RESEARCH DATA SHARING AMONG DOCTORAL STUDENTS IN KERALA

Thesis submitted to the University of Calicut in
partial fulfilment of the requirements for the award of the degree of
DOCTOR OF PHILOSOPHY IN LIBRARY AND INFORMATION SCIENCE

by

JAMSHEER N P

Under the Guidance of
Dr. Mohamed Haneefa K

DEPARTMENT OF LIBRARY AND INFORMATION SCIENCE
UNIVERSITY OF CALICUT
2025

Declaration

I hereby declare that the thesis entitled **Research Data Sharing among Doctoral Students in Kerala** is an authentic record of research work carried out by me, for my Doctoral Degree under the supervision and guidance of Dr. Mohamed Haneefa K., Professor, Department of Library and Information Science, University of Calicut. This has not been previously submitted for the award of any diploma, degree, title, or recognition. The contents of the thesis have undergone plagiarism check using iThenticate software at C.H. Mohammed Koya Library, University of Calicut, and the similarity index found within the permissible limit. I also declare that the thesis is free from AI generated content.

University of Calicut Jamsheer N. P.
Date : 31-12-2025

Dr. Mohamed Haneefa K.
Professor
Department of Library and Information Science
University of Calicut

DEPARTMENT OF LIBRARY AND INFORMATION SCIENCE
UNIVERSITY OF CALICUT

Dr. Mohamed Haneefa K.
Professor

Mob: +91 9895622811
e-mail: dr.haneefa@gmail.com

Certificate

I, Dr. Mohamed Haneefa K, do hereby certify that the thesis entitled **Research Data Sharing among Doctoral Students in Kerala** submitted to the University of Calicut is a record of the bonafide study and research carried out by Mr. Jamsheer N. P. under my supervision and guidance. The report has not previously formed the basis for the award of any degree, diploma, associateship, fellowship or any other title, of similar recognition in this University or any other University or institution of higher learning.

Calicut University Campus

Date: 31-12-2025

Dr. Mohamed Haneefa K.

(Supervising Teacher)

Acknowledgements

I extend my deepest appreciation to all those who contributed to the realisation of this research work.

First and foremost, I am extremely grateful to my research supervisor, Professor Dr. Mohamed Haneefa K. whose support, insightful guidance, scholarly wisdom, and constant encouragement have been the guiding light throughout the entire PhD journey.

I am indebted to the esteemed faculty members of the department of Library and Information Science at the University of Calicut, whose thoughtful feedback, and supportive environment have enriched my work in countless ways. I express my heartfelt gratitude to Dr. Abdul Azeez T. A., former University Librarian and Dr. Vinod V. M. for providing me with advice and support in the research journey. I am deeply thankful to all my dear friends, particularly Sreelatha K., Mohanan A., Manu C., and Dr. Geethanjali M., whose encouragement and companionship meant so much to me.

I also owe a great debt of gratitude to my present and former colleagues and comrades at the University, whose constant support and motivation helped me successfully completing this work. I sincerely acknowledge the generous assistance provided by the library professionals who made this journey smoother.

Thanks to the many scholarly giants whose work paved the way, and to the free and open-source software community for the fantastic tools that aided my work. Special thanks to all participants and respondents who generously shared their time, experiences, and perspectives.

I also extend my heartfelt appreciation to my beloved life partner Shamseena J whose support, patience, and encouragement have been a pillar of strength during this journey. Her understanding and sacrifices enabled me to pursue my work with focus and confidence.

Above all, my deepest gratitude goes to my beloved late mother Suhara E, my father Koyamoideen T. P., my children and siblings for their love, encouragement, and support over many years.

Table of Contents

List of Tables

List of Figures

Abstract

Academic research produces huge amount of data, but once the research is completed, the datasets are often discarded or lost. Sharing research data has numerous benefits. Data sharing improves transparency, fosters research collaboration and support new discoveries. It also save time, effort, and financial resources. Various national and international bodies have formulated research data management and sharing policies. Academic institutions in Kerala lacks such data sharing practices and policies for research data management. There is a need for policy measures and strategies for establishing research data repositories and motivating research data management.

The study aims to investigate the current research data handling practices and identify factors influencing research data sharing among doctoral students in Kerala. The study also explores the attitudes of the doctoral students towards sharing and proposes policy measures for implementing research data management culture. The study is theoretically based on the Theory of Planned Behaviour of Icek Ajzen and the Institutional Theory of W Richard Scott. A statewide survey is used to collect data from 597 doctoral students from various universities in Kerala. To gain an in-depth understanding of the institutional factors, semi structured interview with 20 research supervisors were also conducted.

The study supports data sharing and recognises it as a means to enhance transparency and integrity in research. The study also finds that data quality and verification remain weak due to the absence of formal mechanisms. Research is often viewed as a requirement for degree completion rather than a scholarly output. The research culture is individual-centred, with reluctance for open discussion and limited collaboration. Career ambitions and the publication pressures affect research priorities and data quality. Ethical concerns, including fears of data misuse and plagiarism, reduce trust and discourages openness. The findings highlight the need for institutional research data repositories, proper infrastructure, clear data sharing policies and trained supporting staff. The study recognises the role of academic libraries in research data management. The study has proposed policy measures for implementing research data sharing culture in Kerala.

Keywords: Research Data, Data Sharing, Research Data Management, Research Data Policy, Doctoral Research, Universities in Kerala.

സംഗ്രഹം

പരീക്ഷണ-നിരീക്ഷണങ്ങളിലൂടെ ശേഖരിക്കുന്ന ഡാറ്റ ഉപയോഗിച്ചാണ് ഗവേഷകർ നിഗമനങ്ങളിൽ എത്തുന്നത്. ഗവേഷണ പ്രവർത്തനങ്ങൾ പൂർത്തീകരിച്ച് ഗവേഷണ ഫലങ്ങൾ പ്രസിദ്ധീകരിച്ചതിനുശേഷം, ഇത്തരം ഡാറ്റ ശാസ്ത്രീയമായി സൂക്ഷിച്ചുവയ്ക്കുകയും മറ്റ ഗവേഷകർക്ക് പുനരുപയോഗത്തിന് സാധിക്കുംവിധം പങ്കുവെക്കുകയും ചെയ്യാവുന്നതാണ് . ഇതിനായി ആഗോളതലത്തിൽ നിരവധി ഗവേഷണ ഡാറ്റാ റെപോസറ്ററികൾ നിലവിലുണ്ട്. അനാവശ്യമായ ആവർത്തനങ്ങളും അധികചെലവും ഒഴിവാക്കാം എന്നുള്ളതിനു പുറമെ നിരവധി അക്കാദമിക നേട്ടങ്ങളും ഇത്തരം ഡാറ്റ പങ്കുവെക്കലിൽ ഉണ്ട് . ഡാറ്റ പങ്കിടൽ ഗവേഷണത്തിലെ സുതാര്യത ഉറപ്പുവരുത്തുകയും ഗവേഷകർക്കിടയിലെ സഹകരണം വളർത്തുകയും ചെയ്യുന്നു. ഒരേ ഡാറ്റ ഉപയോഗിച്ച് വ്യത്യസ്തമായ പഠനങ്ങൾ സാധ്യമാണ്. ഇതുവഴി ഗവേഷണത്തിനായി പൊതു ഖജനാവിൽനിന്നും ചിലവാക്കുന്ന നികുതിപണത്തിന്റെ കാര്യക്ഷമമായ വിനിയോഗം ഉറപ്പുവരുത്തുകയും ചെയ്യാം . നിലവിൽ കേരളത്തിലെ സർവ കലാശാലകളിലും ഗവേഷണ സ്ഥാപനങ്ങളിലുമായി വ്യക്തമായ ഗവേഷണ ഡാറ്റ പങ്കിടൽ രീതികളോ ഔപചാരികമായ ഡാറ്റ പങ്കുവെക്കൽ നയങ്ങളോ നിലവിലില്ല . ഇതിന്റെ ഫലമായി, ഗവേഷണ-ഡാറ്റ ഗവേഷകരുടെ കൈവശം മാത്രമാവുകയും ഭാവിയിൽ പുനരുപയോഗത്തിന് ലഭ്യമാകാതെയും പോകുന്നു .

കേരളത്തിലെ സർവ്വകലാശാലകളിൽ നടക്കുന്ന പി എച്ച് ഡി ഗവേഷണങ്ങളിലെ ഡാറ്റ പങ്കുവെക്കൽ സാധ്യതകളെക്കുറിച്ച് പഠിക്കുന്നതാണ് ഈ ഗവേഷണ പ്രബന്ധം . ഗവേഷകവിദ്യാർഥികൾ നിലവിൽ ഡാറ്റ കൈകാര്യം ചെയ്യുന്ന രീതി മനസ്സിലാക്കുക , ഗവേഷണ ഡാറ്റ പങ്കുവയ്ക്കുക എന്ന ആശയത്തോട് അവർക്കുള്ള മനോഭാവം വിലയിരുത്തുക , പങ്കുവയ്ക്കുന്നതിന് അനുകൂലമായും പ്രതികൂലമായും വർത്തിക്കുന്ന വ്യക്തിപരവും സ്ഥാപനപരവുമായ ഘടകങ്ങൾ പരിശോധിക്കുക എന്നിവയാണ് ഈ പഠനത്തിന്റെ പ്രധാന ലക്ഷ്യങ്ങൾ. കേരളത്തിലെ വിവിധ സർവകലാശാലകളിലെ 597 ഗവേഷക വിദ്യാർഥികളിൽ നിന്നും സർവ്വേവഴി വിവരങ്ങൾ ശേഖരിച്ചു . ഡാറ്റ പങ്കിടലിനെ ബാധിക്കുന്ന സമീപനങ്ങളും സ്ഥാപനപരമായ ഘടകങ്ങളും ആഴത്തിൽ വിശകലനം ചെയ്യുന്നതിനായി ഗവേഷണ മാർഗ ദർശികളായ 20 അധ്യാപകരുമായി അഭിമുഖം നടത്തി. ഐസെക് അജ്സന്റെ "തിയറി ഓഫ് പ്ലാൻഡ് ബിഹേവിയർ", ഡബ്ല്യു റിച്ചാർഡ് സ്കോട്ടിന്റെ "ഇൻസ്റ്റിട്യൂഷണൽ തിയറി" എന്നിവ ഈ പഠനത്തിന് സൈദ്ധാന്തികമായ അടിത്തറ നൽകുന്നു .

പഠനം ഡാറ്റ പങ്കിടലിനെ പിന്തുണയ്ക്കുകയും , ഗവേഷണത്തിലെ സുതാര്യതയും സമഗ്രതയും മെച്ചപ്പെടുത്തുന്നതിനുള്ള ഒരു മാർഗമായി അംഗീകരിക്കുകയും ചെയ്യുന്നു . എന്നാൽ ഔപചാരിക സംവിധാനങ്ങളുടെ അഭാവം മൂലം ഡാറ്റയുടെ ഗുണനിലവാരം ഉറപ്പു വരുത്തുക പ്രയാസമാണെന്ന് പഠനം കണ്ടെത്തുന്നു . ബിരുദം പൂർത്തിയാക്കുന്നതിന് ആവശ്യമായ ഒരു മാർഗമായിട്ടാണ് പലപ്പോഴും ഗവേഷണത്തെ കാണുന്നത് . പരിമിതമായ സഹകരണം, വ്യക്തികേന്ദ്രീകൃത ഗവേഷണസംസ്കാരം , തുറന്ന ചർച്ചയുടെ അഭാവം എന്നിവ ഡാറ്റ പങ്കുവെക്കലിനെ തടസ്സപ്പെടുത്തുന്ന ഘടകങ്ങളായി പ്രവർത്തിക്കുന്നു . പ്രസിദ്ധീകരണ സമ്മർദ്ദങ്ങളും, ഡാറ്റയുടെ ഗുണനിലവാരത്തെ ബാധിക്കുന്നു . ഡാറ്റ ദുരുപയോഗത്തെയും രചനാമോഷണത്തെയും കുറിച്ചുള്ള ഭയം ഉൾപ്പെടെയുള്ള ധാർമ്മിക ആശങ്കകൾ വിശ്വാസ്യത കുറയ്ക്കുകയും പങ്കുവെക്കലിനെ നിരുത്സാഹപ്പെടുത്തുകയും ചെയ്യുന്നു . ഗവേഷണ സ്ഥാപനങ്ങൾ കേന്ദ്രീകരിച്ചുള്ള ഡാറ്റ റെപ്പോസിറ്ററികൾ , അടിസ്ഥാന സൗകര്യങ്ങൾ എന്നിവയ്ക്കു പുറമെ വ്യക്തമായ ഡാറ്റ-പങ്കിടൽ നയങ്ങളും, വിദഗ്ധപരിശീലനം ലഭിച്ച ജീവനക്കാരും ഡാറ്റാ പങ്കുവെക്കൽ സംസ്കാരം വളർത്തുന്നതിൽ പ്രധാനമാണ് . ഗവേഷണ ഡാറ്റാ കൈകാര്യം ചെയ്യുന്നതിലും , ഗവേഷക വിദ്യാർത്ഥികൾക്ക് പരിശീലനം നൽകുന്നതിലും അക്കാദമിക ലൈബ്രറികൾക്കുള്ള പങ്ക് പഠനം തിരിച്ചറിയുന്നു . ഗവേഷണ ഡാറ്റ പങ്കുവെക്കൽ സംസ്കാരം കേരളത്തിൽ നടപ്പാക്കുന്നതിനാവശ്യമായ നയപരമായ നടപടികൾ പഠനം നിർദ്ദേശിച്ചിട്ടുണ്ട്.

സൂചക പദങ്ങൾ: ഗവേഷണ ഡാറ്റ, ഡാറ്റ പങ്കിടൽ, ഗവേഷണ ഡാറ്റാ മാനേജ്മെന്റ്, ഗവേഷണ ഡാറ്റാ നയം, ഡോക്ടറൽ ഗവേഷണം, സർവ്വകലാശാലകളിലെ ഗവേഷണം, കേരളത്തിലെ സർവ്വകലാശാലകൾ.

Chapter 1. Introduction

> In fact, virtually all of the great advances in modern technology are the result of research, both in basic science and technological applications, provided by public institutions or public funds. Without that prior groundwork, much of the current boom in technology would have been impossible (Perelman, 2002).

Scientific research is widely recognised as a cornerstone in the development. The global sustainable development report published by the United Nations reiterate the role of research and development in improving the lives of people worldwide (United Nations, 2023). Research is a laborious and often expensive task, and the data generation or collection is an integral part of any research process. Prior to initiating the research process, it is essential to determine whether reliable and relevant data already exist within established data repositories. This preliminary step helps to prevent unnecessary duplication of data collection efforts and promotes the efficient use of existing research resources (Borgman, 2012). The international registry re3data.org lists more than 3,400 discipline specific and multidisciplinary research data repositories as of May 2025, providing a vast, searchable infrastructure through which researchers can locate and deposit datasets. Consulting established data repositories not only curbs avoidable expenditure of time and money but also amplifies the public return on investments in science by fostering transparency, reproducibility, and secondary discovery.

The potential benefits of the open data sharing is well recognised and documented in the previous studies. Data sharing is considered one of the methods to improve the efficiency of research (Munafò et al., 2017). Data from publicly funded studies, if shared, can inform evidence based policymaking and public services. Scientists worldwide openly shared virus genomic data and other findings, which helped in the rapid development of vaccines during COVID-19 pandemic (Gardner et al., 2021). An open data culture in research improves transparency, enhances reproducibility, trust and fosters research collaboration. It increases the return on investment of public research by enabling others to use and extend scientific knowledge.

The implementation of an institutional level research data repository requires wider discussion with stakeholders. The institutional culture, individual attitude and competency of staff plays a vital role in the implementation of research data repositories. Training programmes and proper academic support are inevitable for motivating researchers to contribute their data towards the proposed research data repository. An analysis of download logs from a national Social Science archive showed that archived datasets are being reused at increasing rates for research questions which was not anticipated by the original investigators (Late & Ochsner, 2024). Research data generated with public fund can be systematically deposited in repositories. Other researchers can reuse these data and save time, effort, and money

Publicly Funded Research

According to Purkayastha (2024) universities are increasingly converting the results of publicly funded research into private commodities. This is done by patenting discoveries that were generated with public money by licensing or selling the rights to industry. Research data generally comprises of factual records generated in studies that are financed by government fund. Such data from projects done by public laboratories and agencies by universities, research institutes, and private organisations that receive grants, or any other form of public sector financial support (OECD, 2017). Since public funds are often used to support research as opined by Purkayastha "The major source of funding for creating new knowledge is still publicly funded, but increasingly, their output is privatised". Many governments and institutions consider the data produced as a public resource. As a result, they promote wide access to such data, while ensuring that ethical, legal, and commercial protections are in place (Tenopir et al., 2011).

International policy frameworks underscore the obligation to make publicly funded data discoverable and reusable. The organisation for Economic Cooperation and Development (OECD) principles and guidelines for access to research data from public funding recommend that governments adopt openness as the default, arguing that accessible data maximise the return on investment, advance transparency, and stimulate innovation. UNESCO recommends sharing data "as openly as possible and as closed as necessary" (UNESCO & Canadian National Commission for UNESCO, 2022). India's National Data Sharing and Accessibility Policy (NDSAP) classifies data produced with public money as a strategic asset that should be made available for public use in machine readable form, while safeguarding sensitive information (Government of India, 2012). By placing publicly funded datasets in open repositories, academic and research institutions can improve research reproducibility and support new discoveries through data reuse. This approach also reduces unnecessary duplication of data collection and helps save time, effort, and financial resources (Piwowar & Vision, 2013).

Research Data Sharing

Academic researches produce huge amount of research data, but once the research work is completed, these data are often discarded or lost forever. Deliberate measures are not taken to preserve these data. A study of 516 ecology papers showed that the probability of retrieving an original dataset falls by roughly 17% per year after publication (Vines et al., 2014). The open access policy, which is popular in paper publication, is yet to gain popularity in data publication. State of Open Data 2023 survey of more than 6000 researchers found that 73% had never received institutional support for making their data publicly available and cited "lack of credit" as the leading disincentive (Hahnel, 2025). Analysis of research data can yield insights and findings that the original investigator may not have anticipated (Borgman, 2012). Despite these benefits, multiple studies have pointed out the reluctance of researchers in sharing their research data for various reasons. One key barrier is the additional time and effort required to organise, document, and publish data in a reusable format (Van den Eynden & Bishop, 2014). Tenopir et al. (2011) found that even when researchers express willingness to share their data, they often lack the time to adequately prepare it. Study has highlighted the need for technical tools and expert assistance to support data management activities (Fecher et al., 2015). In a survey conducted among climatologists, researchers indicated they were more likely to share their data if they received recognition or incentives, such as data citation (Thöle & Wegmann, 2024).

Researchers' attitudes toward data sharing are influenced by multiple factors. These include individual characteristics, institutional culture, available institutional infrastructure, and the level of academic and technical support (Borycz et al., 2023; Kim & Stanton, 2016). In the absence of robust institutional backing and clearly defined data sharing policies, researchers are often reluctant to make their data publicly available in reusable formats (Fecher et al., 2015). Studies show that data reuse experience significantly influenced participants' perception of benefit from data sharing and participants' norm of data sharing (Kim & Nah, 2018). In the Indian context, significant progress has been observed following the University Grants Commission's (UGC) mandate requiring Indian Universities to publish theses and dissertations in Shodhganga digital repository. This policy shift has encouraged scholars to deposit their research outputs in the public domain, particularly when institutional or funding mandates require open data practices.

On the global stage, several organisations have been instrumental in promoting a culture of open data. Initiatives by the Open Data Institute (ODI), re3data.org, and the Digital Curation Centre (DCC) have helped raise awareness and provide guidance on responsible data management practices. The leading international research organisations like European organisation for Nuclear Research (CERN) have shared their datasets for public access. Findable, Accessible, Interoperable and Reusable (FAIR) data is the aim of the European Commission. The principle clearly distinguishes between open data and FAIR data (European Commission. Directorate General for Research and Innovation., 2018, p. 21). Depending on the nature of the data, the FAIR data can be made available under restrictions. Various countries and international organisations have adopted research data sharing policies (CODATA & Uhlir, 2015).

New set of policies, governance mechanism, processes and capabilities are required for implementing research data repositories at institutional level. Capabilities, competency and capacity required for setting up of RDR is different from that of building an institutional repository. Research data repositories are different from institutional repository of thesis, dissertations and other intellectual output. Research data repositories are complex set of factors like behavioural, technological and policies. Sustainable data sharing requires a proper business model (OECD, 2017). Proper funding should also be ensured, and this is possible only if all the stakeholders of the research data repository understand the need for data sharing.

Need and Significance of the Study

Sharing research data openly can yield numerous benefits. Prior studies have documented that data sharing maximises transparency and accountability, allowing other scientists and the public to scrutinise and trust research findings. Open availability of data enables independent validation of results. Open research data can also improve the quality and integrity of the original research through peer reanalysis. Other investigators can reuse and build upon the openly shared data, ask new questions, test new hypothesis without collecting duplicate data. This not only encourage innovation but also saves considerable time, effort and money in the research. By reusing existing datasets, researchers can optimise resource use, which is important in the developing countries where funding is scarce. Data from publicly funded studies, if shared, can inform evidence based policymaking and public services.

The output published in the form of a research report was considered as the end result of the research activity. Large scale research and the cost of such research has made it practically impossible (Joint, 2007) and financially unviable to duplicate the data generation process. This has made various national and international bodies to formulate research data preservation and sharing policies. The literature highlights different benefits of research data reuse, diverse approaches to data management and sharing adopted by academic institutions in various countries. But, a review of existing studies, complemented by direct observation, indicates that research institutions and universities in India are yet to adopt or implement research data sharing practices. Universities in Kerala lacks data sharing practices and mandatory policies, reflecting absence data sharing culture.

Investigator has identified, the underutilisation of research data and the lack of data management practice in universities of Kerala. This gap in knowledge underscores the need for an exploratory study to identify the barriers and enablers research data sharing behaviour. The data sharing attitudes and practices of researchers, in Kerala remain under examined. It remains unclear to what extend the doctoral students and research supervisors embrace research data sharing. This study offers a novel contribution by exploring the attitudes of doctoral students and research supervisors towards data sharing in the academic institutions of Kerala. The study proposes to uncover the perceived barriers and motivations that might influence data sharing behaviour.

Institutional culture plays a critical role in shaping the research data sharing. Effective policy formulation for data sharing must account for the cultural dynamics within institutions that influence researchers' data sharing behaviours (Zuiderwijk, 2024). There is a lack of pre implementation feasibility assessments in the research data sharing policy development context. The study can be used as a guideline by policy makers, funding agencies, researchers and librarians to devise ways to initiate the research data preservation and reuse. The findings of the study will help universities and government agencies to design, refine, and implement policies.

There is a need for policy measures and strategies for establishing research data repositories and motivating data sharing behaviour. The study is theoretically based on the Theory of Planned Behaviour (TPB) of Ajzen (1991) and Institutional Theory of Scott (2014). The use of two frameworks underscores the significance of investigating both attitudinal and system level factors, which have been not been attempted in previous studies. The multiple approach helps in understanding both individual level intentions and institutional factors affecting research data sharing. Using a mixed method research design, the study aims to generate insights into the existing practices, challenges, and perceptions of research data generation and management. There is little empirical work from Kerala or similar Indian states that connects the institutional infrastructural variables with empirical evidences in the research data landscape. This gap motivates the present study, by examining how individual motivations, perceived control, attitudes, and institutional factors together influence intention to share research data. Considering the absence of a formal national policy, the findings can inform the development of institutional and national level research data sharing policies.

Statement of the Problem

Despite the well recognised benefits of research data reuse, the existing research culture and attitudes in universities of Kerala do not favour research data sharing. Due to various individual and institutional barriers researchers' willingness to share data remains inconsistent and low. In India, there is a lack of clear guidelines and policies for data management and sharing, leaving it up to individual discretion. Except when funders or journals mandate data sharing for certain projects, research data remains underutilised and unshared. The absence of research data management policy often leads to data loss, violation of privacy, and unnecessary duplication of work. This situation points to an academic culture where data is viewed as a private asset rather than a public good. The problem, therefore, is a misalignment between the public funding of research and the private ownership research outputs. The study emphasises that understanding the individual and institutional factors influencing data sharing is crucial in bridging this gap.

The study is exploratory in nature, aiming to investigate the current situation and to identify important factors that promote or restrict data sharing. The findings help answer why researchers might hesitate to share data. Articulating and understanding this problem is the first step towards encouraging data sharing culture. By examining researchers' intentions and perceived constraints, the study provides ground level evidence needed to promote research data sharing in Kerala. Therefore the study is entitled as "Research Data Sharing among Doctoral Students in Kerala".

Definition of Key Terms

Research data

OECD has defined research data as "factual records (numerical scores, textual records, images, and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings". Biotech PRIDE (Government of India., 2021) defines research data as "data that is collected or created for purpose of analysis leading to understanding a scientific question". According to DCC, research data are the "evidence that underpins the answer to a research question, covering a wide range of types including spreadsheets, lab notes, questionnaires, transcripts, images, and more, which are generated or collected during the research process". According to Australian National Data Services, research data are "facts, observations or experiences on which an argument or theory is based". European Commission in it's Guidelines to the rules on open access to scientific publications and open access to research data has defined research data as "information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images". The present study has adopted the definition offered by the DCC.

Data Sharing

The OECD defines it as "providing access to research data for use by others, subject to legal, ethical, and commercial constraints, with the aim of enabling reproducibility and further research".

Doctoral Students

According to Oxford dictionary, researcher means "a person who studies something carefully and tries to discover new facts about it". Cambridge dictionary defines researchers as "someone who studies a subject, especially in order to discover new information or reach a new understanding". In the study Doctoral Students means PhD students who are doing research for obtaining Doctor of Philosophy degree from a university in Kerala or who has recently completed PhD degree.

Kerala

Kerala is the southern state of India known for its distinctive social indicators, including high literacy rates, relatively low infant mortality, and progressive health and education systems. As per the 2011 Census, Kerala has a population exceeding 33 million, with a literacy rate of 94%, the highest among Indian states. Kerala is also recognised for its strong public infrastructure, decentralised governance, and emphasis on human development.

Theoretical Framework

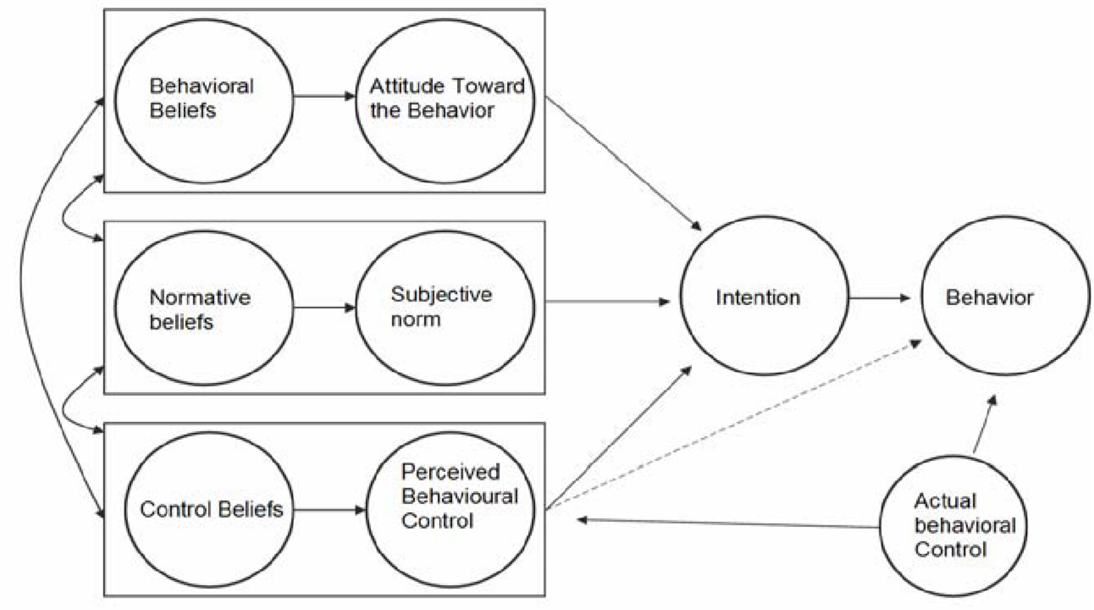
To analyse the determinants of researchers' intentions to share data, the study uses two theoretical frameworks, TPB and Institutional Theory. TPB is used to understand the attitude, intentions and the factors affecting the attitude of researchers towards data sharing behaviour. Research data sharing is also influenced by the institutional culture, available infrastructure and policies of the institution. To investigate these institutional factors, the institutional theory is used.

Theory of Planned Behaviour

The Theory of Planned Behaviour, developed by Ajzen (1991), is a prominent psychological framework used to understand and predict human behaviour. According to this theory, an individual's intention to perform a specific action is the most immediate determinant of that behaviour. This intention is influenced by three major constructs as visible from the Figure 1. Attitude toward the behaviour, subjective norms, and perceived behavioural control are three constructs of TPB. Attitude refers to the individual's overall evaluation of the behaviour as positive or negative. Subjective norms relate to the perceived social pressure from others, such as colleagues, superiors, or institutional expectations. Perceived behavioural control reflects the individual's sense of ease or difficulty in performing the behaviour, which is closely related to their confidence or capacity to carry it out. When these factors align positively, the intention of data sharing increases.

Figure 1.1

Theory of Planned Behaviour (Ajzen, 1991)



The widespread application of the TPB across disciplines confirms its robustness and adaptability. The theory has been successfully used to study behaviours like knowledge sharing and adoption of open data practices. Recent research has demonstrated the usefulness of TPB in understanding how both individual and contextual factors shape the behaviours of academic researchers. By applying the theory to the context of research data sharing in Kerala, the study seeks to uncover the underlying influences on researchers' willingness to share their data.

The theory offers a clear lens through which beliefs about the usefulness or risks of data can be explored. The disciplinary and institutional expectations, the sense of control over the sharing process which influence the willingness to share data. A researcher may be willing to share data, but they may still refrain if they lack the necessary skills or institutional infrastructure to involve in data sharing. In such situations, the TPB model helps capture both motivational and structural barriers to intended behaviour.

Institutional Theory

Institutional Theory is an established framework in organisational and sociological research. Institutional Theory explains how institutions shape the behaviour of individuals and organisations. Organisational change has traditionally been viewed through the lens of altering technologies, structures, or employee competencies (Ogbor, 1990). Institutional Theory shows that behaviours are not merely the result of rational choices but are deeply embedded within and influenced by institutional environments. These environments include formal structures such as laws and policies, as well as informal elements like cultural beliefs, social norms, and professional practices. Organisations and individuals are seen as acting within a social context. The actions are constrained, enabled, or legitimated by the institutional arrangements. Institutional pressures lead to the diffusion of accepted norms and practices, promoting homogeneity across organisations in similar fields.

Individuals respond to change in three different ways. First category comply because it is mandated by formal policies and regulations. Second category adapt because it aligns with prevailing social expectations. Others embrace it because it comply with their personal beliefs (Palthe, 2014). One of the most influential versions of Institutional Theory is presented by Scott, W Richard (2014). According to Scott, institutions consists of three interrelated regulative, normative, and cultural-cognitive pillars. The regulative pillar includes formal rules, laws, and sanctions that guide and constrain behaviour. The normative pillar comprises values, norms, and roles that define what is considered appropriate within a social system. The cultural-cognitive pillar reflects shared beliefs, symbols, and taken-for-granted understandings that give meaning to actions. According to Scott, the stability and power of an institution derive from the alignment of these three pillars. These three pillars together influence how actors perceive their environment and behave within it. This multidimensional framework has made Scott’s version of Institutional Theory one of the most widely cited and applied in contemporary organisational studies (Palthe, 2014).

Figure 1.2

Three Pillars of Institutional Theory

	Regulative	*Normative*	*Cultural-Cognitive*
Basis of compliance	Expedience	Social obligation	Taken-for-grantedness Shared understanding
Basis of order	Regulative rules	Binding expectations	Constitutive schema
Mechanisms	Coercive	Normative	Mimetic
Logic	Instrumentality	Appropriateness	Orthodoxy
Indicators	Rules Laws Sanctions	Certification Accreditation	Common beliefs Shared logics of action Isomorphism
Affect	Fear Guilt/ Innocence	Shame/Honor	Certainty/Confusion
Basis of legitimacy	Legally sanctioned	Morally governed	Comprehensible Recognizable Culturally supported

Institutional Theory encompasses a broad range of concepts, including institutional logics and isomorphism. The present study applies the theory in a focused manner by using Scott's (2008) three-pillar framework. Three pillar framework of Scott examines how institutional structures, values, and shared understandings influence research data sharing practices within academic institutions. The application of Institutional Theory is highly relevant to the present study, which investigates research data sharing in academic institutions in Kerala. In this context, researchers operate not in isolation but within institutional settings that shape their perceptions, opportunities, and actions. The absence of formal policies is part of the regulative pillar in the academic institution. The lack of shared professional norms around data sharing is part of the normative pillar. Limited awareness or understanding of data sharing practices is part of the cultural-cognitive pillar. All these pillars contribute to the limited adoption of data behaviour in the academic institutions in Kerala. Integrating Institutional Theory into the study helps to examine the institutional environments supporting or inhibiting researchers' intentions and capacities to share their data. The framework allows for analysis of the institutional changes which can foster an enabling environment for responsible data sharing. The institutional changes may be policy development, capacity building, and cultural transformation aligning with the three pillars of Institutional Theory.

By using Institutional Theory, the study acknowledges that individual intentions are influenced by a broader institutional culture. As more researchers share data others have a pressure to join the trend to share. Researcher's perceived behavioural control is enhanced by institutional support like access to technical tools and data management infrastructure. The combination of TPB and Institutional Theory helps to examine personal motivational factors and institutional influences on the intention to share research data.

Objectives of the Study

Purpose of the study is to provide pre-implementation insights on research data sharing. The study aims to understand the existing research data management landscape and propose a policy framework tailored to the academic research environment in Kerala. With the theory, subject and population in mind following are the specific objectives of the study:

1. To investigate the research data handling practices of doctoral students in Kerala.
2. To understand attitudes of the researchers towards research data sharing.
3. To assess the perceived motivating factors in research data sharing.
4. To assess the perceived barriers in research data sharing.
5. To explore the influence of institutional factors in research data sharing.
6. To propose policy measures and strategies for establishing research data repositories that address challenges in the light of the objectives 1 to 5.

Scope and Delimitation of the Study

The concept of research data sharing and repository building overlap with various domains and can be studied from different angles. It has technological, ethical and legal, financial, societal and scientific aspects. Policy frameworks, organisational culture, technical infrastructure, procedural workflows, and user attitudes may shape the research data sharing behaviour (Ahmed et al., 2025). Repository creation and management involves additional financial burden. But this is not considered for the present study. In view of the complex and multifaceted nature of the topic, the investigation is limited to a selected number of variables. The study gathers expert insights from research supervisors to understand their experience, conditions and constraints regarding the research data sharing. The study is delimited to investigating the behavioural intention to share research data. The investigation does not intend to measure actual data sharing. The actual sharing behaviours may differ from stated intentions of the researchers, and examining those behaviours or long term data reuse falls outside the scope of the study.

Many stakeholders are involved in the research data landscape, but the study focuses on doctoral students and the research supervisors in publicly funded universities in Kerala. The first stakeholder, the doctoral students are both data generators and users of the data. They are included in the investigation to understand the current research data handling practices and also to understand their attitude towards data sharing. Another stakeholder, research supervisors are both experienced researchers and academicians who are likely to influence the data sharing behaviour at individual as well as at institutional level. The sample represent five different disciplinary categories. They are Engineering, Humanities, Life Sciences, Natural Sciences and Social Sciences to ensure broad coverage of research domains. While multidisciplinary in scope, the study does not conduct granular sub-disciplinary analyses like economics, botany or astrophysics. Conclusions are pertinent to academic researchers in Kerala's general university system and not extended to privately funded research institutions or research institutes other than university research centres.

Conceptually, the study is based in TPB and Institutional Theory, which together frame the individual and institutional factors of interest. By delimiting the theoretical scope to TPB and Institutional Theory, the study intentionally does not incorporate other behavioural or policy models. It does not test technology acceptance models or alternative behaviour change theories. This focused theoretical boundary ensures depth within the chosen frameworks. Additional elements of Institutional Theory, such as isomorphism or institutional logics, which centres on identifying and addressing systemic gaps in institutional practices are not considered in the study.

Finally, the findings and implications are framed for the institutional context of Kerala. The intent of the study is to produce insights which can inform university policy. The study also helps planning institutional data repositories in Kerala and create supportive for data sharing culture. Repository usage statistics or citation metrics were excluded to prioritise qualitative insights into barriers and motivations. While the conclusions may echo broader themes in research data management literature, they are tailored to the cultural and administrative setting of Kerala's public universities. The recommendations are primarily aimed at guiding universities, research institutes, funding agencies, and governments promoting data sharing.

Overview of Methodology

Mixed methodology is used in the study to arrive at understanding of the situation. This approach will provide a comprehensive understanding of various aspects of data sharing and repository development. First, a quantitative survey will be used to understand the existing practices, attitude towards data sharing and the concerns of research scholars in the universities in Kerala. The survey instrument is informed by the TPB constructs like assessing respondents' attitudes toward data sharing, perceived norms, and perceived control such as familiarity with data repositories or data management skills. It also collects information on perceived institutional support and policies, bridging to the institutional factors framework.

Complementing the survey, semi-structured interviews will be carried out with research supervisors who can provide deeper insight into the institutional culture. The interviews allow an in-depth exploration of themes such as what encourages or discourages data sharing in their departments and the institutional challenges they observe. The interview data will be analysed using thematic analysis, to identify common patterns and unique perspectives about data sharing practices and attitudes. The study captures the individual perspectives of doctoral students and viewpoints of senior research supervisors, giving a holistic understanding of the factors.

Thesis Outline

This study seeks to generate meaningful insights into the factors influencing research data sharing and to develop recommendations for policymakers, academic institutions, and other stakeholders engaged in research data management. To achieve this purpose, the thesis is organised into seven chapters, each addressing distinct components of the research process while building toward an integrated understanding of the problem.

Chapter 1: Introduction

This chapter provides introduction to the concept of research data, research data management, benefits and challenges of data sharing. Major stakeholders involved, various components of research data repositories are also discussed in the chapter. Further, the chapter highlights why such a study is conducted, the need and significance of the study and objectives of the study. Scope and delimitation of the study are also discussed in this chapter.

Chapter 2: Review of Literature

This chapter reviews the previous work conducted in the area of research data sharing behaviour. Various sections in the chapter deals with the research data management, policies and the context in India. The chapter also discusses the research gap in the area.

Chapter 3: Methodology

The methodology chapter explains the methodology, ontology and epistemology adopted for the study. The rationale for using mixed methodology research design is discussed in detail in the first section. Following sections discusses the quantitative and qualitative methodologies used and how the results are combined to arrive at comprehensive findings. Development of questionnaire, interview schedule, process of data collection, how the pilot study was conducted and the data analysis tools and procedures are also discussed in the methodology chapter. The methodology chapter specifically addresses the measures taken to confirm validity, reliability, and trustworthiness in both the quantitative and qualitative phases of the research. Methodology chapter concludes with the discussion of measures taken to ensure an ethical study and the style manual used for preparation of the thesis is also explained in the chapter

Chapter 4: Quantitative Analysis

This chapter consists of the analysis of the quantitative data collected using questionnaire from the doctoral students and PhD graduates. Open source software PSPP is used for the quantitative analysis. Research data practices, sharing attitude, major motivating factors, and challenges in data sharing are analysed in separate sections. In addition to the quantitative analysis, analysis and interpretation of the open-ended responses are also provided in this chapter. The chapter ends with the key findings from the quantitative analysis of the survey results.

Chapter 5 Qualitative Analysis

Chapter five analyse the data collected from research supervisors using semi-structured interview. The chapter begins with the profile of the interview participants and discusses various steps involved in qualitative analysis. In-depth analysis of the various factors affecting research data sharing is discussed in the chapter. Chapter is organised in accordance with the themes derived in the thematic analysis. Sub-themes derived in the thematic analysis are discussed under each themes. Chapter ends with major findings from qualitative analysis of the interview data.

Chapter 6 Discussion

This chapter makes inferences of findings from chapter four and five and discusses these inferences in the light of the previous studies. The significance of the result are discussed based on the institutional research culture prevailing in academic institutions in Kerala. Policy suggestions are proposed on the basis of the findings. Various suggestions to improve the data management, preservation and sharing both in the individual and institutional level are provided as the contribution of the research.

Chapter 7: Recommendations and Conclusion

The final chapter synthesises the principal findings by mapping research objective to the empirical evidence obtained. It describes the theoretical, practical, methodological and societal implications of the study. Direct recommendations of the study are also provided in the chapter. Every research has its own limitations, the following sections discuss the methodological, theoretical, and other limitations of the study. The chapter concludes with the scope for future studies in the area.

Style Manual Used

The Publication Manual of the American Psychological Association (APA) 7th edition is used for preparing thesis report. It provides guidelines for paper structure, clarity, inclusive language, and the acknowledgment of sources. APA is strictly followed in headings, body text, tables, citation and reference list entries. In addition to the formatting guidelines, the APA guidelines provided for inclusive and bias-free language is also helped preparing the report.

Chapter Summary

Ideally the science relies on reproducibility and scientific community expect to follow the scientific methods. But for various reasons, all the studies are not reproducible (Baker, 2016). The reasons may vary, from low sample size to modifying hypothesis to match the result (Munafò et al., 2017). Publication of raw data along with the research report helps reproduce the original findings. The study focuses on the current research data practices, motivating and challenging factors among researchers in Kerala. By mitigating the challenging factors and promoting the enabling factors the study is planning to develop policy measures and strategies for establishing research data repositories. The qualitative approach will help to assess the problems and understand possible diagnosis for these problems so that a strategy can be developed to implement institutional research data repositories in Kerala.

References

Ahmed, M., Othman, R., Noordin, M. F., Ibrahim, A. A., & Al-Hussaini, A. I. S. (2025). Factors influencing open science participation through research data sharing and reuse among researchers: A systematic literature review. *Knowledge and Information Systems*, *67*(3), 2801–2853. https://doi.org/10.1007/s10115-024-02284-3

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454. https://doi.org/10.1038/533452a

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

Borycz, J. 1, Olendorf, R. 2, Specht, A. 3, Grant, B. 4, Crowston, K. 5, Tenopir, C. 6, Allard, S. 6, Rice, N. M. 6, Hu, R. 7, & Sandusky, R. J. 8 1 V. U. (2023). *Perceived benefits of open data are improving but scientists still lack resources, skills, and rewards*. 339. https://doi.org/10.1057/s41599-023-01831-7

CODATA, & Uhlir, P. (2015). *The Value of Open Data Sharing: A CODATA Report for the Group on Earth Observations*. https://doi.org/10.5281/zenodo.33830

European Commission. Directorate General for Research and Innovation. (2018). *Turning FAIR into reality :final report and action plan from the European Commission expert group on FAIR data.* Publications Office. https://data.europa.eu/doi/10.2777/54599

Fecher, B., Friesike, S., & Hebing, M. (2015). What drives academic data sharing? *PloS One*, *10*(2), e0118053. https://doi.org/10.1371/journal.pone.0118053

Gardner, L., Ratcliff, J., Dong, E., & Katz, A. (2021). A need for open public data standards and sharing in light of COVID-19. *The Lancet Infectious Diseases*, *21*(4), e80. https://doi.org/10.1016/S1473-3099(20)30635-6

Government of India. (2012). *National data sharing and accessibility policy (NDSAP)*. https://dst.gov.in/national-data-sharing-and-accessibility-policy-0

Government of India. (2021). *Biotech-PRIDE: promotion of research and innovation through data exchange guidelines.* https://dbtindia.gov.in/sites/default/files/uploadfiles/Biotech%20PRIDE%20Guidelines.pdf

Hahnel, M. (2025, January 20). Are we ready to start rewarding researchers for open data? *Figshare*. https://info.figshare.com/are-we-ready-to-start-rewarding-researchers-for-open-data/

Joint, N. (2007). Data preservation, the new science and the practitioner librarian. *Library Review*, *56*(6), 451–455. https://doi.org/10.1108/00242530710760337

Kim, Y., & Nah, S. (2018). Internet researchers' data sharing behaviors: An integration of data reuse experience, attitudinal beliefs, social norms, and resource factors. *Online Information Review*, *42*(1), 124–142. https://doi.org/10.1108/OIR-10-2016-0313

Kim, Y., & Stanton, J. M. (2016). Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*, *67*(4), 776–799. https://doi.org/10.1002/asi.23424

Late, E., & Ochsner, M. (2024). Re-use of research data in the social sciences. Use and users of digital data archive. *PLOS ONE*, *19*(5), e0303190. https://doi.org/10.1371/journal.pone.0303190

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 1–9. https://doi.org/10.1038/s41562-016-0021

OECD. (2017). *Business models for sustainable research data repositories*. https://doi.org/10.1787/302b12bb-en

Palthe, J. (2014). Regulative, normative, and cognitive elements of organizations: Implications for managing change. *Management and Organizational Studies*, *1*(2), p59. https://doi.org/10.5430/mos.v1n2p59

Perelman, M. (2002). *Steal this idea: Intellectual property and the corporate confiscation of creativity*. Palgrave Macmillan US.

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, *1*, e175. https://doi.org/10.7717/peerj.175

Purkayastha, P. (2024). *Knowledge as commons: Toward inclusive science and technology*. Monthly Review Press.

Scott, W. R. (2014). *Institutions and organizations: Ideas, interests and identities* (4th ed.). Sage.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLOS ONE*, *6*(6), e21101. https://doi.org/10.1371/journal.pone.0021101

Thöle, L. M., & Wegmann, M. (2024). Open climate science is brave climate science. *Environmental Research Letters*, *19*(12), 122001. https://doi.org/10.1088/1748-9326/ad893f

UNESCO & Canadian National Commission for UNESCO. (2022). *An introduction to the UNESCO Recommendation on Open Science*. UNESCO. https://doi.org/10.54677/XOIR1696

United Nations. (2023). Global sustainable development report (GSDR) 2023. *UN Department of Economic and Social Affairs*.

Van den Eynden, V., & Bishop, L. (2014). *Sowing the seed: Incentives and motivations for sharing research data, a researcher's perspective* [Publication]. https://repository.jisc.ac.uk/5662/

Vines, T. H., Albert, A. Y. K., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., Gilbert, K. J., Moore, J.-S., Renaut, S., & Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology*, *24*(1), 94–97. https://doi.org/10.1016/j.cub.2013.11.014

Zuiderwijk, A. (2024). Researchers' willingness and ability to openly share their research data: A survey of COVID-19 pandemic-related factors. *SAGE Open*, *14*(1). https://doi.org/10.1177/21582440241234985

Chapter 2. Review of Literature

The review of literature is a crucial component of any doctoral research, as it establishes the intellectual foundation of the study and situates the research problem within existing scholarly debates. In the case of research data sharing, the body of literature is rapidly expanding, reflecting the growing global emphasis on transparency, reproducibility, and open science. Research data are increasingly recognised as vital scientific assets that, if properly managed and shared, can significantly accelerate discovery, reduce duplication of effort, and enhance the credibility of scientific outputs (Piwowar & Vision, 2013; Zuiderwijk, 2024). Despite this, empirical studies continue to reveal substantial challenges to widespread adoption of data sharing practices, ranging from technical barriers to socio-cultural resistance (Tenopir et al., 2011).

Benefits of Data Sharing

Open access to articles and other reports can increase the visibility and impact of research. Open access to the research data can make even a bigger impact (Borgman, 2017). The strategic opening of both administrative and research data yields significant returns on public investment, particularly as the generation of such data is primarily supported through public funding mechanisms. Open data practices enhance transparency, foster innovation, and enable broader societal and scientific reuse of information, amplifying the value derived from the initial investment (Zuiderwijk, 2024). Governments and academic institutions worldwide are increasingly recognising that openly accessible data can catalyse evidence-based policymaking, stimulate economic activity, and accelerate scientific progress, especially in data-intensive fields (Yoon & Kim, 2020). Open access to these data provides the possibility of new products and services. Open data results to open governance and active interaction (Janssen et al., 2012) with the citizen. Open data can be used to generate value from the data which is originally created for other purposes (Jetzek et al., 2019; Zuiderwijk & Spiers, 2019). The use of open data can trigger innovation and, increase the social and economical value of the open data. Both public and private organisations can create value from the use of open data

(Jetzek et al., 2019). One of the ways that have been suggested for improving research impact is Open Access.

International frameworks treat publicly funded research data as a public good to be made "as open as possible, as closed as necessary" balancing openness with privacy. This framing links open data to public accountability, evidence-informed policy, and the Sustainable Development Goals (UNESCO & Canadian National Commission for UNESCO, 2022). Open data make it easier to verify, analyse and detect errors. The scholarly community's interest in post-publication review and retractions has grown, and broad access to underlying data is an important ingredient for such quality assurance. By uploading their research data, researchers got invitation to collaborate and helped secured funding from agencies (Chikoore, 2016). Research in the private sector get noticed when it becomes a product. When the research fails, the outside world may not even know the existence of such a research, which leads to a wasteful duplication of efforts and resources (Perelman, 2002).

Benefits of research data preservation and sharing is established and widely discussed in the literature (Borgman, 2012). Many countries have promulgated policies for sharing research data (Anger et al., 2022). A consistent, newer finding is the citation advantage associated with sharing research data. The majority of research works are funded by public money. Data has started playing role in policy making though the implementation is slow in developing countries like India (Heeks et al., 2021). Literature shows that stronger data availability improves the conditions for independent checks and meta-analyses (UNESCO & Canadian National Commission for UNESCO, 2022). Sharing well-documented datasets reduces duplication of effort, allowing researchers to build on existing work rather than repeating data collection. In large, multi-country surveys, scientists say they value reuse for saving time and resources, provided the data are findable and documented. There is now a growing number of studies, including those of scholars such as Piwowar (2011), who have argued that journal articles that are published with data that has been made available online receive more citations than those that did not (Chikoore, 2016).

By making datasets and analysis code publicly available, open data enables independent researchers to reproduce published results, confirming their accuracy and reducing issues like selective reporting or p-hacking (Spiegelman, 2021). Data quality is a major factor in the proper utilisation (Cai & Zhu, 2015), and it depends on various steps involved right from the data collection to the preservation steps (De Lusignan et al., 2011). Data curation has a critical role in ensuring data quality. Studies suggest to assess and certify repositories to ensure quality (Downs, 2021). In clinical data, secondary investigators are advised to reproduce the published results using the initial methods to make sure that they understand the original data properly (Lo & DeMets, 2016).

Data Sharing and Academic Research

Standard citation analysis measures the academic impact of research, but not its societal impact (Bornmann, 2013; Dorta-González et al., 2024). Research data sharing proved to help showcasing the research impact and research output of the institutions (Ho et al., 2025) . Different stakeholders have role to play in encouraging research data sharing and fostering a more collaborative and cooperative research environment (Gomes et al., 2022). Over the past few years, the push for open data and transparent research practices has gained increasing momentum globally. Scholars have demonstrated that data sharing augments research visibility, reproducibility, and social utility, especially when supported by robust infrastructure, clear policies, metadata standards, and human support roles. Yet, many studies show that intentions to share data frequently fail to translate into actual practice (Robinson-García et al., 2016), especially in settings where institutional supports are weak or ambiguous (Khurana, 2021; "A Use-Driven Approach to Data Governance," 2021). These findings underscore the importance of exploring not just what factors influence willingness to share, but how institutional and infrastructural arrangements mediate that relationship.

The findings indicate that the utilisation and publication of Open Data exhibit distinct trends across countries and research disciplines (Numajiri & Hayashi, 2024). Study conducted among life scientists (Yoon & Kim, 2020) shows that people who used secondary data are likely to share their data and this helps reducing perceived risks and lowering perceived effort. According to

Sherline (2014), there is no relationship between researchers' funding and sharing behaviour. By applying the Norm of Reciprocity theory on scientific data sharing, the researcher has noticed that scientists who receive grant funding for research are less likely to share data than those who do not receive funding for research.

Data Sharing Behaviour

Various authors have studied the research data sharing behaviour of scholars and scientists (Silber et al., 2022; Zenk-Möltgen et al., 2018). Studies show that data sharing is influenced by age, gender and personality (Linek et al., 2017). The study further argues that data sharing requires further investment of time and effort from the part of the research scholars to make it available publicly, researchers are not likely to share data unless they are awarded with direct benefits. Anticipated communal benefits such as advancing scientific knowledge or fostering collaboration can significantly enhance a researcher's willingness to share data. In contrast, concerns about increased scrutiny, criticism, or misuse of data often act as powerful deterrents to open data sharing (Barczak et al., 2022).

Studies in the area of scholarly practices had impact on the policy formulation regarding the data repositories (Karcher et al., 2016). Chawinga and Zinn (2019) has identified various challenges in data sharing. Authors concludes that to arrive at proper policies, policy makers should consider the opinion of researchers who are the creators and users of research data. There exists no significant relationship between the number of publications of a researcher and their willingness to share the data (Sherline, 2014). She believes that if the requester is a famous person in the field, the person is more likely to get the data.

Factors Influencing Data Sharing

There are many barriers that hinders the sharing attitude of researchers from an individual point of view. These challenges can be categorised as legal and ethical, cultural, financial, and technical (Chawinga & Zinn, 2019;

Figueiredo, 2017). A growing body of work shows that data sharing is shaped by intertwined individual, institutional, social, and disciplinary forces (Borycz et al., 2023; Tenopir et al., 2020)Studies show that both institutional factors and individual factors play significant role in the data sharing behaviour among research scholars (Kim & Stanton, 2016). To promote data sharing, organisations can educate, motivate and train researchers (Azami et al., 2023). The studies indicate that the practice of Open Access is still not common among the university researchers studied in the UK, France, and Turkey. The significant concerns identified as barriers to sharing research data were related to data ethics and legal issues (Unal et al., 2019).

Study conducted by Zuiderwijk et. al. (2024) proposes and tests a conceptual framework of infrastructural and institutional instruments for open research data in epidemiology, combining a systematic literature review with interviews and a cross-disciplinary workshop. Interviews highlight two very high-influence facilitators of sharing and reuse as access to a powerful open-data search engine and hands-on support from data managers. Yoon and Kim (2020) confirm that attitudes, subjective norms, and perceived behavioural control mediate sharing intentions. Meanwhile, Zuiderwijk et al. (2024) underscore the institutional environment, showing how norms and policies at organisational and national levels can incentivise or obstruct sharing.

Motivation for data sharing may vary based on the individual perspectives and disciplinary culture (Van den Eynden & Bishop, 2014). A study conducted among clinical trialists shows that if shared authorship is provided, trialists would be willing to share their data. Data sharing is influenced by perceived benefits (Borycz et al., 2023b). An initiative called Clinical Study Data Request provides a mechanism where clinical data are stored, anonymised to protect privacy and shared among requested people (Rockhold et al., 2016). Patients are willing to share de-identified data with health institutions more with public institutions and less with private organisations. Though there are agreements to protect the identity of the individual who participates in clinical trials, there exists no organised mechanism to check the violation of the data use agreement of the clinical trial patients (Bierer et al., 2016).

Generally, there are two types of factors that influences data sharing, macro level factors and micro level factors (He, 2016). Different factors including rewards, institutional policies, cultural norms, and technological infrastructure may influence research data sharing behaviour (Ahmed et al., 2025). Age, personality, and gender of the researcher also influence the data sharing behaviour (Linek et al., 2017). In addition to this, data sharing behaviour depends on the type of data shared and to whom the data is shared with especially if the data are confidential in nature (Bull et al., 2015; Niño De Rivera et al., 2024).

Barriers to Data Sharing

Researchers have raised concerns regarding data preservation and sharing. Studies find data sharing an expensive, time-consuming and laborious task. Without necessary infrastructure, domain-expertise and proper data preparation guidelines, data sharing may be a difficult task. Researchers are willing to share if their data are properly safeguarded within a regulated context (Chabilall et al., 2024). Deliberate effort is required to maintain a digital repository (Smith, 2002). Today technology provides opportunities to design and develop sustainable digital data repositories (Anderson, 2004). The size of the datasets and the effort required to share them are major hurdles in data sharing (Chabilall et al., 2024).

Despite the growing policy and infrastructural support for data sharing, a substantial gap persists between research and practice. Lyon et al. (2020) identified a key reason for this disconnect, the lack of formal training among researchers in data stewardship and implementation practices. Data sharing is usually not a part of formal training, so many researchers may be unaware of how to properly share their data (Logan et al., 2021). There is a perceived need for increased training and awareness programmes as found significant by other authors (Rafiq & Ameen, 2022; Tierney & Ram, 2021). Proper training and guidance in documentation are essential (Karcher et al., 2016) and research community are interested in data management training (Mancilla et al., 2019).

Lack of control due to unclear policies, lack of training, or poor technical platforms is a central barrier (Xu & Zhang, 2024). Studies conducted among researchers in the UK, France, and Turkey reported that they have not received training in research data management (Kohrs et al., 2023; Unal et al., 2019). Training on metadata standards could enhance data documentation, supporting data sharing and reproducibility (Kim & Stanton, 2016). Training in data management would be beneficial for their own studies Karcher et al. (2016).

Lack of time, fear of misuse and effort required are often found to be the reason for withholding research data (Gomes et al., 2022). Unlike the traditional publications, datasets have it's own characteristics and need special treatments in data publication, data citation and data licencing. Receiving acknowledgment or citation is a motivator for data sharing (Anger et al., 2022; Hahnel, 2025). Karcher (2016) pointed out the citation behaviour of researchers, even though people use secondary data,only a limited percentage of scholars acknowledges and provide proper data citation. It has been argued in many publications that open access to publications increases the citation counts of articles, enhancing the academic impact of a scholars (Antelman, 2004; Norris, Oppenheim & Rowland, 2008; Gargouri et al. 2010). In view of this argument, many universities have sought to influence the publishing behaviour of academics by enforcing open access mandates. Open access ensure that journal articles, and a range of other research outputs such as conference proceedings, book chapters and doctoral theses are openly available in institutional repositories (Chikoore, 2016).

Bouckenooghe et al. (2023) provide a first-hand account of the tensions that arise when scholars receive unsolicited requests to make their data public. Drawing on attribution theory, the authors argue that fear of reputational damage is a primary obstacle. Researchers often attribute non-sharing to "external constraints," thereby protecting their professional self-image (Houtkoop et al., 2018). A second barrier involves competitive risk, where data are viewed as strategic assets in a publish-or-perish environment. From an institutional perspective, scrutiny from editors, funders, and senior academics—amplifies reluctance, especially when journal policies remain optional or inconsistent. Although researchers acknowledge the collective benefits of data sharing, many still hesitate to release their datasets once a study is complete (Bouckenooghe et al., 2023). Qualitative evidence points to four recurrent fears, exposure to criticism, competitive “scooping,” breaches of confidentiality, and intensified scrutiny by disciplinary gatekeepers that reinforce a culture of data withholding (Houtkoop et al., 2018; Nahai, 2019). These anxieties are especially salient in fields where large, multi-year datasets represent significant intellectual and emotional investments. Findings suggest that although researchers perceive barriers to data sharing, many barriers can be overcome with interventions (Houtkoop et al., 2018).

Apart from these barriers of data sharing, studies argue that all data can not be shared and reused. Qualitative data created in a context can not be analysed out of the context (Chauvette et al., 2019). Studies has exposed challenges faced by the researchers in the area of qualitative research. These challenges can be categorised into three, epistemological, ethical and practical challenges. It consists of availability of repositories in the field of qualitative research. Chauvette argue that analysis of qualitative data is influenced by the ontological and epistemological positions of the researcher.

Disciplinary Differences in Data Sharing

Both general and disciplinary data repositories have its own advantages and disadvantages (Castle, 2019). Various disciplines ask different research questions, use different types of data, and employ varying research techniques (Fulop, 2001). Studies have explored discipline wise variation in research data management and the sharing behaviour (Pepler & Callaghan, 2015; Piwowar, 2010). Nature and characteristics of research data varies according to the discipline (Tam, 2016). Research data sharing has been a practice among scholars in climate and geology disciplines (Thöle & Wegmann, 2024) . In astronomy, researchers depended on data collected by others and thus have a long tradition of sharing (Anderson, 2004). Research data sharing practice varies from discipline to discipline, in areas like astronomy and physics, the sharing practices are high (Pepe et al., 2014). In-depth work in astrophysics shows that motivations to share and reuse are shaped by community-specific infrastructures, conventions around credit, and perceived collective benefit (Zuiderwijk & Spiers, 2019).

Research data management practices of research scholars varies according to the disciplines. Previous studies have reported that there exist disciplinary differences in the data management practices (Akers & Green, 2014). Earlier studies have different opinions about data sharing behaviour of scholars on the basis of their disciplines. Different authors have studied the connection between disciplines and their attitude towards data sharing. According to Sherline (2014), data sharing is dependent on personal attitude and is not connected to disciplines. Citation gains vary across disciplines and depends on the strength of the data availability route. Direct repository deposition showing stronger effects than data made available only upon request, but the overall direction of the effect remains robust. The advantage is attributed to easier verification, reuse, and discovery by wider audiences (Colavizza et al., 2020).

Role of Academic Libraries in Data Sharing

As the academic landscape embraces open science and research data sharing, the role of libraries has undergone significant transformation. Traditionally viewed as custodians of books and bibliographic records, librarians are increasingly emerging as key enablers in the research data lifecycle, especially in areas such as metadata management, data curation, and repository development (Cox et al., 2019; Tenopir et al., 2011). The shift from traditional librarianship to that of a data librarian is not merely functional but a strategic, requiring new competencies and institutional support systems (Fuhr, 2022).

One of the major challenges in the transition is the lack of clearly defined roles and responsibilities for data librarians, especially in developing country contexts (Awada et al., 2022; Patterton et al., 2018). Brown et al. (2015) have highlighted the transitional challenges faced by librarians, emphasising the need for both formal and informal training mechanisms to equip library professionals (Wildgaard & Rantasaari, 2022). Kohr et al. (2023) has suggested strategies to make the reproducible research a norm by providing training. It is argued that librarians' expertise in organising, preserving, and sharing information can be directly extended to research data management, bridging gaps between researchers and repository infrastructure (Hombali, 2022; Joint, 2007). This change recognises that the traditional role of a librarian as bibliographic manager is no longer sufficient in data-centric environments. Librarian's core skills in classification, indexing, and maintenance of information structures become foundational to managing digital datasets and making them reusable.

Various libraries have started moving to take up this new challenge of building a research data repository (Hackett & Kim, 2023). Moreover, developing new skills based on the foundation laid out by the Library and Information Science discipline is also required (Joint, 2007). Scientists may not have the technical skills and domain expertise required to manage the data repositories. Librarians have been doing the job of management and exchange of information in the old science and they will be able to continue the role in the new science also. Different arguments have come up regarding the role librarians can play in the research data management (RDM) (Singh et al., 2022). Some argue that librarians should be given the responsibility of data management while others think that librarian with the support of data administrators and domain experts can only manage data repositories (Ho et al., 2025). The role of a librarian in an academic library can be seen as an extension of the traditional services or as a total transformation of existing skills and structure of the organisation (Cox et al., 2019).

Kouper et al. (2017) has categorised various research data services into basic, intermediate and advanced in the study. The level of maturity attained by an organisation can be analysed only in the context of the institution (Kouper et al., 2017). Some libraries provide the research data services in collaboration with other departments or universities. The major research data services offered by the university libraries are research data introduction, data management guideline, data curation and storage service, data management training, and data management reference and resource recommendation (Si et al., 2015).

Data curation is a challenging job where librarians can help researchers. It is also argued that even researchers agree on the expertise of the librarians in metadata management and data sharing. Hence, librarians are more equipped for the task of metadata related jobs of data management (McLure et al., 2014). In addition to this, the libraries can initiate various other research support services to help researchers. These services include support in the standard file formats, storage and backup, data sharing and data citation (Hackett & Kim, 2023). McLure et al. observed that researchers often rely on librarians for metadata creation, consistency checks, and standards compliance. Massinde et al. (2021) studied academic librarians' experiences and found that they are increasingly involved in RDM activities such as advising on data sharing, organising data, and providing reuse support. McLure et al. (2014) found that researchers themselves acknowledge the value of librarians in supporting metadata management and ensuring compliance with funders' data-sharing mandates.

Libraries across the globe are progressively responding to the demands for initiating institutional repositories and offering dedicated RDM services (Ohaji et al., 2019). These include training sessions for researchers, provision of tools for data citation, support in file format standardisation, backup strategies, and compliance with FAIR data principles (Wilkinson et al., 2016). Librarians also play a pivotal role in providing mediated curation services, handling controlled-access datasets, and managing consent, and licensing templates. As such, they not only support but also shape the data sharing culture in their institutions.

The future of libraries depends not only on technology but also on economic and financial sustainability (Baker & Evans, 2013). As Smith (2002) said, "Digital systems are sprinters whereas history is a marathon runner". Digital contents on the internet disappear and preservation of digital content required deliberate efforts (Dellavalle, 2003). Data generated over years of human effort will not be able to use in the future if conscious efforts are not in place. There is a need for evidence based theoretical framework to implement the data preservation and sharing behaviour in academic institutions (Nilsen, 2020).

Theoretical Frameworks Used in RD Studies

Previous studies have used Technology, Organisation and Environment framework (TOE) and Diffusion of Innovation Theory (DOI) in research data related studies. Crowston and Qin (2011) proposed a Capability Maturity Model (CMM) for Scientific Data Management (SDM), developed through content analysis of literature. The goal of CMM was supporting the assessment and improvement of SDM practices to increase their reliability. The resulting model structures SDM processes into specific areas such as data acquisition, description, dissemination, and preservation. The study characterises organisational capability across five maturity levels, ranging from ad hoc to disciplined and mature. Later study (Qin et al., 2017) refines and expands the 2011 model, primarily by broadening its scope, adding a process area, and introducing formal assessment tools. A maturity model for assessing the library research data services was also developed (Cox et al., 2019) and the study provides a picture of evolution of the research data services in libraries.

Sendra et al. (2025) in their study investigated research data management practices among data intensive social sciences and humanities scholars. The study extended an existing data lifecycle model to incorporate their perceived needs for research data support services. In a study, Nahotko (2023) collected data from library websites to analyse the determination of the data services maturity level. Kim and Stanton (2016) has used theory of planned behaviour and institutional theory to investigate the data sharing behaviour. The study investigated individual factors like perceived benefits and career risk in the institutional context. Studies have attempted to investigate research data sharing behaviour using multiple theoretical frameworks, TPB and Institutional Theory combined effort (Kim & Stanton, 2016) to understand the sharing behaviour of scientists in journals. Borycz et al. (2023b) have combined TPB and Technology Acceptance Model to learn the data sharing behaviour.

Research Data Sharing Policies

Mandated archiving improves the open access (Vines et al., 2013) and open access reviews. Based on the OECD guidelines (Phillips, 2018), various countries started formulating research data management policies (Office of Scholarly Communication, 2019). Research data practices should be guided by policies for repository managers and researchers in collecting and providing access to the data (Awada et al., 2022). Governments, Universities, funding agencies, and journals has policies on research data management. Funding agencies have a major role to play in the sharing attitude (Anger et al., 2022). Attempts have been made to test how policies can influence the research reproducibility (Hardwicke et al., 2018). Universities have created specific policy documents and guidelines for research scholars on how to manage research data. The policies of universities focus on five major elements; access, retention, sharing, storage and ownership (Liu et al., 2020). Institutional differences should be taken care of while preparing policies to suit the context of the institution (Patterton et al., 2018). Studies in the area of scholarly practices had impact on the policy formulation regarding the data repositories (Karcher et al., 2016)

Sharing policies advocate for open data sharing as a default principle in academic institutions. These policies often align with the “as open as possible and as closed as necessary” mantra and the FAIR data principles. For instance, the India Data Accessibility and Use Policy (Government of India, 2012) aims to radically transform public sector data for large scale social transformation and states that, all data should be open by default unless otherwise specified. China's Measures for the Management of Scientific Data (General Office of the State Council., 2018) also promote open sharing of scientific data funded by government budgets, with certain exceptions. Similarly, the Biotech-PRIDE Guidelines (Government of India., 2021) in India seek to maximise the benefits of public investment in biological data generation by facilitating its exchange and sharing through a national repository, particularly for high throughput, high volume data. Academic policies from institutions like the Australian National University, Stanford University, and the University of Melbourne (Liu et al., 2020) also outline requirements for the management of research data and primary materials, emphasising the importance of data management plans from the project's inception.

A consistent theme across these policies is the stringent protection and controlled access for sensitive data, alongside clearly defined responsibilities and legal compliance. The Biotech-PRIDE Guidelines explicitly prohibit the sharing of sensitive data, which includes personal, conservation, or national security-related information. The UDISE Data Sharing Policy in India categorises data like Sensitive Personal Information (SPI), Personally Identifiable Information (PII), and Commercially Sensitive Information (CSI) as sensitive. UDISE policy mandates specify authorisation and Non-Disclosure Agreements for the data access. China's Measures for the Management of Scientific Data strictly prohibit the open sharing of data involving state secrets, national security, public interest, trade secrets, and personal privacy. The Digital Personal Data Protection (DPDP) Act of India, 2023 establishes a comprehensive legal framework for personal data processing. DPDP requires free, specific, and informed consent for data processing and imposing obligations on Data Fiduciaries to implement security safeguards and erase data when no longer needed. All policies, including those from universities like Australian National University and the University of Melbourne, stress compliance with relevant national legislation such as copyright and privacy acts. Guidelines like the Biotech-PRIDE guidelines, also note that national policies supersede international agreements in cases of misalignment.

Various journals have the policy of sharing research data. The majority of leading journals in Information Science and Library Science have a policies regarding the publication of research data. There is a correlation between impact factor and public availability of the research data (Aleixandre-Benavent et al., 2016). The authors benefit from sharing the data by increasing the visibility of the research. Certain studies points out that, though the journals have policies regarding the publication of related data (Savage & Vickers, 2009), most authors are not following the practices of data sharing (Alsheikh-Ali et al., 2011). Many studies have tried to analyse the author's willingness to share the research data (Vickers, 2006) and proposed measures to improve the availability of research data for research. The existence of a journal policy on data sharing doesn't guarantee the sharing of the data by the researchers (Savage & Vickers, 2009). It was found that hosting data on a publisher’s website is still a widespread solution for sharing. Journal websites are widely used by researchers to host the data (Tal-Socher & Ziderman, 2020). Journals follow different data sharing policies those are compulsory and full access, compulsory and partial access, voluntary, upon request and full access (Beugelsdijk et al., 2020)

Open Data Initiatives in India

In India, several recent works suggest that while national policy frameworks and digital public infrastructure have improved, significant gaps remain in translation at the institutional level. For instance, India's health information systems are characterised by isolated data collection, multiple overlapping systems, and limited interoperability. This constrain researchers' ability to reuse or share clinical or programmatic data (Khurana, 2021). At the same time, the move towards Digital Public Infrastructure (DPI), including open government data portals and payment systems, has increased awareness.

India's policy landscape has increasingly embraced open science, especially for publicly funded research. The National Data Sharing and Accessibility Policy (NDSAP) (Government of India, 2012) established the first system-wide mandate for proactive release of 'shareable, non-sensitive government data' to maximise reuse for scientific, economic and social development. NDSAP frames openness, legal conformity, privacy protection and interoperability as governing principles for public data access. Building on this foundation, the Department of Biotechnology's Draft Biological Data Storage, Access and Sharing Policy (2019) set out domain-specific rules for high-throughput biological data, defining data classes, access regimes, timelines for releasing raw and processed datasets, and responsibilities for deposition in recognised repositories. While openness is positioned as a system wide goal, translating the vision into operational repository infrastructure, incentives, and workflows at universities will require sustained effort (Koley, 2022).

These policies situate openness within a responsible sharing paradigm. The FAIR guiding principles emphasise that research outputs should be findable, accessible, interoperable and reusable, offering a widely adopted technical and organisational blueprint for repositories and data producers (Wilkinson et al., 2016). In the human participant research, India's ICMR National Ethical Guidelines (2017) set expectations for informed consent, privacy, and safeguards for secondary use, supporting tiered access models where necessary to protect participants while enabling legitimate reuse. Together, these frameworks articulate both the rationale for open data and, the conditions under which sharing should occur. This include curation standards, metadata, consent language, and controlled access pathways for sensitive datasets. In practice, the effective uptake of these provisions depends on institutional capacity, clear deposit workflows, and incentives.

Research Data Policy Landscape in India

Studies shows that retraction of journal articles due to falsification or fabrication of data are increasing (Anil Kumar & Siwach, 2023). Though open data repositories in India are not fully open as per the definitions, more open data repositories are coming up (Anilkumar, 2018; Misgar et al., 2020). The study to evaluate the openness and reusability of the data repositories using Tim Berners-Lee's 5-star model shows that majority of the data provided are not in a reusable format (Jamsheer & Haneefa, 2024). Although the University Grants Commission (UGC) mandates depositing PhD theses in the Shodhganga repository, no corresponding policy exists in the universities for preserving and sharing research data created during PhD research work (Kumar et al., 2025).

India's move towards systematic research data sharing initiative began with the National Data Sharing and Accessibility Policy (2012). NDSAP mandates government departments to release shareable datasets created with public funds, thereby establishing the first national commitment to open data. The Department of Biotechnology (DBT) and the Department of Science and Technology (DST) of India introduced the DBT–DST Open Access Policy in 2014. This policy mandates that all researchers receiving funding from either agency are required to deposit the final accepted version of their manuscripts, along with associated metadata, in institutional or central open access repositories within a stipulated period post-publication. The aim of this directive is to enhance the visibility, accessibility, and impact of Indian scientific output, aligning with broader international trends toward open science and FAIR principles (OpenAIRE, 2020; UNESCO, 2021).

However, while the policy reflects a significant normative shift toward openness in the Indian research ecosystem, formal mechanisms for compliance monitoring and enforcement remain underdeveloped. Recent analysis underscores the inconsistent implementation of open access mandates across Indian institutions (Kumar et al., 2025), often due to a lack of institutional infrastructure, awareness among researchers, and standardised procedures. Without a robust infrastructure for research data management and sustained advocacy, such policies risk remaining aspirational. The effective implementation requires strategic investments in repository infrastructure, researcher training, and institutional mandates.

The Draft Biological Data Storage, Access and Sharing Policy 2019, intended to govern datasets produced by publicly funded Life Science projects. Complementing these policy instruments is a growing ecosystem of publicly financed infrastructures, the iSTEM portal and SATHI programme facilitate shared use of high-end research equipment across institutions, while digital platforms such as NPTEL, the National Digital Library of India (NDLI), and Shodhganga provide open educational resources and thesis repositories that model good curation practice for future data archives. At the same time, sectoral initiatives demonstrate feasibility and value. For example, biodiversity programs have advanced norms for data deposit, curation, and reuse in international venues (Talukdar et al., 2019). Parallel developments in open government data also indicate capacity for public-facing portals and stewardship of large datasets (Chakravarty, 2018).

The draft Science, Technology and Innovation Policy 2020 (STIP 2020) seeks to harmonise these efforts through open data policy for all publicly funded research. To operationalise this mandate, STIP 2020 proposes a national portal, the Indian Science and Technology Archive of Research (INDSTA) designed to store and visualise research outputs at scale. Supporting measures include the Scientific Research Infrastructure for Maintenance and Networks (SRIMAN) Guidelines 2022, which formalise principles for sharing physical and digital research assets. The Scientific Social Responsibility (SSR) guidelines 2022 frames data sharing as an ethical duty to society. Together, these initiatives signal a policy shift from isolated data mandates to an integrated national framework that couples regulatory directives with infrastructure and incentive structures.

A national study of top-ranked higher education institutions reports that RDM practices are still at an emergent stage, with uneven provision of repositories, policies, and services across campuses (Bhoi et al., 2023). Policy language around RDM and data sharing is fragmented and inconsistently communicated, leading to low clarity and variable compliance among investigators (T R & Gala, 2020). These findings are consistent with earlier Indian work showing that libraries and research offices are only beginning to formalise RDM roles and that institutional repositories often focus on theses and publications rather than well-curated datasets (Anilkumar, 2018). Despite these policy milestones, several challenges persist. Compliance remains inconsistent due to lack of awareness among researchers, insufficient repository infrastructure, and absence of incentives or penalties. Balancing intellectual property rights with open access continues to be a contested area. As recent global discussions emphasise, the success of open data policies depends not only on mandates but also on researcher training, institutional support, and recognition mechanisms (Yoon, 2017; Zuiderwijk et al., 2024).

Research Gap

Despite increasing global attention toward open science and research data management, studies in the Indian context, particularly in Kerala remain scarce. Most prior investigations into research data practices in India have focused on policy frameworks and infrastructural readiness (Koley, 2022). But little empirical evidence exists regarding how researchers themselves perceive and engage with data sharing. Moreover, institutional studies often highlight technical and policy-level barriers but overlook the perspectives of early-career scholars and supervisors, who are crucial actors in the research ecosystem. This lack of contextualised evidence limits the formulation of practical and culturally appropriate strategies for data sharing in Indian universities.

Existing research also shows an imbalance between the emphasis placed on the potential benefits of open data and the realities of researcher reluctance. Studies from advanced economies highlight factors such as data citation, recognition, and professional incentives as motivators. Whereas Indian researchers continue to operate in environments where such mechanisms are either absent or weakly enforced. As a result, the broader discussion on data sharing has not yet translated into actionable practices within academic institutions in Kerala. The present study addresses this empirical gap by systematically examining the attitudes, motivations, and barriers experienced by doctoral students and research supervisors across multiple disciplines.

There is a significant theoretical gap in how research data sharing is studied in India. International scholarship has applied behavioural frameworks such as TPB to explore researcher attitudes (J. Kim et al., 2017; Zuiderwijk et al., 2024). Institutional perspectives have been examined through governance and policy approaches (Scott, 2014). However, few studies have attempted to integrate these two perspectives. In the Indian context, there has been no systematic application of TPB or Institutional Theory to examine research data sharing. This omission is critical because the interplay between individual intention and institutional structures remains under explored.

Chapter Summary

This chapter synthesises scholarship on research data and research-data sharing, with a focus on publicly funded research and university contexts in Kerala. A growing literature shows that researchers' intentions to share research data are shaped by both individual factors and institutional arrangements. Majority of the literature concentrates on the benefits of the RD, how it is implemented but fails to articulate the future research and how the practice can be implemented in the developing countries with lesser resources.

References

Ahmed, M., Othman, R., Noordin, M. F., Ibrahim, A. A., & Al-Hussaini, A. I. S. (2025). Factors influencing open science participation through research data sharing and reuse among researchers: A systematic literature review. *Knowledge and Information Systems*, *67*(3), 2801–2853. https://doi.org/10.1007/s10115-024-02284-3

Akers, K. G., & Green, J. A. (2014). Towards a symbiotic relationship between academic libraries and disciplinary data repositories: A dryad and university of michigan case study. *International Journal of Digital Curation*, *9*(1), 119–131. https://doi.org/10.2218/ijdc.v9i1.306

Aleixandre-Benavent, R., Moreno-Solano, L. M., Ferrer Sapena, A., & Pérez, E. A. S. (2016). Correlation between impact factor and public availability of published research data in Information Science and Library Science journals. *Scientometrics*, *107*(1), 1–13. https://doi.org/10.1007/s11192-016-1868-7

Alsheikh-Ali, A. A., Qureshi, W., Al-Mallah, M. H., & Ioannidis, J. P. A. (2011). Public availability of published research data in high-impact journals. *PLOS ONE*, *6*(9), e24357. https://doi.org/10.1371/journal.pone.0024357

Anderson, W. L. (2004). Some challenges and issues in managing, and preserving access to, long-lived collections of digital scientific and technical data. *Data Science Journal*, *3*, 191–201. https://doi.org/10.2481/dsj.3.191

Anger, M., Wendelborn, C., Winkler, E. C., & Schickhardt, C. (2022). Neither carrots nor sticks? Challenges surrounding data sharing from the perspective of research funding agencies—a qualitative expert interview study. *PLOS ONE*, *17*(9), e0273259. https://doi.org/10.1371/journal.pone.0273259

Anil Kumar, A. K., & Siwach, S. (2023). Analysis of retracted publications by indian authors due to falsification or fabrication of data. *KELPRO Bulletin*, *27*(2).

Anilkumar, N. (2018). Research data management in india: A pilot study. *EPJ Web of Conferences*, *186*, 03002. https://doi.org/10.1051/epjconf/201818603002

Awada, L., Phillips, P. W. B., & Bogdan, A. M. (2022). Governance and stewardship for research data and information sharing: Issues and prospective solutions in the transdisciplinary plant phenotyping and imaging research center network. *PLANTS, PEOPLE, PLANET*, *4*(1), 84–95. https://doi.org/10.1002/ppp3.10238

Azami, M., Sadatmoosavi, A., & Chashmyazdan, M. (2023). Research data management frameworks: A systematic literature review. *International Journal of Information Science and Management (IJISM)*, *21*(3), 1–18. https://doi.org/10.22034/ijism.2023.1977759.0

Baker, D., & Evans, W. (Eds.). (2013). *A handbook of digital library economics: Operations, collections and services*. Chandos Publishing.

Barczak, G., Hopp, C., Kaminski, J., Piller, F., & Pruschak, G. (2022). How open is innovation research? – An empirical analysis of data sharing among innovation scholars. *Industry and Innovation*, *29*(2), 186–218. https://doi.org/10.1080/13662716.2021.1967727

Beugelsdijk, S., Van Witteloostuijn, A., & Meyer, K. E. (2020). A new approach to data access and research transparency (DART). *Journal of International Business Studies*, *51*(6), 887–905. https://doi.org/10.1057/s41267-020-00323-z

Bhoi, N. K., Patel, J., & Dutta, B. (2023). State of research data management practices in the top-ranked higher education institutions in India. *The International Information & Library Review*, *55*(4), 1–19. https://doi.org/10.1080/10572317.2023.2167051

Bierer, B. E., Li, R., Barnes, M., & Sim, I. (2016). A global, neutral platform for sharing trial data. *New England Journal of Medicine*, *374*(25), 2411–2413. https://doi.org/10.1056/NEJMp1605348

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

Borgman, C. L. (2017). *Big data, little data, no data: Scholarship in the networked world*. The MIT Press.

Bornmann, L. (2013). What is societal impact of research and how can it be assessed? A literature survey. *Journal of the American Society for Information Science and Technology*, *64*(2), 217–233. https://doi.org/10.1002/asi.22803

Borycz, J., Olendorf, R. 2, Specht, A. 3, Grant, B. 4, Crowston, K. 5, Tenopir, C. 6, Allard, S. 6, Rice, N. M. 6, Hu, R. 7, & Sandusky, R. J. 8 1 V. U. (2023a). *Perceived benefits of open data are improving but scientists still lack* *resources, skills, and rewards*. 339. https://doi.org/10.1057/s41599-023-01831-7

Borycz, J., Olendorf, R., Specht, A., Grant, B., Crowston, K., Tenopir, C., Allard, S., Rice, N. M., Hu, R., & Sandusky, R. J. (2023b). Perceived benefits of open data are improving but scientists still lack resources, skills, and rewards. *Humanities and Social Sciences Communications*, *10*(1), 339. https://doi.org/10.1057/s41599-023-01831-7

Brown, R. A., Wolski, M., & Richardson, J. (2015). Developing new skills for research support librarians. *The Australian Library Journal*, *64*(3), 224–234. https://doi.org/10.1080/00049670.2015.1041215

Bull, S., Cheah, P. Y., Denny, S., Jao, I., Marsh, V., Merson, L., Shah More, N., Nhan, L. N. T., Osrin, D., Tangseefa, D., Wassenaar, D., & Parker, M. (2015). Best practices for ethical sharing of individual-level health research data from low- and middle-income settings. *Journal of Empirical Research on Human Research Ethics*, *10*(3), 302–313. https://doi.org/10.1177/1556264615594606

Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal*, *14*(0), 2. https://doi.org/10.5334/dsj-2015-002

Castle, C. (2019). Getting the central RDM message across: A case study of central versus discipline-specific research data services (RDS) at the university of cambridge. *Libri*, *69*(2), 105–116. https://doi.org/10.1515/libri-2018-0064

Chabilall, J., Brown, Q., Cengiz, N., & Moodley, K. (2024). Data as scientific currency: Challenges experienced by researchers with sharing health data in sub-Saharan Africa. *PLOS Digital Health*, *3*(10), e0000635. https://doi.org/10.1371/journal.pdig.0000635

Chakravarty, R. (2018). Open government data (OGD) initiative in India: An empirical analysis. In *Proliferation of Open Government Initiatives and Systems* (pp. 68–90). IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-5225-4987-1.ch004

Chauvette, A., Schick-Makaroff, K., & Molzahn, A. E. (2019). Open data in qualitative research. *International Journal of Qualitative Methods*, *18*, 1609406918823863. https://doi.org/10.1177/1609406918823863

Chawinga, W. D., & Zinn, S. (2019). Global perspectives of research data sharing: A systematic literature review. *Library & Information Science Research, 41*(2), 109–122. https://doi.org/10.1016/j.lisr.2019.04.004

Chikoore, L. (2016). *Perceptions, motivations and behaviours towards "research impact": A cross-disciplinary perspective* [Thesis, Loughborough University]. https://hdl.handle.net/2134/22942

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

Cox, A. M., Kennan, M. A., Lyon, L., Pinfield, S., & Sbaffi, L. (2019). Maturing research data services and the transformation of academic libraries. *Journal of Documentation*, *75*(6), 1432–1462. https://doi.org/10.1108/JD-12-2018-0211

Crowston, K., & Qin, J. (2011). A capability maturity model for scientific data management: Evidence from the literature. *Proceedings of the American Society for Information Science and Technology*, *48*(1), 1–9. https://doi.org/10.1002/meet.2011.14504801036

De Lusignan, S., Liaw, S.-T., Krause, P., Curcin, V., Vicente, M. T., Michalakidis, G., Agreus, L., Leysen, P., Shaw, N., & Mendis, K. (2011). Key concepts to assess the readiness of data for International research: Data quality, lineage and provenance, extraction and processing errors, traceability, and curation. *Yearbook of Medical Informatics*, *20*(01), 112–120.

Dorta-González, P., Rodríguez-Caro, A., & Dorta-González, M. I. (2024). Societal and scientific impact of policy research: A large-scale empirical study of some explanatory factors using Altmetric and Overton. *Journal of Informetrics*, *18*(3), 101530. https://doi.org/10.1016/j.joi.2024.101530

Downs, R. R. (2021). Improving opportunities for new value of open data: Assessing and certifying research data repositories. *Data Science Journal, 20*(1). https://doi.org/10.5334/dsj-2021-001

Figueiredo, A. S. (2017). Data Sharing: Convert Challenges into Opportunities. *Frontiers in Public Health*, *5*. https://doi.org/10.3389/fpubh.2017.00327

Fuhr, J. (2022). Developing data services skills in academic libraries. *College & Research Libraries*, *83*(3). https://doi.org/10.5860/crl.83.3.474

General Office of the State Council. (2018). *Measures for the management of scientific data.* https://cset.georgetown.edu/wp-content/uploads/t0550_scientific_data_EN.pdf

Gomes, D. G. E., Pottier, P., Crystal-Ornelas, R., Hudgins, E. J., Foroughirad, V., Sánchez-Reyes, L. L., Turba, R., Martinez, P. A., Moreau, D., Bertram, M. G., Smout, C. A., & Gaynor, K. M. (2022). Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1987), 20221113. https://doi.org/10.1098/rspb.2022.1113

Government of India. (2012). *National data sharing and accessibility policy (NDSAP)*. https://dst.gov.in/national-data-sharing-and-accessibility-policy-0

Government of India. (2021). *Biotech-PRIDE: promotion of research and innovation through data exchange guidelines.* https://dbtindia.gov.in/sites/default/files/uploadfiles/Biotech%20PRIDE%20Guidelines.pdf

Hackett, C., & Kim, J. (2023). Planning, implementing and evaluating research data services in academic libraries: A model approach. *Journal of Documentation*, *80*(1), 27–38. https://doi.org/10.1108/JD-01-2023-0007

Hahnel, M. (2025, January 20). Are we ready to start rewarding researchers for open data? *Figshare*. https://info.figshare.com/are-we-ready-to-start-rewarding-researchers-for-open-data/

Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., Hofelich Mohr, A., Clayton, E., Yoon, E. J., Henry Tessler, M., Lenne, R. L., Altman, S., Long, B., & Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal *Cognition*. *Royal Society Open Science*, *5*(8), 180448. https://doi.org/10.1098/rsos.180448

He, Y. (2016). *Data sharing across research and public communities* [Digital Repository at the University of Maryland]. https://doi.org/10.13016/M2Z84M

Heeks, R., Rakesh, V., Sengupta, R., Chattapadhyay, S., & Foster, C. (2021). Datafication, value and power in developing countries: Big data in two Indian public service organizations. *Development Policy Review*, *39*(1), 82–102. https://doi.org/10.1111/dpr.12477

Ho, R. C. Y., Wong, S. N., Chia, P., Tang, C., & Ng, M. T. T. (2025). Research data management services in academic libraries to support the research data life cycle: A systematic review. *Journal of the Association for Information Science and Technology*, asi.70008. https://doi.org/10.1002/asi.70008

Hombali, P. (2022). *The role of librarians in research data management: Challenges and opportunities* (SSRN Scholarly Paper No. 4422179). Social Science Research Network. https://papers.ssrn.com/abstract=4422179

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V. M., 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

Jamsheer, N. P., & Haneefa, K. M. (2024). Measuring openness and reusability of selected Indian data repositories. *KELPRO Bulletin*, *28*(2). https://kelprobulletin.in

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, *29*(4), 258–268. https://doi.org/10.1080/10580530.2012.716740

Jetzek, T., Avital, M., & Bjorn-Andersen, N. (2019). The sustainable value of open government data. *Journal of the Association for Information Systems*, *20*(6). https://doi.org/10.17705/1jais.00549

Joint, N. (2007). Data preservation, the new science and the practitioner librarian. *Library Review*, *56*(6), 451–455. https://doi.org/10.1108/00242530710760337

Karcher, S., Kirilova, D., & Weber, N. (2016). Beyond the matrix: Repository services for qualitative data. *IFLA Journal*, *42*(4), 292–302. https://doi.org/10.1177/0340035216672870

Khurana, N. (2021). Issue analysis: A use-driven approach to data governance can promote the quality of routine health data in india. *Global Health, Science and Practice*, *9*(2), 238–245. https://doi.org/10.9745/GHSP-D-20-00347

Kim, J., Schuler, E. R., & Pechenina, A. (2017). Predictors of data sharing and reuse behavior in academic communities. *Knowledge Discovery and Data Design Innovation*, 1–25. https://doi.org/10.1142/9789813234482_0001

Kim, Y., & Stanton, J. M. (2016). Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*, *67*(4), 776–799. https://doi.org/10.1002/asi.23424

Kohrs, F. E., Auer, S., Bannach-Brown, A., Fiedler, S., Haven, T. L., Heise, V., Holman, C., Azevedo, F., Bernard, R., Bleier, A., Bössel, N., Cahill, B. P., Castro, L. J., Ehrenhofer, A., Eichel, K., Frank, M., Frick, C., Friese, M., Gärtner, A., … Weissgerber, T. L. (2023). Eleven strategies for making reproducible research and open science training the norm at research institutions. *eLife*, *12*, e89736. https://doi.org/10.7554/eLife.89736

Koley, M. (2022). Analysis of Open Science Policy Recommendations Proposed in India's 5th Science, Technology & Innovation Policy Draft. *Journal of Science Policy & Governance*, *21*(02). https://doi.org/10.38126/jspg210208

Kouper, I., Fear, K., Ishida, M., Kollen, C., & Williams, S. C. (2017). *Research Data Services Maturity in Academic Libraries*. University of British Columbia. https://doi.org/10.14288/1.0343479

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. https://doi.org/10.47909/ijsmc.192

Linek, S. B., Fecher, B., Friesike, S., & Hebing, M. (2017). Data sharing as social dilemma: Influence of the researcher's personality. *PLOS ONE*, *12*(8), e0183216. https://doi.org/10.1371/journal.pone.0183216

Liu, G., Zotoo, I. K., & Su, W. (2020). Research data management policies in USA, UK and Australia universities: An online survey. *Malaysian Journal of Library and Information Science*, *25*(2), Article 2. https://doi.org/10.22452/mjlis.vol25no2.2

Lo, B., & DeMets, D. L. (2016). Incentives for clinical trialists to share data. *New England Journal of Medicine*, *375*(12), 1112–1115. https://doi.org/10.1056/NEJMp1608351

Logan, J. A. R., Hart, S. A., & Schatschneider, C. (2021). Data Sharing in Education Science. *AERA Open*, *7*, 23328584211006475. https://doi.org/10.1177/23328584211006475

Mancilla, H. A., Teperek, M., Dijck, J. van, Heijer, K. den, Eggermont, R., Plomp, E., Velden, Y. T. der, & Kurapati, S. (2019). On a quest for cultural change —Surveying research data management practices at delft university of technology. *LIBER Quarterly: The Journal of the Association of European Research Libraries*, *29*(1), 1–27. https://doi.org/10.18352/lq.10287

Masinde, J., Chen, J., Wambiri, D., & Mumo, A. (2021). Research librarians' experiences of research data management activities at an academic library in a developing country. *Data and Information Management*, *5*(4), 412–424. https://doi.org/10.2478/dim-2021-0002

McLure, M., Level, A. V., Cranston, C. L., Oehlerts, B., & Culbertson, M. (2014). Data curation: A study of researcher practices and needs. *Portal: Libraries and the Academy*, *14*(2), 139–164. https://doi.org/10.1353/pla.2014.0009

Misgar, S. M., Bhat, A., & Wani, Z. A. (2020). A study of open access research data repositories developed by BRICS countries. *Digital Library Perspectives*, *38*(1), 45–54. https://doi.org/10.1108/DLP-02-2020-0012

Nahotko, M., Zych, M., Januszko-Szakiel, A., & Jaskowska, M. (2023). Big data-driven investigation into the maturity of library research data services (RDS). *The Journal of Academic Librarianship*, *49*(1), 102646. https://doi.org/10.1016/j.acalib.2022.102646

Nilsen, P. (2020). Making sense of implementation theories, models, and frameworks. In B. Albers, A. Shlonsky, & R. Mildon (Eds.), *Implementation Science 3.0* (pp. 53–79). Springer International Publishing. https://doi.org/10.1007/978-3-030-03874-8_3

Niño De Rivera, S., Masterson Creber, R., Zhao, Y., Eslami, S., Mangal, S., Dugdale, L. S., & Reading Turchioe, M. (2024). Public perspectives on increased data sharing in health research in the context of the 2023 National Institutes of Health Data Sharing Policy. *PLOS ONE*, *19*(8), e0309161. https://doi.org/10.1371/journal.pone.0309161

Numajiri, H., & Hayashi, T. (2024). Analysis on open data as a foundation for data-driven research. *Scientometrics*. https://doi.org/10.1007/s11192-024-04956-x

Office of Scholarly Communication. (2019). *University of cambridge research data management policy framework*. Apollo - University of Cambridge Repository. https://doi.org/10.17863/CAM.10528

Ohaji, I. K., Chawner, B., & Yoong, P. (2019, December 15). *The role of a data librarian in academic and research libraries* [Text]. University of Borås. https://informationr.net/ir/24-4/paper844.html

Patterton, L., Bothma, T., & Van Deventer, M. (2018). From planning to practice: An action plan for the implementation of research data management services in resource-constrained institutions. *South African Journal of Libraries and Information Science*, *84*(2). https://doi.org/10.7553/84-2-1761

Pepe, A., Goodman, A., Muench, A., Crosas, M., & Erdmann, C. (2014). How Do Astronomers Share Data? Reliability and Persistence of Datasets Linked in AAS Publications and a Qualitative Study of Data Practices among US Astronomers. *PLOS ONE*, *9*(8), e104798. https://doi.org/10.1371/journal.pone.0104798

Pepler, S., & Callaghan, S. (2015). Twenty Years of Data Management in the British Atmospheric Data Centre. *International Journal of Digital Curation*, *10*(2), 23–32. https://doi.org/10.2218/ijdc.v10i2.379

Perelman, M. (2002). *Steal this idea: Intellectual property and the corporate confiscation of creativity*. Palgrave Macmillan US. https://www.palgrave.com/gp/book/9781403967138

Phillips, M. (2018). International data-sharing norms: From the OECD to the General Data Protection Regulation (GDPR). *Human Genetics*, *137*(8), 575–582. https://doi.org/10.1007/s00439-018-1919-7

Piwowar, H. A. (2010). *Foundational studies for measuring the impact, prevalence, and patterns of publicly sharing biomedical research data* [University of Pittsburgh]. http://d-scholarship.pitt.edu/7204/

Piwowar, H. A. (2011). Who shares? Who doesn't? Factors associated with openly archiving raw research data. *PLoS ONE*, *6*(7), e18657. https://doi.org/10.1371/journal.pone.0018657

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, *1*, e175. https://doi.org/10.7717/peerj.175

Qin, J., Crowston, K., & Kirkland, A. (2017). Pursuing best performance in research data management by using the capability maturity model and rubrics. *Journal of eScience Librarianship*, *6*(2). https://doi.org/10.7191/jeslib.2017.1113

Rafiq, M., & Ameen, K. (2022). Research data management and sharing awareness, attitude, and behavior of academic researchers. *Information Development*, *38*(3), 391–405. https://doi.org/10.1177/02666669211048491

Robinson-García, N., Jiménez-Contreras, E., & Torres-Salinas, D. (2016). Analyzing data citation practices using the data citation index. *Journal of the Association for Information Science and Technology*, *67*(12), 2964–2975. https://doi.org/10.1002/asi.23529

Rockhold, F., Nisen, P., & Freeman, A. (2016). Data sharing at a crossroads. *New England Journal of Medicine*, *375*(12), 1115–1117. https://doi.org/10.1056/NEJMp1608086

Savage, C. J., & Vickers, A. J. (2009). Empirical study of data sharing by authors publishing in PLoS journals. *PLOS ONE*, *4*(9), e7078. https://doi.org/10.1371/journal.pone.0007078

Scott, W. R. (2014). *Institutions and organizations: Ideas, interests and identities* (4th ed.). Sage.

Sendra, A., Late, E., & Kumpulainen, S. (2025). From data lifecycle to research activity model: Research data management in data-intensive social sciences and humanities research. *Aslib Journal of Information Management*. https://doi.org/10.1108/AJIM-12-2024-0959

Sherline, C. (2014). *Measuring the Norm of Reciprocity on Data Sharing Practices: A Carrot or Stick Approach?* [Thesis, University of Tennessee]. https://trace.tennessee.edu/utk_graddiss/3164

Si, L., Xing, W., Zhuang, X., Hua, X., & Zhou, L. (2015). Investigation and analysis of research data services in university libraries. *The Electronic Library*, *33*(3), 417–449. https://doi.org/10.1108/EL-07-2013-0130

Silber, H., Gerdon, F., Bach, R., Kern, C., Keusch, F., & Kreuter, F. (2022). A preregistered vignette experiment on determinants of health data sharing behavior: Willingness to donate sensor data, medical records, and biomarkers. *Politics and the Life Sciences*, *41*(2), 161–181. https://doi.org/10.1017/pls.2022.15

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. https://doi.org/10.1016/j.acalib.2022.102586

Smith, B. (2002). Preserving tomorrow's memory: Preserving digital content for future generations 1. *Information Services & Use*, *22*(2–3), 133–139. https://doi.org/10.3233/ISU-2002-222-313

Spiegelman, E. (2021). Esteemed colleagues: A model of the effect of open data on selective reporting of scientific results. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.761168

T R, M., & Gala, B. (2020). Research data management: An evaluation of Indian research funding agencies' policies and guidelines. *LIBRARY HERALD*, *58*(4), 42–65. https://doi.org/10.5958/0976-2469.2020.00030.9

Tal-Socher, M., & Ziderman, A. (2020). Data sharing policies in scholarly publications: Interdisciplinary comparisons. *Prometheus*, *36*(2), 116–134.

Talukdar, G., Townsend Peterson, A., & Mathur, V. (2019). Biodiversity data sharing by public funded institutions: Perspectives from India. *Biodiversity Information Science and Standards*, *3*, e37219. https://doi.org/10.3897/biss.3.37219

Tam, W. T. (2016). *Discipline and research data in geography* [Thesis, Loughborough University]. https://repository.lboro.ac.uk/articles/thesis/Discipline_and_research_data_in_geography/9496394

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLOS ONE*, *6*(6), e21101. https://doi.org/10.1371/journal.pone.0021101

Thöle, L. M., & Wegmann, M. (2024). Open climate science is brave climate science. *Environmental Research Letters*, *19*(12), 122001. https://doi.org/10.1088/1748-9326/ad893f

Tierney, N. J., & Ram, K. (2021). Common-sense approaches to sharing tabular data alongside publication. *Patterns*, *2*(12), 100368. https://doi.org/10.1016/j.patter.2021.100368

Unal, Y., Chowdhury, G., Kurbanoğlu, S., Boustany, J., & Walton, G. (2019). Research data management and data sharing behaviour of university researchers'. *Information Research: An International Electronic Journal*, *24*, Article 1. http://informationr.net/ir/24-1/isic2018/isic1818.html

UNESCO & Canadian National Commission for UNESCO. (2022). *An introduction to the UNESCO Recommendation on Open Science*. UNESCO. https://doi.org/10.54677/XOIR1696

Van den Eynden, V., & Bishop, L. (2014). *Sowing the seed: Incentives and motivations for sharing research data, a researcher's perspective* [Publication]. https://repository.jisc.ac.uk/5662/

Vickers, A. J. (2006). Whose data set is it anyway? Sharing raw data from randomized trials. *Trials*, *7*(1), 15. https://doi.org/10.1186/1745-6215-7-15

Vines, T. H., Andrew, R. L., Bock, D. G., Franklin, M. T., Gilbert, K. J., Kane, N. C., Moore, J.-S., Moyers, B. T., Renaut, S., Rennison, D. J., Veen, T., & Yeaman, S. (2013). Mandated data archiving greatly improves access to research data. *The FASEB Journal*, *27*(4), 1304–1308. https://doi.org/10.1096/fj.12-218164

Wildgaard, L., & Rantasaari, J. (2022). *RDA Professionalising Data Stewardship—Data Stewardship Landscape Initial Report*. Zenodo. https://doi.org/10.15497/RDA00076

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Xu, J., & Zhang, R. (2024). Attitudinal, normative, and resource factors affecting communication scholars' data sharing: A replication study. *Media and Communication*, *12*(0). https://doi.org/10.17645/mac.7666

Yoon, A. (2017). Data reusers' trust development. *Journal of the Association for Information Science and Technology*, *68*(4), 946–956. https://doi.org/10.1002/asi.23730

Yoon, A., & Kim, Y. (2020). The role of data-reuse experience in biological scientists' data sharing: An empirical analysis. *The Electronic Library*, *38*(1), 186–208. https://doi.org/10.1108/EL-06-2019-0146

Zenk-Möltgen, W., Akdeniz, E., Katsanidou, A., Naßhoven, V., & Balaban, E. (2018). Factors influencing the data sharing behavior of researchers in sociology and political science. *Journal of Documentation, 74*(5), 1053–1073. https://doi.org/10.1108/JD-09-2017-0126

Zuiderwijk, A. (2024). Researchers' willingness and ability to openly share their research data: A survey of COVID-19 pandemic-related factors. *SAGE Open*, *14*(1). https://doi.org/10.1177/21582440241234985

Zuiderwijk, A., & Spiers, H. (2019). Sharing and re-using open data: A case study of motivations in astrophysics. *International Journal of Information Management*, *49*, 228–241. https://doi.org/10.1016/j.ijinfomgt.2019.05.024

Zuiderwijk, A., Türk, B. O., & Brazier, F. (2024). Identifying the most important facilitators of open research data sharing and reuse in epidemiology: A mixed-methods study. *PLoS ONE*, *19*(2). https://doi.org/10.1371/journal.pone.0297969

Chapter 3. Methodology

The methodology chapter is dedicated for explaining the methodology, ontology, and epistemology adopted for the study. Methodology refers to the systematic study of research methods. It involves outlining the logical plan or design that links the initial research questions to the conclusions reached at the end of the study (Yin, 2003). This chapter provides the conceptual and logical rationale for the overall approach chosen, detailing the decisions and choices made about methods and design. A detailed discussion of the rationale for using a mixed methodology is provided in the chapter. The chapter outlines the practical research procedures. It explains how the questionnaire and interview schedule were developed, describes the data collection process in detail, and discusses how the pilot study was carried out. It then highlights the tools and procedures used for analysing the data. By documenting these elements systematically, the methodology chapter enhances the transparency of the research process (Rossman & Rallis, 2017). The chapter concludes with discussion of the steps taken to ensure credibility and trustworthiness in both phases of the study.

Research Philosophy

Research philosophy refers to the underlying beliefs and assumptions that guide investigators in the research work. These perspectives are primarily built on ontology and epistemology. Ontology explores the nature of existence and reality. Epistemology examines what counts as valid or warranted knowledge. It is important to clearly state these beliefs as it affect the research questions, the methods chosen to collect data, and how results are interpreted.

Epistemological Position

The epistemology of the study, the criteria by which knowledge claims are warranted is defined by its methodological pluralism. Study primarily align with Pragmatism, incorporating elements of both post-positivism and interpretivism to address the research problem comprehensively. This philosophy typically supports mixed methods approaches and focuses on utility. The study aims to provide insights and recommendations for policymakers and stakeholders. This emphasis on producing practical, usable knowledge that addresses a real world problem reflects a pragmatic orientation (Glasgow, 2013).

The first phase of the research structure incorporates quantitative elements, suggesting a leaning toward the postpositivist world view for parts of the inquiry. The post-positivism assumes that an objective reality exists ‘out there’ but recognises that human knowledge is conjectural and that claims cannot be held as absolute truth (Creswell, 2014). The study uses TPB, which is a predictive model that necessitates identifying and assessing variables like attitudes, norms, and control that influence outcomes. This requires observation, measurement, and the testing or verification of theoretical propositions, which are hallmarks of the postpositivist tradition. The interview schedule is used in the second phase suggest an interpretive stance, which views knowledge as a social construction. The qualitative method relies on understanding the subjective meanings, perspectives, and experiences of the participants (Wildemuth, 2016). The aim is to assess institutional research culture prevailing and recognising multiple viewpoints.

Ontological Position

Ontology is a branch of philosophy dealing with the essence of phenomena and the nature of their existence. It involves deeply embedded assumptions about the nature of the social world and what is real (Symon & Cassell, 2012). Ontological questions inquire into the nature of existence. Ontology addresses assumptions about the nature of existence, what is real and, concerns whether phenomena exist independent of human perception (realist/ objectivist) or are socially created (relativist/ subjectivist).

The study assumes the existence of certain external, objective elements concerning structure and observable practice. Seeking generalisable patterns by measuring variables is adopted in the first phase by collecting and analysing data aligning with objectivist assumptions. It assumes that institutional structures and practices can be examined, assessed, and improved in concrete ways (Ellis, 1993). This aligns with a realist perspective that assumes a single reality exists independent of any single observer. The use of Institutional Theory implies a reality where formal structures, regulative rules, and organisational contexts which are the objective of the study exist and can be investigated objectively (Yin, 2018).

Research data practices, motivating and challenging factors are inherently subjective and socially determined. The study investigate behaviour and attitudes of researchers, that are influenced by individual perceptions and human interactions. This requires acknowledging multiple individually constructed viewpoints and recognising that reality is partially understood through these constructions. Seeking contextual uniqueness and socially constructed meanings are part of the qualitative phase. This position, referred to as relativist or constructivist ontology, focuses on how different meanings illuminate the topic of study (Yin, 2018).

Pragmatism

The study adopts a pragmatic research philosophy and draws on both objectivist and critical-realist ideas. The pragmatic approach, deeply relevant to research, particularly in guiding methodological choices and focusing on practical outcomes (Creswell, 2014). The focus is on the research problem and, investigators use all available approaches to understand the research questions. Morgan (2014) and Kaushik and Walsh (2019) argue that pragmatism suited to mixed methods research as it allows integration of diverse approaches to produce practical, problem-centred knowledge. Pragmatism aligns with determining if a research question has significance for an organisation or can lead to tangible situation improvement (Wildemuth, 2016).

The complexity of data sharing behaviour, which involves technological, ethical, societal, and organisational aspects cannot be captured by a single method. The converged findings ensure that the final data sharing policy suggestions and strategies are rooted in a holistic view of both quantifiable needs and contextual barriers (Crist & Berman, 2016). The ultimate test of a pragmatic inquiry is whether the outcomes can be tested through usefulness and workability. The underlying philosophical assumptions of quantitative and qualitative approaches can be seen as incompatible by purists (Symon & Cassell, 2012). Pragmatists, however, argue that utility should govern methodological choice, allowing researchers to draw liberally from both sets of assumptions to solve the research problem (Creswell & Creswell, 2023).

Theoretical Framework

The conceptual framework plays a crucial role in simplifying complex phenomena and serves as a bridge for translating research findings into actionable policy and practice (Paradies & Stevens, 2005). The use of a single theoretical framework often provides only a partial understanding of complex implementation processes. Using one framework privileges certain causal factors while neglecting others (Nilsen, 2020). Integrating diverse frameworks also brings challenges because of their contrasting ontological and epistemological assumptions. In the study for instance, whether behaviour is driven by reflective individual intentions or influenced by the embedded institutional norms is a challenging question.

In addition to the institutional culture, willingness is shaped by disciplinary traditions, national data-sharing mandates and individual intentions and priorities. This make a multi-theoretical perspective essential for a meaningful analysis of the problem under study (Oliveira & Martins, 2011). Combining individual level behavioural theories and institutional perspectives, offers a more comprehensive understanding by capturing both personal motivations and organisational influences. Therefore, to understand the dynamics of data sharing, the study adopts a theoretical synthesis that integrates insights from both Theory of Planned Behaviour (Ajzen, 1991)and Institutional Theory (Scott, 2014).

The methodology draws on the established practice of combining multiple theories and concepts. This approach holds potential for generating new research agendas by offering varied perspectives (Farisani, 2022). In political science and policy studies, three approaches to combining theories are identified. First one is synthesis, which aims to produce a single, new theory. Second approach is complementary, which uses different theories to produce a series of perspectives or explanations. Contradictory approach compares theories to select one over others (Cairney, 2013). An advantage of using multiple theories is the ability to improve the understanding of empirical events by providing different perspectives that prompt researchers to seek detailed, relevant evidence they might otherwise miss.

According to Ajzen, intention to perform behaviour can be predicted from attitude towards behaviour. In the words of Ajzen “Intentions to perform behaviours of different kinds can be predicted with high accuracy from attitudes toward the behaviour, subjective norms, and perceived behavioural control”. The theory is supported by empirical evidences from multiple studies done in different fields. Institutional Theory, considers how regulative, normative, and cultural-cognitive pressures in universities influence researchers’ data-sharing intentions. The combined approach helps to understand the interplay of individual motivations, social influences, and organisational support in shaping data sharing behaviours.

Natural Story of Investigation

The work began with the academic ambition of working on a topic that would help society to reap maximum results from academic research. The investigator is a strong believer in open science and shares the view that the result of the publicly funded research should be shared with the public. Design and development of a national level research data repository was the initial goal of the investigator. To learn research data management and repository building, investigator has attended international workshop jointly organised by Data Science and Documentation Research and Training Centre at Indian Institute of Science, Bangalore, India.

As a practising librarian working in the university for more than 20 years, investigator has witnessed several cases of data loss and instances where data being unused. There is a mismatch of scholars routinely recollecting data that already exist elsewhere but remain inaccessible. Through literature review and direct observation, it is learned that, research data is not properly stored, systematically managed and shared in India. Public investment in research represents a substantial outlay of public resources, but the resulting knowledge, especially the underlying datasets rarely becomes a shared public good. Upon reading, it was learned that, research data can be reused and increase the transparency and quality of the academic research. Considering these facts, an investigation into the factors that might facilitate or hinder the research data sharing was felt necessary. Later, the focus of the study broadened to encompass both individual intentions and institutional factors influencing sharing behaviour. Methodologically, the study is operationalised through a Convergent Parallel Mixed Methods Research design (Creswell & Creswell, 2023).

Research Strategy

This section details the justification for using a mixed methodology and position the study within the broader classification of research designs. Methodology can be either mono, parallel or mixed. The mixed methodology is defined by various scholars differently. Cresswell (Creswell & Creswell, 2023) has identified mixing at various stages like research design, data collection stage or analysis stage. Mixed Research Methodology involves purposefully collecting and integrating both quantitative and qualitative data. Though mixed methodology research is not common in Library and Information Science is increasing popularity. Kaplan and Maxwell (2005) using qualitative and quantitative methods has explained the advantages of both methodologies, especially the benefits of a mixed research methodology. In their opinion, in the organisational studies, the quantitative approach alone may fail to bring light to various interconnected aspects like human involvement and processes. Qualitative methodology phase is used to get rich explanation to the findings from the quantitative data and analysis (Venkatesh et al., 2013).

The Rationale for Mixed Methodology Research

Integration of quantitative and qualitative data maximises the strengths and minimises the weaknesses of each methods. Quantitative methods are strong in measuring how widespread a phenomenon is by measuring prevalence and frequency. On the other hand, qualitative methods excel at explaining what the phenomenon means, detailing the context (Kaplan & Duchon, 1988). It also help understanding the processes of how or why something works. Multiple methods allow researchers to inquire into a question with multiple tools and provide a fuller, more contextualised picture of the unit under study. This is necessary for studying complex social phenomena, such as information systems, where processes related to social and organisational aspects are critical (Patton, 2015). Mixed methods enable researchers to address broader or more complicated research questions than a single method could achieve alone (Yin, 2018).

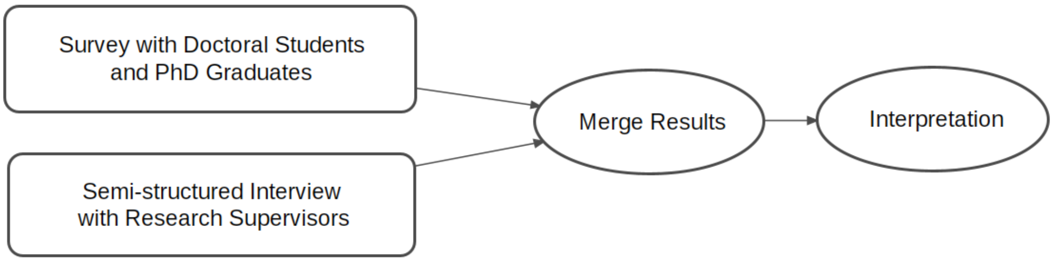
Creswell has categorised the mixed method research into four types. Distinction is based on the sequence of data collection, the priority given to qualitative or quantitative components, and the stage at which integration occurs. Widely used designs are the Explanatory Sequential design, which begins with quantitative data collection and analysis followed by qualitative exploration to explain results. The Exploratory Sequential design, starts with qualitative inquiry to identify themes and variables before moving to quantitative testing. The Convergent Parallel design, collects both data types simultaneously and merged for interpretation. More advanced designs include Transformative Mixed Methods, guided by a social justice lens, and Embedded Mixed Methods, where one type of data plays a supportive role within a larger design (Creswell & Creswell, 2023). Convergent Parallel design is used for the study.

Convergent Parallel Mixed Methods Research

The Convergent Parallel Mixed Method Research design is pragmatically justified for the study. This design involves the simultaneous or concurrent collection and analysis of both quantitative and qualitative data. The goal is integrating the two sets of findings during the interpretation stage to produce a single, comprehensive analysis (Östlund et al., 2011). By integrating these parallel datasets during the final interpretation, the design achieves triangulation, enhancing the credibility and trustworthiness of comprehensive findings (Patton, 2015; Yin, 2018). The convergent design's strength is in producing a final report where both sets of results are explicitly juxtaposed to address the research objectives. The convergent design forces the investigator to explicitly reconcile the contradictions during the interpretation stage, producing a more robust and honest set of conclusions.

Figure 3.1

Convergent Parallel Mixed Method design



Mixed methods offer considerable advantages, along with the inherent challenges. Conducting Mixed Method Research requires the investigator or the research team to possess expertise in two different kinds of methodological approaches. It also require additional time and financial resources compared to studies using a single approach. Combining methods requires the researcher to be aware of the philosophical commitments and tolerate potential ontological oscillation (Symon & Cassell, 2012). Combining the results can lead to conflicting findings that must be reconciled (Patton, 2015). The investigator must be explicit and transparent in describing the entire process especially during the reporting phase (Östlund et al., 2011). The successful use of multiple methods requires a flexible, pragmatic approach where the investigator constantly documents and justifies the decisions made.

Abductive Approach

The inductive and deductive approaches determine which reasoning process is deemed appropriate for generating knowledge and understanding the nature of reality. In general, the deductive approach is historically aligned with positivist and objectivist world views. The inductive approach is aligned with constructivist and interpretivist world views (Rossman & Rallis, 2017). The complex research designs, especially using mixed methods, combine or alternate between these approaches, through the use of abductive reasoning. Abductive reasoning is appropriate when the phenomenon under investigation is not fully understood or systematically studied.

In the Indian context, research data sharing is still at an emerging stage, with limited existing practices and absence of policies. While deductive approaches test pre-established hypotheses, purely inductive approach builds theories. Abduction allows the development of credible explanations by moving back and forth between theory and data. Abductive reasoning is suited to link observed phenomena with theoretical explanations, making it ideal for the use of multiple theories. Abductive reasoning helps reconcile conflicting findings by identifying underlying contextual influences.

Quantitative Study Design

A research design is defined as the logical sequence that connects a study's initial research questions with the empirical data to be collected and, ultimately, with the conclusions drawn. The research design acts as the blueprint for the entire study, and it is shaped largely by the type of question being asked. For instance, "how" and "why" questions often lead researchers to methods such as surveys, experiments, or historical analyses. A strong research design clearly identifies the study questions, outlines the theoretical framework, specifies the cases to be examined. It also explains how the data will be connected to the propositions, and sets out the criteria that will be used to interpret the findings.

Quantitative methods rely on predetermined questions and numeric data, which are subsequently analysed using statistical procedures. The use of online surveys offers several compelling advantages, like efficiency, cost-effectiveness, and data management (Wildemuth, 2016). Online survey tools also give researchers more control over how participants respond. They can specify whether respondents should choose only one option or select multiple answers. This helps maintain data quality and makes the analysis more accurate. Compared to traditional paper-based methods, online surveys offer flexibility in design and anonymity for participants. The automated data aggregation enhance both data quality and response rates. A significant problem associated with survey research in general, including online surveys, is low response rate, which can negatively affect the representativeness of the sample achieved (Evans & Mathur, 2018). To mitigate these weakness, the present study adopted clear instructions, concise question formats, and ensured mobile accessibility to increase response validity and participation.

Population and Sample

The survey population consisted of doctoral students and PhD graduates in various public universities in Kerala. The study considered doctoral students studying in various universities in Kerala during the period January, 2024 to June, 2024 and PhD graduates who have completed their degree during June, 2019 to May, 2024. The study make use of a broad classification of entire subjects or teaching departments under which researchers are pursuing their PhD degree. They are Engineering, Humanities, Life Sciences, Natural Sciences, and Social Sciences. These disciplines uses different types of data and uses varying methodologies. The data requirement, practices varies across disciplines and hence the challenges faced and support requirement may vary substantially. Many students were working on topics of interdisciplinary nature and in such cases they are included on the basis of the department affiliated to. As per the data obtained from the Indiastat database, PhD graduates under various faculty of universities in Kerala is 12000. In order to take the count of the doctoral students, total enrolment of PhD students during June 2018 to May 2023 was considered which account for 14534. The total population for the study consists of both these groups.

Given the total population of 26534 doctoral students and PhD graduates across institutions in Kerala, the sample size for this study was determined using Cochran's formula (Cochran, 1977). Assuming a 95% confidence level, a 5% margin of error, and a population proportion of 0.5, the calculated sample size was 384. Since the population is finite, the formula was adjusted using the finite population correction, resulting in a final sample size of approximately 378 respondents. To ensure representation of all the disciplinary categories, it was decided to select a minimum of 100 participants from each disciplinary categories. Out of 623 responses from doctoral students and PhD graduates, 26 were excluded as they exceeded the set limit or were deemed irrelevant, resulting in a final sample of 597 respondents.

Survey Instrument Development

Questionnaire developed by the investigator is used to collect survey data. There are various tools for measuring readiness, maturity, and usage of research data repositories. But a comprehensive tool to understand the current research data handling practices and measuring the attitude towards data sharing was not available. Survey instrument is constructed based on the Data Asset Framework and Collaborative Assessment of Research Data Infrastructure and Objectives (CARDIO). Factors affecting research data sharing were identified using the existing literature (Cox et al., 2019; Kouper et al., 2017; Zenk-Möltgen et al., 2018). The survey instrument was framed to gather present research data management behaviour and researchers' attitude towards data sharing. The initial questions were focused on the present research data management practices of the research scholars. Later part of the questions focused on attitude of the participants towards data sharing and training on data management. To systematically measure participants' perceptions and attitudes regarding data sharing, the study used Five Point Likert Scale. The instrument developed was tested using pilot study as discussed in the pilot study section of this chapter.

Administration of the Survey Instrument

Online survey link was shared among doctoral students and PhD graduates from various universities and research institutes of Kerala. Survey link was distributed among the respondents using social media communities like Whatsapp and email groups of doctoral students. The link was also shared among the library professional communities in the universities and colleges to distribute among the respondents. The investigator, with over 20 years of experience working in the university library and active engagement in professional library communities, played a pivotal role in facilitating data collection, particularly through the distribution of online questionnaires. The extensive network within academic and library circles enhanced the reach and efficiency of the questionnaire dissemination process. This method was also helpful in avoiding possible bias in the selection of participants. Although the response rate for online questionnaires is lower, they tend to yield more complete responses to all questions.

Analysis of Survey Data

The survey data were processed using a series of simple and transparent steps. All tools used in the workflow were free and open-source, which supported both accessibility and reproducibility. The responses were first imported into OpenRefine (OpenRefine, 2023). This tool was used to clean the data, correct inconsistencies, and standardise formats. After cleaning, the dataset was transferred to LibreOffice Calc for coding and basic checks. The coded file was then imported into the free statistical package PSPP (Free Software Foundation, 2007) for analysis. Percentage analysis formed the core of the quantitative interpretation. This approach helped capture clear patterns in the responses without relying on complex statistical procedures. The chi-square test is also widely used for examining whether there is a significant association between two categorical variables especially for discipline based comparison. The charts in the analysis chapter were prepared using Drawio software.

Qualitative Study Design

Qualitative methods follow emerging research procedures, open-ended questions and involve collecting textual data or images (Marshall & Rossman, 2016). In the words of Polit and Beck (2010), "The goal of most qualitative studies is not to generalise but rather to provide a rich, contextualised understanding of some aspect of human experience through the intensive study of particular cases". Although the inability to generalise findings is a common critique, qualitative method was chosen intentionally. It allowed the investigator to capture the complexities and specific features of the problem in a way that other methods could not. Semi-structured interview schedule is adopted for qualitative data collection. Structured interview though comparatively easy to analyse, often restrict the interviewee from expressing their opinion and sharing their knowledge. Open ended questions allow the interviewee to express their opinion without any restriction. A semi-structured interview is suitable for understanding dimensions and arriving at new dimensions unknown (Creswell & Creswell, 2023; Kvale, 2007).

Interview Population and Sampling

Research supervisors from various universities in Kerala constitute the population for interview. The criteria for selecting interview participants, also known as case selection, are driven primarily by the research question and the overall purpose of the study (Symon & Cassell, 2012). This process usually involves purposeful sampling or theoretical sampling, distinct from statistical sampling. The choice of the sample was in tune with the school system followed in the University of Calicut. Participants from various disciplines were purposefully selected to get a picture of the data practices and sharing attitude across various disciplines. The goal is to select information-rich cases, those from which one can learn a great deal about issues of central importance to the inquiry (Patton, 2015).

Purposive sampling technique was used mainly for two reasons; respondents should be willing to communicate with the researcher, and they should be knowledgeable in the area (Campbell, 1955). Various experts in qualitative research suggests that the number of subjects for the interview should be limited for easier analysis. Kvale's (2007) interview method was used to conduct interviews. The selection of research supervisors was made to ensure representation across disciplines, type of research organisation, gender and age groups. Some of the interview participants has suggested experts for further interviews. The investigator has interviewed 37 participants from five disciplinary categories including the five done in the pilot study stage. The interview process was continued till a state of theoretical saturation, when it was felt that no real new insights were being obtained from further interviews. Considering the richness of the interview, 20 participants were selected for the final analysis.

Interview Instrument

A semi-structured interview schedule is prepared for collecting data from research supervisors. The aim of the interview is to capture first-hand information on the data generation practices under the research supervisor. The attitude of research supervisors towards data sharing, their opinion about various factors of research data sharing and existing policies and practices were included in the interview schedule. The interview questions were framed based partly on Collaborative Assessment of Research Data Infrastructure and Objectives (CARDIO) tool developed by DCC, London and Digital Asset Framework. CARDIO is used to assess the research data repositories, and it is being used as a foundation for the development of repositories. Research data maturity models were also used in the development of interview instrument. Previous literature in the area also helped to develop interview schedule (Azami et al., 2023).

Administration of Interview

The session started with an introduction, where the investigator explained the study's purpose and outlines how anonymity of participants will be assured. Academic profile of all the participants were created prior to the interviews using Google Scholar and Web of Science. Familiarisation was done prior to the actual interview as the concepts like 'research data sharing' or 'research data management' were not familiar with majority of the participants. Semi-structured interview which follows an interview guide containing questions were used and interviews were occurred mainly at the office of the research supervisors. Initially three research supervisors from each disciplines were listed based on the research activities, attitude towards open access philosophy. Before the interview, repositories in various fields and their working was explained to the participants. Many participants found this helpful and the investigator could create a cordial atmosphere in the interview process. Interview were recorded using digital recorder with the prior permission of the participants. In addition to the recorded voice, important points were noted in the notebook. These notes were helpful in the coding stage and identifying themes.

Data Processing and Analysis

Interview data were processed using a systematic, multi-step qualitative approach. The investigator began by preparing the raw data for analysis through transcription of the recorded interviews. Text by text representation is not possible in the interview transcript preparation. Considering the varying disciplinary backgrounds, different terminologies were used by the participants in the interview. Wherever possible exact words are used in the first round. Identifiable information like names and specific locations were removed from the transcript to ensure confidentiality and protect privacy of the participants. To manage and analyse large volume of textual data efficiently, LibreOffice Calc spreadsheet software was used. Once transcription was completed, the material was systematically organised and managed, preparing it for deeper thematic analysis.

To analyse the data, the investigator adopted thematic analysis as the core analytical framework. Thematic analysis is a systematised qualitative approach often used to identify common patterns and unique perspectives, such as those found in interview data (Rossman & Rallis, 2017). It begins after the raw data, such as interview recordings, have been converted into verbatim transcripts. The codes are then constantly compared and grouped together to refine categories and ultimately synthesised into a cohesive and coherent set of broader themes or interpretive concepts.

Coding

Coding is a foundational analytic technique in qualitative research that serves as a systematic process for transforming raw data into meaningful and manageable units. It is the critical link between data collection and the eventual explanation of meaning (Miles et al., 2014). Saldaña (2016) defines coding as "most often a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data." Coding of the interview transcript was done systematically using memos. Initial Open Coding was conducted to identify, label, and categorise the core ideas and incidents represented in the data. Through the iterative process, the investigator derived themes by grouping related codes into broader and more meaningful categories. Secondary-level coding then consolidated these categories into key analytical themes that aligned with the research objectives.

Open Coding, In Vivo Coding, Value Coding and Structural Coding are used in accordance with the nature of the theme. The coding style for the study was aligned with the selected theoretical frameworks. Value coding was utilised to capture individual level perspectives, attitudes, and beliefs, as it is well-suited for qualitative analysis of personal experiences and values (Saldaña, 2016). Structural coding was applied to identify and categorise themes related to institutional processes and frameworks, facilitating analysis of broader systemic patterns. This multiple coding approach ensured that the data analysis reflected the theoretical underpinnings of the study, enabling a comprehensive exploration of both individual and organisational dimensions. The final codes were emerged after six rounds of coding. The findings are subsequently reported according to the emergent themes. Data visualisations were created using the Visual Understanding Environment (VUE) software (Tufts University., 2015). VUE is an open source concept-mapping and visualisation tool developed by Tufts University. A graphical representation of the coding is provided in the appendix.

The Pilot Study

Conducting a pilot study is significant in refining data collection instruments. Pilot study minimise problems by ensuring that data collection procedures are comparable across occasions (Yin, 2018). The investigator has conducted pilot study to make sure that the questions are easy to understand and conveying. In the pilot study, online questionnaire were administered directly in the presence of the investigator to ensure that questions are easy and understandable. Certain terminologies like ‘metadata’, ‘data curation’, and ‘data privacy’ were difficult to understand and, appropriate descriptions were added wherever necessary. One open ended question was also provided to capture unanticipated observation. Pre-testing of the instrument is also done by discussing with experts in the field and fellow research scholars.

Pilot study result has exposed the low level of awareness about research data management and awareness about the existing open data repositories. Hence, the investigator has devised questionnaire with simple questions that capture present data handling practices along with questions related to motivating and challenging factors. Pilot survey and pilot interview were conducted among the participants of University of Calicut. Pilot survey study among twenty doctoral students and ten PhD graduates were conducted. Pilot interview with five research supervisors from five disciplinary categories were conducted. Pilot interview with research supervisors helped refine the survey questions and the responses from the pilot survey also helped improving the interview schedule.

Validity, Reliability and Trustworthiness

Quality of a research study depend on the validity, credibility, reliability, and trustworthiness. This require adopting systematic procedures aligned with the methodology chosen for the study (Wildemuth, 2016). As the present study use mixed methodology, measures are taken to ensure rigour across both quantitative and qualitative approaches. The study has undertaken systematic process, supporting dependability by outlining the research path, conducting a pilot study, and detailing data analysis procedures. For rigorous qualitative and mixed methods research, meticulous analytic documentation is essential, which includes systematically recording procedures, decisions, and interpretations to create an audit trail or chain of evidence (Yin, 2018). The use of a formal protocol or database to store interview transcripts, notes, and analysis procedures, distinct from the final report, markedly increases the reliability of the study.

Triangulation

A primary strategy for enhancing the validity and credibility of the findings, especially in the mixed methodology, is the use of triangulation. Triangulation involves combining multiple data sources or research methods to pursue converging lines of inquiry. By examining the same phenomenon through multiple perspectives, researchers can limit the biases that often come with relying on just one type of data. Triangulation also leads to a richer, more layered grasp of complex research questions (Fetters & Freshwater, 2015).

The present study used two primary sources of data, survey administered to doctoral students and interviews with research supervisors. The two data collection instruments were designed for different respondent groups and therefore did not contain identical questions. But, both instruments explored complementary aspects of research data handling and sharing. This allowed the investigator to compare findings thematically across stakeholder perspectives and achieve a degree of methodological triangulation. In this study, triangulation is not about matching individual items one by one. Instead, it focuses on recognising where the broader themes align and where they differ. As Patton (2015) explains, triangulation enables researchers to "compare the perspectives of people from different points of view", enhancing the credibility and depth of qualitative interpretation. Accordingly, while main-theme triangulation was feasible between the survey and interview results, sub-theme triangulation was limited due to the differing data collection tools and respondent categories.

Member Checking

Another crucial measure for enhancing qualitative credibility is member checking or feedback from participants. In member checking, interpretations are shared with interviewees or key informants to confirm accuracy and comprehensiveness. In the study, investigator has actively examined alternative interpretations and carefully weighed the influence of evidence to strengthen the credibility of the findings. This also helped minimise the potential researcher bias. To further enhance qualitative rigour, investigator has implemented member checking by sharing the interpretations with participants wherever necessary. This allowed to verify the accuracy, clarity, and completeness of the findings. This process helped ensure that the analysis authentically reflected participants' perspectives and reduced the likelihood of misinterpretation (Creswell & Creswell, 2023).

Replicability

To address concerns about the replicability of qualitative research, investigator has adopted a transparent and systematic approach throughout the study. To enhance transparency, the investigator has maintained detailed field notes, a research log, and reflective memos that documented every design decision and the reasoning behind it. This allowed a clear record of the research process to be available for external review. In addition, investigator has organised and stored all collected data in a structured and retrievable format. By ensuring that it could be accessed for verification or reanalysis, it was made possible for other researchers to examine the procedures, decisions, and evidence that informed the findings, thereby enhancing the study's credibility and rigour (Marshall & Rossman, 2016).

Personal Bias

A major challenge in the qualitative research is to make sure that personal bias does not influence the study. Personal bias in qualitative research is recognised as an inherent and complex phenomenon rooted in the core philosophical foundations of the methodology. Traditional positivist ideal strives for complete objectivity by eliminating human factors. According to Rossman and Rallis, qualitative research acknowledges that the researcher is the primary instrument of inquiry, making complete value-freedom unattainable. Therefore, personal bias is viewed less as a flaw to be eradicated and more as an inevitable influence. This influence must be consciously recognised and accounted for to ensure the rigour and credibility of the study (Tong et al., 2007). The possibility of personal bias was minimised by continued discussion with the research supervisor and peers at various stages of the research, like framing research questions, interpretation and analysis (Alvesson, 2003).

Ethical Considerations

Ethical considerations are fundamental to conducting credible and high quality research. This is more relevant when studies involve human subjects. Ensuring the moral conduct of a qualitative study means reasoning through moral principles in every decision taken (Campbell, 1955).

Informed Consent

In the study, investigator has obtained explicit, informed, and voluntary consent from all the participants before involving them in any data collection activities. Each participant was clearly informed about the purpose of the research. The expected nature and duration of their participation, and any potential risks or benefits associated with the study were also discussed. Participation was entirely voluntary, and individuals were assured that they could withdraw at any stage without any negative consequences. Consistent with ethical guidelines, investigator has treated informed consent not as a one-time administrative procedure but as an ongoing process, revisiting and reaffirming consent when necessary, as recommended by Patton and Miles et al. (2014). Investigator has obtained prior permission from the participants for voice recording the interview.

Confidentiality and Anonymity

To protect participant privacy, clear and practical protocols for maintaining confidentiality has been established. In qualitative research, complete anonymity can be challenging because participants may be identifiable even when pseudonyms are used. Therefore, all personal identifiers in transcripts and other records are anonymised. Raw data and analysis files are stored in a secure, password protected folders accessible only to the investigator. Confidentiality agreements outlined where data would be stored, how long it would be retained, and who could access it. These measures were put in place to protect the identity participants and ensure that all shared information remained secure.

Bias and Subjectivity

Since the researcher is the primary instrument of inquiry in qualitative studies, they must proactively recognise and clarify their personal perspectives, beliefs, and potential biases. This is critical because bias poses a threat to the credibility of the findings. To address this, investigator has engaged in continuous self-reflection throughout the study to identify and clarify the assumptions, disciplinary background, and possible biases as recommended by best practices in qualitative research. A reflexive journal documenting critical reflections, decision points, and moments of uncertainty, was also maintained which helped make positionality explicit. This reflexive approach supported the development of credible and balanced findings.

Ethical Reporting

It is acknowledged that poorly designed studies and inadequate reporting can lead to the inappropriate application of research findings. Such findings can lead to poor decision making. To ensure the transparency of research methods and maintain the theoretical possibility that readers could replicate the study methods, Consolidated Criteria for Reporting Qualitative Research (COREQ) is used. COREQ is a 32-item checklist developed by Tong et al. (2007) specifically for the explicit and comprehensive reporting of qualitative studies. The investigator has also documented how sensitive data were handled, stored, and presented to minimise any potential harm to participants. By clearly acknowledging these steps in the reporting process, investigator has aimed to uphold ethical rigour and strengthen the trustworthiness of the study as recommended. In addition, investigator has organised and stored all collected data in a structured and retrievable format, ensuring that it could be accessed for verification or reanalysis if required. By doing so, the investigator has made it possible for other researchers to examine the procedures, decisions, and evidence that informed the findings, thereby enhancing the study's credibility and rigour.

Chapter Summary

Appropriate research methods are used to answer the research questions. Quantitative methodology is used to identify existing research data handling practices. Qualitative methodology is adopted to understand the attitude and perceptions of research supervisors. The TPB constructs are measured quantitatively across a large sample of doctoral students. Institutional Theory used to gather rich, contextual data to identify and address institutional factors. The chapter establishes the necessary foundation of rigour and competence before proceeding to present the empirical results in the following chapters.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Alvesson, M. (2003). Beyond neopositivists, romantics, and localists: A reflexive approach to interviews in organizational research. *The Academy of Management Review, 28*(1), 13–33. JSTOR. https://doi.org/10.2307/30040687

Azami, M., Sadatmoosavi, A., & Chashmyazdan, M. (2023). Research data management frameworks: A systematic literature review. *International Journal of Information Science and Management (IJISM), 21*(3), 1–18. https://doi.org/10.22034/ijism.2023.1977759.0

Cairney, P. (2013). Standing on the shoulders of giants: How do we combine the insights of multiple theories in public policy studies? *Policy Studies Journal, 41*(1), 1–21. https://doi.org/10.1111/psj.12000

Campbell, D. T. (1955). The informant in quantitative research. *American Journal of Sociology, 60*(4), 339–342. https://doi.org/10.1086/221565

Cochran, W. G. (1977). *Sampling techniques*. john wiley & sons.

Cox, A. M., Kennan, M. A., Lyon, L., Pinfield, S., & Sbaffi, L. (2019). Maturing research data services and the transformation of academic libraries. *Journal of Documentation, 75*(6), 1432–1462. https://doi.org/10.1108/JD-12-2018-0211

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4. ed). SAGE.

Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th edn). SAGE.

Crist, E. A., & Berman, E. A. (2016). Mixed methods research in LIS literature: A scoping review. *University Libraries Faculty and Staff Publications*. https://scholarworks.uvm.edu/libfacpub/39

Ellis, D. (1993). Modeling the information-seeking patterns of academic researchers: A grounded theory approach. *The Library Quarterly*. https://doi.org/10.1086/602622

Evans, J. R., & Mathur, A. (2018). The value of online surveys: A look back and a look ahead. *Internet Research*, *28*(4), 854–887. https://doi.org/10.1108/IntR-03-2018-0089

Farisani, T. R. (2022). The impact of social and institutional networks in sustaining South African rural small, medium and micro enterprises during the crisis. *Acta Commercii*, *22*(1). https://doi.org/10.4102/ac.v22i1.1007

Fetters, M. D., & Freshwater, D. (2015). The 1 + 1 = 3 Integration Challenge. *Journal of Mixed Methods Research*, *9*(2), 115–117. https://doi.org/10.1177/1558689815581222

Free Software Foundation. (2007). *GNU PSPP* (Version 2.0.1) [Computer software].

Glasgow, R. E. (2013). What does it mean to be pragmatic? Pragmatic methods, measures, and models to facilitate research translation. *Health Education & Behavior*, *40*(3), 257–265. https://doi.org/10.1177/1090198113486805

Kaplan, B., & Duchon, D. (1988). Combining qualitative and quantitative methods in information systems research: A case study. *MIS Quarterly*, 571–586.

Kaplan, B., & Maxwell, J. A. (2005). Qualitative research methods for evaluating computer information systems. In *Evaluating the organizational impact of healthcare information systems* (pp. 30–55). Springer.

Kaushik, V., & Walsh, C. A. (2019). Pragmatism as a research paradigm and its implications for social work research. *Social Sciences*, *8*(9), 255. https://doi.org/10.3390/socsci8090255

Kouper, I., Fear, K., Ishida, M., Kollen, C., & Williams, S. C. (2017). *Research data services maturity in academic libraries*. University of British Columbia. https://doi.org/10.14288/1.0343479

Kvale, S. (2007). *Doing interviews*. Sage.

Marshall, C., & Rossman, G. B. (2016). *Designing qualitative research* (6th ed.). Sage.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Sage.

Morgan, D. L. (2014). Pragmatism as a Paradigm for Social Research. *Qualitative Inquiry*, *20*(8), 1045–1053. https://doi.org/10.1177/1077800413513733

Nilsen, P. (2020). Making sense of implementation theories, models, and frameworks. In B. Albers, A. Shlonsky, & R. Mildon (Eds), *Implementation Science 3.0* (pp. 53–79). Springer. https://doi.org/10.1007/978-3-030-03874-8_3

Oliveira, T., & Martins, M. F. (2011). Literature review of information technology adoption models at firm level. *Electronic Journal of Information Systems Evaluation*, *14*(1), 110.

OpenRefine. (2023). *Open Refine* (Version 3.8.1) [Computer software]. https://openrefine.org/

Östlund, U., Kidd, L., Wengström, Y., & Rowa-Dewar, N. (2011). Combining qualitative and quantitative research within mixed method research designs: A methodological review. *International Journal of Nursing Studies*, *48*(3), 369–383. https://doi.org/10.1016/j.ijnurstu.2010.10.005

Paradies, Y., & Stevens, M. (2005). Conceptual diagrams in public health research. *Journal of Epidemiology & Community Health*, *59*(12), 1012–1013. https://doi.org/10.1136/jech.2005.036913

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). Sage.

Polit, D. F., & Beck, C. T. (2010). Generalization in quantitative and qualitative research: Myths and strategies. *International Journal of Nursing Studies*, *47*(11), 1451–1458. https://doi.org/10.1016/j.ijnurstu.2010.06.004

Rossman, G. B., & Rallis, S. F. (2017). *An introduction to qualitative research: Learning in the field* (4th ed.). Sage.

Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). Sage.

Scott, W. R. (2014). *Institutions and organizations: Ideas, interests and identities* (4th ed.). Sage.

Symon, G., & Cassell, C. (2012). *Qualitative organizational research: Core methods and current challenges*. Sage.

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, *19*(6), 349–357.

Tufts University. (2015). *Visual Understanding Environment (VUE) (Version 3.3.0)* [Computer software]. https://vue.tufts.edu/

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. *MIS Quarterly*, *37*(1), 21–54.

Wildemuth, B. M. (Ed.). (2016). *Applications of Social Research Methods to Questions in Information and Library Science* (2nd ed.). Libraries Unlimited.

Yin, R. K. (2003). *Case study research: Design and methods*. Sage.

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage.

Zenk-Möltgen, W., Akdeniz, E., Katsanidou, A., Naßhoven, V., & Balaban, E. (2018). Factors influencing the data sharing behavior of researchers in sociology and political science. *Journal of Documentation*, *74*(5), 1053–1073. https://doi.org/10.1108/JD-09-2017-0126

Chapter 4. Quantitative Analysis

This chapter analyses the survey data collected from doctoral students and PhD graduates of various research departments under the universities in Kerala. The quantitative phase aims to investigate current research data management practices and, attitude of researchers towards research data sharing. The analysis of results begin with the question asked in the survey. The survey results reveal the existing data management practices, as well as the major motivating and challenging factors that influence research data sharing.

Population and Sample

The target population for this study comprises current doctoral students enrolled in various universities in Kerala, and PhD graduates who completed their doctoral degrees between 2018 and 2024. The study uses a broad classification of academic disciplines based on the departments in which researchers are pursuing their PhD. These categories are Engineering, Humanities, Life Sciences, Natural Sciences, and Social Sciences. The sample includes a minimum of 100 respondents from five broad disciplinary categories. Stratified sampling technique was used to ensure representation across different academic disciplines.

Data collection was conducted using online questionnaire, which was distributed through multiple digital channels, including email, academic and research-focused WhatsApp groups. In addition to online groups and channels, the questionnaire was distributed through library professionals working in various university libraries and research centres. The questionnaire instrument consisted of 30 items, comprising a combination of multiple-choice questions, 5-point Likert-scale items, and open-ended questions. This mixed item design aligns with best practices in survey research for achieving both depth and breadth of response (Creswell, 2014). A total of 597 valid responses were selected for the analysis. The relatively large sample size not only enhances the statistical power of the study but also increases the reliability and generalisability of the findings. The responses to the questionnaire were systematically coded and categorised in accordance with standard research procedures to facilitate analysis.

Table 1

Demographic Characteristics of respondents

Item	Frequency	Percentage
Gender		
Male	184	30.8
Female	409	68.5
Not disclosed	4	.7
Age		
Below 30	172	28.8
Between 30 and 40	308	51.6
Above 40	117	19.6
Status		
PhD Graduate	288	48.2
PhD Student	309	51.8
Discipline		
Engineering	107	18.0
Humanities	105	17.6
Life Sciences	116	19.4
Natural Science	101	16.9
Social Sciences	168	28.1
Institution		
Central University of Kerala	38	6.4
CUSAT	53	8.9
Kannur University	30	5.0
MG University	113	19.0
NIT Calicut	51	8.5
University of Calicut	186	31.2
University of Kerala	110	18.4
Others	16	2.6

Demographic Characteristics

Details of PhD graduates and doctoral students studying in various universities are listed in table 1. The chart shows the distribution of respondents across different disciplinary categories. Disciplines represented include Social Sciences (28.1%), Life Sciences (19.4%), Engineering (17.9%), Humanities (17.6%), and Natural Science (16.9%). University-wise data indicate that the Social Sciences account for the largest share of research scholars in Kerala. The study sample includes both current doctoral students and PhD graduates drawn from multiple universities, comprising 51.8% PhD students and 48.2% PhD graduates. Within the PhD graduates 37% participants completed their PhD before 2020 and remaining 63% after year 2020.

About half of the respondents belongs to the age group of 30 to 40 (51.9%), followed by below 30 years (28%) and above 40 years age group (20.1%). Researchers in the above 40 age category are mainly PhD graduates. Below 30 age group 88.8% are research students. The respondents from the age group between 30 and 40 has a more balanced distribution of PhD Graduates and Students (46.1% and 53.9% respectively). Female constitute two third of the total respondents (68.5%). In both PhD graduate and doctoral students category female has higher representation. Considering the female to male enrolment ratio in the Universities of Kerala the data collected is proportional in nature. In all the 5 disciplines female scholars are higher than male. Engineering discipline has the minimum with 62.5% and Humanities has maximum with 73.3% of female.

Research Data Practices

This section examines the research data practices of researchers, providing a comprehensive overview of their approaches to data collection, management, and preservation. Specifically, it investigates the volume and types of data collected, the frequency and methods of data backups, and the extent to which data are preserved for future use. These aspects are explored through a series of survey questions designed to capture current research data handling practices.

Data Type

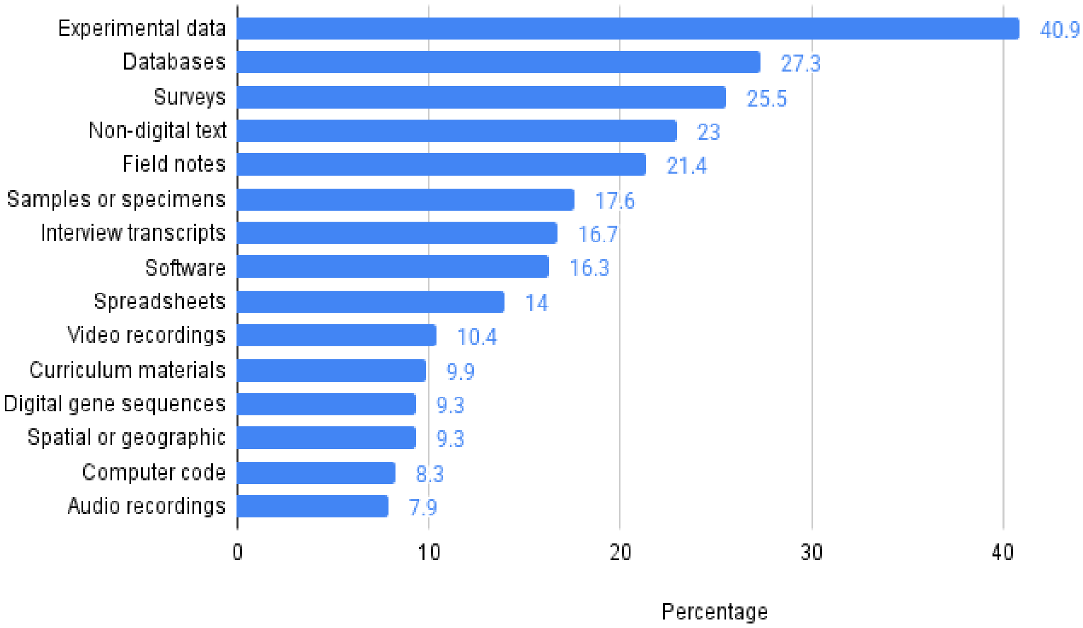
Please indicate the types of data you have collected or generated in your research. If your data type is not listed, choose 'Other' and briefly describe your data. *[Select all that apply]*

The nature of research and the academic disciplines shape the types of data used in scholarly investigations. Drawing insights from prior studies, a comprehensive list of data types was compiled and presented to survey respondents, who could select all applicable options relevant to their research. The resulting table enumerates 26 distinct data types, comprising common research outputs such as experimental data and surveys, as well as specialised forms like digital gene sequences and artistic products. Additionally, the option “I Don’t Produce Any Data” was included to accommodate respondents whose work does not involve data generation, ensuring a complete representation of research practices.

Experimental data is the frequently reported data type used by 40.9% of the respondents, indicating that experimental research is prevalent among respondents. Databases is the second (27.3%) common data type suggesting that, many researchers work with structured data collections, such as public or proprietary databases. Surveys are widely used (25.5%), especially in Social Sciences, reflecting the importance of questionnaire-based data collection. Moderately common data types are Non-digital Text (23%) including manuscripts, archival documents, or other text not in digital form, common in Humanities discipline. Field Notes (21.4%) prevalent in qualitative research, such as anthropology, sociology, and ecology, where observations are recorded during fieldwork. Samples or Specimens (17.6%) and, interview transcripts (16.7%) common in qualitative research, particularly in Social Sciences and Humanities. Software (16.3%) Indicates researchers developing or using software tools, likely in computational or Engineering fields. Spreadsheets with 14% widely used for data organisation across disciplines. Less common data types like video recordings, computer codes and artistic products listed in the table are also used by researchers.

Figure 4.1

Data Type



The dominance of experimental data, databases and surveys suggests that the respondents includes doctoral students from empirical, data-driven fields. These data types are likely to be digital, structured, and reusable, making them easier for data sharing. Qualitative data (e.g., field notes, interview transcripts) and physical data (e.g., samples or specimens) are less common but still significant, highlighting the need for varied data management and sharing strategies. The presence of "I don't produce any data" (3.41%) suggests that the sample includes non-empirical researchers like mathematicians, theorists or literature. This indicate the low level of awareness regarding the research data among researchers.

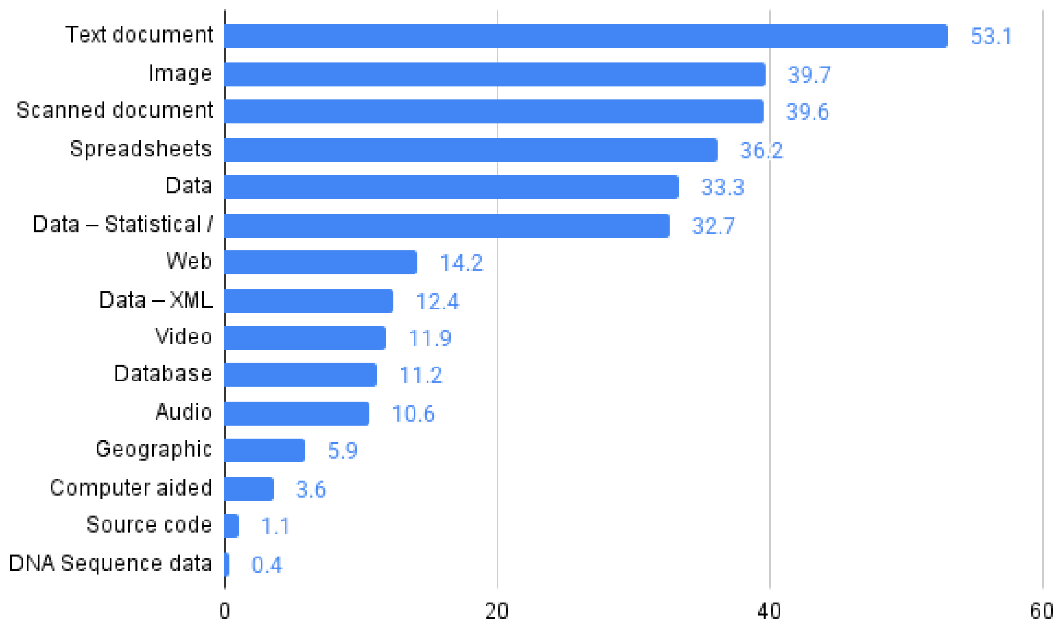
File Type

> Which of the following file formats best describe your digital research data? Examples of specific file extensions are included. [Please choose all that apply].

The survey question provided a list of file formats used in the research, and asked respondents to choose from all applicable formats. The figure 4 shows that, text document is the most prevalent format (53.1%), used by more than half of the respondents. This reflects the significance of text based data in research, such as reports, articles, or notes. Image is common across disciplines (39.7%) and, scanned document (39.6%) nearly as common as images. Spreadsheets are also widely used (36.2%), reflecting structured data in Social Science. XML data (33.3%) and data specific to statistical software like SAS and SPSS (32.7%), shows consistency with the earlier finding (table data type) that experimental data (40.9%) and databases (27.3%) are the most common data types. The dominance of these common formats reflects the widespread use of versatile, accessible formats across disciplines. These structured and shareable formats are easy to create and store making them ideal for data-sharing initiatives. The "don't know" category (3.6%) suggests a small subset of researchers lacking technical awareness of file formats, potentially hindering data sharing. This group may benefit from training in data management.

Figure 4.2

File Types



Data Size

Indicate the estimated amount of digital data the research has created or is likely to generate.

To understand volume of the data produced, an important factor while planning a research data repository, respondents were asked to select from the multiple choice list of data sizes produced. With categories ranging from "can't say" to "above 10 GB." Only 8.5% of the respondents reported generating data above 10 GB, suggesting most researchers handle relatively small datasets.

Table 2

Volume of Data Generated

Discipline	Can't Say	Less than 100 MB	Between 100 MB and 1 GB	Between 1 GB and 10 GB	Above 10 GB
Engineering	13	19	27	30	18
	12.1%	17.8%	25.2%	28.0%	16.8%
Humanities	27	35	19	20	4
	25.7%	33.3%	18.1%	19.0%	3.8%
Life Sciences	25	11	47	26	7
	21.6%	9.5%	40.5%	22.4%	6.0%
Natural Science	18	22	21	26	14
	17.8%	21.8%	20.8%	25.7%	13.9%
Social Sciences	47	41	41	31	8
	28.0%	24.4%	24.4%	18.5%	4.8%
Total	130	128	155	133	51
	21.8%	21.4%	26.0%	22.3%	8.5%
Chi square = 60.30; p-value = 0.000					

Results indicate that 33.2% of the respondents manage datasets between 100 MB and 1 GB, 28.5% between 1 GB and 10 GB, and 10.9% above 10 GB, while 26.6% handle less than 100 MB. Only 0.9% reported no data. This suggests a significant volume of research data is generated, with disciplinary differences evident. Respondents from Life Sciences (51.6%) and Engineering (31.9%) tend to produce larger datasets (between 100 MB and 10 GB or more), while Humanities scholars (42.3%) predominantly manage smaller datasets (Less than 100 MB). The chi-square test with p value less than 0.05 confirms a statistically significant association between discipline and data size. Nearly one fourth of the respondents (21.8%) are not sure of their data size, this may be because respondents who have completed the research work may not have preserved it properly or the low level of awareness.

Backup Frequency

Indicate how often backups are made for the digital files associated with your research.?

The table 3 presents data on the frequency of taking backup for research files across five disciplinary categories. Overall backup frequency shows that, most researchers backup their files at least occasionally and, at the same time very frequent backups are rare. At the aggregate level, 70.3% of respondents back up their research data at least weekly, including daily (23.3%) and weekly (23.5%) backups, suggesting a generally positive orientation towards data safety and continuity. However, a notable proportion of respondents (27.1%) still engage in infrequent backup practices (rarely or never), highlighting persistent gaps in consistent research data management behaviours.

Table 3

Data Backup Frequency

Discipline	Never	Rarely	Monthly	Weekly	Daily	More than once a day
Engineering	0	20	30	28	23	6
	0%	3.4%	5.0%	4.7%	3.9%	1.0%
Humanities	8	26	20	27	22	2
	1.3%	4.4%	3.4%	4.5%	3.7%	0.3%
Life Sciences	8	31	25	26	22	4
	1.3%	5.2%	4.2%	4.4%	3.7%	0.7%
Natural Science	4	19	29	16	25	8
	1%	3.2%	4.9%	2.7%	4.2%	1.3%
Social Sciences	8	38	25	43	47	7
	1.3%	6.4%	4.2%	7.2%	7.9%	1.2%
Total	28	134	129	140	139	27
	4.7%	22.4%	21.6%	23.5%	23.3%	4.5%

Engineering researchers demonstrate moderate-to-high backup regularity, with most respondents reporting monthly (5%) or weekly (4.7%) backups. Daily backups are also relatively common (3.9%). Notably, no respondent reported never backing up data, indicating strong baseline awareness of data protection. Humanities researchers show greater variability in backup practices. While weekly (4.5%) and daily (3.7%) backups are reported, rarely and never responses are comparatively higher than in Engineering. The diversity of data types (textual, interpretive, qualitative) and reliance on personal storage devices may contribute to less standardised backup behaviours. Life Sciences respondents exhibit a balanced distribution across monthly, weekly, and daily backups. However, the presence of rarely and never responses is noteworthy given the data-intensive nature of the discipline. Natural Sciences display strong monthly and daily backup practices. Social Sciences stand out with the highest engagement in frequent backups, particularly weekly, and daily backups.

The result reveals that while regular research data backup practices are prevalent across disciplines, significant disciplinary variations persist. Infrequent backup practices persist, particularly in Humanities, Life Sciences, and Social Sciences. Computationally intensive fields demonstrating more frequent and automated backup behaviours, and humanities oriented disciplines exhibiting comparatively irregular practices.

Data Storage Location

> Where have you stored the data from your research? [Please choose all that apply.]

Table 4 depicts the data storage practice of researchers. Only 3.9% of the respondents use Institutional Storage. This highlight a significant reliance on personal or non-institutional storage methods, which increases the risk of data loss. Personal computer is the prevalent data storage method (83.6%), indicating that most respondents store research data on personal devices. This is convenient but vulnerable to hardware failure and theft. Email is used by half the sample (50.1%). Email is not designed for long-term storage and poses security and capacity risks. Cloud storage (44.1%) is also popular, including services like Google Drive, Dropbox, or OneDrive, offering accessibility and backup features. This reflects growing adoption of cloud solutions. External disk, used by almost 30% of respondents, likely for portability or small datasets. Flash drives (29.6%) are convenient but prone to loss, corruption, or limited capacity.

Table 4

Data Storage Location

Storage	Frequency	Percentage
Personal Computer / Laptop	499	83.6
Email	299	50.1
Cloud storage	263	44.1
External Disk	181	30.3
Flash drive / Pen drive	177	29.6
Lab computer	59	9.9
Institutional Storage	23	3.9

The dominance of Personal Computer and Email highlights decentralised, individual-centric approach to data storage.The moderate use of Cloud Storage, External Disk, and Pen Drive indicates some adoption of backup-oriented methods, but these are still personal rather than institutional solutions. The low use of Lab Computer (9.9%) and Institutional Storage (3.9%) indicate the absence of institutional infrastructure for research data preservation.

Metadata

Do you maintain detailed descriptions about your research dataset which provides context for understanding why and when the data were collected and how they were used. (also known as metadata).

a) Yes, I provide detailed descriptions for my research data.

b) No, I do not provide detailed descriptions for my research data.

c) Partially, I provide some level of description for certain aspects of my research data.

d) Not applicable / I am not sure.

Metadata provides critical context about the purpose, collection, and usage of datasets. Metadata is essential for ensuring data comprehensibility facilitating data reuse. This section analyses the responses to the question designed to ascertain whether doctoral students maintain detailed metadata for their datasets. Table 5 presents the responses, detailing the prevalence of metadata maintenance practices among doctoral students. Nearly half of respondents maintain detailed metadata (45.2%). Many researchers maintain partial metadata only (29.8%).

Table 5

Metadata Preservation Behaviour

Discipline	No	Partially	Yes
Engineering	12	42	30
	14.30%	50%	35.7%
Humanities	5	33	47
	5.90%	38.8%	55.3%
Life Sciences	5	31	65
	5.00%	30.7%	64.4%
Natural Science	9	27	48
	10.70%	32%	57.1%
Social Sciences	18	45	80
	12.60%	31.5%	55.9%
Total	49	178	270
	9.90%	35.8%	54.3%
Chi square = 20.82; p-value = 0.008			

Life Sciences (64.4%) and Social Sciences (47.6%) lead in comprehensive metadata maintenance, likely due to structured data types (e.g., Surveys, 25.5%; Statistical Data, 32.7%) requiring detailed documentation. Engineering has the lowest (35.7%) and highest combined 'unsure' and 'partial' metadata management (64.3%), suggesting less comprehensive metadata. Social Sciences (10.7%) and Engineering (11.2%) recorded the highest rates of no metadata usage, indicating areas for improvement, while Humanities and Life Sciences (5.9% and 5%) have the lowest. The combined share of affirmative and partial responses (75%) reflects a strong tendency among researchers to maintain metadata, an important prerequisite for enabling data sharing. The significant chi-square value ($p = 0.008$) confirms that metadata practices vary by discipline, due to differences in data types, or institutional requirements.

Data Preservation

Have you saved the data generated from your research for long-term storage or archival purposes, and if not, do you plan to do so in the future? (Please select the option that best represents your situation.)

. Yes, I have already preserved my research data

. No, I have not preserved my research data but plan to do so in the near future.

. No, I have not preserved my research data and currently have no plans to do

. I am uncertain about how to preserve my research data.

The preservation of research data for long-term storage or archival purposes is essential not only for enabling future studies but also for facilitating collaboration through peer sharing and enhancing discoverability. This study, focused on understanding research data management practices, places significant emphasis on data preservation as a critical component of effective data stewardship. Respondents selected one of four options that best represented their situation:

Table 6

Data Preservation Practice

Discipline	Uncertain	No	Plan to Preserve	Preserved
Engineering	10	11	17	69
	9.3%	10.3%	15.9%	64.5%
Humanities	11	8	27	59
	10.5%	7.6%	25.7%	56.2%
Life Sciences	6	7	24	79
	5.2%	6%	20.7%	68.1%
Natural Science	5	6	11	79
	5.0%	5.9%	10.9%	78.2%
Social Sciences	15	15	34	104
	8.9%	8.9%	20.2%	61.9%
Total	47	47	113	390
	7.9%	7.9%	18.9%	65.3%
Chi square = 16.84 ; p-value = 0.156				

A notable proportion of researchers (15.8%) remain uncertain or have not preserved their research data, highlighting a significant gap in established preservation practices. In the Humanities, uncertainty levels are high (10.5%), suggesting that researchers face challenges such as lack of institutional guidance, insufficient infrastructure, or inadequate technical support. Similarly, in the Social Sciences, the share of respondents indicating uncertainty or non-preservation (8.9%) reflects the complexities of managing sensitive data, such as interviews or survey responses. Although uncertainty levels are relatively low in the Life Sciences (5.2%) and Natural Sciences (5.0%), the presence of any uncertainty indicates the need for clearer institutional policies and researcher support systems to ensure consistent preservation practices.

An encouraging proportion of respondents (18.9%) expressed their intention to preserve data, reflecting a positive outlook towards improved research data. Respondents from the Humanities category show the highest intention to preserve (25.7%), suggesting willingness to align with open data and preservation norms. The Life Sciences (20.7%) and Social Sciences (20.2%) also reflect moderate levels of intention to preserve, pointing to the emerging preservation cultures in these disciplinary categories. However, the relatively lower figures in Engineering (15.9%) and Natural Sciences indicate that in disciplines where preservation practices are already more established, fewer researchers may be at the planning stage.

The majority of the respondents (65.3%) have already engaged in data preservation, signifying a reasonably strong preservation culture across disciplines. Natural Sciences exhibit the highest preservation rate (78.2%), coupled with the lowest uncertainty, suggesting well-established norms supported by structured funding mechanisms and journal mandates. Similarly, the Life Sciences show a high preservation rate (68.1%), which likely reflects discipline-specific expectations and data management requirements. Engineering (64.5%) and Social Sciences (61.9%) demonstrate moderate but consistent preservation activity, while Humanities show the lowest rate (56.2%), highlighting the need for targeted interventions to strengthen data preservation practices in fields where infrastructural or procedural gaps remain.

Table 7 shows the preservation behaviour of PhD graduates across five academic disciplines. The majority (73.3%) have preserved their work, indicating a strong tendency toward preservation among PhD graduates, and additional 11.1% planning to do so. Natural Science and Life Sciences have the highest preservation rates (88.6% and 83.3% respectively), suggesting stronger cultural or practical norms favouring preservation in these fields. Comparing other disciplines, Engineering and Humanities disciplines have lower preservation rates (62.3% and 65.9%).

Table 7

Data Preservation Behaviour of PhD Graduates

Discipline	Uncertain	No	Plan to preserve	Preserved
Engineering	5	7	8	33
	9.4%	13.20%	15.1%	62.3%
Humanities	4	7	3	27
	9.8%	17.1%	7.3%	65.9%
Life Sciences	1	3	6	50
	1.7%	5%	10%	83.3%
Natural Science	0	4	1	39
	0%	9.10%	2.3%	88.6%
Social Sciences	6	8	14	62
	6.7%	8.90%	15.6%	68.9%
Total	16	29	32	211
	5.6%	10.10%	11.1%	73.3%
	Chi square = 20.94 ; p-value = 0.051			

Data Privacy

> Does your research involve gathering personal information such as addresses, phone numbers, or medical histories?

The ethical handling of personal information is a cornerstone of responsible research, especially when studies involve sensitive data. Ensuring data privacy not only safeguards participants' rights but also maintain trust and compliance with legal and ethical standards, such as data protection regulations. This section examines the privacy measures implemented by researchers, 17.9% of whom reported collecting personal information.

The table 8 reveals that 78.6% of respondents do not collect personal information. This indicate that most researchers do not collect personal information, likely focusing on non-human or non-sensitive data (Experimental Data 40.9% sees Table 1). A notable minority (17.9%) of the respondents collect personal information, requiring careful handling to comply with ethical and legal standards. Engineering (90.7%), Life Sciences (89.7%), and Natural Sciences (87.1%) focusing on non-human data are using non sensitive data. However, 17.9% collect personal information, primarily in Social Sciences (29.2%) and Humanities (27.6%), aligning with data types like Surveys (25.5%) and Interview Transcripts (16.7%). A follow up question explores how these crucial data are handled by researchers.

Table 8

Personal Data Collection Behaviour

Discipline	No	Unsure	Yes
Engineering	97	1	9
	90.7%	0.90%	8.4%
Humanities	65	11	29
	61.9%	10.50%	27.6%
Life Sciences	104	2	10
	89.7%	1.70%	8.6%
Natural Science	88	3	10
	87.1%	3.00%	9.9%
Social Sciences	115	4	49
	68.5%	2.40%	29.2%
Total	469	21	107
	78.6%	3.50%	17.9%

Chi square = 60.88 ; p-value = 0.000

Measures Taken to Safeguard Privacy

What measures have you taken to ensure the privacy of your data? Please select all that apply from the following options. [If your data contains personal information as specified in the previous question]

1. Not applicable; data is not sensitive or confidential
2. Obtained informed consent from the participants to share my data
3. Anonymised/de-identified sensitive or confidential data (Anonymisation is the process of turning data into a form that does not identify individuals.)
4. Destroyed the personal information from research data within a short time of publishing research results
5. Destroy all data within a specified time period
6. No steps are taken to ensure privacy
7. Other:

The follow-up question on privacy sought respondents to select all relevant measures they used to ensure the privacy of the data collected.

Table 9

Measures Taken to Protect Privacy

Privacy measures taken	Count
Not applicable; data is not sensitive or confidential	350
Obtained informed consent from the participants to share my data	106
No steps are taken to ensure privacy	70
Anonymised / de-identified sensitive or confidential data	64
Destroyed the personal information from research data within a short time of publishing research results	16
Destroy all data within a specified time period	8
Data is confidential so system will keep secured always enable all security measures.	2
Keep confidential	1
Kept the collected data in drive	1
Personal information was not published	1
Signed consent from participant to use the data only for the purpose of research, and ensured it won't be published as collected	1

It is important to note that 68 respondents have not taken any steps to ensure the privacy. This indicates the need to raise awareness about handling personal and sensitive data. Respondents has obtained informed consent from the participants to share data, indicating strong adherence to ethical standards. The finding that no privacy measures were taken in 70 instances is concerning, as it indicates that a notable proportion of respondents who collect personal information do not implement any safeguards. Anonymised or de-identified sensitive data accounted for 64 selections (10.3%), showing that some researchers use anonymisation to protect personal information. A small proportion (2.6%), has destroyed personal information from the collected research data within a short time of publishing research results. A minority (1.3%) has destroyed all data within a specified time period, showing data destruction within time limit is not a common practice.

Awareness of Research Data Repositories

> Governments and research institutes have started providing their raw data through open data repositories. Are you familiar with any such online repositories?

Hundreds of discipline-specific repositories and national level administrative data repositories are available. These resources enhance data discoverability, promote collaboration, and ensure long-term preservation. This section examines the awareness and utilisation of such data repositories by researchers. By assessing respondents' familiarity with data repositories, this question aims to identify the extent of awareness.

Table 10

Awareness of Research Data Repositories

Discipline	No	Maybe	Yes
Engineering	49 (8.2%)	13 (2.2%)	45 (7.5%)
Humanities	36 (6.0%)	18 (3.0%)	51 (8.5%)
Life Sciences	54 (9.0%)	12 (2.0%)	50 (8.4%)
Natural Science	46 (7.7%)	12 (2.0%)	43 (7.2%)
Social Sciences	56 (9.4%)	25 (4.2%)	87 (14.6%)
Total	241 (40.4%)	80 (13.4%)	276 (46.2%)

Usage of Data Repositories

If there's useful data available online, would you consider using it for your own research?
• Will use and acknowledge by citation.
• Will use but may not cite.
• Will not use for my study.
• Not decided.
• Other:

The Table 11 presents result on willingness of the respondents to use openly available online data for their research and whether they would cite such data. The table reflects a single-choice question, as each respondent selected one option. The question assesses researchers' openness to data reuse, a critical component of the data use life cycle, and their commitment to ethical citation practices. Use and Cite (80.9%) is the dominant response, indicating a strong willingness across disciplines to use openly available data and acknowledge it through citation.

Table 11

Willingness to Use Online Data Repositories

Discipline	Not decided	No	Use only	Use and Cite
Engineering	11	4	1	91
	1.8%	0.7%	0.2%	15.2%
Humanities	16	1	3	85
	2.7%	0.2%	0.5%	14.2%
Life Sciences	14	6	5	91
	2.3%	1.0%	0.8%	15.2%
Natural Science	18	5	4	74
	3.0%	0.8%	0.7%	12.4%
Social Sciences	18	7	1	142
	3.0%	1.2%	0.2%	23.8%
Total	77	23	14	483
	12.9%	3.90%	2.3%	80.9%
Chi square = 14.34 ; p-value = 0.279				

The question explored whether researchers would use open data in their own research work, and whether they would acknowledge it through the data citation. The results show that most respondents across disciplines are willing to use online data and cite it, with 80.9% selecting this option. This indicates a strong positive attitude toward responsible data reuse.The table shows the usage of data repositories where, Engineering (85.0%) and Social Sciences (84.5%), reflecting openness to reusable data. Respondents from Natural Sciences (17.8%) and Humanities (15.2%) show hesitation, possibly due to low availability of specialised or qualitative data in the repositories. The 80.9% are willing to use and cite data indicates widespread acceptance of data reuse and ethical citation. A very small proportion of respondents (2.4%) indicated that they would use data without citing the creator, highlighting the need for greater awareness of ethical citation practices.

Sharing Attitude

Attitude of respondents towards open access and data sharing is assessed using multiple Likert-scale based questions.

Attitude Towards Open Access Publications

I support Open Access for research articles and other publications, ensuring free public access, especially when publicly funded.

This question intend to assess attitude of participants towards open access publications. Across all the disciplines, 89.1% (support + strongly support) of respondents endorses open access. Only 3.6% has opposed the concept, and 1.0% are unsure, reflecting broad consensus. Social Sciences (97.6%) and Humanities (91.4%) show the highest support, while Natural Sciences (80.2%) is the lowest though still substantial, and higher opposition (11.9%) within disciplines. A small minority of respondents (3.6%), were opposed to open access, and only 6.2% adopted a neutral position towards open access to the publication. The analysis confirms strong support for open access, with significant disciplinary variations (Chi-square = 73.27, p = 0.000). Social Sciences (97.6%) and Humanities (91.4%) lead in support, reflecting higher ideological alignment with open access compared other disciplinary categories.

Table 12

Support for Open Access to Research Publications

Discipline	I Don't Know	Strongly Oppose	Oppose	Neutral	Support	Strongly Support
Engineering	3	0	1	16	33	54
	2.8%	0%	0.90%	15.00%	30.8%	50.50%
Humanities	3	0	3	3	33	63
	2.9%	0%	2.9%	2.90%	31.4%	60.00%
Life Sciences	0	3	2	7	36	68
	0%	2.60%	1.7%	6.00%	31.0%	58.60%
Natural Science	0	5	7	8	39	42
	0%	5.00%	6.9%	7.90%	38.6%	41.60%
Social Sciences	0	0	1	3	46	118
	0%	0%	0.60%	1.80%	27.4%	70.20%
Total	6	8	14	37	187	345
	1.0%	1.30%	2.3%	6.20%	31.3%	57.80%
Chi square = 73.27; p-value = 0.000						

Attitude Towards Data Sharing

> After researchers finish their studies and publish their papers, they often either delete the data they used or keep it stored away. These data can be shared and reused. Sharing this data can have many benefits, It promotes open discussion, transparency and improves the quality.
>
> I believe that research data should be made available through a repository for others to see and reuse?

The objective of the study was to understand the attitude towards research data sharing. A 5-point Lickert scale was used to measure the attitude of the research scholars. To make sure that the respondents have fully understood the context of the question, proper description is also provided along with the question. Based on their level of agreement with the provided statement, this question intend to analyse the attitude of doctoral students towards the sharing and reuse of research data through repositories. Majority of respondents (80%) agree or strongly agree that research data should be made available for others to access and reuse. This reflects a strong positive attitude towards data sharing. Opposition to data sharing is minimal, with most disciplines showing less than 5% combined disagreement.

The majority in each discipline either agree or strongly agree that research data should be made available for reuse via repositories. Respondents from Humanities exhibit the highest level of strong agreement (52.4%) with the statement that research data should be shared through repositories followed by Social Sciences (50.0%). Life and Natural Sciences categories show strong but balanced agreement. These disciplines report a more even distribution between "agree" and "strongly agree". Neutral responses are relatively high in Life Sciences (16.4%) and Humanities (15.2%), indicating the need for training to convert passive agreement into active data sharing behaviour. Majority of the respondents agree that research data should be made available through open repositories.

Table 13

Attitude Towards Research Data Sharing

Discipline	I Don't Know	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Engineering	1	0	2	14	40	50
	0.90%	0.00%	1.9%	13.10%	37.4%	46.7%
Humanities	4	1	5	16	24	55
	3.8%	1.00%	4.8%	15.20%	22.9%	52.4%
Life Sciences	2	1	3	19	46	45
	1.7%	0.90%	2.6%	16.40%	39.7%	38.8%
Natural Science	1	1	4	14	42	39
	1.0%	1.00%	4.0%	13.90%	41.6%	38.6%
Social Sciences	2	5	4	20	53	84
	1.2%	3.00%	2.4%	11.90%	31.5%	50.0%
Total	10	8	18	83	205	273
	1.7%	1.30%	3.0%	13.90%	34.3%	45.7%

Chi square = 23.48; p-value = 0.266

Factors Motivating Research Data Sharing

Please rate how important the following factors are while considering to share your research data openly ?

Effective research data management and the practice of data sharing are recognised as essential components of high-quality scientific inquiry (Tenopir et al., 2011; Pampel et al., 2013). However, researchers' participation in these practices is influenced by a complex interplay of motivating and constraining factors, ranging from perceived benefits and professional incentives to practical challenges such as time, skills, and institutional support. This section examines the motivating factors that encourage researchers to share their data.

Financial Reward

The motivating factor of financial reward for sharing data is examined to understand its influence on data-sharing practices.

Table 14

Financial Reward as a Motivating Factor

Discipline	Not at all Important	Not Important	Slightly Important	Very Important	Extremely Important
Engineering	14	25	24	28	16
	13.1%	23.40%	22.4%	26.20%	15.0%
Humanities	15	26	28	31	5
	14.3%	24.80%	26.7%	29.50%	4.8%
Life Sciences	10	26	36	26	18
	8.6%	22.40%	31.0%	22.40%	15.5%
Natural Science	22	15	21	24	19
	21.8%	14.90%	20.8%	23.80%	18.8%
Social Sciences	24	48	38	29	29
	14.3%	28.60%	22.6%	17.30%	17.3%
Total	85	140	147	138	87
	14.2%	23.50%	24.6%	23.10%	14.6%
Chi square = 29.03; p-value = 0.024					

The responses are fairly distributed, suggesting a range of opinions on the role of financial incentives in data sharing. A combined 47.7% of the respondents (the sum of “very important” and “extremely important”) rate financial benefit as significant. Chi-square = 30.07, p = 0.024 (significant at p = 0.05), indicates a statistically significant association between discipline and financial achievement, suggesting that the perceived importance of financial incentives varies across disciplines.

Low Importance overall, across all disciplines, 37.7% (not at all + not important) rate financial incentives as unimportant, compared to 37.7% (very + extremely) rating them important. Natural Sciences show the highest combined motivation (42.6%) toward financial incentives for data sharing. Humanities have the lowest “extremely important” response (4.8%), indicating lower sensitivity to financial motivation. Humanities, least likely to rate financial incentives as “extremely important” (4.8%). In the opinion of 14.2% of the respondents, the financial reward received for sharing data is not at all an influencing factor in the data sharing decision.

Academic Visibility

Increased visibility through data reuse and citation is a powerful motivator for researchers to share their datasets, enhancing their academic reputation and impact. This section investigates how the prospect of visibility encourages researchers to share data types particularly given the 80.9% willingness to use and cite online data. By exploring this factor, the study seeks to highlight strategies that amplify visibility to promote data sharing and foster a culture of acknowledgment within Kerala’s research community.

A majority of the respondents (85.6%) rate visibility as either very or extremely important, underscoring that academic recognition is a key motivator for data sharing across disciplines. Only 2.3% of the respondents think that visibility is not at all a factor in the data sharing behaviour. Engineering scholars show strong motivation from visibility and in Humanities while visibility matters, there is a greater variation. Life Sciences has highest among all disciplines with a combined recognition of 89.7%. In Humanities the combined value is 75.3 which is lower but covers a two third of the total respondents.

Table 15

Visibility as a Motivating Factor

Disciplines	Not at all Important	Not Important	Slightly Important	Very Important	Extremely Important
Engineering	2	5	7	45	48
	0.30%	0.80%	1.2%	7.50%	8.0%
Humanities	3	5	18	47	32
	0.50%	0.80%	3.0%	7.90%	5.4%
Life Sciences	2	2	8	40	64
	0.30%	0.30%	1.3%	6.70%	10.7%
Natural Science	5	4	5	40	47
	0.80%	0.70%	0.80%	6.70%	7.9%
Social Sciences	2	2	16	77	71
	0.30%	0.30%	2.7%	12.90%	11.9%
Total	14	18	54	249	262
	2.3%	3.00%	9.0%	41.70%	43.9%
Chi square =30.42; value = 0.016					

Visibility is a major motivating factor for data sharing across all disciplines. More than 85% consider visibility either very or extremely important. Life Sciences and Social Sciences show the strongest emphasis on visibility. Humanities show more variation, with some respondents rating visibility as only slightly important. Very few respondents consider visibility unimportant. Differences across disciplines are statistically significant ($p = 0.016$). Overall, visibility clearly emerges as a key driver of researchers' willingness to share data.

Control Over Data

The ability to select individuals or groups permitted to access and use research data is a critical motivator for data sharing. In the survey, control was highly valued, as it ensures researchers retain control over how their data is utilised. This section examines how the desire for control influences researchers' willingness to share sensitive data.

Table 16

Control over data as a Motivating Factor

Disciplines	Not at all Important	Not Important	Slightly Important	Very Important	Extremely Important
Engineering	10	17	24	30	26
	9.3%	15.90%	22.4%	28.00%	24.3%
Humanities	6	13	23	41	22
	5.7%	12.40%	21.9%	39.00%	21.0%
Life Sciences	8	6	17	56	29
	6.9%	5.20%	14.7%	48.30%	25.0%
Natural Science	9	14	14	35	29
	8.9%	13.90%	13.9%	34.70%	28.7%
Social Sciences	7	14	39	66	42
	4.2%	8.30%	23.2%	39.30%	25.0%
Total	40	64	117	228	148
	6.7%	10.70%	19.6%	38.20%	24.8%
Chi square = 24.34 ; p-value = 0.082					

The majority of the respondents (62.9%) rated control over data as either "Very Important" (38.2%) or "Extremely Important" (24.8%), suggesting a strong belief that maintaining control over research data is a key factor when considering data sharing. Only 6.7% of respondents believe that control over data does not influence their decision of data sharing. This is likely to affect their willingness to share data, as scholars may be concerned about retaining ownership and managing who accesses or uses their data. Researchers from the

Life Sciences seek control over their data compared to other disciplines with a highest combined attitude towards control of data (73.3%). Engineering disciplines shows the lowest with 52.3% view control as very or extremely important.

Shared Authorship

Recognition as a co-author on publications arising from shared data serves as a strong incentive for researchers. It not only gives them academic credit but also increases the visibility and reuse of their datasets, creating opportunities for further collaboration. This section explores the motivating factor of co-authorship in data sharing, assessing its influence on researchers' willingness to share data.

Table 17

Shared Authorship as a Motivating Factor

Disciplines	Not at all Important	Not Important	Slightly Important	Very Important	Extremely Important
Engineering	7	15	24	30	31
	6.5%	14.00%	22.4%	28.00%	29.0%
Humanities	10	10	32	26	27
	9.5%	9.50%	30.5%	24.80%	25.7%
Life Sciences	8	6	18	49	35
	6.9%	5.20%	15.5%	42.20%	30.2%
Natural Science	11	12	24	28	26
	10.9%	11.90%	23.8%	27.70%	25.7%
Social Sciences	11	19	36	58	44
	6.5%	11.30%	21.4%	34.50%	26.2%
Total	47	62	134	191	163
	7.90%	10.4%	22.40%	32.00%	27.30%
	Chi square = 20.45 ; p-value = 0.201				

Overall percentage who believe that shared authorship either very important or extremely important is 59.2. Table 17 shows that researchers from Life Sciences discipline has very strong preference for shared authorship as a motivation for data sharing (42.2% “Very Important”, 30.2% “Extremely Important”). This likely reflects the collaborative nature of biomedical research and the need to receive academic credit for data generation and management. In Social Sciences high proportion of respondents valuing shared authorship (60.7% for “Very” and “Extremely Important”). Engineering shows 57% rate co-authorship as "Very" or “Extremely Important”. Researchers from Natural Sciences believe Nearly 53.4% consider it “Very” or “Extremely Important”. Researchers from Humanities believe that 50.5% rate co-authorship as important, but also the highest percentage of “Slightly Important” (30.5%).

Embargo Period

This section investigates the role of embargo periods in encouraging researchers to share data. An embargo period, which allows researchers to delay public access to their data for a specified time, can motivate data sharing by balancing the need for control with the benefits of eventual openness. Embargo periods, are highly valued (68.1% “very important” and “extremely important”), enhancing researchers’ willingness to share data. Across the disciplines, embargoes are generally perceived as important controlling mechanism, with over two-thirds rating them as “Very” or “Extremely Important”. Overall, 68.1% researchers believes that embargo is an important factor in the research data sharing behaviour and, 23.7% researchers has the opinion that embargo is not or slightly important while considering data sharing. For a small group of researchers (8%), embargo does not affect the data sharing behaviour.

Most researchers across all disciplines view embargoes as an important motivator for data sharing. Life Sciences and Natural Sciences show the strongest support. Humanities show moderate but consistent support, with some variation. Only a small minority consider embargoes unimportant. Disciplinary differences are not statistically significant ($p = 0.193$). Overall, embargoes emerge as a highly valued mechanism, giving researchers protected time to produce publications before releasing their data.

Table 18

Embargo as a Motivating Factor

Discipline	Not at all Important	Not Important	Slightly Important	Very Important	Extremely Important
Engineering	6	11	20	35	35
	5.6%	10.30%	18.7%	32.70%	32.7%
Humanities	9	10	11	47	28
	8.6%	9.50%	10.5%	44.80%	26.7%
Life Sciences	10	6	20	36	44
	8.6%	5.20%	17.2%	31.00%	37.9%
Natural Science	10	12	12	26	41
	9.9%	11.90%	11.9%	25.70%	40.6%
Social Sciences	13	9	31	58	57
	7.7%	5.40%	18.5%	34.50%	33.9%
Total	48	48	94	202	205
	8.0%	8.00%	15.7%	33.80%	34.30%
	Chi square = 20.64 ; p-value = 0.193				

Comparison of Motivating Factors

The table number 19 presents a comparison of factors motivating data sharing based on a sample of 597 respondents, with each factor rated on a scale from 1 to 5 (1 being least important, 5 being extremely important). Visibility is the highest-rated motivating factor, with a mean score of 4.22, indicating it is the most compelling reason for sharing data. The low standard deviation (0.90) suggests strong consensus among respondents, with most rating it highly. Embargo Period ranks second, with a mean of 3.78, suggesting that having control over when data becomes publicly available is a significant motivator. The standard deviation (1.23) indicates moderate variability in responses, reflecting differing preferences for embargo durations.

With a mean of 3.64, the ability to control who can access and use the data is also a strong motivator. The standard deviation of 1.16 shows some variability, suggesting that while many value this control, the importance varies across respondents. Being treated as a co-author in publications using the shared data has a mean of 3.60, indicating it is a moderately strong motivator. The standard deviation 1.21 reflects moderate variability, likely due to differences in how much individuals prioritise formal recognition through authorship. Financial Reward is the least motivating factor, with a mean score of 3.00, exactly at the midpoint of the scale. The highest standard deviation (1.27) indicates the greatest variability in responses, suggest that some respondents value financial incentives, while others do not.

Table 19

Comparison of Motivating Factors

Motivating Factor	Mean	Std Dev	Minimum	Maximum
Financial reward I receive for sharing my data.	3.00	1.27	1	5
Increased visibility I get when others reuse and cite my data.	4.22	0.90%	1	5
My ability to select the individuals or groups permitted to access and use my data	3.64	1.16	1	5
If I am treated as a co-author in the publication generated with my research data.	3.60	1.21%	1	5
Embargo period	3.78	1.23	1	5

The data highlights that non-monetary incentives, particularly increased visibility (citations and reuse) and control over data access (embargo periods and control over data), are the strongest drivers for data sharing. These align with academic values like recognition, control, and protecting research integrity. Co-authorship is also important, reflecting a desire for formal acknowledgment in research outputs, though it is slightly less compelling than visibility and control. Financial rewards are the least motivating, with the highest variability. This suggests that while some researchers may be incentivised by money, it is not a universal driver, possibly because academic research is often driven by intrinsic motivations or institutional rewards rather than direct financial gain. The standard deviations indicate that visibility has the most consistent appeal (lowest Std Dev), while financial rewards are the most divisive (highest Std Dev).

Challenges in Data Sharing

Various studies has pointed out that, challenges in research data management significantly influence the sharing behaviour. Statements regarding major challenges expected in the data sharing were prepared with the help of previous studies. Using the 5 point Lickert scale, participants were asked to record their opinion on these statements by choosing from ‘strongly agree’ to ‘strongly disagree’. These responses were coded as described in the section 4.2. The fear of misinterpretation, need for additional effort and time for data preparation, data confidentiality, acknowledging the original data creator and usefulness of the data created are the main challenges identified in the previous studies. This section analyse these challenges in detail. The analysis helps in understanding the scale and depth of challenges and if these are influenced by any other variables or are uniform across variables or have any impact.

Fear of Misinterpretation

My research data could possibly be misinterpreted or misreported.

The concern that research data could be misinterpreted or misreported is a significant barrier in data sharing. Respondents fear that improper use may undermine their findings or reputation. Table 20 shows overall, 51.6% of respondents (agree + strongly agree) expressed concern about misinterpretation while reusing the data by others. Social Sciences show the highest concern with 60.7% of respondents agree to the statement. Humanities follow with 58.1%, suggesting shared interpretive paradigms could increase fear of misuse. Life Sciences (41.3%) and Engineering (44.9%) report comparatively lower levels of misinterpretation concern. Natural Sciences are in the middle range, with only 48.5% expressing concern.

Table 20

Data Misinterpretation

Discipline	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Engineering	6	19	34	31	17
	1.00%	3.20%	5.70%	5.20%	2.8%
Humanities	4	16	24	38	23
	0.70%	2.70%	4.00%	6.40%	3.9%
Life Sciences	17	20	31	33	15
	2.80%	3.40%	5.20%	5.50%	2.5%
Natural Science	2	22	28	31	18
	0.30%	3.70%	4.70%	5.20%	3.0%
Social Sciences	8	25	33	68	34
	1.30%	4.20%	5.50%	11.40%	5.7%
Total	37	102	150	201	107
	6.2%	17.1%	25.1%	33.7%	17.90%

Effort Required for Data Sharing

Major effort and time is required to prepare the data in a shareable format.

The question attempts to understand the relationship between data sharing and the additional effort that researchers are willing to take while preparing the data in a shareable format. Results shows that major effort is required to prepare the data in a shareable format. This variable is particularly important for understanding the behavioural intention to share data. Results from the table 21 shows that 79.1% of respondents believe that data sharing requires effort and time. This results highlight the high perceived effort across disciplines, with Humanities and Life Sciences showing the highest agreement. Very few respondents disagree with the statement, indicating widespread recognition of the workload involved in preparing shareable datasets. Differences across disciplines are not statistically significant ($p = 0.524$).

Table 21

Perceived Effort Required for Data Preparation

Discipline	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Engineering	0	7	15	48	37
	0%	6.50%	14.00%	44.90%	34.60%
Humanities	0	3	13	61	28
	0%	2.90%	12.40%	58.10%	26.70%
Life Sciences	1	7	12	64	32
	1%	6.00%	10.30%	55.20%	27.60%
Natural Science	1	10	14	46	30
	1%	9.90%	13.90%	45.50%	29.70%
Social Sciences	3	15	24	74	52
	1.80%	8.90%	14.30%	44.00%	31.00%
Total	5	42	78	293	179
	0.80%	7.00%	13.10%	49.10%	30.00%
	Chi square = 15.01 ; p-value = 0.524				

Data Confidentiality

> My data is confidential in nature and cannot be shared.

The confidential nature of certain research data poses a significant barrier to sharing, as researchers must prioritise participant privacy and compliance with ethical and legal standards. Researchers who deal with confidential data are often worried about sharing their data publicly. Overall, 36% of the respondents agree or strongly agree that data confidentiality is a concern while sharing data. Social Sciences show the highest level of agreement, suggesting this fear is a more significant factor influencing their sharing behaviour compared to other disciplines. Researchers show mixed attitudes toward sharing confidential data, with no dominant trend across the sample. Neutral responses are the highest, indicating uncertainty or conditional approaches to confidentiality. Humanities and Life Sciences express more caution, whereas Social Sciences are comparatively more supportive. Differences across disciplines are not statistically significant ($p = 0.51$). Overall, the results suggest that confidentiality remains a complex and sensitive issue, requiring clear safeguards, ethical guidelines, and institutional support to enable responsible data sharing.

Table 22

Data Confidentiality as a Challenge

Discipline	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Engineering	15	26	28	27	11
	0.14	0.24	0.26	0.25	0.1
Humanities	6	34	26	26	13
	0.06	0.32	0.25	0.25	0.12
Life Sciences	14	27	37	28	10
	0.12	0.23	0.32	0.24	0.09
Natural Science	9	31	32	19	10
	0.09	0.31	0.32	0.19	0.1
Social Sciences	16	39	43	43	27
	0.1	0.23	0.26	0.26	0.16
Total	60	157	166	143	71
	0.1	0.26	0.28	0.24	0.12
Chi square = 15.18 ; p-value = 0.51					

Acknowledgment for Data Sharing

As a data creator, I fear that the data may be used without due acknowledgment.

The fear that shared data may be used without due acknowledgment is a barrier in data sharing. Researchers seek recognition for their intellectual contributions. This variable touches upon fear of not acknowledging original data creator while sharing data. The fear that data may be used without acknowledging the original data creator is influencing the data sharing behaviour.

Table 23

Acknowledging Data Creator

Discipline	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Engineering	1	15	21	39	31
	0.20%	2.50%	3.5%	6.50%	5.2%
Humanities	3	3	19	45	35
	0.50%	0.50%	3.2%	7.50%	5.9%
Life Sciences	5	16	13	45	37
	0.80%	2.70%	2.2%	7.50%	6.2%
Natural Science	1	8	17	48	27
	0.20%	1.30%	2.8%	8.00%	4.5%
Social Sciences	5	10	27	63	63
	0.80%	1.70%	4.5%	10.60%	10.6%
Total	15	52	97	240	193
	2.5%	8.70%	16.2%	40.20%	32.3%
Chi square = 24.11 ; p-value = 0.087					

A substantial majority of researchers (72.5%) fear that those who reuse data may not acknowledge the original data creator. Social Sciences and Humanities show the highest level of concern, but all fields exhibit similar trends. Only a small minority express disagreement, indicating near-universal recognition of attribution responsibilities. Differences across disciplines are not statistically significant (p = 0.087). Overall, concern about acknowledgment emerges as a major factor influencing attitudes toward data sharing.

Usefulness of the Data

> I do not think that my data could be useful to others.

The results indicate that a majority of the researchers perceive their data as valuable to others, with 57.6% disapproving the statement. Specifically, 29.8% disagreed and 27.8% strongly disagreed, reflecting a strong belief in the utility of their data. Only 19.1% of respondents agreed that their data lacks usefulness, while 23.3% remained neutral. There is a significant association between discipline and the belief that "My data may not be useful to others". Researchers in Life Sciences and Engineering expressed greater confidence in their data's usefulness, likely due to the standardised or broadly applicable nature of data in these fields. Neutral responses are substantial across all fields, reflecting a common uncertainty about whether data generated in one project is relevant for others. This perception is a known barrier to data sharing, as researchers may undervalue the potential for replication, secondary analysis, or educational reuse. Many researchers across disciplines believe their data may not be useful to others, reflecting a notable barrier to data sharing. Overall, 57.6% of the respondents selected Disagree (29.8%) or Strongly Disagree (27.8%), indicating that most researchers do believe their data has value to others.

Table 24

Usability of Research Data

Discipline	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Engineering	5 0.80%	9 1.50%	27 4.5%	43 7.20%	23 3.9%
Humanities	6 1.0%	19 3.20%	18 3.0%	33 5.50%	29 4.9%
Life Sciences	6 1.0%	16 2.70%	22 3.7%	25 4.20%	47 7.9%
Natural Science	8 1.3%	12 2.00%	31 5.2%	27 4.50%	23 3.9%
Social Sciences	11 1.8%	22 3.70%	41 6.9%	50 8.40%	44 7.4%
Total	36 6.0%	78 13.10%	139 23.3%	178 29.80%	166 27.8%
	Chi square = 26.68; p-value = 0.045				

Comparison of Challenging Factors

The table 25 shows that the effort and time required for sharing data is a significant challenge in the data sharing. The mean score of 4.00 for the effort required indicates that researchers find a significant challenge in sharing their data. This is supported by the relatively low standard deviation, suggesting consensus on this point. With means of 3.91 and 3.41 respectively, acknowledgment and misinterpretation concerns highlight the importance of ensuring credit is given where due and of addressing potential misinterpretations of shared data. Ensuring that they receive proper acknowledgment for their data contributions is a strong concern among researchers with mean score 3.91 and standard deviation of 1.03. The moderate mean scores for confidentiality (3.02) and usability (3.61) suggest that while these are concerns, they are not as significant as the effort required to share data. Nonetheless, they are still important areas to address, particularly for sensitive data and data with unclear reuse potential.

Table 25

Comparison of Challenging Factors

Concern	Mean	Std Dev.	Minimum	Maximum
My research data could possibly be misinterpreted or misreported	3.40	1.15	1	5
Major effort and time is required to prepare the data in a shareable format	4.00	0.89%	1	5
My data is confidential in nature and cannot be shared	3.01	1.18	1	5
As a data creator, I fear that the data may be used without due acknowledgment.	3.91	1.03%	1	5
I do not think that my data could be useful to others	3.60	1.19	1	5

Motivations vs Challenges

Among the disciplines, the students of Life Sciences report the highest motivation score (19.0). This may reflect stronger norms of collaboration and data reuse in this field. Social Sciences also show relatively high motivation (18.4), possibly linked to the value of data for secondary analysis and policy research. Engineering (18.1) and Natural Sciences (18.0) report similar levels of motivation. Humanities show the lowest motivation score (17.6), which may be related to discipline-specific practices that emphasise individual authorship and context-specific data.

Table 26

Combined Motivating and Challenging Factors

Item	Motivating Factors	Challenging Factors
Engineering (n = 107)		
Mean	18	17.8
Std. Deviation	4	2.7
Humanities (n = 105)		
Mean	18	18.3
Std. Deviation	4	2.8
Life Sciences (n = 116)		
Mean	19	17.6
Std. Deviation	4	2.6
Natural Science (n = 101)		
Mean	18	17.6
Std. Deviation	5	2.5
Social Sciences (n = 168)		
Mean	18	18.2
Std. Deviation	4	2.8
Total (n = 597)		
Mean	18	17.9

Std. Deviation	4	2.7

In terms of challenges, Humanities report the highest mean score (18.3). This suggests that researchers in this field face more barriers, such as ethical concerns and difficulties in preparing qualitative data for sharing. Social Sciences also report relatively high challenges (18.2), likely due to privacy issues and the effort involved in handling human-subject data. Researchers in Engineering, Life Sciences, and Natural Sciences report slightly lower challenge scores, though challenges remain present across all fields. The standard deviations indicate some variation within disciplines, especially for motivating factors. However, perceptions of challenges are more consistent, suggesting shared concerns across researchers.

Table 27

Mean Value of Motivating and Challenging Factors

Factors	Frequency	Mean	Std Dev	Minimum	Maximum
Motivating	597	18.25	4.01	5.0	25.0
Challenging	597	17.93	2.70	9.0	25.0

The table 27 shows combined values of both challenging and motivating factors. The mean score of 18.25 (average 3.7) indicates that, on an average, the total population perceives motivating factors for preservation as moderately to highly influential. The standard deviation 4.01 suggests considerable variability in perceptions of motivation. The wide range 5.0 to 25 indicates that some individuals experience very low motivation, while others are highly motivated, reflecting diverse contexts or priorities. The mean score of 17.93 suggests that challenging factors are also perceived as moderately to highly influential. The standard deviation of 2.7 indicates less variability in perceptions of challenges compared to motivations. The range 9 to 25 shows that no individual has rated challenges below 9.0, implying that all participants encountered at least moderate barriers to preservation. The mean scores for Combined Motivating and Challenging factors differ by only 0.32 points, suggesting that the perceived strength of motivating factors is nearly equivalent to that of challenging factors.

The standard deviation for Combined Motivating is notably higher than for Challenging Combined, indicating greater heterogeneity in perceptions of motivating factors compared to challenges.

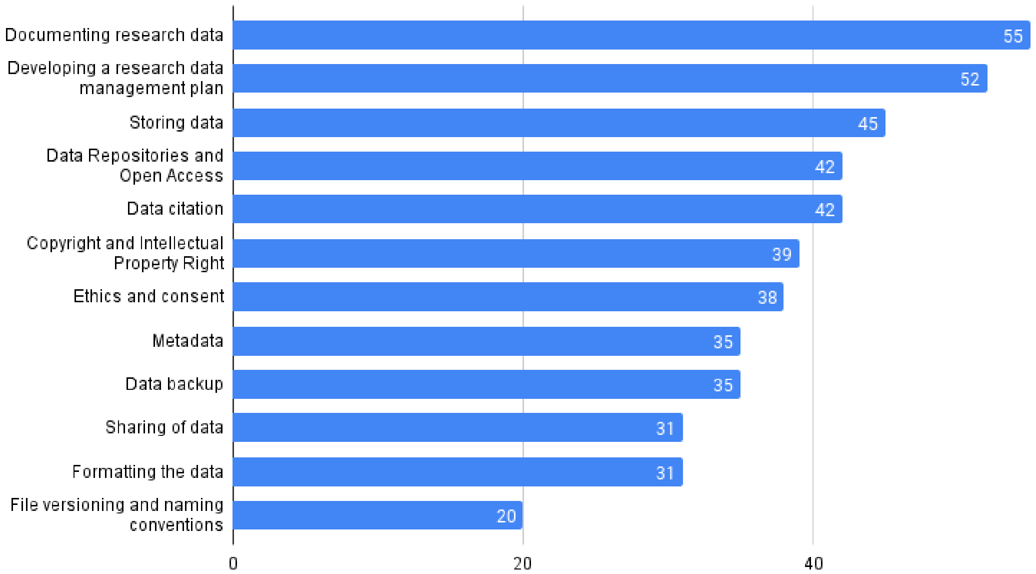
Training and Awareness

Please indicate the areas you would be interested receiving data management training in. [Tick all that apply]

The Figure 3 present data aimed at understanding researchers' training requirements for research data management. Respondents could select multiple options, and the table shows that all respondents chose at least one topic, with 42.2% selecting five or more, highlighting significant interest in RDM training.

Figure 4.3

Training in Research Data Management



Respondents selected multiple training areas, with a total of 2,670 selections, averaging approximately 4.47 topics per respondent (2,670 ÷ 597). Documenting research data (315, 55%) and developing a research data management plan (300, 52%) are the most sought-after topics, reflecting a desire to improve data documentation and strategic planning, critical for data sharing and reuse. Storing data Data Repositories and Open Access, and Data citation are also highly desired, indicating interest in secure storage, repository engagement, and ethical citation practices. File versioning and naming conventions is the least opted, suggesting lower priority for technical aspects of file organisation. The earlier finding that, 80% of researchers across disciplines identify time and effort as major barriers to data sharing can be read along with the training need.

Analysis of Open-Ended Comments

> If you would like to expand on any of the above answers or make further comments, please do so here;

Open ended free text option to record comment is used by 39 (6.5%) respondents. The comment section provided with valuable insights regarding the attitude of researchers towards data sharing. While there's clear support for the potential benefits, concerns about data sensitivity, misuse, and the need for adequate infrastructure and training are prominently voiced. The analysis highlights the importance of addressing these concerns and creating a supportive environment that enables responsible and beneficial data sharing practices. Openly shared data can be reused and cited by other researchers, increasing the visibility and impact of the original research. This wider dissemination of research outputs can lead to more collaborations and contribute to a researcher's career advancement.

Transparency and openness are crucial for scientific rigour. Providing access to the underlying data allows for scrutiny and independent verification of research findings, enhancing the credibility and reliability of research. A PhD graduate from Social Sciences discipline states “Data sharing and open access will beneficial for understanding and finalising problems in a different angle”. Another respondent from Humanities discipline agree that even if there will be challenges these data belongs to the public. “We research scholars receiving public fund from government. So we have to share our findings to the public. There will be lots of problems but it will useful for future generation.”

Table 28

Comments of the Participants

Comment	Theme
Data sharing and open access will beneficial for understanding and finalising problems in a different angle. (Doctoral Student, Social Science) We research scholars receiving pubic fund from government. So we have to share our findings to the public. There will be lots of problems but it will useful for future generation (Doctoral Student, Humanities)	Supportive Data Sharing
Measures for data security should be made aware to Research Scholars and include in the Research Curriculum. Open access is gaining importance such platform should be used for data sharing (Doctoral Student, Social Science)	Need for Training
Depends on the nature of data. Some data may be useful only to address your research question. And if it is qualitative, it may contain lot of personal information. (PhD Graduate, Social Sciences)	Confidentiality of data
Data repository services are provided by major publishing platforms like Elsevier to store data used for writing articles. Likewise, either universities or Research Centres other wise there should be a National Repository Services under central ministry to store and maintain research data in online platforms. (Doctoral Student, Social Science)	Capacity Building
Data collecting methods are very crucial. (PhD Graduate, Social Sciences)	Data Quality
Data sharing is ideally good. The problem is with the creation of ideal contexts. When advancement of knowledge is the aim of research, all these seems good and commendable. Is advancement of the frontiers of knowledge the aim of research here is a million dollar question. (PhD Graduate, Social Sciences)	Research culture
The biggest concern is if the data would be misinterpreted or misused. May be a preview of the data can be shown, but I should be able to authorise who can or can't use the data. And proper acknowledgment should be given in publication (Female, Psychology, Social Sciences)	Controlled Access to data

Many respondents are though supportive of data sharing, have fears or ethical concerns. Concerns are in sharing certain data types, particularly those involving personal or sensitive information or research on marginalised communities. Need to consider privacy issues and potential harm to individuals or groups when making data sharing decisions. Concerns about chances of data misinterpretation, misuse, or plagiarism without proper attribution are the concerns of many respondents.

> Sharing of data and research outputs is most important but so is proper attribution and ways and means to prevent outright misuse, plagiarism and avoiding giving credit where it is due. All publicly funded research needs to be subject to open access except those with strategic value or requiring a reasonable embargo period (PhD Graduate, Life Sciences).

Researchers are hesitant to share their data due to the fear of losing control over how their work is used and potentially losing recognition for their contributions. Respondents wants to have control over their data or restricted access to data as put by a researcher from the Engineering discipline.

> I am not agreeing with the complete sharing of data, since someone had taken a lot of efforts to create such data and others are just consuming it without any effort. So in my opinion, some criteria should be taken before sharing the data. (PhD Graduate, Engineering)

Respondents have emphasised the need for providing researchers with adequate training in data management, sharing practices, and ethical considerations. Researchers understand the need for proper training in data sharing to prepare the data without violating privacy, and at the same time protecting the interest of the data creator. Need for Data Repositories and Policy Frameworks especially a mandatory law to enforce data sharing is also expressed. Need for the incentives and recognition or reward system is also an important factor in data sharing. Creation of a national level framework for data storage and access.

Major Findings from the Survey

1. Majority of the respondents have positive attitude towards research data sharing.
2. Respondents believe that their research data will be useful to others for further analysis and research.
3. Experimental data, databases, and surveys are the common data types, reflecting a structured and shareable data.
4. Text documents, spreadsheets, and statistical data are the common file formats used by respondents.
5. The dominance of personal computer and email for data backup highlights a decentralised, individual-centric approach in data storage.
6. The low use of institutional storage indicates a critical gap in the institutional level research data management facilities in Kerala.
7. The results indicate that majority of the respondents maintain or plan to preserve metadata.
8. The low level of awareness of data repository reveals disconnect between researcher awareness and availability of the data repositories.
9. Strong willingness to use and cite data indicates a widespread acceptance of data reuse and ethical citation.
10. Majority of respondents rated visibility as important, underscoring that academic recognition is a key motivator for data sharing across disciplines.
11. The effort required to prepare data in a shareable format is a major challenge in data sharing.
12. Respondents expressed concern about misinterpretation while others reusing their data.
13. Results suggest that confidentiality of data is a moderate challenge in data sharing, with Social Science category showing higher concerns.
14. Respondents from Life Sciences and Engineering disciplines show relatively higher confidence in their data and have less concern about usability.

15. Results underscores the need for training in Research Data Management. Documenting research data and developing research data management plan are the preferred topics for training.

Chapter Summary

This chapter examined the results of survey data collected to understand the research data practices among PhD graduates and research students from research centres under various universities in Kerala. Survey data was collected using online questionnaire. Research data practices and sharing attitude were analysed across different disciplinary categories. The survey results reveals a complex landscape of research data management in Kerala. A major finding of the quantitative analysis is the strong positive attitude of the respondents towards data sharing. Researchers are open to sharing, but backup practices and metadata use need improvement. Visibility is the major data sharing motive and the effort and time to prepare data is a major concern of the respondents. Inter disciplinary variation are found in certain aspects like data type, file type, data privacy and; this was more evident between science and humanity disciplines.

Chapter 5. *Qualitative Analysis*

> People foolishly imagine that the broad generalities of social phenomena afford an excellent opportunity to penetrate further into the human soul; they ought, on the contrary, to realise that it is by plumbing the depths of a single personality that they might have a chance of understanding those phenomena - Marcel Proust, *In Search of Lost Time.*

This chapter reports results from the qualitative phase of the study. Semi-structured interviews were conducted with research supervisors, whose perspectives as senior academicians and mentors add a crucial interpretive layer. Research supervisors' lived experiences and long term engagement with institutional practices provides an in-depth analysis of the data sharing behaviour.

Selection of Interview Participants

The selection of interview participants was guided by a set of purposefully defined criteria. The criteria ensures the relevance, diversity, and informational richness of the sample. Experience, representation in the academic policy making bodies, international collaboration and number of publications and PhD awarded were considered while selecting interview participants. Research supervisors from different career stages were included. Participants comprises six early career Assistant Professors, six Associate Professors and eight Professors. Table 5.1 summarises the participants of the interview, a total of 20 research supervisors participated in the interview were selected for the final analysis.

Selected participants represent a cross section of the research supervisors from five different disciplines. Many participants are policy makers in various academic bodies like Board of Studies and Academic Councils of the different Universities. For instance, H1 and S1 are holding prominent posts in policy making bodies and are experienced in policy formulation. Participants L1 and N1 have experience of working in various foreign premier institutes. Sample include award winning researchers, prominent authors, reviewers and editors of leading academic journals. Interview participants constitute a wide spectrum considering age, experience, disciplinary categories and institutions. Research supervisors L2 and E1 are experienced in data management and sharing.

Table 29

Profile of Interview Participants

Code	Designation	Gender	Experience in Years
Engineering			
E1	Professor	Male	25
E2	Associate Professor	Male	15
E3	Assistant Professor	Male	13
E4	Assistant Professor	Male	7
Humanities			
H1	Professor	Male	30
H2	Professor	Female	22
H3	Associate Professor	Female	15
H4	Assistant Professor	Female	5
H5	Associate Professor	Male	14
Life Sciences			
L1	Professor	Male	20
L2	Assistant Professor	Male	9
L3	Professor	Male	25
Natural Sciences			
N1	Senior Professor	Male	30
N2	Professor	Male	25
N3	Professor	Female	22
Social Sciences			
S1	Associate Professor	Male	16
S2	Associate Professor	Female	15
S3	Associate Professor	Female	14
S4	Professor	Male	26
S5	Assistant Professor	Male	10

Preparation of Interview Transcript

The interview helped to answer the 'why' questions which was not possible to answer in a quantitative study. Though majority of the research scholars has agreed to share their data, opinion shared by research supervisors were not in tune with that of quantitative result. During the first round of coding, it became evident that drawing meaningful connections between the various elements of the interview data was challenging due to its complexity and volume. As Dey (1993) notes, "we are dealing with complex and voluminous data, diagrams can help us disentangle the threads of our analysis and present results in a coherent and intelligible form". To manage this complexity, Visual Understanding Environment (VUE) software was employed to map concepts and visualise relationships, thereby providing a clearer structure for subsequent analysis.

Describing data sharing

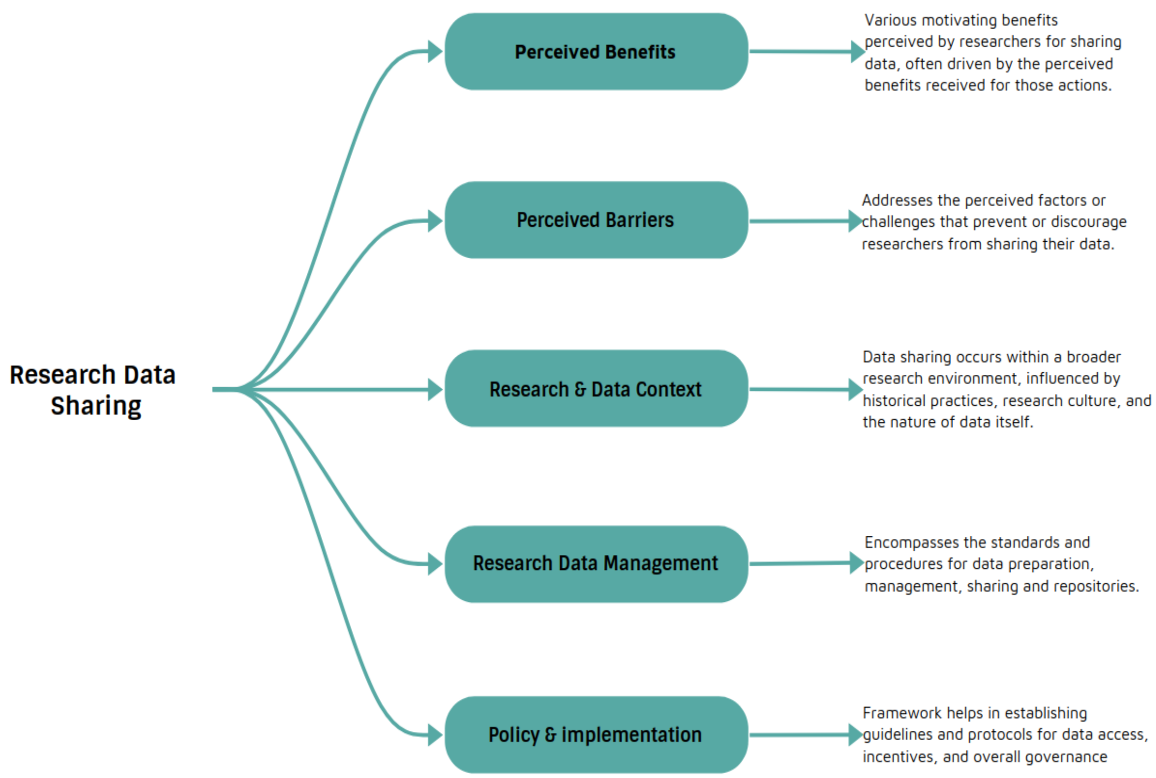
A profile of each participant was prepared prior to the interview process. Participants from Humanities distanced themselves from using the term 'data' as they generally connected the term data with that of experimental data. The term research data sharing was new to some of the participants. Before the interview, the investigator explained the idea and benefits of the research data sharing. Participants from different disciplines used varying terms for 'research data' and 'data sharing'. Each theme discusses, different aspects of research data sharing, and themes are often related. Some of the sub-themes overlap with other themes and hence can not be discussed in isolation. Coding procedure of Saldana (2016) is used for coding, and the final theme is derived after 6 iterations as discussed in the methodology chapter.

Coding

According to Saldana (2016) "most critical outcomes of qualitative data analysis is to interpret how the individual components of the study weave together". Initially a general layout was done using the sub-themes. Then codes were generated by splitting the sentences in the transcript. New sub-themes were generated in this process. Related sub-themes were grouped together to form the themes. Aligning the themes along with the theoretical framework was difficult. As the same code appear in different themes and element of the theories used. For example, the code "data withholding" might come under the 'perceived barrier' theme, and under the 'research and cultural context' theme. The first approach stems from the individual behaviour and addressed by TPB theory, while second comes from the broader cultural and disciplinary background of the organisation and addressed by the subjective and cultural-cognitive aspect of IT theory. This interconnections and dynamics of various aspects is visible in the coding. The investigator has identified 5 themes, 11 sub-themes and 44 codes. Each themes are discussed in detail in the following sections of this chapter.

Figure 5.1

Representation of the Themes Derived

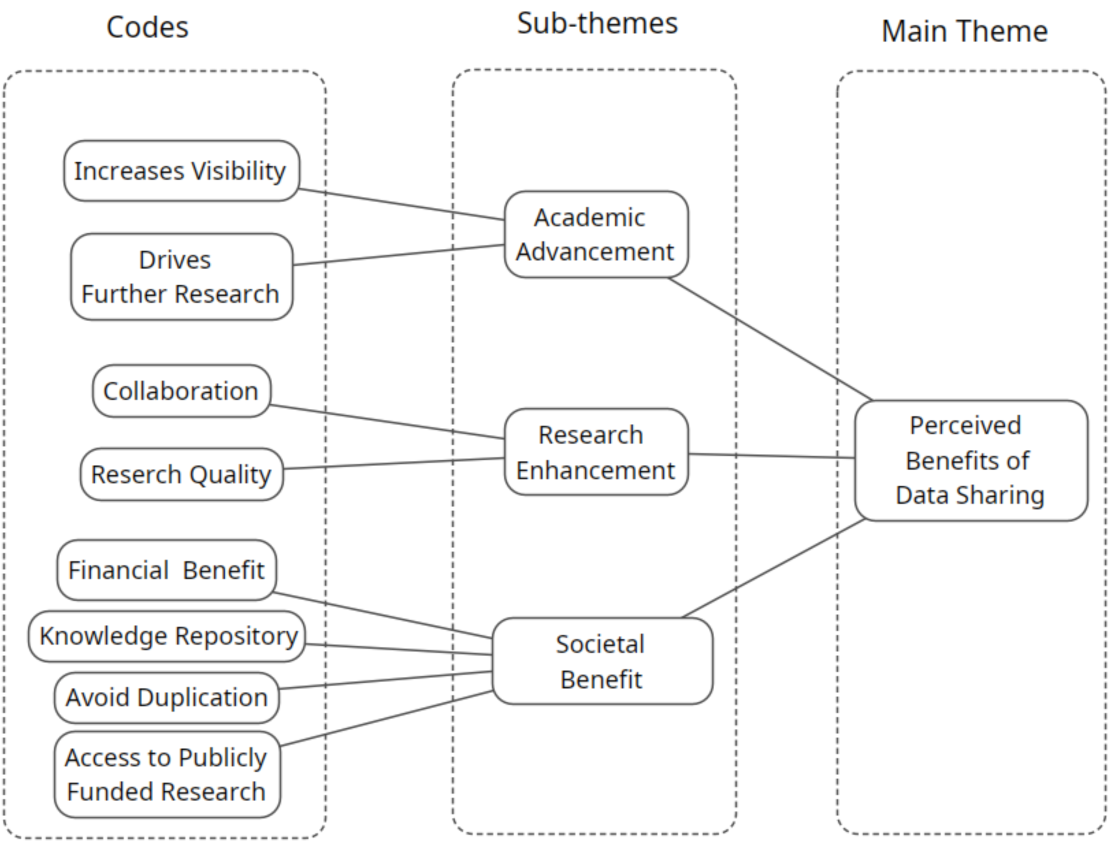


Theme 1: Perceived Benefits from Data Sharing

The primary theme identified in data sharing is the perceived benefits researchers anticipate by sharing their research data. Perceived benefits are rooted in the idea that data sharing behaviour is often driven by the perceived benefits received for those actions. The sub-themes are derived from the different types of benefits that researchers are considered while thinking about data sharing. These benefits consists of academic advancement, research enhancement and, societal benefits data sharing. Since the data preservation and sharing are not practised in the research institutions, only the expected benefits are discussed.

Figure 5.2

Theme 1: Perceived Benefits of Data Sharing



Academic Advancement

The sub-theme academic advancement is derived from three major codes, the increased visibility, the scope for further research and the possibility of secure preservation of own research data, as shown in figure 5.2. Studies underscore the academic benefits research data sharing (Borgman, 2017).

Increases Visibility of Research

> Many journals are demanding data along with the publication and we share data to journals. A major benefit of data sharing is that it increases the citation of the publication and increases the visibility (E2).

Research supervisors believe that data sharing will bring increased visibility to their publication and subsequently improve individual's research career prospects. Data sharing increases the citation rate of publications and enhances overall visibility as evident from the opinion of E2. Previous studies also has identified the increased visibility obtained through sharing research data.

Drives Further Research

In addition to the visibility, there is a possibility of further research in the area. Participant H1 believes in the value of the archival of rare documents which may not be widely available and how this can enhances further research. Participant L1 acknowledges the significant value of the data produced through laboratory experiments. L1 emphasises the potential for other researchers in the field to continue and extend the investigation using the available dataset.

> When conducting research, researchers often use only the information relevant to their specific topic, overlooking other valuable aspects in the original document. To ensure the complete and accurate representation of research, it is essential to publish both the original documents [whenever possible] and the researcher's notes, including their interpretations and comments.[...] This approach allows future researchers to access and utilise the original data, even if the original documents are inaccessible (H1).

> If one student fails to continue research for any reason, for example if researcher lacks an opportunity or not in a position to continue research in the field, any other researcher can continue the research work with created data (L1).

Facilitates Secure Preservation of Own Data

This code discusses the possibility of secure data preservation for one's own research requirement and to prove data authenticity.

> The data collector will be safe in preserving and sharing [to defend the findings in future] (H2).

> As of now, after the research work, we do not store the row data up-to-date. May be after completion of PhD work and, years after the publication, research journals may ask for row data, [...] and we may not have stored the data. There are chances of data loss due to technical or other issues. Data preservation will address problems like authenticity of the work, or challenging our work. We can defend our stand boldly [with the preserved data] (L1).

Data sharing facilitates the secure preservation of one's own research data, and addresses the authenticity concerns (H2). In addition to the this, there is a serious issue of data loss which may happen during the research journey. Data loss happens due to the absence of proper data preservation practices. Data sharing mechanisms can address the problem of data loss due to technical or other issues, which may occur years after the publication when researchers may not have preserved their raw data (L1). To justify the research findings in the long run.

Research enhancement

This sub-theme centres around the concept of research enhancement through data sharing. Shared data improves the overall quality and impact of research. Research enhancement means improving the processes and outcomes of research itself.

Enhances Collaboration

> Our department has a folklore archive which has around 400 hours of video of various performance and related things. Many people have used these archives as materials for further studies (H3).
>
> We have discovered new species of plants in our department, and if we preserve data about these plants, it would become a valuable resource. If a researcher wants to investigate the medical side of plant, they can use our data and need not have to repeat the botany side of the investigation, they need to do only medical side of the plant (L1).
>
> Past research in public health provides us with valuable information. We can rely on this data to make informed decisions. For example, studies have shown an increase in conditions like memory problems, kidney diseases, and autism. We can use this data to investigate whether these health issues are linked to changes in our environment, such as water quality, air pollution, or other factors (N3).

H3 has mentioned the possibility of research collaboration with preserved data. Open access to data improves research opportunities, allowing researchers to build upon existing work. Collaboration from other disciplines is expected through data sharing. Data sharing substantially enhances collaboration in research by fostering interdisciplinary and multi-disciplinary research works. Data sharing enables researchers from different fields to utilise existing data for novel investigations as observed by L1. In addition to the possibility of collaboration in the academic institutions, the shared data will also help build academic and administrative collaboration, especially for decision-making in the public policies.

Research Quality and Integrity

Public availability of research data increases transparency and accountability of research (Chin et al., 2023), and helps maintaining research integrity. N2 has the opinion that the PhD research in India is just for academic degree and does not produce any serious scientific or industrial product. There is a need for structural change in the research culture in our universities. E2 believe that, if researchers make their data publicly available, it is likely to bring more transparency and accountability in research. H4 recognises the possibilities democratisation of knowledge through data sharing.

> Successful industrial collaborations demands accurate and reliable data. In many instances, such as in South Korea [where participant has visited], rigorous verification processes involving multiple individuals are employed to ensure the integrity and accuracy of research findings (N2).

> Open access to data will definitely improve the quality of the research. If you are not confident of your work, [you will not be uploading the data to a public repository] (E2).

> We may be able to bring quality in research through data sharing. Yes , it is true. It would improve the democratisation of knowledge. Knowledge should also be treated as a common good like river or forest. Movement like copyleft are also promoting democratisation of knowledge (H4).

Participants from the Life Sciences disciplines (L1 and L3) believes that, open access to data facilitates the identification of data manipulation or plagiarism, ensuring credit is given where it is due.

> A good research culture will develop with data sharing. Data sharing will definitely enhance the quality of research as well. When data is precise, it will naturally improve the quality of the research. As of now, nobody is checking the row data. A mechanism of data storing, preserving will definitely solve these problems and improve research output. Once students are aware and vigilant about a mechanism of data preservation they will be preparing data methodically (L1).

> When data is publicly available, it becomes easier to identify instances of data manipulation or plagiarism. This increased transparency can help maintain the integrity of research and ensure that credit is given where it is due (L3).

Societal Benefits

This sub-theme deals with the aspect of contribution to scientific progress and public knowledge. Data sharing contributes to scientific progress and public knowledge (Ahmed et al., 2025). Research data is considered a public good, especially when funded by public money, ensuring the public benefits from the investment (Klump, 2017). Furthermore, data sharing can lead to the creation of knowledge repositories. It also reduces the duplication of data collection, saving financial resources and time.

Data is a Public Good

This study centres around the concept that the result of the publicly funded research should be shared with the public, aligning with the Open Science philosophy. E1 has strong positive attitude towards data sharing and shares the view that results of the publicly funded research and all the associated research output should be publicly shared. This philosophy of open science is shared by both L2 and S4.

> I believe that publication is not the only contribution of research. Research contribution comprise datasets, procedure, methodology, and sometime products also. All these together is research (E1)

> According to my knowledge research in all countries is done with the taxpayers' money, all research is publicly funded.[…] We don't want the data to be completely disappeared from the system, because it has a [public] investment. [L2]

> Researchers are collecting data with the government fund, and they have no right to say that I will not deposit the data (S4)

Knowledge Repository

Another societal benefit envisaged by the participant is the possibility of building a knowledge repository through preserving the data. Data sharing can support knowledge building, interacting with both academicians and the public as discussed by participants from the Social Science. Such data preservation may help creating a knowledge repository.

> In Ayurveda we had a treasure of knowledge and we had highly knowledgeable people. But we could not document this information properly, and subsequently we have lost many of this valuable information. We need to preserve such knowledge and provide a controlled access to research purpose only (S3).

Reduces the Duplication of Data Collection

E2 has mentioned advantage of data sharing as the elimination of work duplication. Sharing data avoids unnecessary duplication of work and allows researchers to address different questions using the same data. As pointed out by S1, the shrinking research fund also necessitates sharing data. This approach is useful and helpful across all the disciplines. Researchers will be able to answer different questions with the help of archived records as suggested by S3.

> We don't have to reinvent the wheel. I am doing a work on how to change our public spaces to make it women-friendly. I may be collecting data from various stakeholders to identify the problems faced by women. The data I have collected could be useful to others working in the related area. Such data sharing will avoid unnecessary duplication of work (E2).

> There is no need to repeat the data collection. We do projects in many areas like shipping, primary health centres, usually related with law area. We can do research with the help of available data (S3).

Financial Benefit

Financial benefit means benefit gained by the society or the research institutions by sharing the data in a reusable format. Society and research institutions can benefit financially from data sharing, especially if data is in a reusable format. Universities can generate income from research data, potentially charging a nominal fee for access as suggested by S2. S1 also support the possibility of income generation but only after an embargo period. This mechanism provides researchers with time to analyse and publish their findings before the data is made more widely available. Various government organisations are selling data as a product. The possibility of converting data as a product with financial value is also discussed by the participants. There are also possibility of using the data for commercial purpose within or outside the country as noted by L1.

> We may not assess the business value of a research, if the data is made publicly available, others can reuse the data to create products or services. A person from another country may find this data valuable (L1).
>
> Data is wealth and universities can generate income from research data, at least a nominal charge (S2).
>
> Data can be downloaded after a stage [embargo period]. We can charge for data at a nominal rate. We can create a committee like ethics committee to monitor. (S1)

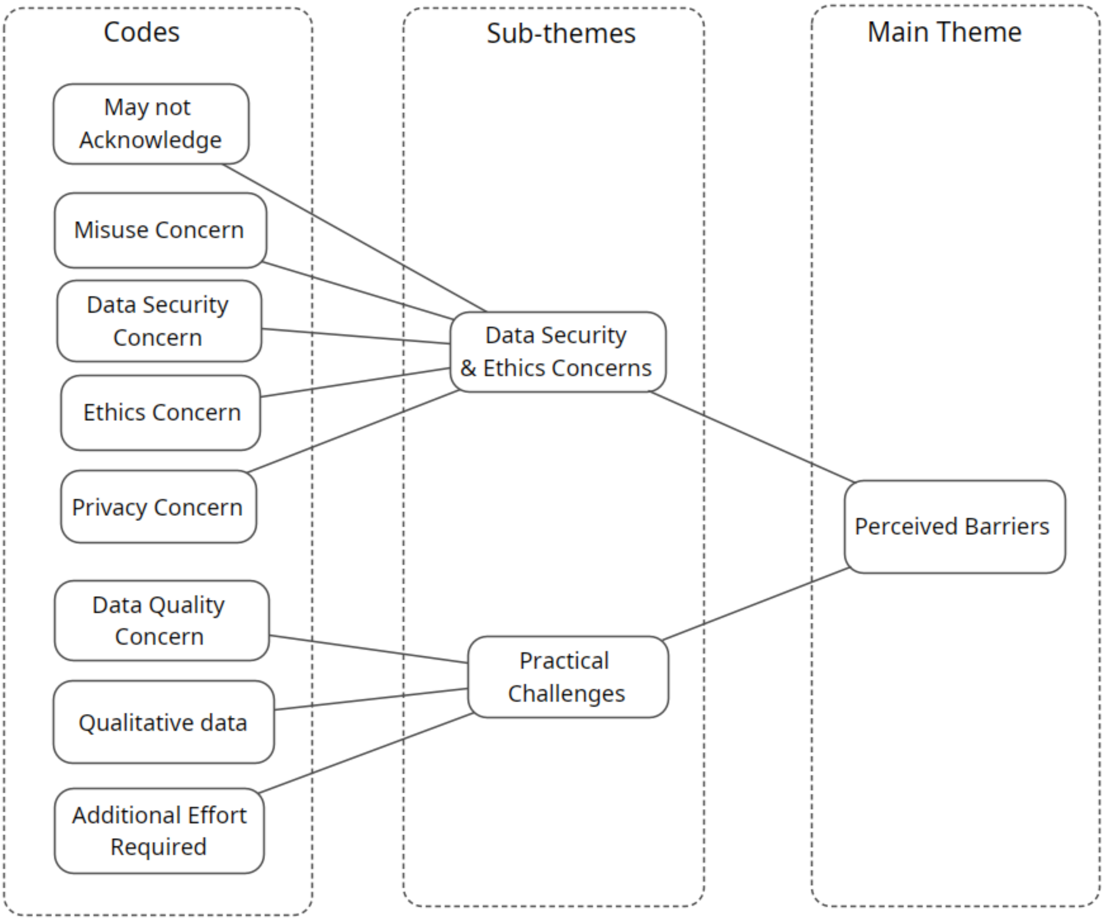
Participants from all the disciplines have positive opinion about the potential to revolutionise research through data sharing. Natural Science, Life Science, and Engineering disciplines consider collaboration as the main benefit of data sharing. Social Science and Humanities participants consider reusable value of the collected data, especially how these data can be used for making policy decision making and enhancing the societal impact of research. As the research is publicly funded, the research data should be contributed to the betterment of all in the society. Most of the participants believes that data sharing will improve the quality of the research through enhanced transparency.

Theme 2: Perceived Barriers of Data Sharing

Research data sharing has advantages and at the same time, interview has revealed concerns that make researchers hesitant in sharing their data. The main challenges are individual privacy, potential to commercial exploitation and intellectual property rights. This theme addresses the perceived factors or challenges that prevent or discourage researchers from sharing their data.

Figure 5.3

Theme 2: Perceived Barriers of Data Sharing



Data Security and Ethics Concerns

This sub-theme encompasses issues like data privacy, potential misuse of data, and the need for ethical considerations in data sharing.

Misuse Concern

A major concern expressed by majority of the participants is the potential to misuse the data and possibility of data falling into the wrong hand.

> Private companies might misuse this data to make money without caring about people's privacy or health.[…] This could accidentally harm the people who provided the data. It's crucial to always prioritise ethical considerations while sharing research data (S2).

> There are many tools and AI software which can be used to generate trend line and one can create n number of publications. There are chances of misuse of data (E3).

> Publishing industry may be behind such initiative as they might get an opportunity to publish data and monetise (H4).

> Data obtained may be misused for other purposes, to what extent we can rely on people (data users). Medical data, other important data with national interest should not be shared openly. […] Data related with a particular population or country may be important and these should be classified (N1).

Participants are aware of the need to protect the privacy of their sample. This is significant when related to vulnerable communities (S2). Participant E3, who is against data sharing, have concerns about the possibility of data misuse. Another suggestion is to avoid data misuse is to classify the data and minimise the possibility of misuse of data which have national importance. Participant H4 is also against the data sharing concept and believe that the very concept may be another opportunity for the publishing houses to make money. N1 who is a strong supporter of data sharing warns about the security of data and need for classification.

Participant Consent

H3 has expressed their concern regarding ethical violation that data sharing is likely to bring. Participant H2 is worried about the ethical violation of data collected from the human subjects especially that of marginalised community. H2 also shared instances where data was taken out of contexts, used to support findings that are directly contradictory to the findings of the initial study. Some participants has expressed their concerns regarding the possible ethical violation from the data sharing and reusing. There are instances of data collection without the consent of the participants of the study as noted by S2. Such data, if published online are likely to invite serious ethical problems.

> People may not like publishing these recordings [sensitive information] as such, it might create problems for them. These conversations and

> rituals are recorded under the condition that all these recordings will be disposed after the study. There are certain customs which is kept only among a limited number of people in the family. Even if you do not disclose their name, respondent identity may be revealed (H3).

> You have to take the consent of the people concerned. You should take an informed consent. […] Data interpretation related problems occurs in research where human subjects are studied. Another researcher may be interpreting the data differently. Collected data is influenced by the subjectivity of the researcher (H2).

> Data collection from tribal community is another problem. NGOs and some researchers do publish data collected from the tribal area without the informed consent of the people under study. This is a serious ethical issue (S2).

Data Security Concern

Data security concerns are significant challenges in the context of research data sharing. These concerns cover the need for secure data preservation, the potential impact of data sharing on national security, and the leakage of personal data.

> I am doing research with spatial data in my project […] I was thinking about sharing these data to the public. [...] A major challenge is that, satellite imagery and remote sense data are crucial data. When we share these data, there are chances of dangers and threats. It may not be possible to share these data due to security reasons (E2).

> We need to classify data as shareable and not shareable. Data related to national security or international politics may not be shared. Ongoing research should not be exposed through data sharing. Data should not be exposed during the period of study, we should wait till the completion of the work. [...] Educational institutions like ours publish work to get the visibility and do not take care of national interest (N1).

> Even after anonymisation, personal health data may be compromised because our personal information including Aadhar data are available with private agencies. Corporate agencies may make use of these data for profit making (S1).

> We get rainfall and other related data from the Indian Meteorological Department, and we can do research and publish the results. But river flow data is not available. State governments in India do not disclose the river flow data (especially of inter-state rivers) due to the inter-state conflicts. Such data are difficult to obtain for research purpose (E4).

Participant S1 expressed concerns regarding the potential compromise of personal data, even after the anonymisation processes. Specifically, S1 highlighted the risks associated with granting private agencies access to such information, which could lead to unintended data leaks or re-identification of individuals. Participant E2 think that research in areas of strategic importance needs protection. N1 has mentioned about how researchers are tempted to publish before the completion of the full project. N1 further suggests data classification as shareable and non-shareable, as a solution. According to E4, these security measures sometimes create problem in accessing data.

Acknowledgement Concern

A significant concern expressed by S5 is the potential for data use without giving proper citations or acknowledging the original researcher as well for the commercial purpose. The need to strike a balance between the individual career prospects and the institutional and national interests is also raised by participants. An Associate Professor from Social Sciences (S2) see this data use without due acknowledgement from another angle, the likelihood of cross boarder data transfer for commercial purpose.

> Will people cite data, is a big question. A major finding in my thesis in the form of a table is reproduced in a government planning board report without my permission or acknowledging me. There may be technology in the future to detect data plagiarism with the help of AI. So, if one does not cite the data, it could create issues, may be in the future. Researchers

> should strictly follow ethics in studies. Nowadays, companies collect consumer preferences primarily through social media platforms like Twitter. These platforms, often controlled by large corporations based in the United States, gather and analyse vast amounts of user data for the purpose of generating revenue (S5).

> Medicinal plants related information may be transferred to foreign countries and commercialised (S2).

Researchers may hesitate to share data due to concerns that others might use data without proper acknowledgment or credit, potentially leading to a loss of recognition for their work.

Privacy Concern

Study participants expressed that data sharing is closely connected to privacy aspects, and privacy violation is considered a serious issue. Researchers are aware of the need to protect the privacy of the participants using anonymity when collecting data (E2). This shows that safeguarding privacy is essential and, neglecting the privacy norms may invite unexpected problems. Researchers should deal these data very carefully. Publication of supporting documents or data creates problems in some research domains. If data is revealed it could be harmful to the subjects of the study. But at the same time without proper evidences, it becomes difficult to establish the findings. For instance, one participant provided the example of publishing letters from migrant employees which might lead to identity disclosure and privacy issues (H5).

> I am doing research using expert opinion. We cannot expose the list of experts to prove our point. Experts may not have problem in stating their opinion, but they may have limitations in disclosing their identity in the public sphere. Only anonymised data is presented. We ensure anonymity clause in collecting such survey and interview data. Whatever data we collect, from human subject we ensure that 'whatever we collect will be kept anonymous ... and will be used for research purpose only' (E2).

> Letters sent by the migrant employees from gulf countries to their family during 90s are used by some researchers to explore the socio economic

> situations exists. Publishing these letters may lead to identity disclosure of the subject and may invite privacy issues. Any data sharing should respect privacy of the people involved in the research (H5).

> We collect documents for digital preservation which do not affect the privacy of people. [...] The owners are requested to remove any materials which may violate the privacy of the individual. To protect privacy, we can restrict access to sensitive information. […] We should create agreements that clearly outline who can access the data and how. If a socially or politically relevant material contains any private matters within it, the data owner can suggest to make it classified. Even if we remove identifying information, someone involved in the event might still be recognisable (H1).

But the restrictions in the name of protecting marginalised people have made research in these areas difficult (S5). Researchers apparently faced delay in getting approval for study from the government departments in study involving human subjects.

> Government departments were not cooperative, because they are not interested in these things. When we approach these departments for data the approval process takes much time. […] A survey planned in tribal areas, the survey questions have to be approved by the department concerned and, they have removed questions from the survey making the study unusable (S5).

Practical Challenges

The interview has identified several practical difficulties that make data sharing challenging, which include the need for additional effort in preparing data, the complexities associated with qualitative data, and the trustworthiness of the research environment.

Additional Effort is required

Participant L1 mentioned that, in addition to the workload of teaching and assigned administrative tasks, research supervisors does not get enough time to verify the authenticity of the research data. S2 has come up with a innovative idea of ‘research club’ to collaboratively solve the problem of additional effort. Research clubs can be seen as a grass root effort to build responsible data sharing. Referring to the additional task of data preservation H5 believes that the data sharing is likely to complicate the research process. And there are suggestions to solve the complexities of additional effort comes with research data management.

> No research supervisor will be able to thoroughly check or verify the data generated by a student. We have limited time even for discussion with the research students, it is not possible to spend time in data verification (L1).
>
> One have to take additional effort to prepare the data in a reusable format. We can set up research data consultancy in the name “Research club” to do such work in the universities (S2).
>
> Creating good descriptions for digital files takes a lot more time and effort than just scanning or copying them. This is a big job and is really important. Just making a digital copy isn't enough. We need to add special information to each document, like keywords, to make it easy to find and use it later (H1).
>
> I think that data sharing initiative is likely to complicate the research process a bit (H5).

Qualitative data

H2 is concerned about the interpretation of qualitative data when a third person reads the data, as collected data may be influenced by the subjectivity of the researcher. Managing these data invites its own problems and difficulties. Managing, preparing these data need special training and care. According to H1, proper metadata is necessary to understand the context of the data collection. Research data are likely to be misinterpreted in the absence of proper documentation (Patton, 2015).

> Data interpretation related problems occurs in research where human subjects are studied. Another researcher may be interpreting the data differently. Collected data is influenced by the subjectivity of the researcher (H2).

> There may be various geographical allusions mentioned in the project, those who do not have knowledge about the particular geographical area, you would not understand the concept. This is the problem with the local history. [...] Only solution is to create a map, listing all the localities and items mentioned in the local history (H1).

This aligns with the understanding that qualitative data analysis is primarily an interpretive act, where meaning is constructed through the investigator's lens. Local history, folklore, and culturally specific practices hold unique insights that may not be generalisable globally but are crucial for understanding specific communities and their needs.

Theme 3: Research and Data Context

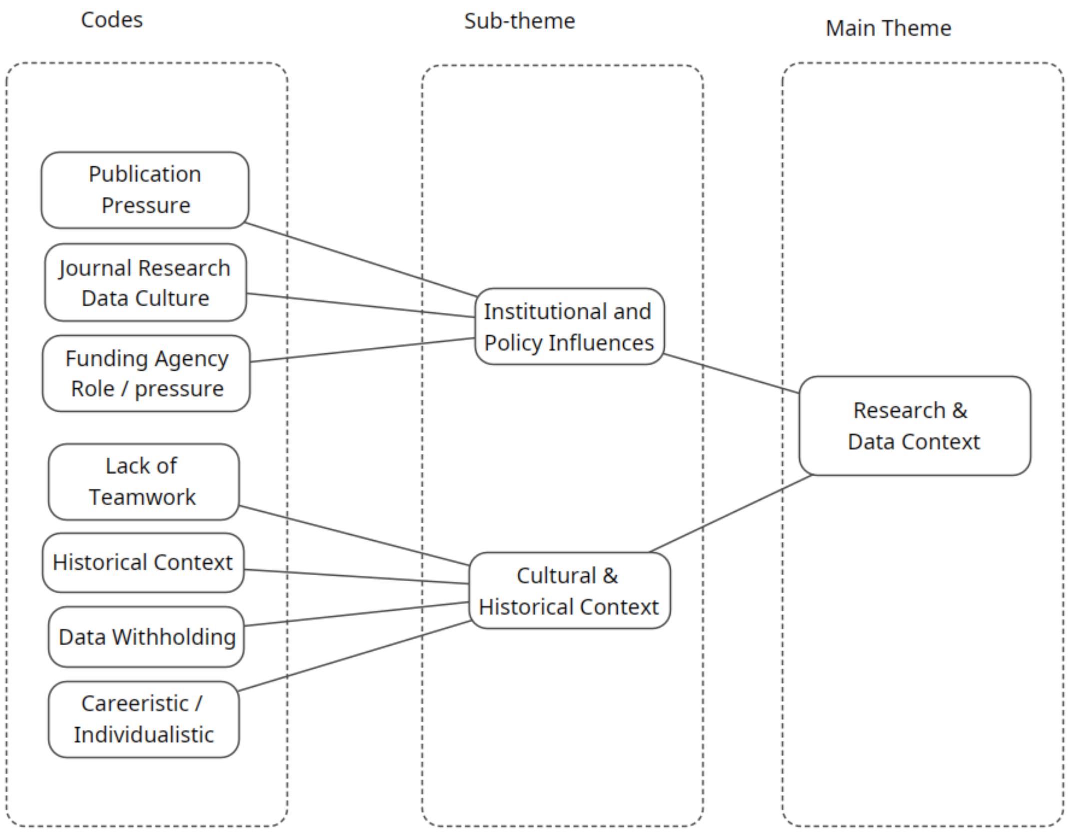
The third theme acknowledges that data sharing occurs within a broader research environment, influenced by historical practices, research culture, and the nature of data itself. Theme discusses the various reasons behind existing research practices and how these are connected with the data sharing behaviour of researchers.

Cultural and Historical Context

The cultural and historical context significantly influence the data sharing behaviour (Kim & Stanton, 2016). This influence stems from established practices within institutions and research disciplines, as well as the overall research culture prevailing in the institution. This encompasses how established practices within the institutions and research disciplines and the overall research culture influence data sharing. This includes how the historical context shapes the research culture and decisions about data sharing.

Figure 5.4

Theme 3: Research and Data Context



Lack of Teamwork

The observation that research is person centric suggests an overemphasis on individual contributions, often driven by personal recognition, career advancement and ego (E1). This can lead to inefficiencies, lack of collaboration, and a focus on individual benefits rather than collective scientific progress (Linek et al., 2017). The lack of teamwork, compounded by reluctance and ego, is identified as a major hurdle in Indian research by N2. This manifests as data withholding, where researchers avoid sharing data due to personal and professional insecurities.

> The main problem in the present research culture is that, our research is person-centric (E1).

> Lack of team work is a major hurdle in Indian research. Reluctance from the part of individuals and ego may be reason behind data withholding (N2).

Careerism

According to L3, the main aim of the research students is to get a degree and find a job. Participant N2 also support this opinion, in his opinion, there should be impactful research rather than academic degree. H1 believes that, the higher rate of unemployment among educated youth and the inability to find a suitable job even after acquiring a doctoral degree make it difficult for the researchers.

> Furthermore, the pressure to publish for career advancement (promotions, securing jobs) incentivised researchers to prioritise quantity over quality. This mentality can lead to an increase in the number of low-quality publications, compromising the overall integrity of scholarly research (L3).

> Professionals from medical field do not enter into research because the earnings and recognition for doctors and engineers are high compared to researchers here (N1).

> This system encourages researchers to focus on publishing frequently rather than conducting high-quality, impactful research. A significant portion of research in India is currently focused on academic degree, with only a small fraction dedicated to Research and Development. (N2)

> People even with a PhD find it difficult to get a job, so their interest is to submit [the PhD thesis] somehow and find a job (H1).

Historical Context

This code discusses how the historical context shapes the research culture and decisions about data sharing. Participant N1, believes that the research quality is connected with the economic condition of India and he is optimistic about the future. The interview participant N1, with over 30 years of experience and international work exposure, highlights the challenges faced by researchers in the country due to global disparities in research ecosystems.

> Majority of the publishers are from the USA and Europe and, when we publish article, it goes to common domain. Individual researcher get benefited from the publications by increased visibility and recognition among peers. Data drain and brain drain happens just like raw materials drained during colonial period. People used to work abroad for financial earnings. Our country need to develop research facilities to attract these researchers and scientists, [but] this development will be very slow. This is not merely the problem with our research culture, because we have other priorities like poverty eradication. We have been living under colonial era for decades. Generally developed countries provide better opportunities and life standards to people from scientific community. They are given due recognition (N1).

Researchers seek better opportunities abroad for financial and professional gains. According to N1, the colonial era has left a legacy of dependence on external sources of knowledge and technology. This can lead to a reluctance to share data, particularly if researchers believe that it might be exploited by others. He emphasised the need for improved domestic research facilities to retain talent, acknowledging that such development is slow due to competing national priorities like poverty eradication.

Data Quality

Participants have shared some noticeable observations about the quality of data produced in the research. Unethical practices in research may influence the research data and quality of research (Yoon, 2017). Data falsification is a serious problem in research, as it can lead to the promotion of false or misleading information (Rao et al., 2024).

> People have started sharing their PhD reports in Shodhganga, only after the mandatory policy by UGC [to share PhD thesis in Shodhganga ETD]. This helped to identify malpractices like plagiarism (E2).

> A majority of research supervisors do not ensure the quality of data involved in the research.[…] It would be good to get a research associate in our department to do these works. This will ensure data quality also. [...] We can also think about a committee at the departmental or research centre level to ensure data quality and prevent data falsification at least by checking randomly (E1).

> As of now, nobody is checking the row data. No research supervisor will be able to thoroughly check or verify the data generated by a student. [...] I have seen that students in our universities are not always meticulous in their research work. […] If there is a policy to preserve and share data, they will store data meticulously right from the beginning of the research. Researchers will prepare data methodically and store it properly, this will improve the quality of the data and research work (L1).

> I've encountered PhD theses with questionable data during my evaluation. [...] These issues often arise from a lack of adherence to proper research procedures and the failure to replicate experiments, highlighting the need for a stronger emphasis on rigorous scientific methodology in our departments (N2).

> Ensuring the research integrity is the sole responsibility of a research supervisor. Research supervisors and evaluators can only understand and ensure the quality of the work. Research supervisor has the responsibility to at least to check how the student has got a result (S4).

There is a lack of mechanisms to verify raw data, as noted by E1. L1 observes that students are often not meticulous in their research practices, which affects data quality and proposes a policy mandating data preservation and sharing. N2 critiques the focus on academic exercises over practical applications and, emphasise the need for rigorous scientific methodology. E1, L1, and S4 underscore the role of supervisors in ensuring the research integrity, but due to the workload and volume of data produced, it is not often happening. E2 believes that regulations have shown results in the past and there is a need for data sharing policies. The proposed solutions reflect a desire for a multi-tiered approach to address the data quality challenges. Research data committees should be established at both departmental and institutional levels within research organisations to check complaints related to the authenticity of research data.

Data Withholding

Data withholding remains a significant challenge in data sharing. The participants have revealed a complex interplay of cultural, strategic, and professional motivations behind restricting access to research data.

> Some individuals may restrict access to documents, claiming them as personal or family property. This is often seen in cases with religious or political significance. […] Concerns about data misuse are frequently cited as a reason for withholding access. Many individuals believe that restricting access to research materials provides them with an intellectual advantage and maintains their dominance within a particular research area. Second reason is hoarding information. Even when access to documents is possible, some researchers may hoard them, treating them as their intellectual property (H1).

> In some papers, researchers do not disclose the actual value, instead they report as 'appropriate amount of'. This is mainly done for future publication or patent generation. Actually some patents are just for showcasing and this is not useful to convert it to a product. The statement 'Data will be available on request' is used withheld the full data (N2).

The observations of H1 suggest that withholding access is often rooted in a desire to maintain control and exclusivity. This is significant in the contexts where documents carry significant cultural value. N2's insights into publication practices further underscore strategic withholding as a common tactic. The use of vague language in papers and the non-disclosure of full datasets, often under the pretext of future publications or patents, point to a prioritisation of personal or institutional gain over transparency and reproducibility.

Institutional and Policy Influences

Institutional pressures and policies can encourage or discourage the data sharing. This sub-theme is about the formal roles, responsibilities, and systems within the institution that directly impacts data management and sharing.

Role of Funding Agencies

Funding agency can play a pivotal role in the development of data sharing culture and repository building. Participants has different opinion about the role that funding agencies can play, from providing guidelines to self publishing the data and related publications.

> Our funding agency (SERB) does not mandate to share data to them or say anything about the data ownership. The intention of research funding is to help knowledge generation, and we just need to acknowledge the funding agency for the funding. SERB does not control the data or have any mandatory regulation to submit the data generated under their funded research (L1).

> We are doing research and someone is selling it for money, I am against that idea. Publication companies are earning a lot. No royalty is given to the authors. Government of India should take an initiative to make sure that these are all published only by the Indian journals. Funding agencies like DST (Department of Science & Technology, India) and DBT (Department of Biotechnology, India) should ensure that all funded projects are published in India and in open access (E3).

> To ensure data quality, we need to check that it meets minimum technical standards and scientific standards. Funding agency usually ensures this, and they will get the assurance that data is generated properly (L2) .

Presently, funding agencies does not mandate data sharing as noted by L1. The criticism raised by E3 underscores the need for intervention of funding agencies to prioritise publishing in Indian journals. E3 suggests a desire to reclaim control over research outputs and make them more accessible. L2 suggests that funding agencies can ensure data quality through technical and scientific standards.

Influence of Publication Pressure

Researchers who want to increase the publications, sometimes engage in unethical practices. Publication pressure refers to a systemic issue within the academic research ecosystem, where external pressures from ranking systems and career incentives distort research practices.

> Pressure to make more publications has deteriorated the quality of the data. NIRF [National Institutional Ranking Framework], India ranking and other ranking systems depends on the quantity of research output [rather than quality] (S1).

> Researchers are tempted to publish maximum number of papers, sometimes one work is split to multiple papers to increase the number of publications. This tendency is to increase the quantity of publications (N1).

> Furthermore, the pressure to publish for career advancement [promotions, securing jobs] incentivised researchers to prioritise quantity over quality. This mentality can lead to an increase in the number of low-quality publications, compromising the overall integrity of scholarly research (L3).

S1 asserts that the pressure to produce more publications has led to a decline in the quality of publications. N1 observes that researchers often split a single piece of work into multiple papers to inflate their publication count. L3 highlights that the pressure to publish for career advancement, such as securing promotions or jobs put pressure on the researchers to focus on quantity rather than quality. The observation of N1 about splitting research report into multiple papers underscores a strategic response to publication pressure, where researchers maximise their output by fragmenting findings. L3's emphasis on career advancement highlights the individual-level consequences of this pressure, where researchers face a trade-off between producing high-quality, impactful work and meeting the publication thresholds required for professional growth.

Research Data Practices of Journals Triggers Data Sharing

Journals can trigger data sharing culture by publishing the data alongside the articles (Zenk-Möltgen et al., 2018). The participant's observations reflect a growing recognition of journals' role in fostering transparency and accountability in research through data-sharing requirements.

> Many journals are demanding data along with the publication, and we share data to journals (E2).
>
> Some journals are providing provision to store data, now it is optional. Journals provide such an opportunity to make sure that, if anyone question the reliability of the study they can verify it by cross-checking the data by analysing the row data. Now the data storage is optional and surely tomorrow they will make it mandatory. I think they are tuning researchers towards a data sharing culture (L1).
>
> Journals and other publications should make data availability a mandatory policy (N2).

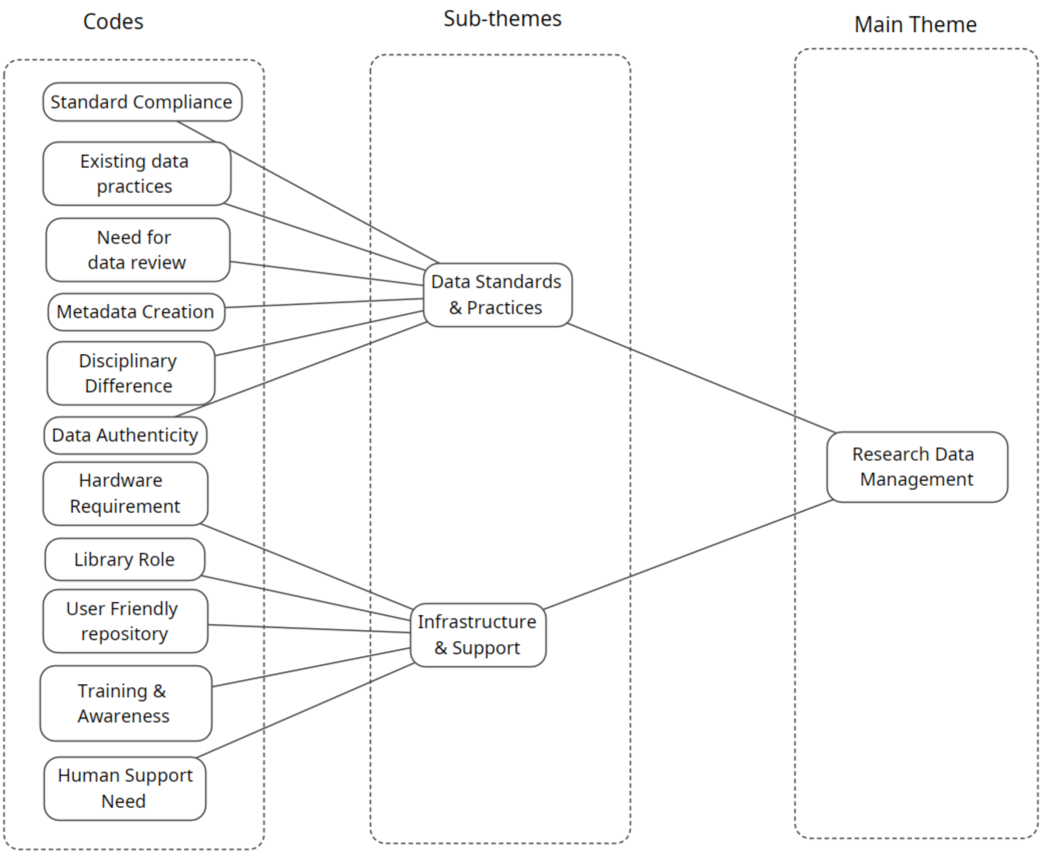
According to L1, journals are slowly bringing the culture of data sharing by creating awareness among researchers and, in the future, journals are likely to implement mandatory data depositing. Many journals demand data alongside publications indicates a shift toward greater scrutiny of research outputs. The insight of L1 into optional data storage provisions highlights a transitional phase in journal policies, where researchers are being gradually acclimated to a data-sharing culture. Call for mandatory data availability policies by participant N2 underscores the urgency of institutionalising transparency to enhance research integrity.

Theme 4: Research Data Management

The theme of research data management covers the standards and procedures for research data management. RDM ensures that data is reliable, usable, and of high quality.

Figure 5.5

Theme 4: Research Data Management



Infrastructure and Human Support

This sub-theme deals with the need for institutional support, training, and rewards for encouraging sharing of research data.

Hardware Requirement

Hardware requirements are crucial in data sharing as they provide the necessary infrastructure for universities to preserve and share research data.

> Universities are knowledge generating centres and there should be institutional level repositories at the universities. We already have data

> centres and support system in universities like University of Kerala, University of Calicut etc. We have the technical infrastructure to create repository platforms, we must upload and share our research data in these servers and make these openly accessible to the public (E1).

> There will be large volume of data, who will invest to preserve these data. Universities can create data repository to preserve research data (L1).

Participant E1 believes that universities need to establish and maintain institutional level repositories with adequate hardware and software platform to accommodate the large volume of data. L1 has mentioned the significant investment required for preserving large volumes of data. Having the appropriate hardware, data centres, and servers enables universities to effectively manage, store, and provide access to research data. University level infrastructure may be utilised for data repository building as suggested by a participant who have managed the digital infrastructure of a major university in Kerala.

Training and Awareness

Participants emphasised the need for training to promote effective research data management. The primary aim of training is to enable researchers to prepare data in a reusable format (Awada et al., 2022).

> University as a whole should take initiative to provide awareness among research scholars. From the very beginning of coursework, students should be trained and equipped in research data management (H2).

> Instead of solely relying on a theoretical examination of research ethics, we should incorporate practical training on research data management into the PhD curriculum. A mandatory ten days hands-on training programme would be more effective than a traditional exam (L3).

> Once students are aware and vigilant about a mechanism of data preservation they will be preparing data methodically. University has included the research ethics in the PQE (Pre-Qualifying Examination) syllabus and, now students are vigilant about the plagiarism and other

> ethical issues. […] Students may find data management training as difficult, but ultimately it will be helpful for them (L1).

One thing that is generally agreed by all participants is that universities should take the initiative to provide awareness among research scholars. L1 and N2 believes that just like research ethics, inclusion of research data management in the curriculum will help the students. These training should be more practical oriented than just theory examinations. Libraries can play role in various aspects of the RDM (S1).

Human Support Need

The management of research data requires significant human support to ensure that, data is well-prepared, maintained, and shared effectively (Ho et al., 2025). Participants highlighted various human support needs, including department level expertise, specialised divisions, research clubs, statistical support, and multidisciplinary teams.

> Each university deals with diverse data and preparing these data in a shareable format would require a huge work. There are multiple disciplines and these disciplines generates diverse data. Managing these data is not an easy task. Maybe we can prepare it at department level (S1).

> We can form a setup for research data consultancy in the name 'Research Club' [...]. This club can provide statistical consultancy, data collection and data sharing. We can build research clubs by employing students and faculty members.[…] There should be a mediator, one who receive and provide data. University research wing can play the role of mediator between the data generator and data re-user (S2).

> Need support from multidisciplinary bodies. We do, research in medical aspects also, how environmental factors lead to systemic or otherwise diseases. In this case we might need help of medical department. […] neither statisticians nor subject experts alone possess the full range of skills required for this complex task (N3).

As discussed in the sub-theme 'Practical Challenges', researchers find the additional task related to data storing, data curation, metadata preparation a huge burden. There is a need for data expertise at the department level to prepare data in a shareable format. Participants understand that, managing diverse data from multiple disciplines is not an easy task and requires domain expertise also. For instance, in economics, a statistician and economist together can understand and prepare the data. Developing a support system at the departmental level is suggested (S1). S2 has proposed research data consultancy called 'research club' to support RDM. Considering the interdisciplinary nature of research work, N3 mentioned that a multidisciplinary team is better for preparing data in a shareable format, as neither statisticians nor subject experts alone possess the full range of skills required for this complex task.

> In GEO repository, there is a data curator who check the data before deposition and the communication with the curator is straight forward and quick, and there are no delays in communication. [...] We get good support from the data curator. […] Data crew should be capable of managing [...] different types of data (L2).

> We cannot accept the data as such, we have to verify the data while depositing the data to ensure the quality, format and content. We should have a team for data management and support (L1).

> In science it would be easy to systematically arrange and publish supporting data. In Social Science and Humanities the very nature of data is complicated. We would need technical as well as academic support to share such data (H5).

> We would require support, funding agency should appoint someone to see that what are the data that can be uploaded. And also to support that how this data can be uploaded. This background work is important, then only we will be able to use this data effectively (E2).

L2, who has experience in data depositing recollects the support provided by the data curator of a repository and emphasis the need for data curator support. H5 reminds that, unlike other disciplines, researchers in the Humanities are not equipped in the use of tools. The nature of data in Humanities is complicated, technical as well as academic support is required to prepare and share these data. E2 thinks that funding agencies can also play a role in providing human support and financial support. There are suggestions to create support team with the help of PhD graduates, this will give them employment opportunities and experience.

User-Friendly Interface

The repository infrastructure and interface of the platform itself may significantly influence data sharing. The accessibility, user experience, and the overall efficiency of data retrieval influence the researchers.

> Data publication procedure should be straightforward and easy. That is super important. Repository interface should be user-friendly. Otherwise, we get lost, and the steps will be very boring and tiring. User-friendly interface is the primary feature of a repository (L2).

> Wikipedia is widely used despite occasional inaccuracies, its user-friendly interface and hyperlinking system make it incredibly accessible. To be truly effective, digital repositories must emulate this user experience. They should allow for easy navigation between related documents, just as Wikipedia links connect different articles. Without this interconnectedness, repositories will be difficult to use and may not be widely adopted. (H1)

Participants like H1 and L2 strongly believe that, to be truly effective, digital repositories must emulate a user-friendly experience. Repositories should enable for easy navigation between related documents. Without this interconnectedness, repositories will be difficult to use and may not be widely adopted. Just like the data users, data uploading task should be easier and supportive. Data publication procedures should be easy and straightforward as mentioned by participant L2. The repository structure should be simple to interact with the common man. Overall, a well-designed repository infrastructure with a user-friendly interface is essential to facilitate data sharing, making data more accessible, retrievable, and useful for a wide range of users.

Role of Library

Studies shows that like various other scholarly repositories, libraries are managing and supporting the research data repositories (Singh et al., 2022). From training to repository management, participants have shared their view on the role that library can play in various steps of research data sharing.

> Effective management of digital repositories relies heavily on the expertise of Library Science professionals. Their knowledge and skills are essential for organising and retrieving documents within these digital environments. There needs to be a strong connection between libraries and research scholars for effective data management (H1).

> Library can provide training to research [doctoral] students in data sharing and other research related tools. From the very beginning of coursework, students should be trained and equipped in Research Data Management (H2).

> Libraries have been keeping documents over a long period, and we have access to these documents because of libraries. Similarly, we should keep data in libraries. Data should be stored in the library along with other documents (N3).

H1 strongly believe that the library has a key role to play in the research data management. Libraries can provide training to doctoral students in data sharing (H2). Effective management of digital repositories relies on the expertise of Library Science professionals. H1 also reiterate the need for a collaborative team, comprising both library professionals and subject-matter experts. Library can play an important role by providing training to the doctoral students. They can conduct data awareness programs for students and promote data literacy. Participants assume that library is a core partner of the research data management. With the decades of experience in document management, libraries can manage research data repositories as stated by N3.

Data Standards and Practices

Data standards and practices are critical for ensuring that research data is reliable, usable, and of high quality (Buyle et al., 2018). These standards encompass various elements which ensure that, data is valid, reliable, and suitable for research purposes and reuse in the future.

Data Authenticity

A recurring concern of the participants is about ensuring the quality and authenticity of the research data.

> How do we check that whether it is an authentic data?. We have to cross-check the authenticity of the data, [ensuring] that the data is authentic should be the responsibility of the depositor. It brings a level responsibility to the data uploading person. This is a double-edged sword, see, if I want to upload the data, I should be convinced that this is an authentic data. Unless and until the depositor is sure about the authenticity of the data, it should be a problem for the data depositor (E2).

> For example, a doctoral student working under the supervision of a colleague at IIT fabricated NMR data for a research study without conducting any actual experiments. The supervisor subsequently discovered the data falsification and removed the student from the research project. This incident serves as a serious reminder that data fabrication can happen. [...]. Instead of rigorously probing the unexpected results, researchers may sometimes resort to minor modifications as a quicker solution (N2).

Ensuring the authenticity of data is important in maintaining research integrity. The depositor bears primary responsibility for ensuring data authenticity, as emphasised by E2. Data sharing, as noted by participant L3, is a powerful tool to enhance transparency and deter unethical practices like data manipulation or falsification. Data sharing, as advocated by L3 reduce risks of manipulation and facilitates checks by making data publicly accessible for scrutiny. The depositor bears significant responsibility for validating data before submission. In the opinion of participants, the concerns raised about data fabrication, manipulation, and the responsibilities of data depositors can be addressed through systematic approaches to verify authenticity and promote transparency.

> Manipulation in primary data is a bigger threat in research. There are chances of data smoothing to arrive at the desired findings. The availability of raw data will reduce such chances. [...] mostly encourage secondary data related studies, because this data would be more authentic compared to primary data. There is a possibility of personal bias in the primary studies (S2).

> Data sharing can help prevent unethical research practices, such as manipulation or falsification of data. [...] When data is publicly available, it becomes easier to identify instances of data manipulation or plagiarism. Increased transparency can help maintain the integrity of research and ensure that, credit is given where it is due (L3).

As mentioned by S2, it is difficult to cross-check the authenticity of data, especially primary data, due to the potential personal biases and the time and effort required for verification.

Metadata Creation

As repeatedly emphasised by participants, comprehensive and well-defined metadata is crucial for making data discoverable and reusable (Wilkinson et al., 2016). Policy should mandate and provide guidelines for metadata creation, including information about data collection, curation, and responsible individuals. H1 reiterates the need for providing metadata but admits that it is a tedious job in his words “without proper metadata there is no repository”.

> If we do not organise the data properly in the repository, it becomes a problem. If we cannot retrieve data, there is no point in having a large data with us. That is what I am saying, data should be in a format to be extracted [retrieve]. […] We should provide correct keywords to get required data (E2).

> What I have learnt is that, digitisation is easy, but describing and preparing metadata is a tedious work […] Documents were scanned and a digital copy of the archive is available, but without proper description of the content [metadata] these documents are nearly useless. I think digitising documents without [comprehensive metadata] is essentially pointless (H1).

> Disciplinary differences should be addressed by discussing with the experts in the area. [...] Data is dependent on the domain and we need to discuss with experts from the area concerned while formulating policies regarding metadata and standards (L1).

Participants E2 and H1 observes that data stored in a repository must be properly organised to be useful. E2 observes that if data cannot be retrieved, its storage is effectively meaningless: “If we cannot retrieve data, there is no point in having a large data with us” This sentiment is echoed and deepened by H1, who draws on experience from a digitisation project where documents were scanned but rendered nearly unusable due to the absence of metadata. H1 emphasises that without accurate and comprehensive metadata, including appropriate keywords, the digital copies are essentially pointless. This reflects the broader view in the literature that metadata is not a supplementary aspect of data management but its semantic core (Borgman, 2017; Mayernik, 2016). These responses also align with the FAIR (Findable, Accessible, Interoperable and Reusable) data principles. Respondent L1 highlights the importance of addressing disciplinary variations in data documentation.

Standard Compliance

A repository could be a viable solution if coupled with clear guidelines for data depositing and sharing. The availabilities of both metadata standards and data repositories significantly affect researchers’ norm of data reuse, which influences their data reuse intentions (Kim, 2021).

> To ensure data quality, we need to check that it meets minimum technical standards and scientific standards. […] We can come up with

> scientific and technical guidelines for each type of data [...]. This standard guidelines will improve the quality of the research also (L2).

> In the case of funded research, there will be similar works under same fund. All these projects have similar genre and there should have a clear framework talking about the process of uploading before we start the research. If you already set up a framework for the data, then it will be easier to upload the data (E2).

Participant L2 has emphasised the need for both technical and scientific standards. E2 suggests that, it will be better to set the standards in advance, especially in the funded research. Participants are aware of the need to ensure quality in data while preparing standard guidelines. They also warn that, any regulatory guidelines which compromises the quality of data will make the repository worthless. This implies that funding agencies can play a vital role in upholding the standards. Additionally, it was also noted that standards also help improve the quality of the research. This highlights the need for a common set of guidelines to ensure that data is valid, reliable, and suitable for research purposes and future reuse as prescribed by FAIR principles.

Need for Data Review

Need for data review is connected with the data verification and closely related to the data quality. Supervisors have felt the need to go beyond the reports and check the raw data to confirm the findings.

> We do not get time to verify row data. In some cases, if there is a major problem in the findings, we reject publication by stating that "your data is not accurate". Most of the time we trust the authors with their data. If we have any doubt about the data, we may ask or use the option to reject the paper. We may not go after the raw data (considering the large volume of data) (L1).

> Peers in the peer review process can check this data while we forward our paper for publication. This peer review process may help detecting the mistakes in the datasets (L2).

> I've encountered PhD theses with questionable data during my evaluation (N2).

> I have often felt the need for checking the raw data behind the work. They only provide result, no row data, we do not get an option to cross-check. [...] Right now there are no mechanisms to verify the quality of data and quality of research (S2).

This code deals with the need to publish raw data along with the research publications to ensure transparency and improve research quality. Participants L1, N3, and S2 admits the need for data review due to the chances of data manipulation discussed earlier. L2 also shares the idea of peer review of data.

Existing Data Practices

At present research data management depends on the skill and attitude of the research students. According to L1, not everyone is able to systematically store and retrieve the data. E1 and L2 have shared their experience in data sharing. Participant E1 is using repository called Geo and participants from Computer Science have used Kaggle data repository. N3 has preserved her data over the years, but not in a systematic manner and, fully understand the need for systematic repository building for future use of these data.

> Right now the data preservation is purely depended on the skill and attitude of students. Some students keep the data systematically but, many are not doing this way. […] Not everyone is able retrieve the data when required. In some foreign universities [...] principal investigator or guide used to store data in their private repository even before the publication of the research report (L1).

> We publish data in the public domain. My works are already available in … (personal website), and larger datasets are shared in Kaggle data repository and Github. We have done a project in Ayurveda plants and all these images are available in open domain. People are using my datasets shared in the repositories. But not everyone cite (I do not care about citing) or acknowledge my work (E1).

> I have data deposited in GEO 4 years ago, and I have not completed the paper publication fully. There is a provision in repository to hold data till the paper is published (L2).

> I have preserved data generated over years of research, but have not compiled them for reuse (N3).

Disciplinary Differences

Disciplinary differences significantly influence data practices and sharing attitudes, reflecting varied norms, beliefs, and values across fields.

> Data types vary on the basis of the type of research. [...] We will be using macro level and micro level data in urban planning (E2).

> There are multiple disciplines and these disciplines generates diverse data. Managing these data is not an easy task (S1).

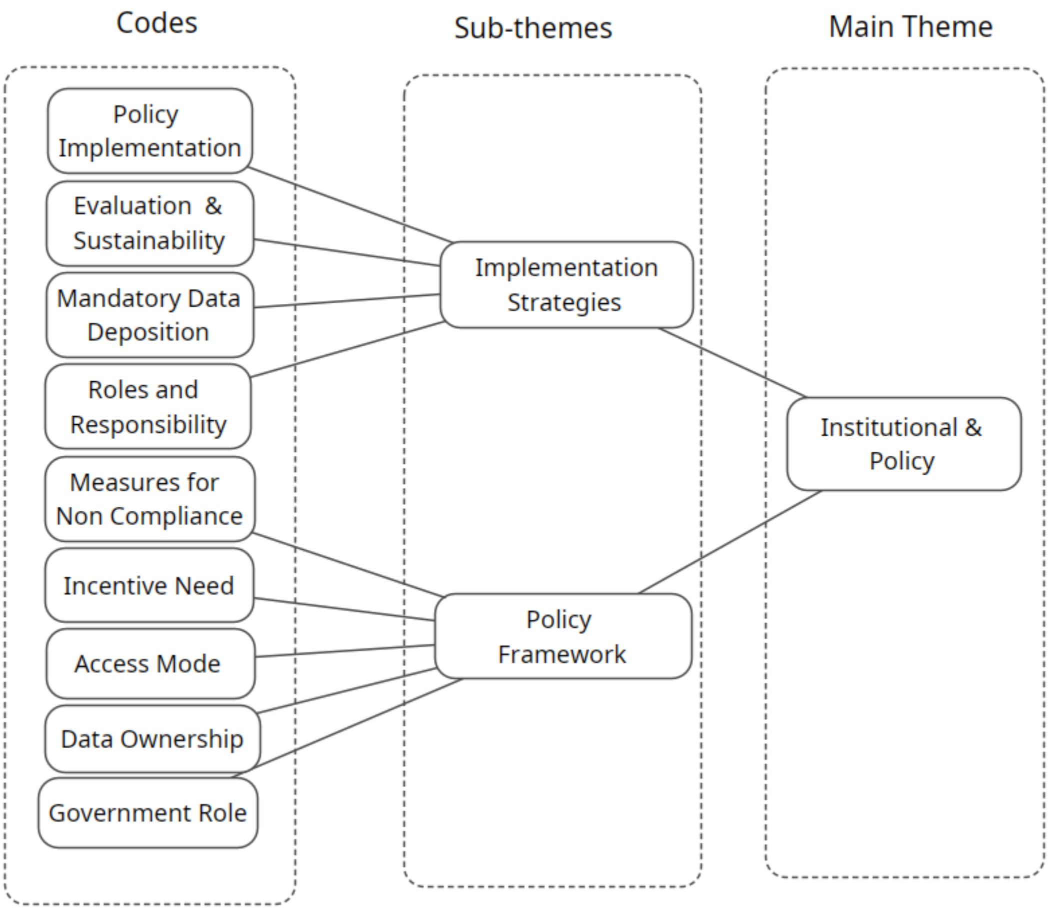
> There are two main types of data in historical research. 1. Original documents: These are the actual sources of information, such as original manuscripts, artefacts, or data sets. 2. Researcher's notes: These are notes and summaries created by the researcher while analysing the original documents. A significant portion of PhD research in history relies heavily on existing published materials. Data publication is primarily relevant when researchers discover or utilise new and previously unpublished documents (H1).

Participants E2 and H1 wanted to make sure that, any proposal to create a data repository should address the diversity in disciplines and the variety of the data generated in the research activities. Overall, data standards and practices ensure that data is reliable, usable, and of high quality. Metadata creation, standard compliance, and data review are crucial components to ensure data authenticity.

Data curation is highlighted as a key factor in ensuring the reusability of data. The need for a team dedicated to data management and support was highlighted by the participants. As L2 has mentioned, “The data generation has a purpose and the purpose fulfils only when the data is properly processed and prepared for reuse". This emphasises the need for active management of data to ensure that it is not only preserved but is also prepared for future use by different users. Data should be verified for quality, format, and content before depositing in a repository. Libraries are identified as a potential agency for data management. This refers to the need for institutional support, training, and rewards to facilitate data sharing. Universities should have institutional level repositories, data centres, and support systems to create repository platforms and make research data openly accessible.

Figure 5.6

Theme 5: Policy and Implementation



Theme 5: Policy and Implementation

The final theme policy and implementation deals with the need for regulations to establish research data sharing and repository at institutional level. Participants have mentioned the need for policy and suggested measures to encourage data sharing by addressing the challenges. A policy framework helps in establishing guidelines and protocols for data access, incentives, and overall governance.

Policy Framework

One of the key themes that emerged in the interview is the need to formulate policy for data sharing. Funding agency policies and journal policies have received attention, institutional level data policies are less well known (Briney et al., 2015).

Data Access Mode

There is a spectrum of opinion about the data access control. Some participants support full access to the data. L2 argues that all publicly funded research data should be freely available to the public. Participant H1 suggests controlled access due to the confidential nature of the data in discipline. L1 suggest that the access right should be decided by the original researcher. While H3 recommended seeking the permission from the data providers, N2 is doubtful of the attitude of the researchers and he recommends to create a data sharing committee which can resolve disputes when any researcher withholds data intentionally. S2 believes that data should be available only to eligible people doing research in the same field.

> We can release only basic details about the information [metadata] and keep the full documents confidential. [...] We should create agreements that clearly outline who can access the data and how. If a socially or politically relevant material contains any private matters within it, the data owner can suggest making it classified (H1).

> Researchers should have the right on the data, and they should be able to decide the access mode.[…] Details about data can be published, and people can contact the original researcher (L1).
>
> With the permission of the original data creator or collector (H3).
>
> While author request mode can be effective, its success heavily relies on the individual researcher's willingness to cooperate. To ensure a fair and consistent approach, a data sharing committee should be established according to established policies (N2).
>
> We need to control the access to data. Also ensure that the data is accessed and used for research purpose and societal betterment only (N3).
>
> Access should be granted only after carefully reviewing the purpose of the research and the researcher's credentials. Institutions should have clear guidelines and procedures in place to oversee data sharing and ensure responsible use [S2].

The results highlight the need for different levels of data access based on the sensitivity and potential risks related to the data. First category consists of data that requires restricted access. This category includes data containing sensitive personal information, data with potential security implications and, data where intellectual property rights need to be managed. According to the participants, access might be granted based on specific criteria, such as researcher affiliation, ethical approval, and data use agreement. The second category consists of data that may not be shareable at all. These data might be sensitive, poses significant risks if misused, or be subject to agreements that prevent sharing.

Incentives and Promotions

Researchers can be motivated to share data through incentives. Data citation is advised by majority of the participants. There are also opinions like providing financial benefit to the data generators to encourage sharing. Previous studies shows that data citations has advantages (Colavizza et al., 2020). The need for a mandatory data citation policy was also suggested to ensure that original researchers receive credit for their work.

> We cannot compel anyone. Data sharing is an individual decision. People may share data if they get any benefits from sharing data. I don't think ... [that people will share it voluntarily] (H3).

> With a discussion with the initial researcher one can arrive at an agreement about incentive for sharing data it could be authorship or acknowledgement (L1).

> We do not have a precedence of co-authorship in Malayalam, and I do not agree with the concept of co-authorship. Because the secondary user is the one who apply the logic and design the tool to analyse the shared data. Co-authorship is normally seen in the field of science (H5).

> We need to think about some sort of incentives [like Access Benefit Sharing]. People will naturally share when they receive something (S3).

Researchers may not always voluntarily share the data. H3 strongly believe that incentive mechanisms can encourage data sharing culture. It could be data citation, co-authorship or financial assistance as proposed by the participants. L1 suggests shared authorship but participants from the Humanities and Social Sciences find it not so feasible as mentioned by H5. In the case of co-authorship there is a clear variation of opinion between Science disciplines and Humanities and Social Sciences disciplines. Lack of well defined shared authorship norms may create uncertainty about credit and recognition. Formulating clear authorship and attribution policies is essential for clear, transparent, and equitable guidelines on credit allocation for secondary data use.

Government Role

This code discusses the role of government departments in generation and management of data.

> While working in the water authority department, we were directed to mark the maximum flood level during the flood 2018. We get rainfall and

> other related data from the Indian Meteorological Department, and we can do research and publish the results (E4).
>
> We may not always depend on government data sources, being a system they might have their hidden agenda and there may be manipulations in data. We may depend on NGOs and private agencies in sharing data (H4).
>
> While working on a project to monitor the Periyar river, I discovered that the Kerala Water Authority lacks sufficient data on water conditions across different seasons and years. Although they expressed willingness to accept data from research institutions, this data sharing isn't happening. [...] It's our responsibility to share data [generated by research institutions] with local authorities and [government] departments. These authorities should preserve this data and recognise its importance for future water management and decision-making purposes (N3).
>
> However, [majority of] government departments do not always collect and manage data effectively. Many government employees lack the training to work with data. […] We need to hire people with research skills for important government jobs. [...] A team of people with research skills would be better than relying on one person to collect data (S5).

Participants emphasis the need for collaboration between research institutes and government departments in data sharing. Participant N3 has the opinion that though government departments are collecting or generating valuable data and these data are not properly stored or processed. She is optimistic that the government agencies and research institutes can collaborate in data sharing. H4 has taken a cautious stand on government role, he believes that, government political agenda may reflect in the data. Data collection needs to be systematic right from the beginning for better result. Systematic data management would need skilled government staff with research background. A structured mechanism for academic and administrative data sharing should be established to enable collaboration between government departments and research institutions.

Fixing Data Ownership

Data ownership relates in some way to meaningful control, ownership, and other claims to data articulated by a variety of agents ranging from individuals to countries (Hummel et al., 2021). Data ownership is a much debated issue and participants have different opinion about data ownership.

> Initially for the short period the inventor or researcher shall be given some sort of benefits, but ultimately the output should go to the society. After a certain stage this should be converted to the societal knowledge. We need a policy that converts IPR to societal (S1).

> Our funding agency [SERB] does not mandate to share data to them or say anything about the data ownership. The intention of research funding is to help knowledge generation, and we just need to acknowledge the funding agency for the funding (L1).

> Individual who is working on the area is the owner and prime source. Though it is transferred from generations, he might have enriched it from his own experience. We provide list of all who have provided information for authenticity. That itself is a valuable contribution to research community (S4).

S1 believes that, researchers should receive initial benefits but ownership of research outputs should eventually be transferred to society. According to L1, funding agencies does not mandate data sharing or specify data ownership for funded research. Participant S4 recognises individuals as owners of knowledge as they enrich through their experience. Balancing individual contributions with societal benefits in research requires clear policies on intellectual property rights, data ownership, and traditional knowledge.

Measures for Addressing Non-compliance

Institutions need to prepare guidelines to ensure that researchers comply the research data guidelines to ensure authenticity and fair usage of data. In the context of government funded research, data sharing could be strengthened by introducing penalties for withholding data (Sherline, 2014).

> 'If [data is] not authentic then what is a major question. At a later stage, if it is found that someone has uploaded wrong data, it should be a black mark for the researcher. […] We should make sure that, there shall be a declaration that 'this data is authentic to the best of my knowledge', then only we should upload the data. If you are not confident of the data or if you have a level of ethical aspects with you, you will not be putting up [wrong or manipulated] data in the open forum (E2).

> As a prerequisite for thesis submission, all PhD students should be required to submit the complete set of raw data used in their research to their respective departments.[…] this would ensure that all research data generated during the course of the PhD program is properly documented and preserved (L3).

According to E2, the data depositor should ensure the authenticity of the data and any violation shall badly affect data depositors. To ensure mandatory data deposition, L3 suggests to connect data deposition with PhD degree awarding.

Implementation Strategies

Implementation strategies are vital in translating research data sharing policy framework into a practical action. Effective implementation depends on clearly defined roles, responsibilities, and specific measures. This ensures that data sharing becomes an integral part of the research culture. The strategies required to implement the research data sharing, various suggestions for data sharing and repository implementations are discussed under this sub-theme.

Roles and Responsibilities

According to E2, data depositors should be held accountable for ensuring the authenticity of uploaded data, which introduces both responsibility and potential challenges. H2 suggests that the role of managing data repositories shall be assigned to library. Providing training in research data management shall be the duty of library professionals. L3 believes in the role of institutions in research data management by ensuring sufficient human support and infrastructure. In addition to the support, N2 has suggested to formulate data sharing committees. S4 advocate separate division for data management and does not promote library for this role.

> We need to fix a level of responsibility to the uploader. This is a double-edged sword. If I want to upload data, I should be convinced that, the data is authentic. Ensuring the authenticity of data should be the responsibility of the depositor. It brings a level responsibility to the data uploading person. (E2)

> Library can provide training to research students in data sharing and research related tools. From the very beginning of coursework, students should be trained and equipped in research data management (H2).

> We could consider redesignating existing positions within the university to provide support to researchers. This service could be integrated with the library with the help of computer experts, providing researchers with convenient access to statistical expertise. This will undoubtedly enhance the quality of research conducted within the institution (L3).

> To ensure a fair and consistent approach, a data sharing committee should be established according to established policies (N2).

> A special division shall be created under the research wing, exclusively for the data management. The responsibility of creating the data repository does not have to be entrusted to the library only. If the library can do it, then the library or the IT department can be entrusted [...] Data sharing is the responsibility of the university, not [that of] individual (S4).

Mandatory Data Deposition

A mandatory data policy in research can significantly influence data sharing practices. A majority of the participants agrees to mandatory data depositing. Participants S4 and L3 have suggested linking the data deposition to the PhD thesis submission process ensuring data deposition. Though majority support the mandatory policy to deposit data, one participant has suggested to bring the voluntary culture of data depositing, instead of mandatory policies (H5).

> Publishing PhD thesis in the Shodhganga repository and plagiarism checking norms has changed the research culture. Researchers started publishing the soft copy of the research report only after the UGC mandate to publish it. PhD degrees are awarded only after the submission of the soft copy of the thesis to the library. Likewise, there should be mandatory policy to deposit raw data to the university. (S4)

> University should make data preservation mandatory. I was asked to produce data by the journal after 4 years of completion of the research work. I was only the corresponding author, my research students were first and second authors, the first author has completed research and left the department to join a teaching post. Unfortunately we could not produce all datasets as one among many datasets were missing. We lost it somehow. We could not send data to the journal publisher, as we did not store the data in the department. If we had a culture of data preservation this would not have happened. This retraction issue has affected me [my academic reputation]. (L1)

> Since this is a personal decision, 'unless and until it is mandatory, I think no one will give you data'. Our research community is not open to even a discussion on one's research topic. Research supervisors do not encourage such discussions. Human nature is self centric. We can frame a policy at the UGC level or university level to publish raw data in the new researches (S2).

> I do not support mandatory policy, can't we bring such a data sharing practice ? [voluntarily] (H5).

Policy Implementation

When developing a new system, especially one requiring cultural change, resistance is common as observed by L1. Participants have suggested several strategies to address these challenge. Data sharing culture can be implemented gradually in a phased manner (L3) as it involve a cultural change. Participants recommends linking the data sharing with the PhD thesis submission process. A policy of this nature could be implemented at the departmental level, commencing with the preservation of data under the supervision of the research supervisor, as recommended by Participant L3. The data submission can be linked with mandatory soft copy thesis submission as suggested by S2 and S4. Research departments can play a significant role in this as suggested by H1.

> We should create a framework for data repository. There should be a framework to upload data like 'Shodhganga' [ETD repository]. (E2)

> PhD students should submit all their research materials to their respective departments instead of retaining them personally. This practice ensures that, valuable research outputs are preserved and made accessible for future use. The department can then select and publish relevant documents, with the help of original researchers. (H1)

> Initially university need to create awareness about the need for data management and sharing. […] We need to convince them that the data preservation practice will be good for the students. […] If you bring a mandatory policy all of a sudden, there will be resistance [...]. Some universities are providing their thesis in electronic format for evaluation. If there is a mechanism for sharing raw data along with the thesis, the thesis evaluation will be more genuine. (L1)

> As a first step, we can encourage each department within the university to create its own repository for research data. […] It's important to recognise that a culture of collaboration and data sharing is still developing within our universities. Alternatively, we could require research supervisors to maintain a repository of all publications, records, and data generated by their research students. This approach may encourage greater participation, as researchers may be more comfortable sharing data with their trusted supervisors. (L3)

> Complete raw data should be submitted along with the PhD thesis. Data depositing process should be completed before awarding the PhD degree,

> once the degree is awarded it will be difficult to get the data from research students (S4).

Evaluation and Sustainability

Research data preservation and sharing is an ongoing process, as the ever-increasing complexities would require changes in the policies from time to time. Ensuring a sustainable repository is a challenging job. Repositories should ensure its long term availability, especially when researchers use secondary data for research as pointed out by participant E4.

> If research is done using the secondary data, how do we ensure the long term availability for later verification. Maybe we need to download and keep the data for later enquiries. (E4)

> This collaborative approach would not only benefit individual researchers but also contribute significantly to the overall understanding of the unique biodiversity of the Malabar region. To make this system more effective, a network connecting all university libraries across Kerala should be established (H1)

Major Findings from the Qualitative Analysis

1. The existing institutional infrastructure is not fully equipped to handle research data preservation and sharing.
2. The study emphasises the need to implement systematic data preparation and preservation practices.
3. Ensuring the authenticity of the research data is a challenge in doctoral research.
4. Open access to research data reduces the data manipulation and encourage quality and transparency in research.
5. Libraries are viewed as central agency for facilitating research data management, offering technical and organisational support.
6. Participants recognises research data as a public good, emphasising its collective value and the importance of making it openly accessible.

7. Driven by individual benefits and careerism, research in institutions in Kerala are primarily person-centric and lacks teamwork.
8. The pressure to produce more publications, is seen as having deteriorated the quality of data and the output of research.
9. Major effort required for preparing data in a shareable format is a barrier in the research data sharing.
10. There is a need for framing clear policies on accessibility to research data.
11. The implementation of research data sharing policies should follow a gradual and well-structured approach, preceded by sensitisation and capacity-building programmes.
12. Research data should be classified to determine the level of accessibility, ensuring the protection of national interest.
13. Research data should be mandatorily preserved in designated research centres.
14. Doctoral students should be required to submit Data Management Plan as part of their research proposal.
15. Doctoral students should submit their research data as a prerequisite for the award of PhD.
16. Research data repositories should be created and managed at institutional level.
17. A trustworthy research environment plays an important role in fostering positive attitudes toward data sharing among researchers.
18. Journal policies play a significant role in the research data repository and sharing awareness.
19. Funding agencies can encourage data sharing culture by mandating data sharing in the publicly funded research.
20. Research data repositories should prioritise the inclusion and preservation of geographically and locally significant data.

21. Government departments can play an important role by preserving and sharing the data generated by government departments.
22. Government departments shall appoint trained experts to oversee the collection and management of research data.

Chapter Summary

This chapter presents the result from the analysis of the semi-structured interview data collected from the research supervisors. Interview conducted with twenty research supervisors from five disciplinary categories. Qualitative data has provided suggestions to formulate policies in Indian context. Supervisors proposes policy oriented recommendations based on their dual role as academicians and policymakers. These views helped in understanding the historical and cultural aspects behind the research data management and attitude towards data sharing.

References

Ahmed, M., Othman, R., Noordin, M. F., Ibrahim, A. A., & Al-Hussaini, A. I. S. (2025). Factors influencing open science participation through research data sharing and reuse among researchers: A systematic literature review. *Knowledge and Information Systems*, *67*(3), 2801–2853. https://doi.org/10.1007/s10115-024-02284-3

Awada, L., Phillips, P. W. B., & Bogdan, A. M. (2022). Governance and stewardship for research data and information sharing: Issues and prospective solutions in the transdisciplinary plant phenotyping and imaging research center network. *PLANTS, PEOPLE, PLANET*, *4*(1), 84–95. https://doi.org/10.1002/ppp3.10238

Borgman, C. L. (2017). *Big data, little data, no data: Scholarship in the networked world.* The MIT Press.

Briney, K., Goben, A., & Zilinski, L. (2015). Do You Have an Institutional Data Policy? A Review of the Current Landscape of Library Data Services and Institutional Data Policies. *Journal of Librarianship and Scholarly Communication*, *3*(2), Article 2. https://doi.org/10.7710/2162-3309.1232

Buyle, R., Compernolle, M. V., Vlassenroot, E., Vanlishout, Z., Mechant, P., & Mannens, E. (2018). *"Technology Readiness and Acceptance Model" as a Predictor for the Use Intention of Data Standards in Smart Cities*. https://doi.org/10.17645/MAC.V6I4.1679

Chin, J., Zeiler, K., Dilevski, N., Holcombe, A., Gatfield-Jeffries, R., Bishop, R., Vazire, S., & Schiavone, S. (2023). The transparency of quantitative empirical legal research published in highly ranked law journals (2018–2020): An observational study. *F1000Research*, *12*, 144. https://doi.org/10.12688/f1000research.127563.1

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

Ho, R. C. Y., Wong, S. N., Chia, P., Tang, C., & Ng, M. T. T. (2025). Research data management services in academic libraries to support the research data life cycle: A systematic review. *Journal of the Association for Information Science and Technology*, asi.70008. https://doi.org/10.1002/asi.70008

Hummel, P., Braun, M., Tretter, M., & Dabrock, P. (2021). Data sovereignty: A review. *Big Data & Society*, *8*(1), 2053951720982012. https://doi.org/10.1177/2053951720982012

Kim, Y. (2021). A study of the roles of metadata standard and data repository in science, technology, engineering and mathematics researchers' data reuse. *Online Information Review*, *45*(7), 1306–1321. https://doi.org/10.1108/OIR-09-2020-0431

Kim, Y., & Stanton, J. M. (2016). Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*, *67*(4), 776–799. https://doi.org/10.1002/asi.23424

Klump, J. (2017). Data as social capital and the gift culture in research. *Data Science Journal*, *16*, 14. https://doi.org/10.5334/dsj-2017-014

Linek, S. B., Fecher, B., Friesike, S., & Hebing, M. (2017). Data sharing as social dilemma: Influence of the researcher's personality. *PLOS ONE*, *12*(8), e0183216. https://doi.org/10.1371/journal.pone.0183216

Mayernik, M. S. (2016). Research data and metadata curation as institutional issues. *Journal of the Association for Information Science and Technology*, *67*(4), 973–993. https://doi.org/10.1002/asi.23425

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). Sage.

Rao, K. N., Mair, M., Arora, R. D., Dange, P., & Nagarkar, N. M. (2024). Misconducts in research and methods to uphold research integrity. *Indian Journal of Cancer*, *61*(2), 354–359. https://doi.org/10.4103/ijc.ijc_4_23

Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). Sage.

Sherline, C. (2014). *Measuring the Norm of Reciprocity on Data Sharing Practices: A Carrot or Stick Approach?* [Thesis, University of Tennessee]. https://trace.tennessee.edu/utk_graddiss/3164

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. https://doi.org/10.1016/j.acalib.2022.102586

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Yoon, A. (2017). Data reusers' trust development. *Journal of the Association for Information Science and Technology*, *68*(4), 946–956. https://doi.org/10.1002/asi.23730

Zenk-Möltgen, W., Akdeniz, E., Katsanidou, A., Naßhoven, V., & Balaban, E. (2018). Factors influencing the data sharing behavior of researchers in sociology and political science. *Journal of Documentation*, *74*(5), 1053–1073. https://doi.org/10.1108/JD-09-2017-0126

Chapter 6. Discussion

This chapter synthesises the findings from the survey and interview. This integrated approach helps the study to meet its aims by showing why certain attitudes and behaviours exist. The discussion chapter is structured around the major findings emerged. Study offers basis for developing policy measures and practical strategies that can encourage a research data sharing.

Data as a Public Good

The concept of data ownership in the academic research is a complex and debated topic, involving various stakeholders and different perspectives. The results from the study indicate a tension between individual researcher ownership and institutional ownership. A significant viewpoint from the qualitative phase of the study is the idea of 'data as a public good', especially when research is publicly funded. As one participant stated, "Knowledge should also be treated as a common good like river or forest". Study suggests a model in which data is initially protected, giving the original researcher time to conduct additional work and publish further outputs. After this period, the data can gradually be made public through an embargo mechanism. The discussion around data ownership reflects different epistemological orientations and ethical considerations. While individual researchers have a vested interest in their data, the argument for 'data as a public good' emphasises the societal benefit of shared knowledge. The complexities surrounding indigenous knowledge further highlight the need for culturally sensitive and ethically sound approaches to data ownership.

With lesser amount spent for research, research institutions have to compromise on the purchase of scientific equipments and instruments for the laboratories. In what Varoufakis (2024) describes as an age of techno-feudalism, large corporations control most tools and capacities for big data analysis. In these context, citizens should have access to data generated with public funds. The shrinking budgetary allocation emphasises the need for publicly sharing available research data. Scholars like Appadurai (2013) believes that public should have access to the tools for research, and research should be the right of the citizen. Data preservation is the critical link that enables the creation of these repositories. It will also ensure that, publicly funded research contributes to a lasting and accessible body of knowledge for the benefit of society.

> The right to the tools through which any citizen can systematically increase the stock of knowledge that they consider most vital to their survival as human beings and to their claims as citizens (Appadurai, 2013).

Person-centric Research Culture

A strong sense of self interest and individual ownership of research findings often leads to data withholding. There is a tendency within the research community to avoid open discussions about the topics under research. This reluctance manifests as data withholding, where researchers avoid sharing data due to personal and professional insecurities. Many research supervisors does not promote collaboration, further reinforcing this individualistic attitude. The absence of encouragement from supervisors in data sharing is reported in the previous studies (Kumar et al., 2025). The research environment in Kerala is predominantly person-centric, emphasising individual contributions driven by personal recognition and career advancement. The lack of teamwork is identified in the study as a major barrier in the research collaboration. Researchers avoid sharing data to maintain their control or exclusivity.

Publication Pressure and Research Quality

Results from the quantitative phase shows that acknowledgement and visibility are the two major influencing factors in data sharing behaviour. This finding corroborate with previous studies (Chikoore, 2016; Kim & Adler, 2015; Murtaza Ashiq et al., 2022; Navarro-Molina & Melero, 2019; Saeed & Ali, 2019). Participants from both phases of the study agree that data sharing can help promote collaborative research. Increased visibility through data sharing is perceived as a major benefit, potentially leading to more citations and a higher chance of publication. However, the pressure to publish for career achievements influence the researchers in prioritising the quantity of publications over quality of research work. Careerism significantly influences research practices. Doctoral students prioritise academic degrees and job prospects over producing impactful research.

The pressure to produce publications for institutional rankings further intensify this issue, often compromising data and research quality. This was noted in the interview with a senior supervisor as he notes “The pressure to publish for career advancement incentivised researchers to prioritise quantity over quality. This mentality can lead to an increase in the number of low-quality publications, compromising the overall integrity of scholarly research” (L3). Various institutional ranking criteria and career promotional mechanisms in India depends on the quantifiable scales leading to publication pressure.

> The quantification of output in terms of teaching hours, students’ performance, research papers, contributions to the income of the university from business and students, or ‘key performance indicators’, have become the yardstick for judging teachers, departments, and institutions (Purkayastha, 2024, p. 23).

Data Quality as a Barrier in Sharing

The quality of data generated in doctoral research are often perceived as inadequate due to concerns about accuracy and reliability in data. The lack of institutional mechanisms to ensure quality and data authenticity of data remains a significant challenge. Observations of questionable data in PhD observed by the interview participants suggest that it is treated as an academic exercise rather than practical contribution. If the researcher find that data might not be useful to others, they are unlikely to invest time and effort in sharing. For data to be seen as reusable, researchers must trust its reliability and contextual clarity. This further boosts their confidence in sharing as evident from the qualitative phase of the study. Concerns about data misuse and misinterpretation remain significant barriers to data sharing among researchers from the Engineering, Life Science and Natural Science categories. Ethics and violation of privacy are major concerns among researchers in Humanities and Social Sciences disciplines.

Evidence from the study shows the absence of proper systems to verify raw data and the inconsistent research practices followed by doctoral students. Research supervisors are often hindered by workload pressures. This can be solved by providing reviewers with access to raw data for verification and prevent data manipulation. The study point out the need for data quality checks. The supervisors' suggestions indicate growing concern about data quality issues in doctoral research. They also suggest need for a cultural shift towards stronger and more consistent data management practices. According to a supervisor from the Engineering discipline "A major challenge in data repositories is ensuring authenticity and quality of data". The suggestion of establishing a committee to randomly check data quality could be effective in preventing falsification. The recommendations point to the need for a multi-pronged approach that brings together individual researcher responsibility, strong institutional support, and clear policy frameworks.

Data Drain from Publicly Funded Research

Research funded with public money often ends up being published in foreign journals. As a result, the intellectual property and potential benefits are captured by international publishers and corporations rather than remaining within the country. This aligns with the concerns raised by Purkayastha (2024) in his book *Knowledge as Commons*. He argues that publicly funded research is increasingly being privatised and enclosed through intellectual property rights. Such practices allow private capital to benefit while limiting the wider societal gains that public research is meant to support. Perelman (2002) has discussed how knowledge generated using the public money is privatised. The issue of cross-border data transfer is a complex one, driven by a combination of factors like publication pressure, career motives, and the globalised nature of research.

While the free flow of knowledge is crucial for scientific progress and collaboration, there are concerns about the data drain. Research supervisors fear that valuable data, including culturally and strategically important data, might be transferred and potentially exploited without adequate safeguards. Participants have shared a critical concern regarding cross-border data transfer as discussed in other studies (Chaudhuri & Kang Joseph, 2024). This is driven by the reliance of Indian researchers on foreign publications. The reliance is often influenced by the publication pressure and career motives, pushing researchers to publish their findings (González-Alcaide et al., 2017; Odeny & Bosurgi, 2022), even before the completion of their projects. Though international collaboration and dissemination are vital, the results from the interview shows the need to foster a research culture that balances openness and protect national interest.

Cultural and Behavioural Barriers to Data Sharing

Variations in data management practices and data sharing attitudes across different academic disciplines (Numajiri & Hayashi, 2024) contribute to the challenges in establishing universal data sharing policy. According to the survey results, participants from the Humanities reported higher effort as a barrier in data sharing compared to other disciplines. This could be attributed to the lack of standardised data management practices and tools tailored to the field's needs as noted by Oliver et al. (2023). Researchers from the Humanities mainly follow qualitative kind of research method and produce data which is complex in nature, demanding additional effort to share. The effort required for data sharing are perceived differently across disciplines (Thorogood & Knoppers, 2017). This highlights how institutional support, can influence attitudes of researchers by affecting their perceived ability to share data.

The results echo global surveys that identify the time and effort required to prepare data as a major obstacle (Digital Science et al., 2017). In addition to the busy schedule with teaching and administrative responsibilities, research supervisors have expressed their inability to spend additional time for data verification and management. Despite strong motivations like visibility to share data, the effort involved in formatting, documenting, and depositing data discourages researchers from sharing. Moreover, studies suggest to invest more effort in datasets with high probability of reuse (Pronk, 2019). Therefore, if scientists believe that data sharing requires their effort, they are less likely to share their data with others (Kim & Stanton, 2016).

Trustworthy Research Environment

The unethical practices and the absence of trustworthy research environment are factor preventing open discussion in academic research in Kerala. This environment influences the data sharing behaviour of researchers. As one participant from Engineering has pointed "If you are not confident of the data or if you have a level of ethical aspects with you, you will not be putting up [wrong or manipulated] data in the open forum [data repository]". When researchers are committed to ethical principles and are confident in the integrity of their data, they are more likely to share their data. As a supervisor from engineering discipline (E3) noted 'There are some unfortunate things like reviewers holding article for few days, understands the trend line and he or she publishes the same work. Such things are also happening". The commercial exploitation of the publication industry also reported to have influences the sharing behaviour (Dijck, 2014). Result reveals that the existing research culture itself can be a barrier to the data sharing.

Study recognises data sharing as a mechanism to identify unethical practices, thereby improving research quality. The finding from the qualitative phase reveals the existence of unethical research culture and its influence on the quality of data and the research. Data fabrication, data manipulation and the lack of methodical data preparation are some of the research misconducts identified in the study. Unscientific data management and weak data preservation practices are evident in the quantitative phase also. These unethical practices in research could influence the quality and authenticity of the research data (Yoon, 2017). The research integrity issues are reported in previous studies (Evans et al., 2022; Gorman et al., 2019). The results emphasise that a collaborative and inclusive approach is essential for systematic research data management and sharing.

Infrastructure for Research Data Management

The absence of adequate infrastructure, technical skills, and support systems are the major barriers of research data sharing identified in the study. The lack of teamwork leads to reliance on individual skills and attitudes. The result suggest that addressing these challenges could involve collaborative approaches, like developing “research clubs”. Research clubs can provide researchers with practical support and opportunity for inter disciplinary collaboration. It will also provide an opportunity to learn ethical research data practices. However, this should be supported by institutional mechanisms for quality checks (Peng et al., 2016). The study among clinical trialists show the need for resources and support for data sharing as a primary need (Lo & DeMets, 2016). Results also show weak infrastructure, lack of resources, and absence of positive attitude for effectively collecting and managing data in the government departments. Appointing employees with research aptitude in Government departments to support and monitor data collection process can solve this problem.

The survey results reveal heavy reliance on personal and decentralised storage methods. This dependence on less secure methods, coupled with weak backup practices increase the risk of data loss. Despite strong data sharing motivations, the weak infrastructure limit the potential for data sharing. The need for reliable storage is evident from the backup practices of doctoral students reported in the survey results. This need is further underscored by the data loss incident described by one of the interview participants from Life Sciences. The development of centralised storage facilities and enforcement of backup and storage policies are critical to enhance data security. The study also shows that having a user-friendly interface is an important factor. It makes repositories easier to use and can motivate researchers to share their data. Earlier studies have highlighted the need for apt repository platforms (Gomes et al., 2022; Wilkinson et al., 2016). The data repository should be able to interact with the common man also. All the output related to the research should be interconnected for easier navigation and optimum use.

Libraries as Agents in Research Data Management

The study has identified libraries as a potential agent in providing training and managing repositories. According to one participant from Social Science, “Library and Information science can play a role in the data awareness program. Library staff can provide data awareness programmes for the students” (S1). In the opinion of another participant, “To adequately prepare future library professionals for this critical role, library science curricula must include comprehensive training in digital preservation and metadata management of archival materials" (H1). This would help library professionals develop the skills needed to serve researchers and doctoral students through libraries. However, this also calls for a shift in the current work culture and in the attitudes of staff involved in research data services, as (Anderson, 2004) suggests. Libraries are widely used for research data training and repository management in various universities.

Incentive for Data Sharing

Data citation is a widely recognised and acknowledged primary incentive for data sharing. Apart from data citation, opportunities for collaboration, increased visibility to the research are the benefits perceived by the participants. According to Lo and DeMets (2016) "secondary investigators could be collaborators rather than antagonists, with shared authorship as an incentive and an ideal". Survey result shows that financial rewards is the least preferred benefit by researchers. Shared authorship and financial rewards are not universally accepted by participants of the interview. Given the low mean value and high variability of financial rewards found in the survey, offering monetary incentives may not be broadly effective and could be targeted to specific groups who value them. Participants from the Humanities do not agree with the concept of shared authorship for secondary data users. In the cases where data benefits specific communities, incorporating principles of benefit-sharing could ensure that data is used ethically, with communities benefiting from the research outcomes (de Souza & Bhardwaj, 2024; Laird et al., 2020).

Training in Research Data Management

The survey results indicate that majority of the researchers across disciplines view time and effort as major barrier in data sharing. This could be aligned with the demand of the participants for training in research data management. Lack of awareness on secure storage solutions, the poor back up practice, high rate of respondents who have not taken any privacy protection measures suggests uncertainty, necessitating targeted education. Supervisors propose inclusion of research data management in postgraduate coursework to equip students with necessary data literacy skills tailored to the disciplinary needs. While the quantitative phase identified the areas where training is required, the qualitative phase emphasised the need for including the training as part of the doctoral research. Hands on training is expected to provide the researchers with proper support (Heise et al., 2023) and training should be imparted to different stakeholders involved (Rousi et al., 2024).

Cultural Shift for Open Research Data

Institutional pressures and policies can encourage or discourage research data sharing. From the institutional point of view, the transition to mandatory data sharing policies may face initial resistance from the doctoral students. Journals and funding agencies can support this shift by providing infrastructure for data storage, clear guidelines, and incentives for open science practices (Park & Greene, 2018). Such measures would complement suggestions like mandatory sharing policy of the funding agency discussed by interview participants (E3 and L2) and institutional reforms to prioritise quality over quantity (S5 and L3), fostering a more transparent and collaborative research culture. Journal data policies and peer review processes encourage data sharing (Zenk-Möltgen & Lepthien, 2014). The findings suggest that journal policies play an important role in driving change in doctoral students. Data sharing initiative would face initial resistance from the doctoral students due to the deeply embedded cultural practices. It is suggested that implementation should take place in phases and only after adequate sensitisation and training. This approach recognises the need for a gradual cultural shift in how research data is managed and shared.

Policy Needs for Data Sharing

Research data management policy can significantly influence data handling practices. Majority of the interview participants agree to mandatory data depositing policy. Deposition of research data along with the PhD thesis report is notable suggestion derived from the study. Study highlights the need for researchers to meticulously prepare data in a standardised and reusable format. Currently, data preservation practices depend on skills and attitudes of the individual. According to participant from the Life Sciences "Once students are aware and vigilant about a mechanism of data preservation they will be preparing data methodically" (L1). Authority control, consistent use of terms and identifiers, should also be integrated to the policy to enhance discoverability of the data and integration (Spiegelman, 2021).

> Researchers should submit raw data along with the PhD thesis. It will be safe for the researcher. Data preservation will address problems like authenticity of the work, or challenging our work. We can defend our stand (L1).

The challenges in data quality were discussed and measures are proposed to address the challenges. There shall be policies to ensure accountability in data sharing (Anger et al., 2022). The proposed solutions reflect a desire for a multi-tiered approach to address these challenges. Policies mandating data sharing and preservation (L1) could incentivise better data management practices, while standardised guidelines (L2) could provide a framework for consistency across disciplines. The suggestion to establish research associates and departmental committees (E1) indicates the need for distributed responsibility to alleviate the burden on supervisors and enhance accountability. The primary responsibility for ensuring quality and authenticity of deposited data should rest with the data depositors. Researchers who violate the guidelines should be blacklisted from further research funding as recommended in the interview.

University level data repositories can be created by using the digital infrastructure that already exists. This approach was recommended by a supervisor with experience in managing university digital infrastructure. “We have the technical infrastructure to create repository platforms, we must upload and share our research data in these servers and make these open accessible to the public”. Given the large volume of data and the complexities involved, institutional repositories would provide a practical way to manage research data. A national-level consortium connecting these repositories would further help in findings data preserved in various institutional repositories.

Theoretical Interpretations

The section interpret the results through two theoretical frameworks guiding the study. The integration serves to generate a theoretically grounded synthesis that reconciles micro-level behavioural dynamics with macro-level structural influences.

Theory of Planned Behaviour

The first component of TPB is the attitude towards data sharing. The positive attitude grows out of a set of behavioural beliefs about what data sharing can achieve. The study identify academic and societal benefits of data sharing. These perceived advantages shape a favourable attitude towards data sharing. At the same time, negative behavioural beliefs also shape how researchers judge data sharing. Though the overall attitude towards sharing is positive, the specific behaviour of sharing research data is influenced by these positive and negative beliefs and evaluations.

Subjective norms in TPB reflects the perceived social pressure to perform a behaviour based on what important referents like research supervisors and peers approve or disapprove of. The study points to the need for clear institutional guidance and policies. Yet the findings reveal unclear, conflicting, normative landscape. The person-centric culture and the lack of teamwork imply that the dominant social norm may not favour collaboration and data sharing in practice. The concern about acknowledgment reflects a perceived norm where credit is not given. A trustworthy research environment is seen as encouraging sharing. The low trust, related to weak subjective norms around collaboration and integrity acts as a barrier in sharing.

The third component, Perceived Behavioural Control (PBC), refers to how easy or difficult a person thinks a behaviour will be. It is shaped by the beliefs about the resources, skills, and opportunities they have to carry it out. The finding that effort and time required for data preparation is a direct manifestation of low PBC. The identified need for training and institutional support directly corresponds to factors that would enhance PBC. The decentralised storage practices also point to a lack of easily accessible institutional resources, lowering PBC for centralised deposition. Limited supervisor support reflect gaps in perceived behavioural control and subjective norms.

Institutional Theory

The demand for mandatory policies is a major finding of the study. Need for policies and institutional frameworks reflects the understanding that formal regulations are necessary to enforce data sharing. These regulative elements aim to create the necessary regulative environment for sharing data as the regulative pillar of Institutional Theory.

The finding that research is primarily person-centric and driven by individual benefits and careerism reveals a strong prevailing norm in the Normative pillar of Institutional Theory. The lack of collaboration and team work is an outcome of this individualistic normative structure. Researchers focus on their own progress rather than collaborative efforts like data sharing. This norm contributes to data withholding. While there is an emerging positive attitude towards “data as a public good”, this appears to be an aspirational or ideal norm rather than the dominant, practised one. The suggestion of “research clubs” and multidisciplinary teams can be seen as bottom-up attempts to cultivate new, more collaborative norms. Disciplinary differences suggest that fields with established data-sharing norms, may foster this trust.

Challenges like the fear of misinterpretation or misuse, concern about acknowledgment, and beliefs about data confidentiality represent shared cognitive elements. The reluctance and ego contributing to the lack of teamwork can be understood as cognitive factors shaping how researchers perceive their data and their interactions. The uncertainty of participants in research data management reflect gaps in shared cognitive understanding. The need for “sensitisation and training” directly targets changing these cognitive elements.

Theoretical Synthesis

Social pressure to share data connects closely with the subjective norms described in the TPB. Researchers often look to the expectations of their supervisors and peer group creating a pressure. The normative pillar of Institutional Theory helps make sense of this. It shows how the norms of a discipline, the shared expectations, and the everyday practices within a research community slowly shape how people behave. When the broader community views data sharing positively it becomes another source of normative pressure. The lack of teamwork is a manifestation of the individualistic norm. The reluctance and ego contributing to the lack of teamwork are part of the cognitive framework within which researchers operate. The shared cognitive understanding is that, the individual effort is paramount and collaboration is low. In such an environment researchers might not trust others. Researchers may believe that they do not want to share data due to these cognitive barriers.

Promoting data sharing requires a shift towards more collaborative practices and a less person-centric culture. This change is challenging because it confronts deeply embedded normative and cognitive elements of the existing institutional context. The resistance to adopting data sharing policies and practices can be understood as resistance to changing these ingrained norms and beliefs. Addressing the lack of teamwork and fostering collaboration is thus important in the successful implementation of policy. Simply imposing regulative measures may not be sufficient without addressing the underlying cultural factors. Collaborative solutions like “research clubs” or multidisciplinary support teams are suggested as potential ways to build better practices. In essence, within the framework of Institutional Theory, the lack of teamwork is presented as a significant normative and cognitive barrier that reinforces a person-centric culture. This leads data withholding and resistance to change initiatives aimed at promoting data sharing.

Fostering data sharing culture requires a deliberate effort to address the interplay of individual psychological factors (TPB) and the broader institutional context (Institutional Theory). Simply implementing regulative policies without addressing normative and cognitive elements is likely to face resistance. A gradual implementation of data sharing, introduced after proper sensitisation and training, would be more effective and sustainable. This implicitly recognises the need to address the normative and cognitive dimensions for successful institutional change. Policies must not only mandate sharing (regulative) but also actively cultivate collaborative norms (normative) and build trust and understanding (cognitive). It should also be supported by adequate resources and training (PBC). It will help overcome the barriers identified, especially the deeply rooted lack of teamwork and person-centric culture. Drawing on institutional theory, effective change requires addressing the regulative, normative, and cognitive reasons for behaviour. While regulative changes can be implemented relatively quickly, fostering a culture of data sharing involves gradual shifts in norms and internalised values. These steps help strengthen subjective norms and attitudes in TPB terms. They also address well-known barriers related to credit, recognition, and institutional support.

Interpreted through TPB, the results show that perceived behavioural control (skills, tools, time) and subjective norms (supervisory and departmental expectations) strongly shape intention to share date. Institutional Theory complements this by highlighting the regulative (policies, mandates), normative (recognition and credit), and cultural-cognitive (assumptions about data ownership and sharing) conditions that either enable or suppress sharing. Universities are embedded within their institutional contexts, and a significant change is difficult not only because of the internal resistance but also because of this normative embeddedness.

Policy Suggestions Based on the Study

The findings leads to the development of a policy for building research data sharing ecosystem. The success of the Shodhganga ETD repository, driven by a mandatory UGC policy, demonstrates the power of top-down regulative measures in encouraging data deposition. Implementing effective research data sharing requires significant infrastructural and policy changes. It is evident from the study that lack of guidelines and weak infrastructural support raise perceived risk and workload for researchers. Clear national mandates paired with resources address barriers consistent with Institutional Theory's emphasis on the power of regulative pillars.

Establish a Clear Policy for Research Data Sharing

- **Mandatory Data Deposition:** The study strongly suggests that a mandatory policy be formulated to implement research data sharing. This should be linked to the thesis submission process, in which doctoral students are required to deposit their research data with their department before the PhD degree is awarded. The implementation of the mandatory policy should be in a gradual pace and only after proper sensitisation and training.

- **Data Classification and Access Mode:** Framing clear data access policies is essential. This includes classifying data as shareable and not shareable, especially considering sensitivity of the data. Access should ideally be controlled and primarily for research purposes and societal betterment.

- **Data Ownership:** Addressing fears through clear communication and controlled access mechanisms can foster a positive attitude towards sharing. Converting Intellectual Property Rights to societal knowledge after an initial period through embargo mechanism is recommended.

- **Addressing Non-compliance:** Policies should include measures for non-compliance with data sharing regulations. The data depositor should be held accountable for the authenticity of the uploaded data. An internal data committee should be established to handle complaints related to data quality and authenticity.

- **Technical and Scientific Standards**: Establishing minimum technical and scientific standards, guidelines, templates, and training for researchers are essential. Fixing responsibility for data quality on the uploader can improve accountability. Appointing research associates and departmental committees to monitor the data can strengthen this further.

Institutional Level Repository

- Various Universities already have digital infrastructure to set up institutional level repositories. By setting up an institutional level research data repository, the universities can make use of the data generated from the public fund.

- At present, data storage is mostly decentralised and based on individual preferences. There is a significant gap in reliable institutional storage. Hence universities and research institutions need to invest in centralised, secure, and scalable storage facilities.

Human Support

- Human support is essential in research data management. Specialised roles, such as data curators are also essential for successful implementation of the repositories. A support system can be created at the departmental level to guide students and researchers. At the institutional level, multidisciplinary teams and ‘research clubs’ can strengthen collaboration and skill development. In addition, involving recent PhD graduates in these support structures could create meaningful employment opportunities.

Training and Awareness Programs

- **Mandatory Training:** Research data management training should be integrated into the PhD curriculum. A practical, hands-on training program is suggested as more effective. Libraries can play a major role in providing training.

Foster Supportive Research Culture and Provide Incentives

- **Promote Collaboration and Teamwork:** The person-centric research culture can be addressed by creating systems that actively encourage collaboration and knowledge sharing. Promote awareness about responsible data management, and implement clear policies for data ownership and intellectual property rights. Sensitisation programs and mechanisms for controlled access can be used to address fears of misinterpretation or confidentiality breaches.

- **Incentives and Recognition:** Incentive mechanisms should be introduced to encourage data sharing. These may include promoting data citation, clarifying norms around shared authorship to motivate researchers to share their data, and offering formal recognition or rewards.
- **Mechanisms for Building Trust:** A trustworthy research environment encourages data sharing. Policies should aim to build trust by addressing researcher concerns transparently.

Address Researcher Concerns and Challenges Directly

- **Privacy and Ethics:** To address concerns about privacy, ethics, and potential misuse, clear policies on data access, confidentiality, and ethical practices should be implemented.
- **Acknowledge and Cite:** Proper acknowledgment and data citation should be ensured. Training on ethical citation practices is also required.

Ensure Data Quality and Prevent Data Manipulation

- **Data Quality and Misinterpretation:** Policies should address concerns about misinterpretation and aim to improve the quality of data. Mandatory preservation and sharing policies would encourage researchers in methodical data preparation. Mechanisms to verify raw data and a data review process should also be established.
- **Assigning Responsibility for Data Authenticity:** Policies should make it clear that the data depositor should be held accountable for ensuring the authenticity of the uploaded data. Policies should include measures for non-compliance with data sharing regulations to ensure authenticity and fair usage.
- **Mandating and Supporting Metadata Creation:** Policies should mandate and provide guidelines for metadata creation. This helps address concerns about data misinterpretation and understand the context of the data.

Recognise Disciplinary Differences

- Policies should be flexible enough to accommodate the variations, and consider disciplinary needs in aspects like metadata standards and policy formulation. Discipline specific training will help prepare and manage and share their data responsibly.

Actors in the Research Ecosystem

- **National Level:** Government of India should provide the regulatory framework for research data sharing. A facility of federated search at the national level could also be established to provide visibility and access to the institutional repositories.

- **State Level:** State higher education authorities and funders should translate the national regulation into grant conditioned Data Management Plans (DMP), time-bound deposit for thesis datasets, and budget lines for FAIR-compliant repositories and training.

- **Funding Agency:** Funding agencies should encourage research data sharing by mandating it for publicly funded research. At the national level, funders shall constitute a monitoring committee for open-access datasets to oversee usage and adjudicate disputes.

- **University Level:** University leadership and research offices shall convert system mandates into institutional policy, incentives, and accountability. Universities should therefore (i) recognise datasets as citable outputs in evaluation, (ii) integrate data citation tracking in research information systems, and (iii) publish repository performance dashboards.

- **Libraries and IT services:** These are the operational agencies for raising perceived behavioural control. They should operate a FAIR-compliant data repository and provide training to the doctoral students in RDM.

- **Research Departments:** Within departments, research supervisors and doctoral committee shape the immediate research environment in which data practices are learned. Departments should appoint RDM champions, and integrate short in-house data sessions into research seminars.

- **Management Committee:** For managed-access data, a data access and management committee should review requests and assess researcher competence, proposed use of the data, and the duration of access. The committee should also publish periodic reports and address any breaches. This approach maintains transparency in restrictions while still supporting legitimate data reuse.

- **Doctoral students:** Doctoral students should prepare DMPs and use standard consent language permitting secondary use. Doctoral students shall deposit datasets on a time-bound schedule and cite others' data and use controlled access where ethics require it.

- **Feedback Mechanism:** Policy makers and stakeholders should implement feedback mechanisms. Various bodies shall publish indicators such as percentage of projects with approved DMPs, numbers of dataset DOIs, reuse and citations. Other indicators such as controlled access requests, training coverage shall be published.

Chapter Summary

The findings provide a basis for broader discussions and targeted interventions to promote research data sharing among doctoral students. The findings explain why deliberate policy and infrastructure interventions are required for realising the public value of research data. Having established the major empirical outcomes, this chapter interprets the findings using TPB and Institutional Theory frameworks, which together explain individual intentions and behaviour within a broader organisational and cultural context.

References

Anger, M., Wendelborn, C., Winkler, E. C., & Schickhardt, C. (2022). Neither carrots nor sticks? Challenges surrounding data sharing from the perspective of research funding agencies—a qualitative expert interview study. *PLOS ONE*, *17*(9), e0273259. https://doi.org/10.1371/journal.pone.0273259

Appadurai, A. (2013). *The future as cultural fact: Essays on the global condition*. Verso.

Chaudhuri, R., & Kang Joseph, A. (2024). Living in a fragmented world: India's data way. *India Review*, *23*(2), 154–176. https://doi.org/10.1080/14736489.2024.2324638

Chikoore, L. (2016). *Perceptions, motivations and behaviours towards 'research impact': A cross-disciplinary perspective* [Thesis, Loughborough University]. https://hdl.handle.net/2134/22942

de Souza, S. P., & Bhardwaj, K. (2024). India's Conception of Community Data and Addressing Concerns for Access to Justice. *Digital Society*, *3*(1), 16. https://doi.org/10.1007/s44206-024-00102-5

Digital Science, Hahnel, M., Treadway, J., Fane, B., Kiley, R., Peters, D., & Baynes, G. (2017). *The state of open data report 2017*. Digital Science. https://doi.org/10.6084/M9.FIGSHARE.5481187

Dijck, J. van. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, *12*(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776

Evans, N., Buljan, I., Valenti, E., Bouter, L., Marušić, A., de Vries, R., Widdershoven, G., & the EnTIRE consortium. (2022). Stakeholders' experiences of research integrity support in universities: A qualitative study in three european countries. *Science and Engineering Ethics*, *28*(5), 43. https://doi.org/10.1007/s11948-022-00390-5

Gomes, D. G. E., Pottier, P., Crystal-Ornelas, R., Hudgins, E. J., Foroughirad, V., Sánchez-Reyes, L. L., Turba, R., Martinez, P. A., Moreau, D., Bertram, M. G., Smout, C. A., & Gaynor, K. M. (2022). Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1987), 20221113. https://doi.org/10.1098/rspb.2022.1113

González-Alcaide, G., Park, J., Huamaní, C., & Ramos, J. M. (2017). Dominance and leadership in research activities: Collaboration between countries of differing human development is reflected through authorship order and designation as corresponding authors in scientific publications. *PLOS ONE*, *12*(8), e0182513. https://doi.org/10.1371/journal.pone.0182513

Gorman, D. M., Elkins, A. D., & Lawley, M. (2019). A Systems Approach to Understanding and Improving Research Integrity. *Science and Engineering Ethics*, *25*(1), 211–229. https://doi.org/10.1007/s11948-017-9986-z

Heise, V., Holman, C., Lo, H., Lyras, E. M., Adkins, M. C., Aquino, M. R. J., Bougioukas, K. I., Bray, K. O., Gajos, M., Guo, X., Hartling, C., Huerta-Gutierrez, R., Jindrová, M., Kenney, J. P. M., Kępińska, A. P., Kneller, L., Lopez-Rodriguez, E., Mühlensiepen, F., Richards, A., … Weissgerber, T. L. (2023). Ten simple rules for implementing open and reproducible research practices after attending a training course. *PLOS Computational Biology*, *19*(1), e1010750. https://doi.org/10.1371/journal.pcbi.1010750

Kim, Y., & Adler, M. (2015). Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. *International Journal of Information Management*, *35*(4), 408–418. Scopus. https://doi.org/10.1016/j.ijinfomgt.2015.04.007

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. https://doi.org/10.47909/ijsmc.192

Laird, S., Wynberg, R., Rourke, M., Humphries, F., Muller, M. R., & Lawson, C. (2020). Rethink the expansion of access and benefit sharing. *Science*, *367*(6483), 1200–1202. https://doi.org/10.1126/science.aba9609

Lo, B., & DeMets, D. L. (2016). Incentives for clinical trialists to share data. *New England Journal of Medicine*, *375*(12), 1112–1115. https://doi.org/10.1056/NEJMp1608351

Murtaza Ashiq, Ashiq, M., Nosheen Fatima Warraich, & Warraich, N. F. (2022). Librarian's perception on data librarianship core concepts: A survey of motivational factors, challenges, skills and appropriate trainings platforms. *Library Hi Tech*. https://doi.org/10.1108/lht-12-2021-0487

Navarro-Molina, C., & Melero, R. (2019). Motivación, barreras e incentivos para la compartición y reutilización de los datos de investigación. Visión de los investigadores. *El Profesional de La Información*, *28*(5). https://doi.org/10.3145/epi.2019.sep.16

Numajiri, H., & Hayashi, T. (2024). Analysis on open data as a foundation for data-driven research. *Scientometrics*. https://doi.org/10.1007/s11192-024-04956-x

Odeny, B., & Bosurgi, R. (2022). Time to end parachute science. *PLOS Medicine*, *19*(9), e1004099. https://doi.org/10.1371/journal.pmed.1004099

Oliver, G., Cranefield, J., Lilley, S., & Lewellen, M. (2023). Data Cultures: A scoping literature review. *Information Research an International Electronic Journal*, *28*(1), 3–29. https://doi.org/10.47989/irpaper950

Park, Y., & Greene, C. S. (2018). A parasite's perspective on data sharing. *GigaScience*, *7*(11), giy129. https://doi.org/10.1093/gigascience/giy129

Peng, G., Ritchey, N. A., Casey, K. S., Kearns, E. J., Prevette, J. L., Saunders, D., Jones, P., Maycock, T., & Ansari, S. (2016). Scientific stewardship in the open data and big data era roles and responsibilities of stewards and other major product stakeholders. *D-Lib Magazine*, *22*(5/6). https://doi.org/10.1045/may2016-peng

Perelman, M. (2002). *Steal this idea: Intellectual property and the corporate confiscation of creativity*. Palgrave Macmillan US.

Pronk, T. E. (2019). *The Time Efficiency Gain in Sharing and Reuse of Research Data | Data Science Journal*. https://doi.org/10.5334/dsj-2019-010

Purkayastha, P. (2024). *Knowledge as commons: Toward inclusive science and technology*. Monthly Review Press.

Rousi, A. M., Boehm, R. I., & Wang, Y. (2024). Data stewardship: Case studies from North American, Dutch and Finnish universities. *Journal of Documentation*, *80*(7), 306–324. https://doi.org/10.1108/JD-12-2023-0264

Saeed, S., & Ali, P. M. N. (2019). Research data management and data sharing among research scholars of life sciences and social sciences. *DESIDOC Journal of Library and Information Technology*, *39*(6), 290–299. https://doi.org/10.14429/djlit.39.06.14997

Spiegelman, E. (2021). Esteemed colleagues: A model of the effect of open data on selective reporting of scientific results. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.761168

Thorogood, A., & Knoppers, B. M. (2017). Can research ethics committees enable clinical trial data sharing? *Ethics, Medicine and Public Health*, *3*(1), 56–63. https://doi.org/10.1016/j.jemep.2017.02.010

Varoufakis, Y. (2024). *Technofeudalism: What killed capitalism*. Melville House.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Yoon, A. (2017). Data reusers' trust development. *Journal of the Association for Information Science and Technology*, *68*(4), 946–956. https://doi.org/10.1002/asi.23730

Zenk-Möltgen, W., & Lepthien, G. (2014). Data sharing in sociology journals. *Online Information Review*, *38*(6), 709–722. https://doi.org/10.1108/OIR-05-2014-0119

Chapter 7. Recommendations and Conclusions

> If I cannot develop a theory, then I will be satisfied with my construction of a key assertion, a summative and data-supported statement about the particulars of a research study, rather than the suggested generalisable and transferable meanings of my findings to other settings and contexts. (Erickson, 1986)

The chapter summaries the study, which used a mixed methodology approach to explore the factors involved in research data sharing among doctoral students in Kerala. Qualitative study findings enriched the survey findings, and provided further theoretical insights into the research data sharing behaviour. This chapter discusses how the research objectives of the study are achieved. Based on the analysis from the previous chapters, this chapter attempts to make actionable insights to formulate targeted recommendations for policymakers, academic institutions, funding bodies, and practitioners. In addition, this chapter also presents limitations of the study, research contributions and areas for future study.

Achieving Objectives of the Study

The study was guided by six interrelated objectives that sought to examine both individual and institutional dimensions of research data sharing in Kerala. This section revisits each objective and demonstrates how it was addressed through the study's methodology and findings.

Objective 1: To investigate the research data handling practices of doctoral students in Kerala.

The objective was addressed primarily through the survey questionnaire distributed to doctoral students across universities in Kerala. The respondents described how they currently handle their research data. This included the ways they store it, how they document it, the backup methods they use. The results showed that while most researchers rely on personal storage devices and institutional servers, practices for metadata creation, long-term preservation, and structured documentation remain limited. Interviews with supervisors corroborated these findings, emphasising that institutional guidelines for Research Data Management are absent. These findings highlight that Kerala's institutions are still at an emerging stage of formal RDM practice, underscoring the need for institutional support.

Objective 2: To understand attitudes of the researchers towards research data sharing

Researchers' attitudes were explored through Likert Scale based survey items aligned with TPB constructs. The survey revealed positive attitude towards data sharing. Majority of the respondents expressed willingness to make their datasets available if proper recognition and infrastructure are in place. However, attitudes are also shaped by concerns about misuse, plagiarism, and lack of academic credit. Interviews with supervisors generally support openness and data sharing. But they are concerned about the data quality and existing research culture. This divergence highlights the tension between normative openness and individual self-protection, consistent with earlier studies on data sharing reluctance.

Objective 3: To assess the perceived motivating factors in research data sharing

This objective was attained through the analysis of survey data and thematic analysis of interview. Findings from the quantitative study revealed that key motivators include formal recognitions like data citations, credit and opportunities for collaboration. Supervisors echoed these motivators but also stressed the role of institutional incentives, such as considering datasets as valid research outputs in promotion and assessment. The findings show that motivation to share is highest when personal benefits align with the recognition structures of the institution and the discipline. This pattern supports TPB's view that attitudes and subjective norms play a central role in shaping intentions.

Objective 4: To assess the perceived barriers in research data sharing

Barriers were identified through both quantitative and qualitative phases of data collection and analysis. Findings from the quantitative phase shows that the effort and time required for sharing data is a significant challenge in the data sharing. Research supervisors frequently cited lack of time, inadequate skills for preparing datasets, and unclear guidelines as obstacles. They further pointed to infrastructure limitations, concerns over confidentiality, and the absence of clear ownership rules. These barriers were mapped in relation to TPB as lowering perceived behavioural control, and in relation to Institutional Theory as reflecting weak regulative and normative supports. The findings suggests that without simplified processes, training, and institutional assurance, the intention to share cannot translate into action.

Objective 5: To explore the influence of institutional factors in research data sharing

This objective was addressed by examining survey data on access to tools and training, supplemented by supervisors' narratives on institutional factors. As far as the fifth research question is concerned, findings indicate that lack of technical skills and inadequate institutional infrastructure were challenges in data sharing. Analysis demonstrated that practical enablers such as training, user-friendly repositories, and mediated support significantly enhance researchers' perceived control. The enhanced perceived control will eventually strengthen their intention to share data.

Objective 6: To propose policy measures and strategies for establishing research data repositories that address challenges in the light of the objectives 1 to 5

Drawing on the findings, the study proposes policy measures in the Discussion chapter. The policy measures and strategy includes mandating Data Management Plans, creating institutional repositories aligned with FAIR principles. Policies suggest recognising datasets as citable outputs and offering capacity building programs for students and supervisors. Interviews with experienced supervisors added a practical dimension for implementation. Suggested policies support context-sensitive, phased implementation of research data sharing. It also provides access control mechanism that help maintain a balance between openness and confidentiality of research data. Policies includes roles and responsibilities of various actors in the research data sharing landscape. These proposals directly respond to the systemic gaps identified in Objectives 1 to 5 and offer practical and sustainable solutions.

Research Contributions

The study makes a significant contribution in understanding the supporting and challenging factors in research data sharing in Kerala. Major research contributions are discussed under various sections.

Theoretical Implications

The combined use of TPB and Institutional Theory provided a robust dual framework to analyse the phenomenon of research data sharing from both individual and institutional perspectives. TPB offered a base to understand the formation of individual intentions in shaping researchers' willingness to share data. At the same time, Institutional Theory covers the organisational and systemic factors including regulative rules, normative expectations, and cultural-cognitive pillars that frame and constrain individual choices.

By applying these two perspectives together, the study was able to capture the complex and multi layered nature of data sharing behaviour. While TPB clarified *how* individual researchers form their intentions, Institutional Theory explained *why* institutional environments either enable or discourage such intentions. It further provided guidance for policy formulation in the preservation and sharing of research data. Together, TPB and Institutional Theory provided complementary insights. They also offered a strong basis for practical recommendations that link individual behaviour with institutional policy-making.

Methodological Implications

Although prior research has examined data sharing and associated behaviours, no empirical studies have investigated this phenomenon within the specific context of Kerala. This study therefore makes a methodological contribution by applying a mixed methods design to explore a subject that has received little attention in the Indian academic landscape. The study used a two-phase approach to capture perspectives from different stakeholder groups. A questionnaire was administered to doctoral students across multiple universities in Kerala. The qualitative insights from semi-structured interviews were invaluable in contextualising the survey findings. This methodological design demonstrates the value of triangulating evidence from distinct but complementary groups. The study offers a model that can be replicated to explore similar topics in other Indian contexts.

Practical Implications

At a system level, the study indicates that implementing research data repositories in India would yield both economic and academic gains. Open access to research data increases the return on public investment. It can also expand the volume and quality of publishable work by reducing duplication and improving methodological transparency. The findings therefore support policy recommendations and strategies to guide design and development of institutional repositories.

At the institutional level, the findings points to a strategy that couples governance, capacity, and service. Introducing clear institutional frameworks for data sharing will ease burden of research supervisors and promote high-quality, reproducible research. Repository services which streamline preservation, sharing, and reuse will lower the practical cost of participation. At the researcher level, the results underscore the importance of enhancing perceived behavioural control. Simplified processes and user-friendly interfaces reduce the time and effort barriers identified in this study.

Societal Implications

All research should be beneficial to the society, though assessing the same is difficult (Bornmann, 2013). This study addresses the disconnect between public funding and public benefit in research. Consistent with body of literature, the present study found broad normative support for data sharing among research students. This suggests a strong potential for translating public investment in research into publicly accessible knowledge. The second societal benefit is the efficiency and cost savings. The study shows that without deliberate action, research data are often lost or become inaccessible. Addressing the barriers in research data sharing would lead to clear societal benefits. When datasets are available, researchers in smaller institutions, NGOs, start-ups, and administrative departments can participate more fully in evidence based problem solving.

Recommendations

Building on the detailed policy framework proposed in Chapter 6, the following recommendations are put forward for immediate action.

1. Increase Perceived Benefits and Value

- Highlight the personal and professional benefits of data sharing.
- Demonstrate the real world examples where data sharing increases visibility and innovative ideas.

2. Foster a Supportive Research Culture

- Encourage senior and influential researchers in the field to advocate for open data sharing to shift the overall culture.
- Universities should provide incentives and awards in data sharing for researchers who exemplify good data-sharing practices.

3. Enhance Researchers' Confidence and Control

- Provide training and resources on research data management.
- Simplify the process by implementing user-friendly research data platforms and tools for data submission.

4. Strengthen Institutional Policies and Incentives

- Integrate clear data-sharing policies at the institutional level along with mandates from funding agencies and publishers.

- Institutions should offer tangible rewards such as research grants, funding opportunities, or career incentives for those who share data.

5. Ensure Data Quality and Trustworthiness

- Provide clear guidelines on data documentation, metadata standards, and quality checks to ensure that shared data is reliable and usable.
- Adopt FAIR principles in research data management.

6. Address Concerns over Data Ownership and Misuse

- Develop clear guidelines and contracts that outline data ownership, authorship, and credit-sharing arrangements.
- Offer support in navigating legal, ethical, and privacy concerns, especially in fields involving sensitive data.

7. Encourage Collaboration and Data Sharing Networks

- Establish platform or networks where researchers can easily share and access data within their field.
- Encourage interdisciplinary research projects that necessitate data sharing across different domains, making it a natural and integrated practice.

Limitations of the Research

- The study examined behavioural intention rather than actual behaviour. As a result, any conclusions about data sharing should be interpreted with caution.
- Despite efforts to ensure coverage, the design remains vulnerable to self-selection and non-response bias typical of online questionnaires and voluntary interviews.
- There may be uneven representation across disciplines and institutions in the study. This imbalance limits the precision of any subgroup

comparisons and the study does not aim to conduct detailed subgroup analyses.

Recommendations for Further Research

- Future studies could extend the investigation to stakeholders who influence data sharing practices but were not included in the present study. These include collaborators, commercial partners, repository managers, journal publishers, and funding agencies.
- Assess the capacity and willingness of library professionals in taking up the additional role of data librarianship and related tasks of research data and repository management.
- Research data sharing is not a homogenous topic, detailed studies to understand the data sharing behaviour of different disciplinary and subject groups can be attempted with the cultural background of Indian academic institutions.
- Though many have showed positive attitude towards data sharing, the actual behaviour is not assessed. A study to assess the actual behaviour of researchers can be attempted.

References

Bornmann, L. (2013). What is societal impact of research and how can it be assessed? A literature survey. *Journal of the American Society for Information Science and Technology*, *64*(2), 217–233. https://doi.org/10.1002/asi.22803

Erickson, F. (1986). Qualitative methods in research on teaching. In *Handbook of research on teaching* (3rd ed., pp. 119–161). Macmillan.

Appendix I

Questionnaire

Dear Doctoral Student / PhD Graduate,

I am doing PhD on the topic "Research Data Sharing among Doctoral Students in Kerala" under the supervision of Prof. Dr. Mohamed Haneefa K. The study tries to understand the present situation and also how research scholars view sharing their research data.

Research data means any materials you've created or gathered for your studies. This could be tables of numbers, pictures, audio recordings, video clips, computer code, or written documents to validate original research findings. Sharing this data has several advantages, like saving time and resources, and it can also enhance the overall quality of research.

Thank you for taking the time to contribute to the survey.

1. Your Name [Optional] ....

2. Gender : Female ☐ Male ☐ Other ☐

3. Belongs to the following age group:

   Below 30 ☐

   30 - 40 ☐

   Above 40 ☐

4. Choose Research status (Mark any one of the choices).

   Completed PhD ☐ PhD Ongoing ☐

5. If you have completed PhD, indicate the year of thesis submission.

   ....

6. Please select the university with which your research centre is affiliated for your PhD studies.

- Central University of Kerala ☐
- Cochin University of Science and Technology ☐
- Kannur University ☐
- Kerala Agricultural University ☐
- Kerala Kalamandalam ☐
- Kerala Veterinary and Animal Sciences University ☐
- Kerala University of Health Sciences ☐
- Mahathma Gandhi University ☐
- NIT Calicut ☐
- Sree Sankaracharya University of Sanskrit ☐
- Thunchath Ezhuthachan Malayalam University ☐
- University of Calicut ☐
- University of Kerala ☐

Other: ---

7. Please specify the name of your research department or institute. If affiliated with a college, kindly provide the name of the college and the respective department.

----

8. Your subject area of research (Broad subject area, examples are given in bracket) Mark only one

- Humanities (History, Fine arts, Linguistics , Phiosophy etc) ☐
- Social and Behavioural Sciences (Education, Psychology etc) ☐
- Life Sciences (Biology, Medicine, Agriculture, Medicine etc) ☐
- Natural Sciences (Chemistry, Mathematics, Geosciences etc) ☐
- Engineering Sciences (Computer Sciences and Engineering etc) ☐

9. Please record the number of publications you have produced so far (Mark only one per row).

Particulars	None	1 to 5	6 to 10	More than 10
Books /Book Chapters				
Peer Reviewed Journal article				
Conference Papers				
Patents				

10. I support Open Access for research articles and other publications, ensuring free public access, especially when publicly funded. (Mark only one)

Strongly Support ☐

Support ☐

Neutral ☐

Oppose ☐

Strongly Oppose ☐

I don't know ☐

11. Please indicate the types of data you have collected or generated in your research. If your data type is not listed, choose 'Other' and briefly describe your data. [Select all that apply]

Non-digital text (e.g. hand-written notes, sketches, paper laboratory notebooks) ☐

Artistic products ☐

Audio recordings ☐

Computer code ☐

Crowdsourcing data ☐

Curriculum materials ☐

Databases ☐

Digital gene sequences or similar digital renditions ☐

of biological /organic/ inorganic

samples or specimens ☐

Experimental data ☐

Field notes ☐

Interview transcripts ☐

Patient records ☐

Samples or specimens ☐

Software ☐

Spatial or geographic data ☐

Spreadsheets ☐

Surveys ☐

Video recordings ☐

I don’t produce any data ☐

Other: ....

12. Which of the following file formats best describe your digital research data? Examples of specific file extensions are included. [Please choose all that apply.]

Audio (.aif, .iff, .mp3, .wav) ☐

Computer aided design / CAD (.dwg, .dxf, .pln) ☐

Data (.csv, .dat) ☐

Data – Statistical / SAS, SPSS (.sav, .sdq, .spv) ☐

Data – XML (.xml) ☐

Database (.db, .mdb, .pdb, .sql) ☐

Geographic Information Systems / GIS (.gpx, .kml) ☐

Image (.bmp, .gif, .jpg, .png, .ps, .psd, .svg, .tif)

Scanned document (.jpg, .jpeg, .pdf)

Spreadsheets (.wks, .xls, .ods)

Text document (.doc, .docx, .odt, .log, .rtf, .txt)

Video (.avi, .mov, .mp4)

Web (.html, .xhtml)

Don't know

Other ....

13. Indicate the estimated amount of digital data the research has created or is likely to generate. (Mark only one)

Less than 100 MB (studies involving text-based documents, simple datasets, or low- resolution images.)

Between 100 MB and 1 GB (those involving larger datasets, multimedia files, or more extensive documentation.)

Between 1 GB and 10 GB (significant data generation, analysis, or storage requirements.)

More than 10 GB (research activities generating vast amounts of digital data like bigdata)

Can't Say.

14. Indicate how often backups are made for the digital files associated with your research. (Mark only one)

More than once a day Daily

Weekly Monthly Rarely Never

15. Where have you stored the data from your research? [Please choose all that apply.]

- Personal Computer / Laptop ☐
- Lab/ Depart computer ☐
- Cloud storage (Google Drive, dropbox) ☐
- Email ☐
- Institutional Storage ☐
- Flash drive / Pendrive ☐
- External Disk ☐

16. Does your research involve gathering personal information such as addresses, phone numbers, or medical histories?

Yes ☐ ☐ Un☐

17. What measures have you taken to ensure the privacy of your data? Please select all that apply from the following options. [If your data contains personal information as specified in the previous question]

- Not applicable; data is not sensitive or confidential ☐
- Obtained informed consent from participants to share my data ☐
- Anonymised sensitive or confidential data (Anonymisation is the process process of removing or changing personal information from data so that nobody can reasonably identify) ☐
- Destroyed the personal information from research data within a short time of publishing research results ☐
- Destroy all data within a specified time period ☐
- No steps are taken to ensure privacy ☐
- Other ☐

18. Do you maintain detailed descriptions about your research dataset which provides context for understanding why and when the data were collected and how they were used. (also know as metadata) (Mark only one)

Yes, I provide detailed descriptions for my research data. ☐

No, I do not provide detailed descriptions for my research data. ☐

Partially, I provide some level of description research data. ☐

Not applicable/I am not sure. ☐

19. Have you saved the data generated from your research for long-term storage or archival purposes, and if not, do you plan to do so in the future? (Mark only one)

Yes, I have already preserved my research data. ☐

No, I have not preserved my research data but plan to do so in the near future. ☐

No, I have not preserved my research data and currently have no plans to do so. ☐

I am uncertain about how to preserve my research data. ☐

Other ....……………..

20. Governments and research institutes have started providing their raw data through open data repositories. Are you familiar with any such online repositories?

Yes ☐ No ☐ Maybe ☐

21. If there's useful data available online, would you consider using it for your own research? (Mark only one)

Will use and acknowledge by citation. ☐

Will use but may not cite. ☐

Will not use for my study. ☐

Not decided.. ☐

Other ....

22. Please indicate the areas you would be interested receiving data management training in [Tick all that apply]

Developing a research data management plan ☐

Documenting research data ☐

Storing data ☐

Formatting the data ☐

Sharing of data ☐

Creating proper description for data (metadata) ☐

Ethics and consent ☐

Copyright and Intellectual Property Right ☐

Data Repositories and Open Access ☐

Data citation ☐

Data backup ☐

File versioning and naming conventions ☐

23. After researchers finish their studies and publish their papers, they often either delete the data they used or keep it stored away. These data can be shared and reused. Sharing this data can have many benefits, It promotes open discussion, transparency and improves the quality. (Mark only one)
I believe that RESEARCH DATA should be made available through a repository for others to see and reuse?

Strongly agree ☐

Agree ☐

Neutral	
Disagree	
Strongly disagr	
I Don't know	

24. Please rate how important the following factors are while considering to share your research data openly.?

Statements	Extremely Important	Very Important	Slightly Important	Not Important	Not applicable
Financial reward I receive for sharing my data.					
Increased visibility I get when others reuse and cite my data					
My ability to select the individuals or groups permitted to access and use my data					
If I am treated as a co-author in the publication generated with my research data.					
Embargo period (Time when research findings are kept private before public release)					

25. Please indicate to what extend you agree or disagree with the following statements. [Please choose all that apply.]

Statements	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
My research data could possibly be misinterpreted or misreported					

Major effort and time is required to prepare the data in a shareable format					
My data is confidential in nature and cannot be shared					
As a data creator, I fear that the data may be used without due acknowledgment.					
I do not think that my data could be useful to others					

26. If you would like to expand on any of the above answers or make further

comments, please do so here;

....
....
....

I sincerely appreciate your time and cooperation. Thank you so much for your contribution.

Jamsheer N. P.

Appendix II

Interview Schedule

Research is usually publicly funded. With the help of new technology; data generation, management, data analysis and reuse is easier than before. At the international level, funding agencies, publishers and universities are advocating open access to research data. Various funding agencies and countries have policies for sharing research data. Government of India publish open government data through data.gov.in and has policy for sharing data (NDSAP 2012).

1. What types of research data are produced in your research area ?
2. How do you or your students manage research data? Can you describe any challenges you or your research students faced in managing research data?
3. Have you ever thought of storing these data for future use?
4. Can you tell me about any instance of substandard data noticed in thesis review or any other occasion ? / Tell me about any occasion when you witnessed data that can not be considered up to standard practice.
5. How such situation can be avoided ?
6. How do you ensure the quality of the data produced or collected for research.
7. What is your thoughts on open access Philosophy?
8. Can you imagine any scenarios where sharing research data would be particularly valuable in your field? If so, what types of data do you think are suitable for sharing, and why?
9. In what ways do you think sharing research data could impact your work, either positively or negatively? Have you ever came across the practice of data sharing ?
10. What do you see as the main benefits of making your research data openly accessible?

11. What concerns or challenges come to mind when you think about sharing your data with others?

12. What are your primary concerns or reservations, if any, about sharing research data openly? What factors would do you consider before you share your data?

13. Have you ever thought that it would have been better if you had the access to datasets along with the publication so that you could test and analyse if necessary?

14. Have you observed any trends or shifts in how your research community views data sharing?

15. How do you perceive the data sharing initiative in your research community ? Do you think that this idea would be acceptable to the scholars in your area of research.

16. How do you see the attitude of research community towards data sharing initiative ?

17. Have you ever thought that it would have been better if you had the access to datasets along with the publication so that you could test and analyse if necessary?

18. Generally data sharing is expected to bring increased quality and progress in research through transparency, collaboration, innovation and reproducibility. How do you consider this in your research area?

19. What kind of incentives or institutional support systems would encourage researchers to engage in data sharing? Do you think data sharing practices should be integrated into research evaluation processes? If so, how?

20. What specific skills or knowledge would you need to feel more comfortable with sharing your data?

21. Can you elaborate on the human resource support required for preparing research data in a shareable format. ?

22. Based on your experience, what advice would you offer to institutions aiming to establish or improve research data repositories?
23. Considering the interests of the data creators, privacy and open access philosophy, what mode of access you would prefer in your research area?
24. How do you perceive the role of your institution in promoting or discouraging data sharing?