Vol. 5, No. 2, 2025, 1-17. DOI: 10.47909/ijsmc.192

ORIGINAL ARTICLE

Barriers and need for dataset sharing in the publishing of research thesis

Atul Kumar[1], **Amol Gawande**[2], **Jaiprakash M. Paliwal**[3], **Vishwas Pendse**[4], **Shailendrakumar Kale**[5], **Akash Agarwal**[6], **Vinaydeep Brar**[7], **Manesh Palav**[8], **Sagar Nimbalkar**[9], **Amandeep Saini**[10], **Gauri Rathi**[11], **Shirish S. Raibagkar**[12]

[1] Dr. D. Y. Patil B-School, India.
[2] Dr. D. Y. Patil B-School, India.
[3] Symbiosis Center for Skill Development, Symbiosis International (Deemed University), India.
[4] Sanjay Ghodawat University, India .
[5] Marathwada Mitra Mandal's College of Engineering, India..
[6] Rajkiya Engineering College, India.
[7] Dr. D. Y. Patil B-School, India.
[8] Global Business School and Research Centre, Dr. D. Y. Patil Vidyapeeth, India.
[9] Vidya Pratishthan's Institute of Information Technology, India.
[10] Dr. D. Y Patil School of Management, India.
[11] Mahatma Gandhi Vidyamandir Samajshree Prashantdada Hiray College of Management and Technology, India.
[12] Dr. D. Y. Patil B-School, India.
Corresponding author
Email: ssrssr696@gmail.com. **ORCID:** https://orcid.org/0000-0002-8017-7740.

ABSTRACT

Objective. This study aimed to understand the reasons for the non-sharing of datasets by Indian social science doctoral researchers and to determine whether researchers fully understand the significance of dataset sharing.

Design/Methodology/Approach. A quantitative methodology was employed, entailing the administration of a questionnaire to 361 recent Indian doctoral recipients in the social sciences. The questionnaire comprised two sections: one ascertaining barriers to non-sharing of datasets and the other concerning the need of dataset sharing. Each section contained 10 statements, and responses were collected using a five-point Likert scale. Finally, a *t*-test was employed to ascertain if the sample means differed significantly from the population mean.

Results/Discussion. Several barriers were identified that hinder the sharing of research datasets. These included the absence of specific provisions in the regulations of the Indian apex body (University Grants Commission [UGC]) for PhD research, the lack of encouragement from research supervisors and centers to share data, the exclusion of datasets during the final defense viva-voce, challenges in sharing datasets for reasons such as their non-existence or ad-hoc compilation, and the limited peer practice of

Received: 26-01-2025. **Accepted:** 20-02-2025. **Published:** 21-02-2025.

How to cite: Kumar, A., Gawande, A., Paliwal, J., Pendse, V., Kale, S., Agarwal, A., Brar, V., Palav, M., Nimbalkar, S., Saini, A., Rathi, G., & Raibagkar, S. (2025). Barriers and need for dataset sharing in the publishing of research thesis. *Iberoamerican Journal of Science Measurement and Communication; 5*(2), 1-17. DOI: 10.47909/ijsmc.192

dataset sharing. The average agreement for these barriers was 82%, which was statistically significant. Concurrently, researchers concurred with the need of dataset sharing, including enhanced transparency in the research process, improved reliability of research findings, facilitation of peer researchers' comprehension of its structure and other salient details, harmonization of Indian research with international practices of data sharing, adherence to COPE norms on research ethics, and enhancement of thesis presentation during various forums, including the final defense viva-voce. The average agreement among researchers for these needs was 87%, which was statistically significant.

Conclusions. Despite the global endorsement of data sharing as a highly desirable research practice, Indian researchers from the social sciences domain often face various challenges that prevent them from sharing their datasets. The non-sharing of datasets could raise concerns about the authenticity and reliability of research. Concurrently, the study indicated that researchers were aware of the advantages associated with data sharing. Collective action from the apex regulatory authority (UGC), research centers, research supervisors, and research scholars could lead to a desired improvement, particularly in the sharing of underlying datasets, which would enhance the reliability and transparency of research.

Originality/value. This study represented a novel effort to comprehend the significant barriers to data-sharing practices in research within the context of Indian doctoral education in the social sciences and demonstrated that researchers exhibit a favorable disposition toward data sharing. This study contributed to the advancement of policy and practice regarding data sharing, with the objective of enhancing transparency, openness, and accountability in scientific research.

Keywords: datasets; PhD theses; data sharing; repository; social sciences; open science.

1. INTRODUCTION

The practice of sharing research data with the public is widely recognized as a valuable component of modern research methodologies (Corti *et al.,* 2019). Martone *et al.* (2018) emphasize the significance of making datasets accessible as a fundamental aspect of contemporary scholarship. In cases where the integrity of scientific research is examined, the dissemination of data, transparency in research methodologies, and the ability to reproduce results are considered essential elements of scientific publications. Alter and Gonzalez (2018) offer practical insights into addressing challenges to data sharing for reuse in the context of psychology research. The authors examine key concerns related to data sharing, including research data ownership, preserving the anonymity of respondents, proper attribution to authors who created the data, handling codebooks and metadata, addressing sources of origin, as well as other issues such as file formats and managing various versions of data. This study drew on global best practices and policies for sharing research data as a foundational reference. Leading agencies such as the Committee on Publication Ethics (COPE), the World Association of Medical Editors (WAME), and the Open Access Scholarly Publishing Association (OASPA) recommend the sharing of research data. In addition, the data-sharing policies of several prominent global academic research publishers (Emerald, 2025; Springer Nature, 2025; Taylor and Francis, 2025) were examined to identify recommendations for data sharing. A collective consideration of the policies of COPE, WAME, and OASPA, as well as the three publishing houses, led to the following conclusions:

- All the leading agencies and publishers are encouraging researchers to practice data sharing by making research datasets accessible to the public.
- However, this encouragement is not mandatory; it is not a policy that datasets must be made accessible to the public.
- The leading agencies and publishers are providing the required support to the researchers to access a repository for depositing their datasets.
- While encouraging researchers to share their datasets, they caution researchers about the non-violation of research ethics, such as disclosing confidential information.

As a result of these policies, researchers at the international level are responding by depositing their datasets in a repository and providing an access link as part of their research publication (Kumar *et al.* 2022a,b; Kumar *et al.,* 2023, 2025; Singh *et al.,* 2024).

1.1. Data sharing in the Indian context

Data sharing is relatively infrequent in the Indian PhD social science research domain, and the sharing of research datasets has received limited focus. This notable difference between global and Indian policies and practices warrants further investigation. The University Grants Commission (UGC), as the primary regulatory body for higher education in India, plays an important role in overseeing doctoral research and establishing regulations. However, it is important to note that the UGC currently does not have a policy specifically addressing the sharing of research data (University Grants Commission, 2022). In addition, the underlying dataset remains undisclosed, despite the full disclosure of the thesis. This finding was supported by a review of the "Shodhganga," a central repository for national theses in India. An analysis of 100 uploaded theses from 2024 highlighted the absence of dataset sharing. This study suggests the inclusion of provisions in UGC regulations that could encourage the sharing of research datasets. The large size of the Indian research community provides the context for this study. Doctoral enrollment in India has grown at an annual rate of over 7% during the period from 2014-2015 to 2018-2019 (Department of Higher Education, n.d.). This period also witnessed the admission of over 700,000 doctoral students, resulting in an annual average of approximately 150,000 enrollments. When compared to leading OECD nations, this figure is particularly noteworthy. In 2017, the United States had the highest number of doctoral graduates, with slightly less than 75,000 degrees awarded. Germany and the United Kingdom had approximately 30,000 doctorates each, ranking second and third, respectively (OECD, 2019). The three countries with the highest number of doctorate recipients in 2017 collectively accounted for approximately 125,000 doctoral degrees, which is slightly lower than the average of approximately 150,000 admissions per five-year period in India.

Currently, there is no standardized protocol for disseminating research datasets to the public. Upon referencing a thesis repository known as Shodhganga, one can typically observe the availability of various contents, including the title, table of contents, certificates and declarations, chapters, and references, in the form of PDF files. The public has access to the full thesis, including its chapters and references. However, the underlying dataset of the research is not readily available. The practice of sharing datasets is not commonly observed throughout the thesis process. A review of international research papers (Kumar *et al.,* 2022; Kumar *et al.,* 2022; Kumar *et al.,* 2023, 2025; Singh *et al.,* 2024) indicates that international research publishing practices encourage researchers to share their dataset with the public by depositing it in a repository and providing access links to the dataset. Considering the recognized needs of dataset sharing as highlighted by international researchers (Alter & Gonzalez, 2018; Corti *et al.,* 2019; Martone *et al.,* 2018) and the practices encouraged by leading agencies such as COPE, WAME, and OASPA and publishing houses (Emerald, 2025; Springer Nature, 2025; Taylor and Francis, 2025) to share their research datasets, this study explores the barriers to dataset sharing in India. A review of the UGC PhD regulations indicates the absence of provisions regarding the sharing of research datasets (University Grants Commission, 2022), even though the regulations do stipulate the sharing of theses through deposition in the national repository "Shodhganga."

The research questions for this study are as follows:

- RQ1: What are the barriers to non-sharing of datasets associated with the research?
- RQ2: What is the need for sharing datasets associated with the research?

The study is organized into four sections: review of literature, methodology, results and discussion, and conclusion. In the literature review, the authors examine substantial research on data sharing and establish the hypotheses for the study. The methodology section delineates the framework for testing the hypotheses. Subsequently, the results of the data analysis are presented and discussed. Finally, conclusions are derived based on the study's findings.

1.2. Review of literature

A review of the extant literature on data sharing was conducted. With the advent of large-scale and high-throughput data analysis techniques, fundamental facts are increasingly executed electronically, entailing the transmission and reception of information from scientific databases. For projects necessitating a substantial volume of data, it has become customary for all pertinent data to be readily available on a public website upon the publication of an article (Toronto International Data Release Workshop Authors, 2009). Data sharing can present a range of challenges (Borgman, 2012). The reasons for disseminating data were explored through examples from various fields of study. The process of research can be verified or replicated; the findings of studies that have received public sponsorship can be disseminated to the public; novel inquiries can be posed regarding existing data; and the advancement of research and innovation can be promoted. The requirement of data sharing is not a novel concept. In a 1985 report (Fienberg *et al.,* 1985), the Committee on National Statistics posited that the sharing of scientific data contributes to the maintenance of open scientific inquiries. Recognizing the potential financial challenges for original researchers, the committee advocated for scientific practices that support data sharing to the extent feasible.

The practice of data sharing in research is associated with funding, with funding agencies requiring consideration of data sharing in all genomics funding proposals. This requirement has influenced scientific practice, particularly in the domains of publishing and conducting research. The authors identify challenges that need to be addressed for the potential benefits of data sharing to be fully realized, as envisioned by funding bodies (Kaye *et al.,* 2009). Over the past decade, governments, funders, and publishers have implemented open-data policies. Although existing studies suggest that scientists increasingly value sharing and reusing data, actual practices do not always align with these emerging policies (Tenopir *et al.,* 2018). Tedersoo *et al.* (2021) have suggested that data availability statements (DASs) provided upon reasonable requests may not be as effective and could benefit from reconsideration by journals. To encourage more data sharing upon manuscript acceptance, researchers might find it beneficial to receive additional incentives, such as acknowledgments or extra points in grant and job applications. Data sharing is the process of making data accessible to others (Michener, 2015). Environmental scientists compile and share a substantial volume of data. These data supplement existing datasets and can be applied in meta-analyses, model parameterization, and the validation of research findings.

Data sharing remains infrequent for a variety of reasons. Organizations may regard their data as a source of competitive advantage because, akin to individual researchers, they may perceive themselves as either collaborators or participants in the broader context of securing funding (Chaplin, 2004). Open science confers numerous benefits to researchers, including increased cooperation, trust in results, and goodwill (Popkin, 2019). Data sharing is imperative for providing data resources for geographic modeling and simulation (Wang *et al.,* 2020). Current data-sharing strategies primarily entail the publication of data using file-level metadata, a process that requires the identification, organization, and synthesis of raw data files for future utilization. This approach contrasts with several successful examples of data sharing through web-based platforms. According to Kamikubo *et al.* (2021), the discovery and dissemination of accessible datasets offer both opportunities and challenges, which call for institutional, legal, and technical privacy frameworks that better address the concerns of these communities. A scoping review of the literature on data sharing (Woods & Pinfield, 2021) identified the need to build on existing cultures and practices, reach people where they are, and customize interventions to support them. Other salient messages include the imperative for comprehensive disclosure and elucidation of the policy or service, the necessity of having disciplinary data advocates to exemplify optimal practices and catalyze cultural transformation, the importance of suitable resource interventions, and the need for providing substantial technical infrastructure and protocols, such as data file labeling, the utilization of DOIs, data standards, and data repositories.

Polanin and Terzian's (2019) study, adjusted for demographics and pretest scores, found that individuals who received a data-sharing agreement were more inclined to disclose their dataset than those in the control group. The receipt of a data-sharing agreement increased the likelihood of a control group member disclosing their dataset by approximately 25%. Despite its potential to hasten scholarly advancement in the field of psychological science, public data sharing remains relatively uncommon (Houtkoop *et al.,* 2018). A survey was conducted among 600 authors of psychology publications to identify perceived obstacles to public data sharing and potential solutions. The findings highlighted the limited prevalence of data sharing, with respondents from industries where sharing was considered uncommon expressing a preference for on-demand data sharing. They also viewed sharing as requiring additional effort and noted the absence of training on data-sharing practices. To address these challenges, the authors put forth recommendations, including the provision of training resources that demonstrate effective data-sharing methodologies, as well as the establishment of robust support systems from organizations, journals, and sponsors.

Studies have emphasized the necessity and advantages of sharing data in research. However, scholars have recognized that data sharing still needs to obtain the necessary traction in real-world scenarios (Chaplin, 2004; Houtkoop *et al.,* 2018). Thus, it is critical to comprehend the obstacles impeding the practice of data sharing in research. Studies specifically examining the sharing of datasets about PhD research are scarce and have not yet been conducted. White *et al.* (2024) posit that, akin to researchers in other fields, educational science scientists are increasingly confronted with incentives and requirements to make data that support empirical studies publicly accessible. However, the preparation and dissemination of research data is a formidable task. The analysis provides a comprehensive overview of the current state of data sharing among education and special education researchers, delving into the benefits of data sharing and offering a curated list of resources already available to assist researchers in education and related fields. Additionally, it outlines the steps to data sharing throughout the research life cycle and provides answers to frequently asked questions about sharing. The objective of this analysis is to support researchers who are beginning to consider incorporating data sharing into their research workflows. According to the analysis of Obiora *et al.* (2024), Africa is progressing toward the establishment of a network of health databases, with the potential for some data harmonization. Additionally, endeavors are being made to ensure the security of data in repositories and biobanks, strengthen intellectual property protection, and safeguard the ethical rights of study participants, whether in communities or individuals. However, the analysis highlights several challenges to progress, including limited technological infrastructure, inconsistent access to electricity, and difficulties in conducting research in areas such as poverty, governmental processes, and corruption. These factors have contributed to the underrepresentation of certain regions in Africa in the study. Furthermore, the necessity of clarity regarding the timely dissemination of data during global health emergencies, such as the coronavirus disease 2019 (COVID-19) pandemic, has given rise to novel concerns regarding the potential economic and scientific repercussions that may ensue from the disclosure of novel pathogens or variants of preexisting pathogens. This issue warrants careful consideration, as not addressing it may impact the collection and dissemination of valuable data that could aid in future diagnoses and treatments, potentially slowing the global exchange of such discoveries.

Sanabria-Z *et al.* (2024) underscore the imperative to nurture complex thinking skills within higher education. The authors posit that this necessity arises in response to the rapid advancements in technology and the increasing accessibility of Open Educational Resources (OER). Drawing from UNESCO's guidelines and the evolving conceptualizations of complex thinking, this study delineates OER platforms that facilitate the cultivation of such competencies. The study's primary research questions are addressed through a systematic literature review (SLR), focusing on three key areas: the identification of open licenses compatible with 5R activities, the examination of technical choices available based on the ALMS

Framework, and the analysis of competencies in complex thinking that can be fostered through OER. The study identified and categorized key platforms that support the development of complex thinking abilities, specifically inventive, critical, scientific, and systemic thinking. These platforms were then assessed for their technological capabilities and licensing models. The analysis revealed that most of the platforms under study support 5R activities under Creative Commons licenses, demand sophisticated editing abilities, and more than half cover several sub-competencies of complex thinking. The study's findings indicate that the adaptation, reuse, and redistribution of OER can significantly contribute to the cultivation of complex thinking abilities and the alignment of educational methodologies with the demands of the 21st century. The study further proposes the implementation of future studies aimed at evaluating user behavior and the sustainability of these platforms. Additionally, the study recommends the establishment of frameworks by higher education institutions to standardize OER sites and address the needs of students.

To promote scientific discovery and alleviate the burden of illness, other scientists engaged in a variety of studies were motivated to voluntarily disseminate their data, as stipulated in the agreement that was reached when the human genome was initially sequenced, thereby making the data available to the public domain (Terry & Littler, 2024). However, challenges remain in the flow of raw data, where resources often move from less developed countries to wealthier nations and corporations that develop medical countermeasures. This situation is compounded by ongoing obstacles to data sharing, such as concerns about patient privacy, intellectual property management, and addressing inequalities in data reuse. When infectious disease outbreaks occur, it is essential to ensure that research data are shared efficiently to support timely responses to emerging pathogens, especially during a potential pandemic. The chapter reviews the COVID-19 pandemic and the challenges in developing a global data-sharing system that accommodates the needs and interests of diverse stakeholders, including individuals worldwide, alongside the fundamentals of open access research. McLeod and O'Connor (2022) explored the challenges associated with preserving and disseminating qualitative data in educational research by critically examining methods and discussions from several social science fields. The politics of knowledge creation, open-access agendas, and changes in research techniques in the data management era are all discussed about ethical, epistemological, and methodological issues. Using Australia as a case study, the authors first examined pragmatic and interpretive choices regarding the archiving of qualitative data before mapping the legal and policy frameworks that govern research data management. The authors contend that qualitative research's unique ethical, methodological, and knowledge claims may not have been fully incorporated into governance and data-sharing policies. Instead, the focus is shifted toward the expansion of methods related to quantitative data, sometimes leading to the creation of data that may lack full context. The authors further emphasize the importance of carefully examining the complexities and potential challenges of openness and data sharing, as well as the conception, creation, and curation of data archives. This entails a more profound comprehension of the temporal and affective dynamics inherent in data archiving. It is imperative to conceptualize data archiving as a process of (re) invention that curates "archives for the future" and contributes to shaping a historical perspective in educational research.

Tedersoo *et al.* (2021) asserted that data sharing is a fundamental component of contemporary science, facilitating reproducibility and large-scale investigations. The authors conducted a comprehensive assessment of data availability in research articles from nine fields published in the *Nature* and *Science* journals. The study also documented the corresponding authors' concerns, demands, and justifications for their decision to withhold data. Despite notable advancements over the past decade, particularly in recent years, data availability and willingness to share data remain significantly heterogeneous across different fields. It was suggested that journals reconsider allowing claims that data are available upon request, as such claims may often be seen as ineffective. Researchers may be more encouraged to share their data when their manuscripts are accepted, particularly if they can offer genuine rewards, such as recognition or additional points on grant and employment applications. The authors suggest that funding agencies consider supporting data maintenance, that applications be evaluated using publicly available research data, and that funders and academic publishers explore options for data-sharing oversight.

The plutoF repository contains cross-disciplinary survey data, and Vlahou *et al.* (2021) noted that the General Data Protection Regulation (GDPR) became enforceable in EU member states in 2018 to standardize personal data protection laws across the EU. Nearly all forms of personal data processing, including those related to biomedical research, are governed by regulations. The authors conducted consultations with ethics and legal experts to identify the primary practical concerns that biomedical researchers frequently encounter when sharing data and biological samples. The authors then explained how these concerns are addressed within the framework of the GDPR. The authors identify specific sections of the GDPR that necessitate elucidation, particularly regarding the consent requirements for study participants. The objectives of the proposed amendments are threefold: (1) to enhance harmonization and restrict exclusions based on national legislation, (2) to substantiate the concept of broad consent, and (3) to formulate a framework for the secondary utilization of data. To achieve these enhancements, prominent societies within the field will spearhead the creation of a document that will serve as a guide for the optimal interpretation of the GDPR. This document will be finalized following a period of comments from a diverse range of stakeholders, concurrently fostering public participation and education on pertinent subjects, including various forms of consent and residual risk for re-identification, at the local, national, and global levels. The authors anticipate that this post will encourage a comprehensive dialogue with all relevant stakeholders, helping to realize the potential benefits of data protection ethics.

Markiewicz *et al.* (2021) posit that the sharing of research data is imperative to ensure reproducibility and optimize the return on public investments in scientific research. They have developed OpenNeuro, a BRAIN initiative data archive that facilitates the exchange of data from various brain imaging data formats in accordance with FAIR data-sharing guidelines. The authors underscore the significance of the Brain Imaging Data Structure standard in enabling efficient data exchange, curation, and reuse. The archive currently disseminates more than 600 datasets, encompassing information from over 20,000 individuals, encompassing various phenotypes, species, and measurement methods. A growing body of published reuses, amounting to over 150 papers, has underscored the impact of shared data. Koch *et al.* (2021) underscored the pivotal role of the benchmark dataset in the framework of machine-learning research. They serve to gauge progress toward common objectives and help bring scholars together around shared research issues. The dynamics of benchmark dataset use and reuse within or between machine-learning sub-communities have been explored to a lesser extent, despite the importance of benchmarking methods in this field. In this study, the authors explored these dynamics by examining the differences in dataset utilization patterns between machine-learning sub-communities and between 2015 and 2020. In the context of task communities, authors have observed a growing emphasis on a select number of datasets, a notable adoption of datasets from other tasks, and a focus on datasets introduced by academics from a limited number of prestigious institutions in the field.

Walter *et al.* (2021) assert that Native American Information Data is recognized as a cultural and financial asset by Sovereignty, which asserts the right of Indigenous peoples to control their gathering, ownership, and use. The advent of Big Data and the concomitant emphasis on enhancing the accessibility of publicly owned data (Open Data) have amplified the influence of data. In the prevailing data landscape, indigenous communities persist in their marginalization from data utilization and the exercise of policy influence, despite their substantial representation in data pertaining to social disadvantages and their anticipated overrepresentation in the application of these novel technologies. The prevailing data architecture has not fully incorporated the distinct perspectives and worldviews of Indigenous communities, and their unique data requirements are not yet fully addressed by the emerging Open Data infrastructure. This gap is reflected in the limited consideration of Indigenous data challenges in open-data discourse and publications. While the potential benefits of this data revolution are widely acknowledged, not all will benefit equally, given the marginalization of social, cultural, and political positions. In addressing this important issue, the authors delve into the potential of Indigenous Data Sovereignty to not only mitigate risks but also to foster shared benefits. Additionally, they explore the unintended consequences of the pervasive influence of open and big data.

Gabelica *et al.* (2022) conducted a study to analyze researchers' adherence to their DAS from publications published in open-access journals. The authors established the required DAS as the study's objective, and they examined every article from 333 open-access publications published by BioMed Central in January 2019. The authors classified the DAS types and surveyed the corresponding authors who indicated a willingness to share data in the DAS. All the included manuscripts were asked for consent to participate in the study, and they verified that the data were accessible, allowing for reanalysis after gaining access to the raw datasets. The DAS was present in 3,416 of the 3,556 publications examined. The datasets are accessible upon reasonable request according to the most common DAS category (42%). However, of the 1,792 papers where the DAS indicated that the authors were willing to provide their data, 1,669 (93%) did not respond or chose not to share the data. Only 123 (6.8%) of the 1,792 authors who answered the request for data sharing out of 254 (14%) provided the requested information. The compliance percentage remained consistent for authors who did not provide the DAS, indicating that the DAS may not always ensure data sharing, even when authors state in their work that they will release data upon request.

Considering the extant literature and within the context of Indian PhD scholars, the following hypotheses are posited:

- H1: There are multiple barriers for researchers not sharing their research datasets.
- H2: The sharing of research datasets is needed.

2. METHODOLOGY

The primary data for this study was collected from Indian research scholars across the social science domain. The details of such researchers were obtained from the Indian thesis repository, Shodhganga (2024). A review of the contents of approximately 100 theses from the repository was conducted, and it was observed that none of the theses provided access to the PhD research dataset. At a confidence interval of 5% and confidence level of 95% for a substantially large population of approximately 20,000, commonly used tables for sample sizes, such as those by Krejcie and Morgan (1970), indicate that a sample size of 377 is appropriate. The sample size of 377 was rounded to 400 to improve representation of the population and minimize potential sampling errors. Google Forms was utilized to create and administer the survey, which was disseminated to approximately 800 research scholars, with an anticipated response rate of 50%. A convenience sampling approach was used to select these 800 individuals, given the practical constraints of the study. The survey was conducted in December 2024 and yielded 361 responses, resulting in a response rate of 45%. The participants provided explicit consent through the questionnaire, and their names were not collected to maintain confidentiality. The study received review and approval from the ethics committee of a local university, ensuring adherence to ethical standards. The dataset is available in a repository and can be accessed at https://www.openicpsr.org/openicpsr/project/188081/version/V2/view. The survey questionnaire was composed of three sections: profile information, factors influencing participation in data sharing, and advantages of data sharing. The profile section contained 9 questions, while the barriers and need sections each included 10 questions.

The 10 reasons for the absence of data sharing were primarily derived from the recent UGC regulations regarding PhD degrees released in 2022 (University Grants Commission, 2022) that do not include provisions for data sharing by PhD scholars. Concurrently, a list of 10 statements was incorporated to evaluate the need of sharing data, primarily based on extant literature (Borgman, 2012; Cheah *et al.,* 2015; Colavizza *et al.,* 2020; Fienberg *et al.,* 1985; Houtkoop *et al.,* 2018; Michener, 2015; Piwowar *et al.,* 2007; Popkin, 2019; Tedersoo *et al.,* 2021; Toronto International Data Release Workshop Authors, 2009; Wang *et al.,* 2020; Woods & Pinfield, 2021). Annexure 1 contains the questionnaire format. Responses were collected on a five-point Likert scale, with the following options: strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, and strongly disagree. The validity of the questionnaire was examined using the Brown *et al.* (2015) checklist, yielding satisfactory results. Furthermore, the reliability of each section of the questionnaire was assessed using Cronbach's alpha method, yielding scores of 0.97 for section I (barriers), 0.95 for section II (need), and 0.97 for the overall questionnaire. These scores indicated the survey was reliable, as all scores exceeded the acceptable threshold of 0.70.

The methods employed for hypothesis testing were derived from the recommendations set forth by Kumar *et al.* (2022a,b), Khedkar *et al.* (2022) and Kumar *et al.* (2023, 2025). The responses for each section were categorized into two groups: those who expressed agreement and those who expressed disagreement. To differentiate between somewhat agree and somewhat disagree responses, stronger responses were assigned a weight of 2 for both strongly agree and strongly disagree answers, while responses that indicated "neither agree nor disagree" were assigned a weight of 0 and were excluded from the computations. To ascertain whether the group scored higher in agreement or disagreement, an average was computed for each sub-question. The larger of the two average score percentages for each section, as recognized by the Likert scale (Brown, 2011), was then compared against a hypothesized population mean of 50%, indicating no significant difference. A *t*-test, commonly used when the population's standard deviation is unknown, was applied to make this comparison.

3. RESULTS

This section is organized into three subsections. The initial two subsections present the results regarding the descriptive features of the profile data and the testing of the two hypotheses. A discussion of the results follows.

3.1. Descriptive features of the profile section

Frequency count and percentages sample ($n = 361$) are shown in Table 1

The proportion of female researchers (52%) exceeded that of male researchers (48%) within the sample, showing a 4% difference. Although this difference is relatively small, it is interesting to observe that female researchers make up a slightly larger proportion than male researchers. The age distribution of the sample revealed that 3% of the participants were under 30 years of age, 33% were between 30 and 39 years old, 31% were in the 40-49 age range, and 33% were 50 years or older. The proportion of participants aged 50 years or older is of interest, as it may suggest that they began their doctoral studies at a later stage or required additional time to complete their degrees. With respect to regional distribution, 23% of the respondents were from northern India, 24% from eastern India, 25% from western India, and 28% from southern India, indicating an even distribution across the four regions. With respect to the attainment of doctoral degrees, 27% of respondents completed their programs in 2024, 21% in 2023, 27% in 2022, and 25% prior to 2022. A significant proportion of respondents (93%) possessed sample sizes of 400 or less, with a mere 7% having samples that exceeded 400. This indicates that a substantial number of researchers have worked with relatively large sample sizes. (A sample size of 400 is frequently employed, as many researchers find a sample size of 377 sufficient at a 95% confidence level with a 5% margin for a large population of 20,000, rounding to 400.) With respect to survey administration, 35% of respondents employed Google Forms, 30% conducted manual administration, and 35% utilized alternative methods, including MS Excel and MS Word. It is evident that most researchers have utilized digital tools for primary data collection. Most respondents (92%) personally collected data, while 8% collected it through another party. A significant majority (99%) of respondents reported upholding the confidentiality of survey participants, with a small minority of five respondents not fully adhering to this practice. The sample characteristics suggest that ethical standards are generally being followed. Following the completion

Sr. No.	Variable	Category	Frequency	Percentage (%)
1	Gender	Male	174	48
		Female	187	52
2	Age	<30 years	11	3
		30-39 years	120	33
		40-49 years	111	31
		≥50 years	119	33
3	Region	North	82	23
		East	85	24
		West	91	25
		South	103	28
4	PhD year	2024	98	27
		2023	76	21
		2022	97	27
		Earlier than 2022	90	25
5	Sample size	<400	164	45
		Around 400	173	48
		>400	24	7
6	Mode of data collection	Google Forms	125	35
		Manually	108	30
		Other	128	35
7	Primary data collection by self	Yes	333	92
		No	28	8
8	Identity anonymized	Yes	356	99
		No	5	1
9	Dataset retention	Yes	172	48
		No	189	52

Table 1. Samples profile characteristics. Source. Authors primary data.

of their doctoral studies, 48% of respondents retained their dataset, while 52% did not maintain it subsequently. It is important to mention that the UGC has yet to issue formal guidelines in this regard.

3.2. Hypothesis testing

Tables 2 and 3 present a straightforward count summary of responses related to the questionnaire's 10 barriers and need section.

Responses	ST.1	ST.2	ST.3	ST.4	ST.5	ST.6	ST.7	ST.8	ST.9	ST.10
Strongly agree	145	138	145	146	154	145	132	147	154	146
Somewhat agree	150	132	145	157	163	145	137	150	152	152
Neither agree nor disagree	3	2	3	4	3	3	4	2	3	5
Somewhat disagree	36	58	31	29	25	29	39	27	22	25
Strongly disagree	27	31	37	25	16	39	49	35	30	33
Total	**361**	**361**	**361**	**361**	**361**	**361**	**361**	**361**	**361**	**361**

Table 2. Results of 361 replies to the causes section of the questionnaire.
Source. Authors primary data.

The data analysis involved the allocation of a weight of two to the strongly agree and strongly disagree responses, with the objective of calculating the response counts and percentage, subsequently divided into two categories: disagreement and agreement. Thereafter, the percentages of disagreement and agreement for the barriers and need were summed, yielding an average percentage of 82% among those in agreement with the barriers, and 18% among those in disagreement. The hypothesis was then subjected to a 95% confidence

Responses	ST.1	ST.2	ST.3	ST.4	ST.5	ST.6	ST.7	ST.8	ST.9	ST.10
Strongly agree	187	158	160	169	162	162	158	164	172	160
Somewhat agree	134	136	147	149	165	149	131	145	155	153
Neither agree nor disagree	3	2	3	4	3	3	4	3	2	5
Somewhat disagree	20	41	22	19	17	19	35	23	19	22
Strongly disagree	17	24	29	20	14	28	33	26	13	21
Total	**361**	**361**	**361**	**361**	**361**	**361**	**361**	**361**	**361**	**361**

Table 3. Results of 361 replies to the need section of the questionnaire. Source. Authors primary data.

level test, given that the observed agreement percentage exceeded the expected population mean of 50%, thereby suggesting that the level of agreement is statistically significant. The average percentage of individuals who expressed agreement with the need was 87%, while the average percentage of those who expressed disagreement was 13%. The hypotheses were tested at a 95% confidence level, and the observed agreement percentage of 87% was higher than the theoretical population mean of 50%, which could suggest that the agreement occurred randomly. The results are listed in Table 4.

Parameter	H1	H2
Average = Ho (sample mean)	82%	87%
SD (standard deviation)	1.25	1.13
H1 (hypothesized mean of population)	50%	50%
n (sample size)	361	361
t-value (Ho−H1) / ((SD) / √n)	4.93	6.39
p-value	<0.0001	<0.0001

Table 4. Testing of the two hypotheses. Source. Primary data calculations.

The two null hypotheses —namely, that there are no significant barriers for researchers to refrain from sharing datasets and that there is a neutral perception of the need— were not supported by the data, which favored the alternate hypotheses ($p < 0.0001$). In summary, this study highlights the primary barriers that may hinder researchers from adopting international data-sharing policies endorsed by leading agencies such as COPE, WAME, OASPA, and prominent global publishers. The result of the second hypothesis indicates that researchers hold the policy and practice of sharing datasets in high regard. The authors suggest that the UGC could consider mandating the sharing of research datasets, and that universities, research centers, supervisors, and scholars should work toward implementing this policy.

4. DISCUSSION

The initial hypothesis posited several reasons why researchers might opt against publicly disseminating their research datasets. The survey, which queried 10 distinct reasons, showed a notable level of agreement among the surveyed researchers. The most frequently cited reason was —"I am not aware of a policy for sharing data with the public" (89%), while the least agreed upon reason was —"I do not think sharing data is necessary" (75%). The average agreement across all 10 reasons was 82%, which was significant. Of note, all reasons received an agreement rate of at least 75%. Given that the reasons were Likert items, the study did not prioritize individual reasons. Instead, the study considered the overall Likert scale of the 10 reasons and concluded the average agreement of 82% (Brown, 2011). This level of agreement was statistically significant and notably higher than the random chance agreement of 50%. Some additional reasons for the non-sharing of datasets included: "there is no such provision in the UGC norms for PhD," "my research supervisor never asked about the sharing," "the research center never asked about the sharing," "no questions are asked about it during the final defense viva-voce," "the dataset was not sharable for various reasons ranging from it not being there to it being compiled in an ad-hoc manner," "in my knowledge, no one else has shared the dataset," "the data collection process is described elaborately in the methodology chapter of my thesis," and "the data analysis section in the thesis elaborately describes the dataset."

Upon examination, the second hypothesis is supported by the findings. Researchers demonstrated a comprehensive understanding and validation of the need associated with the sharing of datasets. The researchers exhibited a robust response to all 10 needs enumerated in the second section of the questionnaire. The need that received the highest level of agreement was —"A metric counting the number of accesses to the dataset can be developed and would indicate the use of the dataset by other researchers" (92%). The need that received the lowest level of agreement was —"Readers of the thesis can use the dataset for various purposes" (82%). When averaged, the overall agreement for the 10 needs was 87%, suggesting strong agreement. Other needs that received considerable agreement included: "it would lead to transparency in the research process," "it would help in improving the reliability of the research," "it can help current researchers understand its structure and other important details," "the sharing of the dataset will bring Indian research at par with international practices of data sharing," "it will lead to compliance with the Committee on Publication Ethics (COPE) norms on research ethics," "it will improve the presentation of the entire thesis during various presentations including the final defense viva-voce," "it will lead to increased citations of the research work," and "the datasets can be used for further research."

While the secondary data sources (Alter & Gonzalez, 2018; Borgman, 2012; Cheah *et al.,* 2015; Colavizza *et al.,* 2020; Corti *et al.,* 2019; Fienberg *et al.,* 1985; Houtkoop *et al.,* 2018; Martone *et al.,* 2018; Michener, 2015; Piwowar *et al.,* 2007; Popkin, 2019; Tedersoo *et al.,* 2021; Wang *et al.,* 2020; Woods & Pinfield, 2021) and the data-sharing policies of prominent publishing houses (Emerald, 2025; Springer Nature, 2025; Taylor and Francis, 2025) consistently advocate for the sharing of research datasets, it is important to note that the most recent UGC 2022 PhD guidelines do not address the topic of dataset sharing. The guidelines stipulate that the complete thesis should be uploaded to the central depository (University Grants Commission, 2022), but they do not include any mention of sharing or depositing the associated datasets. The demand for data sharing in the academic domain has existed for some time, with research in this area dating back to 1985 (Fienberg *et al.,* 1985). The absence of guidance on this matter by the Indian apex body, the UGC, is concerning. The datasets play a pivotal role in research studies, as findings and conclusions are derived from these datasets, and hypotheses are evaluated through statistical analysis. However, there appears to be limited engagement with data sharing within the Indian research community. Research indicates a significant number of respondents expressing agreement with statements such as, "there is no such provision in the UGC norms for PhD," and so on, mentioned previously. This suggests a lack of awareness or understanding of the issue at the institutional or macro level. At the micro level, there is considerable agreement with the statement, "I need to be made aware of such a policy of sharing the dataset with the public."

It is difficult to identify the exact reasons why some researchers may not be sharing their datasets. It appears that there are no major obstacles that significantly hinder this practice. In the absence of clear explanations at both the institutional and individual levels, one might hypothesize that the reason could be the one mentioned earlier in barriers —"The dataset was not sharable for various reasons ranging from it not being there to it being compiled in an ad-hoc manner." This hypothesis is supported by the findings of Gabelica *et al.* (2022), who observed that 1669 (93%) of the 1792 papers where the DAS indicated that authors who indicated willingness to share their data did not receive a response or chose not to provide the data. Further, a retraction notice from *The TQM Journal* indicates challenges in obtaining the dataset upon request *(The TQM Journal,* 2025). While constructing a database can be a manageable task, especially with tools like Google Forms, the process of sharing the file in various formats is typically straightforward. When referring to the present content uploaded as part of depositing the PhD thesis in the central repository, more than 15 items were identified, from the title to the references. Interestingly, the underlying dataset on which all the remaining things are dependent is missing from the list. On a positive note, it is encouraging that a significant majority of Indian PhD holders recognize the need of sharing datasets. They overwhelmingly endorsed the advantages of sharing the datasets.

5. CONCLUSION

This study yielded two primary conclusions. First, there are various barriers that research scholars face in sharing their datasets. Second, researchers acknowledge the need of sharing datasets. Despite international support for depositing datasets, Indian research scholars in the social sciences face several challenges in sharing their research data. The most recent PhD regulations do not mandate the sharing of the research dataset, which forms the foundation of the research results and conclusions. The hesitation of Indian PhD researchers to share their research datasets at different levels may be attributed to several factors. It is imperative for them to recognize the importance of data sharing. The findings of this study suggest that research supervisors may not prioritize this issue, highlighting the need for local higher education institutions (HEIs) and the apex institution (UGC) to show greater attention to this matter. The availability of datasets generally does not require excessive effort to share them with the public. This approach would ensure the protection of the respondents' confidentiality. However, this need remains insufficiently addressed, and it is crucial for all stakeholders, including leading agencies such as COPE, WAME, and OASPA, as well as publishing houses like Springer Nature, Taylor & Francis, and Emerald, to acknowledge the significance of this issue. They could consider encouraging researchers to share their datasets by depositing them in a publicly accessible repository or using other available methods. India may also consider taking steps to align with this practice to support the integrity of research while protecting individual privacy.

Given the pressing need for sharing research datasets, it is suggested that the UGC encourage the upload of these datasets to the central repository, Shodhganga, by PhD research scholars, along with a minimum of 15 additional items. The decision not to share associated research datasets may not align with recommended research practices and could raise concerns about the accessibility of datasets for PhD research scholars. The study cites a recent retraction notice as support for this suggestion,

> "It has come to our attention that there are statistical errors present within the article. As part of an investigation into the article's findings, the authors were requested to provide a copy of the data so that the editorial team could verify the findings. *The authors were unable to provide the requested data.* As a result, the findings of the article cannot be relied upon." *(The TQM Journal,* 2025)

The findings of this study carry implications for all stakeholders involved in the research process. Primarily, PhD research scholars should recognize the importance of sharing the research dataset (unless there is a valid reason for not doing so) to improve the dependability, reliability, validity, and credibility of their research work. It is important for research scholars to understand the process of depositing datasets in open repositories. Alternatively, they should deposit the dataset and the other research material uploaded to the central repository, Shodhganga. Supervisors play a role in this process; they should ensure they have reviewed the underlying research datasets. After such certification, a PhD viva should be conducted. It would be beneficial for the apex body, UGC, to consider making it mandatory for PhD researchers to upload their research datasets along with all other content. It is noteworthy that PhD researchers often share the acknowledgment file of their thesis but may not upload the research dataset. These implications apply not only to India but also to other nations that may not yet prioritize this aspect. Indian doctoral research could benefit from aligning with international data-sharing policies, and in practice, doctoral theses might be shared along with the underlying dataset. Without this, questions about the availability of the dataset might emerge *(The TQM Journal,* 2025). The process of sharing is generally straightforward when the dataset is available, similar to the other items, ranging from the title to the reference list, that are currently uploaded with the central thesis repository.

This study is subject to certain limitations, including those related to the sampling method. Furthermore, the non-probability technique was employed for the selection of the sample, which constitutes an additional limitation of the study. It is imperative to encourage extensive research in this pivotal domain across all pertinent levels to ensure the mandatory sharing of research datasets under typical circumstances.

Conflict of interest

The authors declare that there is no conflict of interest or competing interest of any sort with any individual or organization.

Contribution Statement

Atul Kumar: Data curation and visualization.

Amol Gawande: Conceptualization.

Jaiprakash M. Paliwal: Funding and conceptualization.

Vishwas Pendse: Methodology.

Shailendrakumar Kale: Conceptualization.

Akash Agarwal: Investigation.

Vinaydeep Brar: Visualization.

Manesh Palav: Methodology and Validation.

Sagar Nimbalkar: Conceptualization and methodology.

Amandeep Saini: Funding and validation.

Gauri Rathi: Data curation and validation.

Shirish Raibagkar: Supervision, writing original draft, review and editing.

Data availability statement

The dataset has been deposited with a repository and can be accessed from https://www.openicpsr.org/openicpsr/project/188081/version/V2/view

Funding statement

No external funding was used for this study.

Ethics approval statement

The authors declare that they have carried the work ethically and further assure to comply with ethics related to processing and publication of the article. ●

REFERENCES

Alter, G., & Gonzalez, R. (2018). Responsible practices for data sharing. *American Psychologist, 73*(2), 146. https://doi.org/10.1037/amp0000258

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology,* 63(6), 1059-1078. https://doi.org/10.1002/asi.22634

Brown, J. D. (2011). Likert items and scales of measurement. *Statistics, 15*(1), 10-14. https://hosted.jalt.org/test/bro_34.htm

Brown, S. M., McBride, G., Collingridge, D. S. (2015). Validation of the Intermountain patient perception of quality (PPQ) survey among survivors of an intensive care unit admission: A retrospective validation study. *BMC Health Services Research, 15,* 155. https://doi.org/10.1186/s12913-015-0828-x

Chaplin, M. (2004). A challenge to conservationists. *World Watch Magazine, 17*(6). www.worldwatch.org/node/565

Cheah, P. Y., Tangseefa, D., Somsaman, A., Chunsuttiwat, T., Nosten, F., Day, N. P., Bull, S., & Parker, M. (2015). Perceived benefits, harms, and views about how to share data responsibly: A qualitative study of experiences with and attitudes toward data sharing among research staff and community representatives in Thailand. *Journal of Empirical Research on Human Research Ethics, 10*(3), 278-289. https://doi.org/10.1177/1556264615592388

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PLoS One, 15*(4), Article e0230416. https://doi.org/10.1371/journal.pone.0230416

Corti, L., Van den Eynden, V., Bishop, L., & Woollard, M. (2019). *Managing and sharing research data: A guide to good practice*. Sage.

Department of Higher Education. (n.d.). *All India Survey on Higher Education (AISHE) Annual Reports from 2013 to 2020-21.* https://aishe.gov.in/aishe/home

Emerald. (2025). *Quality assurance in education. Emerald Publishing.* https://www.emeraldgrouppublishing.com/journal/qae

Fienberg, S. E., Martin, M. E., & Straf, M. L. (1985). *Sharing research data* (pp. 8-225). National Academy Press.

Gabelica, M., Bojčić, R., & Puljak, L. (2022). Many researchers were not compliant with their published data sharing statement: A mixed-methods study. *Journal of Clinical Epidemiology, 150,* 33-41. https://doi.org/10.1016/j.jclinepi.2022.05.019

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V., Nichols, T. E., & Wagenmakers, E. J. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science, 1*(1), 70-85. https://doi.org/10.1177/2515245917751886

Kamikubo, R., Dwivedi, U., & Kacorri, H. (2021). Sharing practices for datasets related to accessibility and aging. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 1-16). https://doi.org/10.1145/3441852.3471208

Kaye, J., Heeney, C., Hawkins, N., de Vries, J., & Boddington, P. (2009). Data sharing in genomics – Re-shaping scientific practice. *Nature Reviews Genetics, 10,* 331-335. https://doi.org/10.1038/nrg2573

Khedkar , E. B., Kumar, A., Ingle, A., Khaire, R., Paliwal, J. M., Bagul, D., Warpade, S., Londhe, B. M., Malkar, V., Huddedar, S. P., Jambhekar, N. D., & Raibagkar, S. S. (2022). Study of the causes and consequences of cloned journal publications. *Publishing Research Quarterly, 38*(3), 558-572. https://doi.org/10.1007/s12109-022-09907-z

Koch, B., Denton, E., Hanna, A., & Foster, J. G. (2021). Reduced, reused and recycled: The life of a dataset in machine learning research. *arXiv preprint arXiv:2112.01716.* https://arxiv.org/abs/2112.01716

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement, 30*(3), 607-610. https://doi.org/10.1177%2F001316447003000308

Kumar, A., Brar, V., Chaudhari, C., & Raibagkar, S. S. (2022a). Discrimination against private-school students under a special quota for the underprivileged: A case in India. *Asia Pacific Education Review,* 1-10. https://doi.org/10.1007/s12564-022-09815-z

Kumar, A., Gawande, A., & Raibagkar, S. (2022b). Quality complacency in Indian higher education institutions between the second and third cycles of accreditation. *Quality Assurance in Education, 30*(4), 431-445. https://doi.org/10.1108/QAE-01-2022-0019

Kumar, A., Gawande, A., Agarwal, A., Kale, S., Brar, V., & Raibagkar, S. (2023). Sealing the gaps: Enhancing quality control in business school dissertations. *Quality Assurance in Education, 31*(3), 469-484. https://doi.org/10.1108/QAE-12-2022-0207

Kumar, A., Gawande, A., Kale, S. K., Agarwal, A., Brar, V., & Raibagkar, S. (2025). The impact of an effective academic audit on accreditation performance. *Quality Assurance in Education*. Ahead-of-print. https://doi.org/10.1108/QAE-12-2024-0275

Markiewicz, C. J., Gorgolewski, K. J., Feingold, F., Blair, R., Halchenko, Y. O., Miller, E., Hardcastle, N., Wexler, J., Esteban, O., Goncavles, M., Jwa, A., & Poldrack, R. (2021). The OpenNeuro resource for sharing of neuroscience data. *Elife, 10,* Article e71774. https://doi.org/10.7554/eLife.71774

Martone, M. E., Garcia-Castro, A., & VandenBos, G. R. (2018). Data sharing in psychology. *American Psychologist, 73*(2), 111. https://psycnet.apa.org/doi/10.1037/amp0000242

McLeod, J., & O'Connor, K. (2022). Ethics, archives and data sharing in qualitative research. In *The university as an ethical academy?* (pp. 105-117). Routledge.

Michener, W. K. (2015). Ecological data sharing. *Ecological Informatics, 29,* 33-44. https://doi.org/10.1016/j.ecoinf.2015.06.010

Obiora, O. L., Shead, D. A., & Olivier, B. (2024). Data sharing considerations and practice among health researchers in Africa: A scoping review. Digital Health, 10, 20552076241290955. https://doi.org/10.1177/20552076241290955

OECD. (2019). *Education at a glance 2019.* https://doi.org/10.1787/f8d7880d-en

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS One, 2*(3), Article e308. https://doi.org/10.1371/journal.pone.0000308

Polanin, J. R., & Terzian, M. (2019). A data sharing agreement helps to increase researchers' willingness to share primary data: Results from a randomized controlled trial. *Journal of Clinical Epidemiology, 106,* 60-69. https://doi.org/10.1016/j.jclinepi.2018.10.006

Popkin, G. (2019). Data sharing and how it can benefit your scientific career. *Nature, 569*(7756), 445-447. https://doi.org/10.1038/d41586-019-01506-x

Sanabria-Z, J., Alfaro-Ponce, B., Gonzalez-Perez, L. I., & Ramirez-Montoya, M. S. (2024). Reusable educational resources for developing complex thinking on open platforms. *Education and Information Technologies, 29*(1), 1173-1199. https://doi.org/10.1007/s10639-023-12316-0

Shodhganga. (2024). *Shodhganga: A reservoir of Indian theses @ Inflibnet.* https://shodhganga.inflibnet.ac.in/

Singh, M., Paliwal, J., Rao, M. K., & Raibagkar, S. (2024). Impact of goal congruence on higher education institutions' performance quality. *Quality Assurance in Education, 32*(3), 387-400. https://doi.org/10.1108/QAE-12-2023-0215

Springer Nature. (2025). *Journal of Academic Ethics.* https://link.springer.com/journal/10805/submission-guidelines

Taylor and Francis. (2025). *Journal of Business & Finance librarianship.* https://www.tandfonline.com/action/authorSubmission?show=instructions&journalCode=wbfl20#data sharing-policy

Tedersoo, L., Küngas, R., & Oras, E. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Science Data, 8,* 192. https://doi.org/10.1038/s41597-021-00981-0

Tenopir, C., Christian, L., Allard, S., & Borycz, J. (2018). Research data sharing: Practices and attitudes of geophysicists. *Earth and Space Science, 5*(12), 891-902. https://doi.org/10.1029/2018EA000461

The TQM Journal. (2025). Retraction notice: Achieving organizational performance by integrating industrial Internet of things in the SMEs: A developing country perspective. *The TQM Journal, 37*(1), 288-288. https://doi.org/10.1108/TQM-11-2024-0438

Toronto International Data Release Workshop Authors. (2009). Prepublication data sharing. *Nature, 461,* 168-170. https://doi.org/10.1038/461168a

University Grants Commission. (2022). *Ph.D. regulations.* https://www.ugc.gov.in/pdfnews/0909572_Minimum-Standards-and-Procedure-for-Award-of-PhD-Degree.pdf

Vlahou, A., Hallinan, D., Apweiler, R., Argiles, A., Beige, J., Benigni, A., Bischoff, R., Black, P. C., Boehm, F., Céraline, J., Chrousos, G. P., Delles, C., Evenepoel, P., Fridolin, I., Glorieux, G., van Gool, A. J., Heidegger, I., Ioannidis, J. P. A., Jankowski, J. & Vanholder, R. (2021). Data sharing under the General Data Protection Regulation: Time to harmonize law and research ethics? *Hypertension, 77*(4), 1029-1035. https://doi.org/10.1161/HYPERTENSIONAHA.120.16340

Walter, M., Lovett, R., Maher, B., Williamson, B., Prehn, J., Bodkin-Andrews, G., & Lee, V. (2021). Indigenous data sovereignty in the era of big data and open data. *Australian Journal of Social Issues, 56*(2), 143-156. https://doi.org/10.1002/ajs4.141

Wang, J., Chen, M., Lü, G., Yue, S., Wen, Y., Lan, Z., & Zhang, S. (2020). A data sharing method in the open web environment: Data sharing in hydrology. *Journal of Hydrology, 587,* Article 124973. https://doi.org/10.1016/j.jhydrol.2020.124973

White, C. M., Estrera, S. A., Schatschneider, C., & Hart, S. A. (2024). Getting started with data sharing: Advice for researchers in education. *Research in Special Education, 1.* https://doi.org/10.25894/rise.2604

Woods, H. B., & Pinfield, S. (2021). Incentivising research data sharing: A scoping review. *Wellcome Open Research, 6,* 355. https://doi.org/10.12688%2Fwellcomeopenres.17286.2

ANNEXURE 1

QUESTIONNAIRE		
Barriers and Need for Dataset Sharing in the Publishing of Research Thesis		
I expressly record my consent to participate in the survey (Yes, No)		
PROFILE INFORMATION		
1	Gender (Male, Female)	
2	Age (<30 Years, 30-39 Years, 40-49 Years, ≥50 Years)	
3	Region (North, East, West, South)	
4	Year of completion of PhD (2024, 2023, 2022, Earlier Than 2022)	
5	What was the size of your sample for PhD research? (<400, 400, >400)	
6	How had you administered the data collection? (Google forms, Manually, Other Ways)	
7	Did you collect the entire primary data on your own? (Yes, No)	
8	Was the respondent's identity anonymized? (Yes, No)	
9	Have you retained the dataset with you after awarding of the degree? (Yes, No)	
I. Barriers for non-sharing of the dataset		
Rate the following statements on a scale of Strongly agree, Somewhat agree, Neither agree nor disagree, Somewhat disagree, Strongly disagree		
No.	**Statement**	**Response**
1	There is no such provision in the UGC norms for PhD	
2	My research supervisor never asked about the sharing	
3	The research center never asked about the sharing	
4	No questions were asked about it during the final defense viva-voce	
5	I am not aware of such a policy of sharing the dataset with the public	
6	The dataset was not sharable for various reasons ranging from it not being there to it being compiled in an ad-hoc manner	
7	I do not think that there is a need to share the dataset	
8	In my knowledge no one else has shared the dataset	
9	The data collection process is described elaborately in the methodology chapter of my thesis	
10	The data analysis section in the thesis elaborately describes the dataset	
II. Need of sharing of the dataset		
Rate the following statements on a scale of Strongly agree, Somewhat agree, Neither agree nor disagree, Somewhat disagree, Strongly disagree		
No.	**Statement**	**Response**
1	It would lead to transparency in the research process	
2	It would help in improving reliability of the research	
3	It can help current researchers in understanding its structure and other important details	
4	The sharing of the dataset will bring Indian research at par with international practices of data sharing	
5	It will lead to compliance with the Committee on Publication Ethics (COPE) norms on research ethics	
6	It will improve the presentation of the entire thesis during various presentations including the final defense viva-voce	
7	Readers of the thesis can use the dataset for various purposes	
8	It will lead to increased citations of the research work	
9	A metrics counting number of accesses to the dataset can be developed and would indicate the use of the dataset by other researchers	
10	The datasets can be used for further research	