Research Journal of Library Sciences ISSN 2320–8929
Vol. **14(2),** 1-6, July **(2026)** *Res. J. Library Sci.*

Research Data Management Practices, Data Sharing Concerns and Training Needs among Academics and Research Scholars: Evidence from Banaras Hindu University, India

Bhawana[1*] and Satya Prakash Singh[2]
[1]Department of Library and Information Science, Banaras Hindu University, Varanasi, Uttar Pradesh, India
[2]Satya Prakash Singh, Junior Research Fellow, Department of Library and Information Science, Banaras Hindu University, Varanasi, Uttar Pradesh, India
bhawanaapandey07@gmail.com

Available online at: www.isca.in, www.isca.me
Received 05th June 2026, revised 10th July 2026, accepted 20th July 2026

Abstract

Research data management has emerged as an integral part of research integrity, transparency, preservation and reuse. This study aims to analyse the RDM practices, data-sharing concerns, and training needs of academics and research scholars at Banaras Hindu University (BHU). It was designed as a descriptive survey, and an online questionnaire was used during March-April 2024. One hundred and seven valid responses were received from the Faculties of Arts, Science and Social Science. The analysis reveals that standard office documents are common across disciplines, whereas Science respondents produce a more diverse range of machine-generated, structured and software-related data. Some respondents use central servers and repositories, but most storage remains dependent on personal devices. Although many respondents are willing to share data upon request, concerns about misuse or misinterpretation, legal and ethical issues, and the risk of losing competitive advantage limit wider sharing. Awareness of metadata, DOI and data citation is comparatively higher than awareness and use of institutional RDM policy and data management plans. The paper recommends discipline-specific RDM training, repository-based storage, institutional data-sharing guidelines and library-led advisory services to foster a stronger culture of RDM at the university.

Keywords: Research Data Management; Data Management Plan; Metadata; Data Sharing; Academic Libraries; Banaras Hindu University.

Introduction

Research data have become a critical scholarly asset because contemporary research increasingly depends on digital evidence, reproducible workflows and responsible sharing. Research data management (RDM) refers to the organisation, description, storage, preservation, access and reuse of data throughout the research lifecycle. In universities, effective RDM improves the credibility of research, reduces data loss, supports collaboration and enables future verification of findings[1,2]. Poorly managed data, by contrast, may remain inaccessible, undocumented or unusable even when the underlying research is valuable[1,2].

Academic libraries are closely connected with RDM because of their long-standing expertise in metadata, repositories, information organisation, preservation and user training. International studies have shown that libraries are increasingly involved in RDM services, including data management planning, data literacy, repository support, data citation and research consultation[3,4]. However, the development of RDM remains uneven across institutions and disciplines. Science researchers often handle larger and more technical datasets, while humanities and social science researchers may face different challenges relating to documentation, ethical use, confidentiality and interpretation[5,6].

In the Indian context, RDM services are still emerging, and many institutions are in the early stages of developing formal policies, infrastructure and training mechanisms[7,8]. This creates a need for institution-level empirical studies that identify how researchers actually create, store, document and share data. The present study focuses on Banaras Hindu University (BHU), a large multidisciplinary university, and examines RDM practices across the Arts, Science and Social Science faculties. The study contributes to the LIS literature by providing discipline-specific evidence to guide university libraries and administrators in designing practical RDM support services.

Review of Literature: The literature on RDM highlights the growing pressure on universities to manage large quantities of digital research data. Pinfield, Cox and Smith observed that universities need policies, infrastructure and services to support the creation, storage, preservation and sharing of research datasets[2]. Cox and Pinfield further argued that RDM should be treated as a strategic institutional function rather than an isolated technical activity[1].

The data-intensive research environment has also led to capability and maturity models that help institutions assess policy, IT infrastructure, support services, metadata and training[5,6].

Several studies place academic libraries at the centre of RDM service development. Tenopir et al.[3] found that librarians recognise RDM as an emerging area but require stronger skills and institutional support. Yoon and Schultz identified service, information, education and network support as major areas of library RDM websites[4]. Lee and Stvilia emphasised that repository staff view data curation as a collaborative activity encompassing metadata, preservation, appraisal and reuse[9]. Recent literature connects RDM with Open Science, arguing that good data management supports transparency and reproducibility even where data cannot be fully open[10].

The broader literature has also discussed RDM in medical libraries, the human factor in digital curation, RDM as a wicked institutional problem, good practices in university research information management and data management framework guidance[11-15]. These works collectively show that RDM is not only a technical repository issue but also a policy, skills, ethics and institutional coordination issue.

Indian studies indicate that RDM services in academic libraries are developing but remain limited. Tripathi, Shukla and Sonkar reported that university libraries have a potential role in RDM policy, training and repository support[7]. Singh, Bharti and Madalli found that Indian academic libraries face barriers such as a lack of RDM policy, limited implementation and inadequate institutional collaboration[8]. These studies provide a broad national perspective, but there is still limited discipline-wise evidence from individual universities. The present study addresses this gap by examining the perceptions and practices of researchers at BHU.

Objectives: The study was conducted to recognise the type and source of research data generated and utilised by academics and research scholars, assess the amount of work needed for processing acquired data and documenting data, explore where data is stored and how it is shared or used in collaboration, analyse the major concerns that restrict data sharing, assess awareness and use of RDM policy, data management plans, metadata, DOI and data citation practices, and identify the RDM training already received as well as the areas in which further training is required.

Methodology

The study used a descriptive survey method. An online questionnaire was distributed to teaching staff and research scholars from the Faculties of Arts, Science, and Social Science at Banaras Hindu University. The questionnaire covered respondent profile, types and sources of data, data processing, documentation, storage, collaboration, data-sharing behaviour, concerns about data sharing, knowledge of RDM components and training requirements. A nonprobability sampling method, specifically voluntary response/convenience sampling, was employed, as only those who received the questionnaire link and agreed to participate were included. The survey ran from March-April 2024, and reminders were sent after 3 weeks. A total of 107 completed and usable responses were received. Data were analysed using frequency distributions, percentage analysis and discipline-wise cross-tabulations. To add inferential support, the Pearson chi-square test of independence was applied to selected two-by-two comparisons between Arts and Science respondents for major data-generation categories. Since several questions allowed multiple responses, each response category was interpreted independently, and no causal inference was drawn.

Results and Discussion

Respondent profile: The distribution was almost equal between research scholars and teaching staff in the sample. Science formed the largest disciplinary group, followed by Arts and Social Science. The representation of younger respondents was strong, while senior academics were also present in the sample. Therefore, the data reflect both early-career and experienced voices on RDM practices.

Table-1: Profile of respondents (n = 107).

Category	Frequency	Percentage
Research scholars	53	49.53
Teaching staff	54	50.47
Arts	36	33.64
Science	51	47.66
Social Science	20	18.69
Age 21-30 years	41	38.32
Age 31-40 years	29	27.10
Age 41-50 years	13	12.15
Age 51-60 years	24	22.43
Male	83	77.57
Female	24	22.43

Data generation, acquisition and preparation: Standard office documents were the most common data type in Arts and Social Science, while Science respondents reported a more diverse data profile, including structured statistical data, raw machine-generated data, software-related data and archived

files. This confirms that RDM support cannot be uniform across all disciplines. Arts and Social Science users require support for documentation, file naming and qualitative data organisation, whereas Science users require stronger support for structured datasets, software, metadata and preservation workflows. To test whether these descriptive differences were merely numerical or statistically meaningful, item-wise Pearson chi-square tests were conducted for Arts and Science respondents using yes/no responses for each item. Significant disciplinary differences were found for standard office documents (chi-square = 6.15, df = 1, p = .013), software applications (chi-square = 7.64, df = 1, p = .006) and archived data (chi-square = 5.88, df = 1, p = .015). The difference for raw or machine-generated data approached significance (chi-square = 3.52, df = 1, p = .061). These results provide inferential support for the claim that Science has a more technically varied data profile, while the Arts rely more strongly on standard office documents.

Table-2: Major types of research data produced by discipline%.

Type of data	Arts	Science	Social Science
Standard office documents	80.56	54.90	80.00
Images	44.44	54.90	40.00
Structured statistical data	27.78	41.18	5.00
Raw/machine-generated data	13.89	31.37	15.00
Source code	5.56	13.73	0.00
Software applications/ modelling tools	5.56	29.41	15.00
Archived data files	5.56	25.49	10.00
Internet and web-based data	38.89	35.29	10.00

Field data collection was common in all three faculties. Arts respondents were more dependent on multiple known sources, whereas Science respondents showed a relatively balanced use of field data, known sources and team-based collection. Most respondents reported that externally acquired data required either some cleaning or substantial effort before use. Arts and Social Science respondents particularly indicated that making acquired data usable often consumes considerable time, suggesting the need for training in data cleaning, organisation and documentation tools.

Documentation and storage of research data: Data file descriptions were the most frequently assigned additional information across all faculties, indicating basic awareness of documentation. However, discovery metadata such as creator, funding body, project title and keywords was reported by only a small proportion of respondents. The low use of discovery metadata is a major concern because it directly affects retrieval, citation and reuse. A considerable share of Science and Social Science respondents also reported not assigning additional information to their data.

Data storage practices show another important risk. Personal devices were the most common storage medium, especially in Science, while cloud storage was more common in Social Science. Use of central servers, university repositories and outside repositories remained very low. This dependence on personal devices may lead to data loss, weak version control and poor long-term accessibility. It also indicates that researchers require clearer institutional pathways for secure storage, backup and repository deposit.

Table-3: Data sources and processing effort (percentage).

Item	Arts	Science	Social Science
Collect new data from the field	58.33	58.82	50.00
Collect from own research team/group	30.56	25.49	20.00
Collect from one known source	16.67	13.73	5.00
Collect from multiple known sources	63.89	41.18	40.00
Collect from personal /professional network	27.78	21.57	10.00
A little cleaning/modification required	25.00	33.33	20.00
A lot of effort required to make data usable	61.11	45.10	55.00
Data usable as received	2.78	3.92	5.00
Do not use outside data	11.11	17.65	20.00

Table-4: Documentation and storage practices (percentage).

Item	Arts	Science	Social Science
Administrative information assigned	25.00	33.33	20.00
Data file description assigned	55.56	43.14	40.00
Technical information assigned	27.78	3.92	30.00
Discovery information assigned	16.67	15.69	15.00
No additional information assigned	16.67	25.49	25.00
Store data on own devices	25.00	33.33	20.00
Use cloud storage	13.89	17.65	25.00
Use a central server/university repository	8.33	3.92	5.00
Use outside repositories	2.78	5.88	0.00

Collaboration, data sharing and concerns: There was some collaboration and data sharing, though not extensive. The most common method of sharing data among Science respondents was within their teams, and they were also more likely to share data with those who requested it. Compared with other groups, respondents in the Arts had a relatively higher level of external institutional collaboration. Social Science respondents had the highest non-sharing rate. The most common mode of sharing was availability upon request across all faculties, indicating that researchers are not completely opposed to sharing but prefer controlled access rather than fully open access.

The concerns reported by respondents explain this cautious behaviour. Data misuse was the most prominent concern in Arts and Science, whereas misinterpretation and loss of competitive advantage were more visible in Arts and Social Science. Legal and ethical issues were also important, especially for Science respondents. These results show that researchers need not only technical storage systems but also trust-building mechanisms such as data-sharing agreements, attribution norms, ethical guidelines and intellectual property protection.

Table-5: Data sharing behaviour and major concerns (percentage).

Item	Arts	Science	Social Science
Share with the same research team	25.00	33.33	20.00
Share with the same university	19.44	15.69	10.00
Share with other institutions	22.22	3.92	15.00
Share with everyone who asks	16.67	29.41	25.00
Do not share data	11.11	9.80	25.00
Data is openly available upon request	55.56	60.78	45.00
Legal and ethical issues	25.00	33.33	20.00
Misuse of data	47.22	39.22	20.00
Misinterpretation of data	38.89	3.92	30.00
Fear of losing scientific edge	30.56	17.65	25.00
Lack of policy and rights protection	27.78	9.80	5.00

RDM knowledge and training needs: Awareness of institutional RDM policy and data management plans was weaker than awareness of metadata, DOI and data citation styles. About one-third of the total respondents reported awareness of an RDM policy, and only 27.10 per cent confirmed the existence of a DMP mechanism. A significant portion was unsure, which is an important result for institutional policy communication. The lack of communication and visibility appears larger than the lack of willingness. In other words, respondents appear receptive to RDM, but institutional expectations, services and policy mechanisms have not been communicated strongly enough across the academic community. Nevertheless, 78.50 per cent believed that a DMP helps researchers manage data. This indicates that researchers recognise the value of RDM even when formal institutional practices are not yet strongly visible.

The training data show a clear gap between the training received and the training desired. The largest gaps appear in DMP, metadata, file naming and version control. The strongest demand was for DMP training, followed by value-addition in research data, version control, and storage. Data citation had the highest level of training already received, but interest in further training remained moderate. The findings justify a library-led RDM training programme delivered in modules and customised by discipline.

Table-6: RDM knowledge and training needs (percentage).

Item	Yes/ Received	No/Not received	Uncertain/ Interested
Institution has an RDM policy	34.58	15.89	49.53 uncertain
Institution has a DMP mechanism	27.10	19.63	53.27 uncertain
Ever used a DMP for research	34.58	47.66	17.76 uncertain
Have DMP for current project	44.86	36.45	18.69 uncertain
DMP helps in managing data	78.50	6.54	14.95 uncertain
Familiar with metadata	67.29	24.30	8.41 uncertain
Familiar with DOI	68.22	23.36	8.41 uncertain
Received DMP training	31.78	68.22	83.18 interested
Received metadata training	26.17	73.83	69.16 interested
Received version control training	36.45	63.55	73.83 interested
Received research data storage training	48.60	51.40	68.22 interested

Recommendations: The findings suggest the need for a practical and phased RDM support model at the university level. First, the university should define or highlight an institutional RDM policy that encompasses ownership, storage, retention, metadata, access conditions, data citation, rights protection and ethics of sharing. Second, the university library should establish an RDM advisory service to address metadata, repository management, preservation, and user education. Third, a secure institutional repository or central storage pathway should be

promoted to reduce dependence on personal devices. Fourth, training should be discipline-specific: Science researchers need advanced support for structured datasets, metadata, version control and software-related data, while Arts and Social Science researchers need stronger support for data documentation, qualitative data management, file naming, data citation and ethical sharing. Finally, DMP templates should be introduced for research scholars and faculty projects so that data planning becomes part of the research process from the beginning.

Conclusion

The study reveals that RDM practices at BHU are evolving and unevenly distributed across disciplines. Researchers collect various kinds of data, and many of them use personal devices to store their data and apply only limited discovery metadata. Data sharing is more controlled than open, primarily because of concerns about misuse, misinterpretation, legal and ethical problems and loss of competitive advantage. There is encouraging awareness of metadata, DOIs, and citations, but significant uncertainty remains about institutional RDM policy and DMP mechanisms. What is most needed is structured training that is discipline-responsive and supported by visible institutional mechanisms. Based on the findings, LIS professionals and academic libraries have a clear opportunity to move from information support towards research data stewardship. A library-led RDM programme, supported by institutional policy and repository infrastructure, can strengthen research integrity, preservation and reuse in the university. The study is limited by its relatively small sample of 107 valid responses and by the use of a non-probability, voluntary-response/convenience sampling technique. Therefore, the findings should be interpreted as an exploratory institutional assessment rather than a statistically representative account of all academics and researchers at BHU. Future studies may use stratified random sampling, larger faculty-wise samples and more advanced statistical modelling to validate the patterns identified in this paper.

Acknowledgement

Bhawana was responsible for the conceptualisation of the study, development of the research design, preparation of the questionnaire, data collection, data analysis, interpretation of findings and drafting of the manuscript. Satya Prakash Singh contributed to data organisation, statistical verification, literature support, interpretation of results and critical revision of the manuscript. All authors reviewed the final version of the manuscript and approved it for submission.

References

1. Cox, A. M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, 46(4), 299-316.

2. Pinfield, S., Cox, A. M., & Smith, J. (2014). Research data management and libraries: Relationships, activities, drivers and influences. *PLOS ONE*, 9(12), e114734.

3. Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2014). Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research*, 36(2), 84-90.

4. Yoon, A., & Schultz, T. (2017). Research data management services in academic libraries in the US: A content analysis of libraries' websites. *College & Research Libraries*, 78(7), 920-933.

5. Jones, S., Pryor, G., & Whyte, A. (2012). Developing research data management capability: The view from a national support service. In R. Moore, K. Ashley, & S. Ross (Eds.), Proceedings of the 9th International Conference on Preservation of Digital Objects (iPRES 2012) (pp. 142-149). University of Toronto Faculty of Information.

6. Lyon, L., Ball, A., Duke, M., & Day, M. (2012). Developing a community capability model framework for data-intensive research. In R. Moore, K. Ashley, & S. Ross (Eds.), Proceedings of the 9th International Conference on Preservation of Digital Objects (iPRES 2012) (pp. 9-16). University of Toronto Faculty of Information.

7. Tripathi, M., Shukla, A., & Sonkar, S. K. (2017). Research data management practices in university libraries: A study. *DESIDOC Journal of Library & Information Technology*, 37(6), 417-424.

8. Singh, R. K., Bharti, S., & Madalli, D. P. (2022). Evaluation of research data management (RDM) services in academic libraries of India: A triangulation approach. *The Journal of Academic Librarianship*, 48(6), 102586.

9. Lee, D. J., & Stvilia, B. (2017). Practices of research data curation in institutional repositories: A qualitative view from repository staff. *PLOS ONE*, 12(3), e0173987.

10. Borghi, J. A., & Van Gulick, A. E. (2022). Promoting open science through research data management. *Harvard Data Science Review*, 4(3).

11. Surkis, A., & Read, K. (2015). Research data management. *Journal of the Medical Library Association*, 103(3), 154-156.

12. Poole, A. H. (2015). How has your science data grown? Digital curation and the human factor: A critical literature review. *Archival Science*, 15, 101-139.

13. Awre, C., Baxter, J., Clifford, B., Colclough, J., Cox, A., Dods, N., Drummond, P., & Fox, Y. (2015). Research data management as a wicked problem. *Library Review*, 64(4/5), 356-371.

14. Davidson, J., Jones, S., Molloy, L., & Kejser, U. B. (2014). Emerging good practice in managing research data and research information within UK universities. *Procedia Computer Science*, 33, 215-2

15. Australian National Data Service. (2017). Creating a data management framework: ANDS guide. Australian National Data Service.