International Journal Research Publication Analysis

AUTOMATING THE INVISIBLE LABOR: A STUDY ON THE ADOPTION AND IMPACT OF AI TOOLS IN ACADEMIC RESEARCH DATA MANAGEMENT WITH SPECIAL REFERENCE TO PLAGHAR DISTRICT

*Vivek M More

Loknete Ramsheth Thakur ASC College, Mokhada.

ABSTRACT

Research Data Management (RDM) is a critical pillar of scholarly integrity and impact, yet it remains a labor-intensive and often undervalued process, particularly in resource-constrained environments. This study investigates the potential of Artificial Intelligence (AI) tools to automate key RDM tasks within the specific context of Plaghar District, a region facing challenges like limited funding, infrastructure, and specialized staff. Using a mixed-methods approach—a survey of 85 researchers and librarians, followed by 15 in-depth interviews—this research maps the current RDM landscape and identifies barriers to AI adoption. Findings reveal a significant awareness gap regarding AI solutions, compounded by pervasive barriers including unreliable internet, high costs, a pronounced skills gap, and a lack of institutional RDM policies. However, a strong undercurrent of motivation exists among researchers to improve their research impact. The study concludes by proposing a strategic, multi-level framework for sustainable AI-RDM integration, emphasizing the role of open-source tools, targeted capacity building, and institutional policy development. This research offers a model for similar regions seeking to harness AI to alleviate administrative burdens and enhance research data quality.

KEYWORDS: Research Data Management, Artificial Intelligence, Academic Libraries, Plaghar District, Technology Adoption, Digital Divide, Scholarly Communication.

1. INTRODUCTION

Contemporary academic inquiry has increasingly embraced data-driven methodologies, propelled by the worldwide Open Science initiative. Robust Research Data Management (RDM)—which includes structuring, archiving, safeguarding, and disseminating datasets—has become indispensable for upholding scholarly rigor, enabling replication, and amplifying research influence (Borgman, 2015). Yet, the labor-intensive aspects of data stewardship, such as crafting metadata and arranging files, represent a type of "hidden effort" that is vital but frequently overlooked in university evaluation and incentive structures (Poole & Garwood, 2020).

As organizations in the developed world advance intricate RDM assistance programs, often incorporating cutting-edge innovations, a pronounced disparity persists in the developing world and particular locales. Palghar District, home to dynamic educational centers generating research with both local and international significance in areas like agronomy, social studies, and community health, illustrates this disparity vividly. Scholars and information specialists in the region navigate the core requirements of RDM while facing hurdles like unreliable and costly broadband access, insufficient budgets for digital tools, and the absence of specialized personnel for data assistance.

At the same time, Artificial Intelligence (AI) applications are gaining traction, offering the prospect of streamlining everyday RDM activities. These solutions can produce explanatory metadata, categorize documents, recommend strategies for protecting privacy in data, and enhance findability (He et al., 2023). The opportunity for these innovations to bypass outdated, labor-heavy RDM workflows in areas such as Palghar is particularly appealing. Nevertheless, implementing these technologies extends beyond mere availability; it hinges on a multifaceted dynamic involving knowledge levels, available assets, technical proficiency, and organizational commitment.

This investigation, accordingly, aims to explore the subsequent research inquiries:

1. How are RDM protocols and resources currently implemented and supported within Palghar District's higher education settings?
2. To what extent are researchers and librarians aware of and inclined toward AI applications in RDM?
3. Which precise enablers and obstacles influence the integration of AI-enhanced RDM solutions in this environment?

Through tackling these inquiries, this article intends to deliver a substantiated model for purposefully embedding AI within Palghar District's scholarly infrastructure, ultimately bolstering its broader research capabilities.

2. Literature Review

2.1. The Imperative of Research Data Management: RDM is no longer an optional add-on but a core component of the research lifecycle. Funding agencies and publishers increasingly mandate data management plans and data sharing, making RDM a prerequisite for global scholarly participation (Tenopir et al., 2020). Proper RDM ensures data longevity, prevents data loss, and maximizes the return on research investment by enabling data reuse.

2.2. AI Applications in RDM:

- **Automated Metadata Generation:** Using natural language processing (NLP) to analyze datasets and generate relevant keywords and descriptions (Candela et al., 2021).

- **Data Classification and Organization:** Machine learning models can identify file types, group related datasets, and detect personally identifiable information (PII) for anonymization.

- **Data Quality Enhancement:** AI tools can help clean datasets by identifying outliers, inconsistencies, and missing values (He et al., 2023). These advancements signal a shift from librarians and researchers performing manual curation to overseeing and validating AI-assisted processes.

2.3. Technology Adoption in Resource-Constrained Environments

Studies on technology adoption in the Global South highlight challenges beyond mere tool availability. Warschauer (2004) conceptualizes the "digital divide" not just as a gap in access, but in the resources, skills, and support needed to use technology effectively. In academic settings, this translates to barriers like unreliable infrastructure, lack of localized training, and insufficient institutional policy, which are critical to understanding the potential rollout of complex tools like AI for RDM (Ashiq et al., 2021).

2.4. The Research Gap

While literature on AI in RDM is growing, it is predominantly cantered on well-resourced universities. There is a conspicuous lack of studies investigating the readiness and specific challenges for adopting these technologies in localized, resource-constrained contexts like Plaghar District. This study aims to fill this gap.

3. Methodology

A sequential mixed-methods design was employed to provide both breadth and depth of understanding.

3.1. Population and Sampling: The target population included academic researchers (faculty and postgraduate students), librarians, and IT administrators from the three major tertiary institutions in Plaghar District. A purposive sampling technique was used to ensure representation across disciplines and roles.

3.2. Data Collection:

- **Phase 1 (Quantitative):** An online survey, distributed via institutional mailing lists, gathered 85 complete responses. The survey used Likert scales and multiple-choice questions to assess: current RDM practices, awareness of AI for RDM, and perceived barriers and benefits.

- **Phase 2 (Qualitative):** Following survey analysis, 15 semi-structured interviews were conducted with a stratified sample of respondents (8 researchers, 5 librarians, 2 IT staff). Interviews explored themes of institutional readiness, trust in automation, and specific capacity-building needs. Interviews were transcribed and anonymized.

Data Analysis: Survey data were analyzed using descriptive statistics (frequencies, percentages) with SPSS software. Qualitative interview data were analyzed using thematic analysis, following the process outlined by Braun & Clarke (2006) to identify, analyze, and report patterns (themes).

4. Findings and Analysis

4.1. The Current State of RDM in Plaghar District

The survey revealed a fragmented RDM landscape. Over 75% of researchers reported storing active data primarily on personal hard drives and USB sticks. Only 15% had ever created a formal Data Management Plan (DMP). While 80% of librarians were aware of RDM concepts, only 30% reported their library offering any form of structured RDM support, typically limited to basic advice on data storage.

4.2. Awareness and Perception Gap

A stark awareness gap was identified. While 90% of respondents had heard of AI in a general sense, less than 10% were aware of its specific applications for RDM tasks. However, when the concept was explained, 88% of researchers expressed strong interest, primarily citing "saving time" (95%) and "improving data organization" (87%) as key motivators.

4.3. Key Barriers to Adoption

Thematic analysis of the interviews crystallized four major barriers:

1. **Infrastructure:** "The internet is too slow and expensive to rely on cloud-based AI tools," noted one IT administrator. This was the most frequently cited obstacle.

2. **Economic:** The subscription costs for commercial AI software were universally perceived as prohibitive. As one researcher stated, "Our research grants are small; there is no budget for such advanced software."

3. **Skills Gap:** A double deficit was identified: a lack of foundational RDM skills, compounded by no exposure to AI concepts. A librarian commented, "We ourselves need training before we can guide others."

4. **Policy and Institutional Support:** The absence of mandatory RDM policies or institutional data repositories was seen as a major demotivator. "There is no top-down push for this, so it remains a low priority," explained a senior faculty member.

4.4. Potential Facilitators and Opportunities

Despite the barriers, strong facilitators emerged. Researchers were highly motivated by the prospect of increasing the visibility and citation of their work. There was also significant interest in using AI to manage and preserve unique local data, such as oral histories and agricultural surveys. The concept of a collaborative, district-wide approach to secure resources was met with enthusiasm.

5. Discussion: A Strategic Framework for AI-RDM Integration

The findings indicate that a direct, tool-centric approach to AI adoption in Plaghar District is likely to fail. Instead, a holistic, multi-level framework is required to build a sustainable ecosystem.

5.1. Interpreting the Findings

The low awareness but high interest suggests a "latent readiness" that can be activated with demonstration and education. The barriers are systemic and interlinked; solving the infrastructure issue alone, for instance, would not address the critical skills gap.

5.2. Proposed Multi-Level Framework:

- **Individual Level:** Focus on advocacy and "capacity building for adoption." Workshops should introduce low-cost, open-source AI tools (e.g., tools for batch file renaming, basic

metadata extraction) that offer immediate "quick wins." This builds confidence and demonstrates tangible value.

- **Institutional Level:** Libraries must be empowered as change agents. This requires:
- o **Policy Development:** Creating simple, mandatory DMP templates for all research projects.
- o **Capacity Building:** Training librarians to become "AI-RDM facilitators."
- o **Infrastructure Investment:** Prioritizing the setup of a local institutional repository, which can later be integrated with AI plugins.
- **Regional/Consortia Level:** To overcome resource limitations, institutions in Plaghar District should form a consortium. This body could:
- o Advocate for improved internet infrastructure with local government.
- o Negotiate group licenses for affordable AI tools.
- o Develop a shared "Plaghar District Research Data Hub" to host training materials and provide a platform for collaboration.

5.3. Recommendations for Tool Developers

For AI to be viable in contexts like Palghar, tools must be designed with offline functionality, low bandwidth requirements, and intuitive, multi-lingual user interfaces.

6. CONCLUSION AND FUTURE WORK

This study has demonstrated that while the path to integrating AI into RDM in Plaghar District is fraught with challenges, the potential benefits for research efficiency and impact are substantial. The barriers are not insurmountable but require a strategic, collaborative, and capacity-building-focused approach, as outlined in the proposed framework. The success of such an initiative hinges on viewing AI not as a silver bullet, but as a tool to augment human expertise within a supportive ecosystem.

A key limitation of this study is its focus on a single district, which may affect the generalizability of the findings. Future research should include a pilot implementation of a specific open-source AI tool within one institution in Plaghar to gather empirical data on its effectiveness and user experience. Furthermore, a comparative study with another district would be valuable for refining the proposed framework.

7. REFERENCES

1. Ashiq, M., Usmani, M. H., & Khan, A. (2021). A systematic review of research data management and library services in developing countries. *Journal of Academic Librarianship, 47*(5), 102398.

2. Borgman, C. L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. The MIT Press.

3. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.

4. Candela, L., Castelli, D., & Pagano, P. (2021). The role of AI in data infrastructures. *Journal of Documentation, 77*(5), 1073-1090.

5. He, S., Kadiyala, A., & Zeng, M. L. (2023). Artificial intelligence for data management: A systematic literature review. *Journal of the Association for Information Science and Technology, 74*(4), 451-469.

6. Poole, A. H., & Garwood, D. A. (2020). The invisible labor of digital data curation. *Library Trends, 68*(4), 654-677.

7. Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., ... & Page, R. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLOS ONE, 15*(3), e0229003.

8. Warschauer, M. (2004). *Technology and Social Inclusion: Rethinking the Digital Divide*. The MIT Press.