministries of education and higher education departments should fund and support their R&D unit and institutions to take these initiatives and develop the RDRs.

8.3.3 National Level Central RDR

The national level centralized the RDR is required country like India in which academic and research institutions faculty members, scientists and researchers have been involved in the research activities regularly and generated extremely large size of data in the various types and different formats. As presented in figure 2 centralized RDR will need to harvest the metadata of research data from the individual institutions and as well as disciplined or ministries and higher education level RDRs. The nodal office, a national agency or network under the UGC, Government of India can host the central server for RDR with required OAI-PMH systems. Since UGC is a statutory body of the Government of India through an Act of Parliament for the coordination, determination, and maintenance of standards of university education in India (UGC, 2022). It is a right unit at the top level to take responsibility to developing centralized RDR and make enabling each institution and ministries / higher education departments will follow the strategic framework in developing the RDR and offering RDM services. The users can access the research data of various institutional RDRs through central search interface without any registration, restriction, and limitation. The users will follow the links to source of institutional RDR to access the full data set(s). This kind of centralized national RDR increases the academic and research prestige of the country and helps in disseminating research data worldwide.

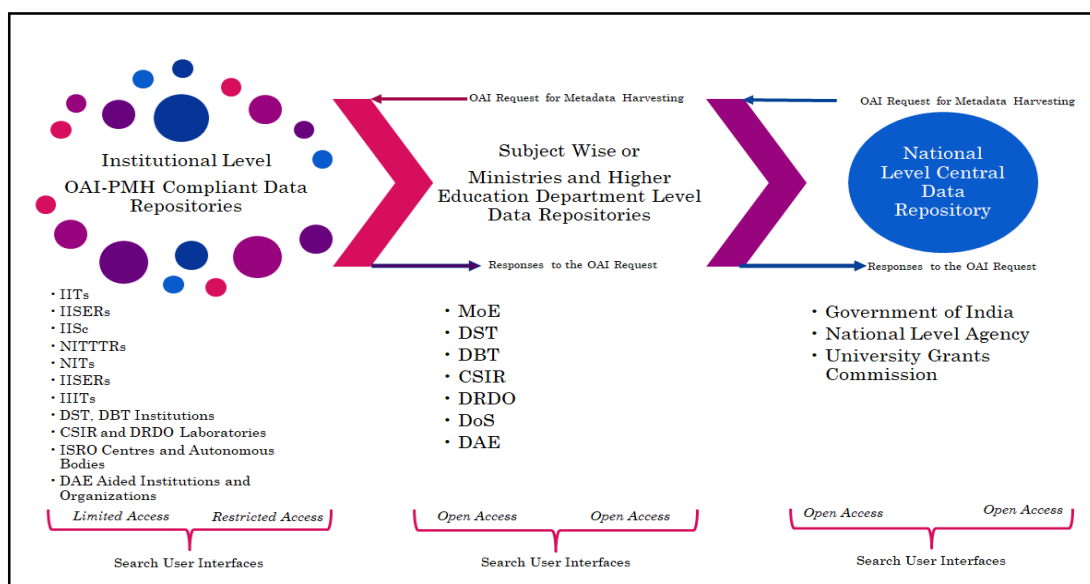


Figure 2: Workflow of RDRs system in India

9. Discussion

Major findings, including individual institutions and organizations, ministries & higher education departments, national level, and central level government of India are major pillars in developing the interoperable RDR system in India. As presented in above figure 1, each of the pillars must work collaboratively

and have interoperability to establish the sustained RDR system in India. Individual institutions or organizations are responsible for developing OAI-PMH-compliant repositories in their respective institutions. The facilitator will play an essential role among RDM stakeholder (Figure 3) in developing policy compliance, guidelines, data quality control and research data curation activities, including the research dataset collection/receiving, cleaning, classification, metadata creation, assigning license conditions, and build trusted data repositories to publish research data at the institutional level. The institutional-based RDR will provide registered and restricted (limited) access within the campus; only processed data, published data, and analyzed data will be available for access to other researchers, and confidential data will not be made available. The Government funding agencies, including the Ministry of Education and Higher Education Departments, can host the RDRs of the group of disciplined institutions or institutions funded by the Ministries of Education and Higher Education Departments, Government of India. These second level RDRs will host the research data their all-respective institutions and researchers generated.

Research data will be harvested through the individual OAI-PMH-compliant institutional RDRs regularly. The coordinating institution of these second-level RDR needs to be central server setup, install the open metadata harvest system, including the configuration of OAI base URL and then issue OAI-PMH requests for collecting metadata from their respective institutional RDR (Lagoze & Sompel, 2015). These are value-added repositories and institutional RDRs that enable the research communities to view these repositories to search and retrieve the research data by discipline wise and data generated (research results) by funded ministries and higher education departments at once search without any restrictions. However, users must follow the source URL of institutional RDR to download full datasets. Ministries of education and higher education departments should fund and support their R&D unit and institutions to take these initiatives and develop the RDRs.

A national research data services centre by the Government of India or a national-level agency/network is required to support the RDR initiatives by the individual institutions, ministries, and higher education departments. Therefore, centralized RDR will need to harvest the metadata from the individual institutions and disciplines or ministries and higher education level RDRs. The nodal office, a national agency or network under the UGC, Government of India can host the central server for RDR with required OAI-PMH systems since UGC might be the suitable unit at the top level to take responsibility to develop a centralized RDR and enable each institution and ministry / higher education departments to follow the strategic framework in developing the RDR and offering RDM services. The users can access the research data of various institutional RDRs through a central search interface without any registration, restriction, and limitation. The users will follow the links to the source of institutional RDR to access the full data set(s). This kind of centralized national RDR increases the academic and research prestige of the country and helps in disseminating research data worldwide.

10. Conclusion

India is lacking in establishing the national level RDM services and repositories like National research data services of other countries like 'Australian Research Data Commons (ARDC), UK Data Service, Research

Data Canada, Research Data Management - Science Europe, Research Data Netherlands, National Coordination Point Research Data Management Netherlands, National Research Data Infrastructure, Germany,' etc. have been working on developing the road map towards implementing a national RDRs at their respective national level and working with RDM stakeholders of parent institutions to ensure research data is openly available to support innovation that benefits to all research community. Hence, the proposed conceptual framework in this research study may give way to developing national-level RDM services and repositories to enable more discoverable research data and enhance data usage. The cumulative research data will be accessed from the disciplined repositories as well as central national level RDR, where the researcher will get the benefit of retrieving research data across these three levels of repositories i.e. institutional level OAI-PMH compliant data repositories, subject-wise or ministries or higher education development level data repositories and national level central data repository. The major ground-level stakeholders are data creators, facilitators, government at large, and users.

References

- Agadi, K.B., (2017). Research data management services in library in academic libraries perceptions. In World Data System Asia-Oceania Conference 2017. Kyoto: Kyoto University.
- Ali, A. K. (2019). Research data management support. The 14th Munin Conference on Scholarly Publishing. Tromsø: UiT University.
- Allaway, R.J., et al., (2019). Engaging a community to enable disease-centric data sharing with the NF Data Portal. *Scientific Data*, 6(319). <https://dx.doi.org/10.1038/s41597-019-0317-x>
- Alsaad, A., O'Hara, K. & Carr, L. (2019). Institutional repositories as a data trust infrastructure. *WebSci'19 Companion*. Boston: ACM Digital Library.
- Amorim, R. C., Castro, J. A., Silva, J. R. & Ribeiro, C. (2017). A comparison of research data management platforms: architecture, flexible metadata and interoperability. *Universal Access in the Information Society*, 16(4), 851–862. <https://dx.doi.org/10.1007/s10209-016-0475-y>
- Anilkumar, N. (2018). Research data management in India: a pilot study. *EPJ Web of Conferences*, 186, 03002. <https://dx.doi.org/10.1051/epjconf/201818603002>
- Arguillas, F., Heslop, J. & Whyte, A. (2015). Reviewing research data platform capabilities at Cornell Institute for Social and Economic Research (CISER). Edinburgh: A Digital Curation Centre Case Study.
- Baker, K. S. & Duerr, R. E. (2017). Research and the changing nature of data repositories. In L. R. Johnston, *Curating research data, volume 1: practical strategies for your digital repository*. Chicago: Association of College and Research Libraries.

Barsky, E., Laliberté, L., Leahey, A., Trimble & Leanne. (2015). Collaborative research data curation services: a view from canada. In L. Johnston, *Curating Research Data, Volume 1: Practical Strategies for Your Digital Repository*. Chicago: Association of College and Research Libraries.

Bhardwaj, R. K. (2019). Content analysis of indian RDRs: prospects and possibilities. *DESIDOC Journal of Library & Information Technology*, 39(6), 280-289. <https://dx.doi.org/10.14429/djlit.39.6.15137>

Bhoi, N. K. (2018). Mendeley Data repository as a platform for research data management. In B. Rautaray, D. K. Swain, & C. Swain, *Marching Beyond Libraries: Managerial Skills and Technological Competencies* (pp. 481–487). Bhubaneswar: Kalinga Institute of Industrial Technology.

Chiang, P. N., Lee, J.S. & Jeng, W. (2020). Who's in charge? discovering the autonomy in an institutional data repository for research data curation and sharing. *iConference 2020 Proceedings*. Boras: iSchools.

Cudre-Mauroux, P. (2020). Design considerations on SWITCH's connectome vision. Zorich: A SWITCH Innovation Lab.

Curdt, C. & Hoffmeister, D. (2015). Research data management services for a multidisciplinary, collaborative research project: Design and implementation of the TR32DB project database. *Program*, 49(4), 494–512. <https://dx.doi.org/10.1108/PROG-02-2015-0016>

Fazal, F. A. & Chakravarty, R. (2019). Role of library in research support: a study of Bharathiar University. *Library Philosophy and Practice*(2780).

Francis, D. J. & Das, A. K. (2019). Open data resources for clean energy and water sectors in India. *DESIDOC Journal of Library & Information Technology*, 39(6), 300-307. <https://dx.doi.org/10.14429/djlit.39.6.15227>

Gerwig, K. J. (2017). Current outreach and marketing practices for RDRs. In L. Johnston, *Curating Research Data, Volume 1: Practical Strategies for Your Digital Repository*. Chicago: Association of College and Research Libraries.

Gerwig, K. J. (2017). Current outreach and marketing practices for RDRs. In L. Johnston, *Curating Research Data, Volume 1: Practical Strategies for Your Digital Repository*. Chicago: Association of College and Research Libraries.

Glöckner, F. O. & Diepenbroek, M. (2019). NFDI4BioDiversity: biodiversity, ecology and environmental data. *Biodiversity Information Science and Standards*, 3(e37282). <https://dx.doi.org/10.3897/biss.3.37282>

Gordon, A. S., Millman, D. S., Steiger, L., Adolph, K. E. & Gilmore, R. O. (2015). Researcher-library collaborations: data repositories as a service for researchers. *Journal of Librarianship and Scholarly Communication*, 3(2), P1238.

Government of India. (2022, 09 16). Open Government Data (OGD) platform India. Retrieved from Data.gov: <https://data.gov.in/>

- Gramstadt, M. T. (2012). *Kultivating kultur: increasing arts research deposit*. ARIADNE.
- Gunjal, B. & Gaitanou, P. (2016). *Research data management: a practical approach to overcome challenges to boost research*. IASSIST Conference 2016. Bergen: IASSIST.
- Gutam, S. (2017). *Research data management in national agricultural research systems of Bangladesh and India*. *Geoinformatics Research Papers*, 5(BS1002). <https://dx.doi.org/10.2205/CODATA2017>
- Hiom, D., Fripp, D., Gray, S. & Snow, K. &. (2015). *Research data management at the University of Bristol: Charting a course from project to service*. *Program*, 49(4), 475–493. <https://dx.doi.org/10.1108/PROG-02-2015-0019>
- Ishak, I. et al., (2019). *Influencing factors in determining RDR infrastructure for research data management*. *International Journal of Engineering and Advanced Technology (IJEAT)*, 9(1). <https://dx.doi.org/10.35940/ijeat.A2645.109119>
- Jacobs, N., Thomas, A. & McGregor, A. (2008). *Institutional repositories in the UK: the JISC approach*. *Library Trends*, 57(2), 124-141. <https://dx.doi.org/10.1353/lib.0.0035>
- Johnston, L. R. & Coburn, L. (2020). *Data sharing readiness in academic institutions*. Data Curation Network.
- Johnston, L. R. (2017). *Introduction to data curation*. In L. R. Johnston, *Curating research data, volume 1: practical strategies for your digital repository* (pp. 1-29). Chicago: Association of College and Research Libraries.
- Karcher, S., Kirilova, D. & Weber, N. (2016). *Beyond the matrix: repository services for qualitative data*. Moynihan Institute of Global Affairs, 1.
- Krier, L., & Strasser, C. (2014). *Data management for libraries: a LITA guide*. Chicago: ALA Tech Source.
- Lagoze, C., & Sompel, H. V. (2015, 01 08). *The Open Archives Initiative Protocol for Metadata Harvesting*. Retrieved from Open Archive: <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- Macgregor, G. (2020). *Enhancing content discovery of open repositories: an analytics-based evaluation of repository optimizations*. *Publications*, 8(1), 8. <https://dx.doi.org/10.3390/publications8010008>
- Mahato, S. S. & Gajbe, S. B. (2018). *A Comparative study of open source data repository software: Dataverse and CKAN*. *Library Herald*. <https://dx.doi.org/10.5958/0976-2469.2018.00005.2>
- Mallery, M. (2014). *Book Reviews: A Review of Data Management for Libraries: A LITA Guide*. *Journal of Electronic Resources Librarianship*, 26(4).
- Manu, T. & Gala, B. (2021). *Growth and development of research data management in India: a visual analysis of published literature indexed in Scopus*. *Library Philosophy and Practice (e-journal)*.
- Müller, J., Heiss, K. I. & Oberhoffer, R. (2017). *Implementation of an open adoption research data management system for clinical studies*. *BMC Research Notes*, 10(252). <https://dx.doi.org/10.1186/s13104-017-2566-0>

Pal, B. & Singh, S. K. (2019). Indian academic RDR (iandr) with inflibnet: a futuristic plan. 12th International CALIBER-2019 (pp. 36-46). Bhubaneswar: Kalinga Institute of Industrial Technology.

Pampel, H. et al., (2013). Making RDRs visible: the re3data.org registry. PLOS ONE. <https://doi.org/10.1371/journal.pone.0078080>.

Patel, D. (2016). Research data management: a conceptual framework. *Library Review*, 65(4/5), 226-241. <https://dx.doi.org/10.1108/LR-01-2016-0001>

Patterson, L., Bothma, T., & Deventer, M. V. (2018). From planning to practice: an action plan for the implementation of research data management services in resource-constrained institutions. *South African Journal of Libraries and Information Science*, 84(2).

Payal, Awasthi, S. & Tripathi, M. (2019). A selective review of literature on research data management in academic libraries. *DESIDOC Journal of Library & Information Technology*, 39(6), 338-345. <https://doi.org/10.14429/djlit.39.06.14451>

Simons, N. & Richardson, J. (2012). New roles, new responsibilities: Examining training needs of repository staff. *Journal of Librarianship & Scholarly Communication*, 1(2), P1051. <https://dx.doi.org/10.7710/2162-3309.1051>

Singh, N. K., Monu, H. & Dhingra, N. (2018). Research data management policy and institutional framework. 2018 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services (ETTLIS). Noida: IEEE Xplore.

Treloar, A. & Wilkinson, R. (2008). Rethinking metadata creation and management in a data-driven research world. 2008 IEEE Fourth International Conference on eScience (pp. 782–789). IEEE. <https://dx.doi.org/10.1109/eScience.2008.41>

Tripathi, D. P. & Pandey, S. R. (2018). Developing a Conceptual Framework of Research Data Management for Higher Educational Institutions. 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services (ETTLIS-2018). Noida: IEEE Explore. <https://dx.doi.org/10.1109/ETTLIS.2018.8485193>

Tripathi, M., Chand, M., Sonkar, S. K. & Jeevan, V. J. (2017). A brief assessment of researchers' perceptions towards research data in India. *IFLA Journal*, 43(1), 22–39. <https://dx.doi.org/10.1177/0340035216686984>

Tripathi, M., Shukla, A. & Sonker, S. (2017). Research data management practices in university libraries: a study. *DESIDOC Journal of Library & Information Technology*, 37(6), 417-424. <https://dx.doi.org/10.14429/djlit.37.11336>

UGC. (2022, 05 09). Genesis. Retrieved from University Grant Commission: <https://www.ugc.ac.in/page/Genesis.aspx>

Witt, M. (2008). Institutional repositories and research data curation in a distributed environment. *Library Trends*, 57(2), 191-201. <http://dx.doi.org/10.1353/lib.0.0029>