

Strategies and Challenges in Archiving and Sharing Research Data: Few Practicing examples from JNU

By

Dr. Shiva Kanujia Sukula, Deputy Librarian*

Dr. Parveen Babbar, Deputy Librarian*

Dr. Neelam Thapa, Assistant Professor**

&

Dr. Somesh Vishwakarma, Assistant Librarian*

*Dr B R Ambedkar Central Library, Jawaharlal Nehru University, New Delhi, India

**Dept of LIS, Dr. Harisingh Gour Vishwavidyalaya, Sagar, Madhya Pradesh, India

Abstract:

The strategies and practices in archiving technologies have been considered digital solutions to support the researchers in education and learning. The inclusion of digital technologies for archiving and facilitating the research in various streams has been in vogue in most of the academic libraries worldwide. The awareness of archiving tools and technologies among the library staff to provide support in delivering the information services to the researchers and other educational users has been analyzed from time to time. The present study shares the current practices in context with archiving techniques and higher education through a literature review (rapid-review method) and a case study method. Incorporating accessibility in various information products and services with a wide range of learning tools among the researchers is highly in demand. The vast amount of data generation, sharing, storing and exchanging, etc., are in practice. The inclusion of data collection and storing technologies has been inevitable and necessary to allow easy access to vast amounts of information resources for research students.

There are experiences for sharing best practices, and various open tools support partnerships with communities of support for all aspects of data archiving preservation workflow. These processes involve the transfer, processing, and various activities related to data description and preservation. The current practices and support mechanisms have evolved with the moving times. Significant examples of collaborative initiatives are observed: digitization, maintaining digital collections, virtual collection repatriation, etc. A growing community is available for data-related collaboration to provide data resources for planning and executing digital preservation programs. The processes, such as transcription, including authority research, have been developed rigorously with community support. The paper will also discuss the archival policies and data management tools used in libraries for archiving and for the research support to its users.

Keywords: research needs, research data, library support, data archiving, data preservation.

Introduction

Academic and research libraries are witnessing the proliferation and significance of Research Data. There have been speculations and guidelines from the funding agencies to enforce various policies regarding the archiving of data and its availability. Enhancing the knowledge, it is essential to understand the intricacies of data as:

Defined by the United States Office of Management and Budget (<http://www.whitehouse.gov/omb>) "Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This 'recorded' material excludes physical objects (e.g., laboratory samples)".

It is significant to discern that the research data do not include a few items such as Trade secrets, commercial information, personnel, medical information, etc. For a deeper understanding and practical context in libraries, it is essential to decide the nature and set of files and information related to validating the research findings. Due to limitations in storage and technological aspects, prioritizing the subset of r data is necessary as everything cannot be saved or stored. Sometimes, few raw data files may be useful at a particular step, which may be redundant later on. It is crucial to estimate the kind of data that would be useful to reproduce the research results.

Data Archiving & Preservation

With the overlapping steps and practices, sometimes, backing up and archiving seems similar to many people. It is significant to differentiate the concepts such as "backup" and "archiving" which are often applied interchangeably in saving the file or some information. It is necessary to realize that both are entirely different processes as the 'backup' process is meant for making multiple copies of various files keeping the variation in mind. The archiving is intended for " preservation of files based on 'as-is'. , as a final) record" (source: DataONE education module).

Planning of preserving the data involves the local archive storage (local server, network or any digital repository). There are various public data repositories meant for multiple subject areas in www.re3data.org, for example. Most of the time, digital repositories in universities (called "institutional repository") function as an archive and a platform to share the data/information. This data/information is often accompanied by DOI, and discoverable through Google, etc.

Significant components for archiving data include; understanding the file formats for long term access, the data and research documentation at various steps, data retention policy for no less than the minimum retention period, the use of human subjects, and device research. The additional data sharing and archiving necessities are comprised of issues such as Ownership and privacy in the context of copyright and IP ownership and ethical requirements like privacy and confidentiality. There are few other significant issues such as retention, distribution, and control of the data.

Data Archiving, Integrity and Sharing at University Level Environment

The fragile and susceptible nature of digital data always demands precautions and measures like refreshment and replication. Software Obsolescence is another major issue while archiving the final version of dataset(s) to fully access the future. It is advisable to store the data in a proprietary format with the provisions to ensure access to older datasets, followed by migration of older datasets or emulation using a suitable virtual machine. The best practice may include converting data to an open format for extended preservation, exchange, and sharing. For data storage and backup, for example, The UK Data Archive (<http://www.data-archive.ac.uk/create-manage/storage>) has provided the guidelines:

- Usage of data formats for storage strategy with long-term availability and at least two other storage and location forms by maintaining three copies (original copy, external local copy and a remote copy)
- Checking data integrity and updated knowledge of institutional strategy, data retention policies in the context of funding agencies, publisher/ aggregators, and the organization

Data management plans in university libraries contemplate expanding increased public access for government-funded research. To improve research data sharing, digitally formatted data should be stored and made available for research and retrieval. This is possible by applying various strategies such as improving public access by protecting privacy and confidentiality etc., ensuring the creation of data management plans for long-term preservation, including cost-effectiveness and profitability. It is crucial to keep the merits, strategies, and mechanisms of data management plans consistent with organizational policies to promote data submission to increase public access. This also includes aspects such as interoperability, standardization of data citation/attribution approvals, training for library staff, relevant depositors, and long-term preservation goals. There are conditions under which data must be shared! Some community standards and funding agency guidelines, and data-sharing mandate express few variations. These variations are governed by discipline type and data type, and the procedures include: data sharing should take place "within a reasonable time

frame," depending on the nature of the research project, for further policy development. This further envisages the following terms of embargo periods in the context of factors such as political nature/ educational purposes / commercial ventures / intellectual property rights.

The above discussion leads to observation that the research data and other resources need reliable digital storage at various higher education scenarios. In this context, archiving process, Data Management Plan can include the documentation of Ownership and responsibility for the data with the support of "Memoranda of Understanding" (MOUs); authorized access and restriction measures, along with backup copies, distributed over a length of time at decided locations in relevant formats with adequate descriptive metadata (Smithsonian Libraries and Archives - Natural History Building, 2021).

Background

The considerable amount of data, sharing, storage, and sharing etc., is in practice. The content of data collection and storage technology is inevitable and necessary to provide research students with easy access to large amounts of data. There is an experience of sharing best practices and various open tools working with communities that support the entire data collection and retention workflow. These processes include multiple operations related to the processing, description, and protection and transportation of data. This study shares current archiving methods and practices in university libraries through the literary review (rapid review method) and the case study method. It is complicated to adopt, and Incorporating access to a wide range of information products and services, including a wide range of learning tools, is highly demanding among researchers. This study is divided into two aspects: review of the literature and a case study of JNU Library.

Review of Literature

The present study has applied the Rapid review method due to time limitations and observes the most recent practices (literature from the past two years' 2020-2021 publications). Rapid reviews are a form of knowledge synthesis in which the systematic review process components are simplified or omitted to produce information promptly (Khangura, Konnyu, Cushman, Grimshaw, & Moher, 2012). Yet quick reviews might be susceptible to biased results due to streamlining the systematic review process (Kumar & Sharma, 1998)(Ganann, Ciliska, & Thomas, 2010) (Watt et al., 2008). The following working definition, "a rapid review is a type of knowledge synthesis in which components of the systematic review process are simplified or omitted to produce information in a short period"(Khangura et al., 2012).

Areas for the rapid reviews: the present paper has focused upon the following areas:

- Data Archiving & Preservation processes and knowledge content
- Research data and information sharing
- Knowledge users, techniques, and services

Some of the important studies under the rapid reviews are elaborated below:

T Balogun and T Kalusopa published how web archiving of indigenous knowledge (IKS) can be done through institutional repositories in South Africa. The authors provided a framework for Web archiving IKS-related websites in South Africa (Balogun & Kalusopa, 2021)

A Mkadmi in his book *Archives in the Digital Age: Preservation and the Right to be Forgotten* published the key concepts of digital archives, its methods and Strategies, how archives are done in the age of Digital Humanities. The author also explained the role of big data in digital archives and finally how right to be forgotten relates to the preservation of archives (Mkadmi, 2021).

A Slavic, R Siebes and A Scharnhorst in their book *Linking Knowledge* explained what the Need for Knowledge Organization is and how to classify Open Data Linked for Knowledge Organization. What are the Challenges and Opportunities of Linked Open Data and publishing a Knowledge Organization System as Linked Data (Slavic, Siebes, & Scharnhorst, 2021)

Faten Hamad, Maha Al-Fadel and Aman Al-Soub in their paper *Awareness of Research Data Management Services at Academic Libraries in Jordan* discussed the roles, responsibilities and challenges. The author elaborated that RDM services requires Library professionals with new skills and collaboration for working with researchers and library users so that they can manage their research data. The paper elaborated the requirements and challenges in academic libraries in Jordan to provide RDM services (Hamad, Al-Fadel, & Al-Soub, 2021)

A Young Yoon, Angela P. Murillo and Paula Anders McNally examined the digital preservation course content through an analysis of course syllabi to understand what is taught in library schools related to the Data Archiving and Preservation Process (Yoon, Murillo, & Anders McNally, 2021)

LC Gleim and S Decker in their paper *Open Challenges for the Management and Preservation of Evolving Data on the Web* discussed the challenges and propose data persistence layer for data management and preservation, paving the way for increased interoperability and compatibility (Gleim & Decker, 2020)

Anjana R. Bunkar and Dhaval D. Bhatt discuss the role of library in towards Research Data Management System and identified the perception of Researchers & Academicians of Parul University. The paper concluded that researchers and academicians in library are more concerned about their intellectual property rights while sharing the data on the public domain (Bunkar & Bhatt, 2020)

M Ashiq, MH Usmani and M Naeem did a systematic literature review on research data management practices and services. In the literature they found that an active collaboration is required and combination of measures is required to better manage research data among stakeholders and university services departments to identify the challenges and issues (Ashiq, Usmani, & Naeem, 2020).

How libraries are support the Research data management is explained by Fernanda Gomes Almeida and Beatriz Valadares Cendon. The authors have surveyed the literature to propose taxonomy of services to support research data management. The paper gives the guideline for institutions or professionals who wish to develop RDMS (Almeida & Cendón, 2020).

Case study of JNU Library

Dr B R Ambedkar Central Library, Jawaharlal Nehru University (JNU) promotes Data sharing so that the research data is available to a broader audience, and thus further promotes the long-term preservation. The staff of the JNU Central Library provides the help to faculty, researchers and students with support on long term management of data and curation services. This helps the users to preserve their research data into the future in a trusted repository. The suggested repository can be discipline specific data repository or centre of it can be National or International Repository. Similarly data preservation and archiving are also need by the publishers for publication in journal or other publications, but these are generally added in non-curated repositories by the users.

The concept of data sharing started with users sharing the data on personal or lab websites, Electronic Lab Notebooks (ELNs), wikis, and similar tools. Since these platforms and sharing mode are not secured and authentic, the libraries came in forefront to provide the best solution that ensures that data is discoverable, accessible, and preserved over the long term. JNU Central Library is regularly helping and guiding its researchers and users to select an appropriate repository, data journal, or other strategy for sharing data.

Discussion

The repository option to deposit the data depends on the Repository and its policies and the term and conditions like how the data will be accepted, what are the requirements of submission and is there any cost or fees for depositing the research data. The detailed long term preservation policy needs to be studied by the user before submitting the data.

The role of data management for archiving and sharing is not limited to providing the data collaboration and sharing tools to the users and researchers but also involves helping them with the data collection and analysis. The libraries around the world are providing the data management services which also includes support of high performance computing and guiding with Copyright and Intellectual Rights of the data. Data management is incomplete without Metadata Services, Security Support, Storage, Backup and Data Recovery Services.

Conclusion

The Strategies in Archiving and Sharing Research Data need to be ethical and there is legal obligation when archiving and sharing research data of the researchers and the users. The data need to be secured and access need to be authentic. All the data at receiving and giving end need to be as per the license agreement. Sharing of data helps the researchers to validate their data and reuse if for further research. The data can be reused for teaching and even commercial purposes. Since the data can be part of research project and can have the commercial and financial value. While sharing the data libraries need to secure the data and sort the challenges of ethics, legal, financial, cultural and technological nature.

References:

- Almeida, F. G., & Cendón, B. V. (2020). Library support for research data management. *Qualitative and Quantitative Methods in Libraries*, 9(2), 109–126.
- Ashiq, M., Usmani, M. H., & Naeem, M. (2020). A systematic literature review on research data management practices and services. *Global Knowledge, Memory and Communication*.
- Balogun, T., & Kalusopa, T. (2021). Web archiving of indigenous knowledge systems in South Africa. *Information Development*, 02666669211005522.
- Bunkar, A. R., & Bhatt, D. D. (2020). Perception of researchers & academicians of Parul University towards research data management system & role of library: A study. *DESIDOC Journal of Library & Information Technology*, 40(3).
- Ganann, R., Ciliska, D., & Thomas, H. (2010). Expediting systematic reviews: methods and implications of rapid reviews. *Implementation Science*, 5(1), 56.
<https://doi.org/10.1186/1748-5908-5-56>

- Gleim, L. C., & Decker, S. (2020). Open Challenges for the Management and Preservation of Evolving Data on the Web. In *MEPDaW@ ISWC* (pp. 64–70).
- Hamad, F., Al-Fadel, M., & Al-Soub, A. (2021). Awareness of research data management services at academic libraries in Jordan: Roles, responsibilities and challenges. *New Review of Academic Librarianship*, 27(1), 76–96.
- Khangura, S., Konnyu, K., Cushman, R., Grimshaw, J., & Moher, D. (2012). Evidence summaries: the evolution of a rapid review approach. *Systematic Reviews*, 1(1), 10. <https://doi.org/10.1186/2046-4053-1-10>
- Kumar, A., & Sharma, R. (1998). *Marketing Management*. Atlantic Publishers & Distributors (P) Limited. Retrieved from <https://books.google.co.in/books?id=CJeNPwAACAAJ>
- Mkadmi, A. (2021). *Archives in the Digital Age: Preservation and the Right to be Forgotten*. John Wiley & Sons.
- Slavic, A., Siebes, R., & Scharnhorst, A. (2021). Publishing a Knowledge Organization System as Linked Data. *Linking Knowledge*.
- Smithsonian Libraries and Archives - Natural History Building. (2021). Best Practices for Storing, Archiving and Preserving Data. Retrieved from <https://library.si.edu/research/best-practices-storing-archiving-and-preserving-data>
- Watt, A., Cameron, A., Sturm, L., Lathlean, T., Babidge, W., Blamey, S., ... Maddern, G. (2008). Rapid reviews versus full systematic reviews: An inventory of current methods and practice in health technology assessment. *International Journal of Technology Assessment in Health Care*, 24(2), 133–139. <https://doi.org/DOI:10.1017/S0266462308080185>
- Yoon, A., Murillo, A. P., & Anders McNally, P. (2021). Digital Preservation in LIS Education: A Content Analysis of Course Syllabi. *Journal of Education for Library and Information Science*, 62(1), 61–86.