CHAPTER-2

SUSTAINABLE RESEARCH DATA MANAGEMENT: BALANCING LONG-TERM PRESERVATION WITH IMMEDIATE RESEARCH NEEDS

Chaithra A.M[1] and Dr. Ramesha B[2]

[1]Research Scholar, Department of Library and Information Science, Bangalore University, Bengaluru

Email: chaithramlisc1993@gmail.com

[2]Senior Professor, Department of Library and Information Science, Bangalore University, Bengaluru

Email: bbramesha@gmail.com

ABSTRACT

Sustainable Research Data Management (RDM) is crucial for ensuring that data generated through scientific research is both accessible in the short term and preserved for long-term reuse and discovery. As the volume and complexity of research data continue to grow, researchers and institutions face significant challenges in balancing immediate data needs-such as rapid analysis and dissemination-with the long-term goals of data preservation and integrity. This paper explores the principles and practices that underpin sustainable RDM, emphasizing the importance of planning, effective storage solutions, metadata and adherence to International standards such as the FAIR principles (Findability, Accessibility, Interoperability and Reusability). It examines the key challenges faced by researchers, including resource constraints, technological obsolescence and disciplinary variations and discusses strategies for overcoming these obstacles. The paper also highlights case studies that demonstrate successful approaches to RDM, showcasing how institutions and researchers can implement frameworks that serve both immediate and long-term research needs. Ultimately, this work calls for a comprehensive approach to RDM that ensures the longevity, accessibility and usability of research data for future generations.

E Book ISBN: 978-81-981814-6-6 | DOI: https://doi.org/10.34256/vadlibs.25.02.19

Keywords: Sustainable Research Data Management, Long-term Preservation, Immediate Research Needs, Data Lifecycle, FAIR Principles, Data Sharing, Metadata, Research Integrity

INTRODUCTION

In today's data-driven research landscape, the importance of effective Research Data Management (RDM) cannot be overstated. As the volume, variety and complexity of research data increase across disciplines, ensuring that this data is appropriately managed, preserved and made accessible for both immediate and future use has become a key challenge. Sustainable RDM refers to the integration of strategies, practices and tools that allow research data to be efficiently managed throughout its lifecycle-ensuring both short-term usability and long-term preservation. Achieving this balance is critical not only for meeting the immediate needs of research teams but also for maximizing the potential for future discovery, verification and reuse.

The growing recognition of the value of data as a critical research asset has led to an increased emphasis on responsible data management, especially in the context of reproducibility, transparency and collaboration. Researchers and institutions must ensure that their data is organized, documented and stored in ways that support ongoing and future research. However, this task is fraught with challenges. Research data must often be made accessible and actionable quickly to meet the immediate needs of project timelines and publication schedules. Yet, these needs can sometimes conflict with the long-term goals of preserving data in formats and repositories that ensure future usability, compliance with ethical guidelines and alignment with institutional or funding agency requirements.

The tension between immediate research needs and long-term data preservation arises from several factors, including resource constraints, technological obsolescence and differing standards across academic disciplines. Additionally, the rapid pace of technological change and the increasing use of big data and complex datasets further complicate efforts to balance these competing priorities. While immediate access to data is essential for researchers, ensuring the longevity of data for future generations of scholars, policymakers and the public is equally important.

This paper explores the concept of sustainable RDM by examining the principles, strategies and best practices that can be employed to create a cohesive framework that addresses both short-term research needs

and long-term preservation goals. The paper discusses the critical elements of data management planning, including data collection, documentation, storage, sharing and curation and offers insights into overcoming the challenges inherent in maintaining this balance. Ultimately, the goal is to provide a comprehensive understanding of how researchers and institutions can adopt sustainable RDM practices that not only safeguard the immediate usability of their data but also ensure its enduring value for future academic, scientific and societal progress.

Research Data Management (RDM) is an essential aspect of the modern research ecosystem, ensuring that data generated through scientific inquiry is organized, accessible and preserved for future use. As the volume of research data continues to grow exponentially, managing these data sets becomes increasingly complex. One of the key challenges faced by researchers, institutions and policymakers is balancing the immediate needs of research projects with the long-term requirements for data preservation. Sustainable Research Data Management (RDM) addresses this challenge by providing frameworks and strategies that ensure data is not only usable in the short term but also preserved for future discovery, reproducibility and reuse.

THE IMPORTANCE OF SUSTAINABLE RDM

Sustainable RDM involves the careful planning, organization, and preservation of data throughout its lifecycle from creation to long-term archival. This approach is critical for several reasons:

- **Reproducibility and Transparency**: For research to be trusted, it must be reproducible. Proper data management ensures that others can access and validate the results of a study, fostering transparency and scientific integrity.
- **Collaboration and Sharing**: Research today is increasingly collaborative and interdisciplinary. Effective RDM facilitates the sharing of data across research groups, institutions and even countries enhancing the speed and quality of scientific discovery.
- **Long-Term Impact**: Many data sets have long-term value that extends beyond the original research project. Preserving data allows future researchers to build on existing knowledge, test hypotheses or even uncover new insights that were not previously possible.

- **Compliance with Policies and Regulations**: Funding agencies, governmental bodies and academic institutions often mandate certain standards for data management. Ensuring data is properly managed and preserved helps meet legal, ethical and funding requirements.

However, achieving sustainability in RDM is not straightforward. Researchers and institutions must balance immediate data needs-such as quick access and analysis-against the long-term objectives of data preservation and archiving.

KEY CHALLENGES IN BALANCING IMMEDIATE AND LONG-TERM NEEDS

- **Resource Constraints**: Implementing effective data management strategies requires both financial and human resources. Research teams may prioritize the immediate needs of their projects (e.g., analysis, publications and presentations) over long-term preservation, which can incur additional costs for storage, curation and metadata management.
- **Data Formats and Technologies**: The rapid pace of technological advancements means that data formats, software and hardware used in research can quickly become obsolete. Ensuring that data is stored in formats that will be accessible in the long term while also being usable for immediate analysis can be a significant challenge.
- **Data Volume**: With the increasing use of big data, IoT and other high-volume data-generating technologies, the sheer amount of data produced by research projects is often difficult to manage. Balancing the needs of immediate access with the infrastructure required for long-term storage can overwhelm available resources.
- **Changing Research Needs**: The dynamic nature of research often means that data requirements change rapidly. Immediate needs may include fast processing or temporary access, while long-term requirements demand structured data with metadata, documentation and preservation strategies that might not be feasible in a time-constrained project.
- **Disciplinary Variations**: Different research fields have varying expectations and practices regarding data management. While

disciplines such as genomics or environmental science may require highly structured and standardized data formats for long-term preservation others like the humanities may have less stringent requirements for preservation and reuse.

PRINCIPLES OF SUSTAINABLE RESEARCH DATA MANAGEMENT

A sustainable approach to RDM incorporates several key principles designed to address both short-term and long-term needs:

1. Data Lifecycle Management

The concept of the data lifecycle is central to sustainable RDM. The data lifecycle covers the stages from planning and creation through processing, analysis, sharing and preservation. Researchers must plan for each of these stages, considering both immediate and long-term needs.

Planning: Developing a Data Management Plan (DMP) at the outset of the research project is crucial. The DMP should outline strategies for data collection, organization, sharing and preservation. It should also address issues like ethical considerations, metadata standards and long-term storage solutions.

Creation and Collection: During data collection, it's important to follow consistent protocols that will allow the data to be easily organized, annotated and prepared for long-term storage.

Processing and Analysis: Immediate needs often focus on processing and analyzing data quickly. This requires efficient tools and platforms that allow for rapid data manipulation without compromising on the documentation and organization needed for preservation.

Sharing and Reuse: Once the research is completed, data should be made available for sharing. Repositories, such as institutional repositories, subject-specific archives and general-purpose platforms like Zenodo or Dryad allow for data sharing while ensuring compliance with preservation standards.

2. Data Storage and Preservation

Long-term data preservation is a critical component of sustainable RDM. It ensures that data is maintained in a manner that preserves its usability and accessibility over time. The following practices are important for balancing short-term accessibility with long-term preservation:

Choosing Appropriate Storage Solutions: Immediate research needs often require fast, local storage solutions such as cloud platforms or external drives. However, for long-term preservation, data should be stored in trusted repositories that adhere to international standards and ensure data integrity and security.

Format Selection: Data should be stored in open, non-proprietary formats that are widely supported and unlikely to become obsolete. For example, storing data in formats like CSV for tabular data or NetCDF for multidimensional data ensures future accessibility.

Metadata and Documentation: Effective metadata is essential for long-term usability. Proper metadata ensures that data is understandable to future users with details on its origin, structure, format and how it was created. This documentation should be integrated during the creation and analysis stages of data management.

Regular Backups and Validation: To ensure the long-term preservation of data, regular backups should be taken and the integrity of stored data should be periodically validated. This helps prevent data loss due to hardware failures or unforeseen circumstances.

3. Interoperability and Standards

Sustainable RDM requires that data be interoperable across platforms, tools and systems. This is especially important when dealing with data from diverse sources or disciplines. Adhering to international standards (e.g., the FAIR principles - Findability, Accessibility, Interoperability and Reusability) helps ensure that data can be easily shared, understood and reused over time.

FAIR Principles: Applying FAIR principles in the management of research data ensures that data is structured and described in a way that is both human- and machine-readable, increasing it's potential for reuse and collaboration.

Data Standardization: Standardizing data formats, naming conventions and metadata schemas within a research discipline or project facilitates data sharing and reduces barriers to long-term access and reuse.

4. Training and Support

One of the key factors in the success of sustainable RDM is the availability of training and support for researchers. Researchers, particularly those in the early stages of their careers, may lack the knowledge or skills to implement effective data management practices.

Institutions should provide training on RDM best practices, tools and strategies for ensuring both immediate and long-term data needs are met.

CASE STUDIES AND EXAMPLES

Several institutions and projects have demonstrated successful models of sustainable RDM:

The UK Data Service: This service provides access to a wide range of high-quality, reusable data sets. By adhering to stringent data management practices and preserving datasets for long-term access, the UK Data Service ensures that valuable data remains available to researchers long after the original study has concluded.

The Data Conservancy at Johns Hopkins University: The Data Conservancy is a long-term archiving and preservation service that provides both immediate and future access to datasets. By applying best practices in data curation, metadata standards and digital preservation, it facilitates the long-term usability of data.

CONCLUSION

Sustainable Research Data Management is an ongoing challenge that requires a careful balance between immediate research needs and long-term preservation goals. By implementing sound data management strategies, adhering to international standards and providing adequate resources and training, researchers and institutions can ensure that valuable data is both useful in the short term and preserved for future generations. This holistic approach not only supports the reproducibility and transparency of scientific research but also fosters collaboration, innovation and long-term scientific progress.

REFERENCES

[1] Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., & Read, E. (2015). "Data sharing by scientists: Practices and perceptions." *PLOS ONE*, 10(6), e0127726.

[2] Trust, T., & Wilkerson, J. R. (2017). "Sustainable data management for research." *College & Research Libraries News*, 78(5), 258-264.

[3] Borgman, C. L. (2012). The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059-1078.

[4] Beagrie, N., & Houghton, J. (2014). The Value of Digital Preservation: A Study of the Benefits of Long-Term Data Storage and Access. JISC.

[5] National Institutes of Health (NIH). (2015). *Data Management and Sharing Policy*.

[6] Data Management Working Group (2019). *Digital Preservation Handbook*. Digital Preservation Coalition.

[7] Groth, P., & Bizer, C. (2011). "Publishing and Using Linked Data on the Web." *Data Science Journal*, 10, 213-231.

[8] Gibbons, S. (2017). "Data management and sharing: Current trends and future prospects." *Library Trends*, 65(3), 352-367.

[9] European Commission (2020). *European Open Science Cloud (EOSC) Strategic Implementation Plan*.

[10] Sweeney, L., & Ohm, P. (2019). *Ethics of Research Data Management and Sharing*. Springer.

[11] Houghton, R. J., & Cooper, C. S. (2016). *Ensuring Long-term Access to Research Data: A Guide for Researchers*. Research Libraries UK (RLUK).

[12] Van den Eynden, V., Corti, L., Woollard, M., & Bishop, L. (2011). *Data Management and Sharing in Social Science Research*. SAGE Publications.

Published by Vyom Hans Publications,
ISBN (Digital Download and Online): 978-81-981814-6-6
ISBN (Book): 978-81-981814-5-9
Year: April, 2025
DOI: https://doi.org/10.34256/vadlibs.25.02.19

How to Cite

Chaithra, A. M. & Dr. Ramesha, B. (2025). Sustainable Research Data Management: Balancing Long-Term Preservation with Immediate Research Needs. Ed. by Kumar, J. & Sedam, M. V. in "Advancing Library and Information Science: Innovations, Practices, and Future *Directions*" (pp. 19-26). Vyom Hans Publications. https://doi.org/10.34256/vadlibs.25.02.19 ; ISBN: 978-81-981814-6-6
