Perceived Barriers to Open Science among Researchers in Mathematics, Natural Sciences, and Cognitive Sciences

Christian Riedel [1], Shafayet Hossen Chowdhury [1], Ralf Engbert [2], and Ulrike Lucke [1]

Abstract: This study identifies the intensity of perceived barriers to integrate open science practices among researchers in natural sciences, cognitive sciences, and mathematics. Barriers include knowledge gaps in data and software sharing, reuse concerns, and disincentives. The study finds that each barrier limits researchers from integrating open science practices, but the perception of barriers, research data, and software varies with the research field. The study suggests that targeted educational programs, improved IT infrastructure, and revised policy frameworks are needed to better integrate open science practices across diverse research cultures.

Keywords: Open Science, Perceived Barriers, Cultural Change

1 Introduction

Open science is increasingly recognized as a transformative way to enhance the transparency, reliability, and efficiency of research [St20]. It calls for a paradigm shift where scientists openly share their methodologies, data, software, and findings, fostering a culture of collaboration and transparency that could reshape how knowledge is produced and distributed [Ba18; Pe13]. By emphasizing practices such as data and software management and sharing, study pre-registration, and open-access publishing, open science seeks to improve research integrity, reproducibility, and knowledge transfer [Ba18; Fe20; Le14; Te21].

Due to a variety of data acquisition, data analysis, and methodologic approaches across the scientific disciplines, there is a diversity of practices and perspectives on open science [Ba18; Fe20; St20]. Therefore, considerations and sensitivity to the characteristics of the individual research fields and research cultures are crucial to implementing such practices [Ba18]. Subject-specific initiatives such as the National Research Data Infrastructure (NFDI) consortia underscore the growing importance of open science and the universal push towards it concerning the perspectives within individual research fields [Gr19].

Open science practices have been acknowledged as beneficial across various disciplines [Be17; Ho18]. However, researchers face challenges in engaging in these practices. Some of these challenges are structural and are imposed by specific research cultures and

[1] University of Potsdam, Institute of Computer Science, An der Bahn 2, 14476 Potsdam, Germany, christian.riedel@uni-potsdam.de, https://orcid.org/0000-0001-5154-4153; md.shafayet.hossen.chowdhury@uni-potsdam.de, https://orcid.org/0000-0001-5536-9807; ulrike.lucke@uni-potsdam.de, https://orcid.org/0000-0003-4049-8088

[2] University of Potsdam, Department of Psychology, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany, ralf.engbert@uni-potsdam.de, https://orcid.org/0000-0002-2909-5811

doi:10.18420/inf2024_186

institutional frameworks. For example, technical hindrances, such as interoperability issues, the reliance on closed and proprietary tools, and challenges associated with large data volumes, impose significant barriers [Fe20]. Additionally, lacking institutional support structures and regulatory frameworks do not adequately promote or facilitate open science, resulting in low adoption rates [Ba18; Be17]. Strategic concerns, such as privacy, security, and legal constraints, further complicate the willingness of providers to publish open data [Be17]. Misaligned career incentives, where the current academic reward system does not sufficiently recognize or reward open science efforts, are also a significant concern [Go22b]. Researchers, therefore, perceive insufficient career benefit to engaging in these practices despite their potential to enhance research quality and impact.

Further challenges can occur on an individual level, where researchers face various challenges that discourage them from adopting open science practices. These challenges can include the additional effort required to prepare and share data [Ho18], personal attitudes [Go22b], a lack of formal training in data and software management and open science practices [Fe20; Ho18; PM19], as well as concerns over intellectual property and the potential misuse of shared data [Ho18]. These personal obstacles are compounded by the prevalent practice in many fields to only share data upon specific requests rather than proactively [Ho18]. Moreover, the quality of data and the complexity of navigating legal and licensing issues can further impede effective utilization by researchers [Be17].

The main objective of this study is to investigate the barriers to open science perceived by members of the Faculty of Science and the intra-faculty unit 'Cognitive Science' at the University of Potsdam and the Collaborative Research Center (CRC) 1294 - 'Data Assimilation.' The study aims to determine their individual relevance, including knowledge barriers, reuse concerns, and disincentives, as categorized by Gomes et al. [Go22a]. We want to better understand barriers and subject-specific characteristics that may affect perceptions and practices related to data and software. The study hypothesizes that the relatively consistent research culture within these fields avoids confounding factors related to strongly varying research methodologies and requirements. The research aims to develop measures to better support the implementation of open science practices. Ultimately, the approach is to integrate the actual concerns and perceptions of researchers into strategies designed to enhance the openness, integrity, and reproducibility of research.

2 Investigating Barriers

Gomes et al. [Go22a] provide a detailed breakdown of the barriers hindering the adoption of open science practices within the biological sciences community – barriers that are also common across various other research fields. The authors categorize these barriers into three groups: knowledge barriers, reuse concerns, and disincentives. They analyze the specific barriers for each category and propose strategies to overcome them. The identified barriers and categories are summarized below.

Knowledge Barriers are rooted in a lack of information or skills for adequate data and software management and sharing:

- **Unclear Value:** A lack of perceived benefits from sharing data or software
- **Unclear Process:** A lack of data and software sharing information, including choosing a suitable repository
- **Complex Workflows:** Research methodologies that involve complex or manual steps that are challenging to repeat
- **Large Files:** The size of some datasets poses challenges in terms of storage and accessibility
- **Insecurity:** Fear of exposure to criticism for potential errors or imperfections in data or software

Reuse Concerns reflect concern about the potential negative consequences of open sharing, both ethically and regarding data integrity:

- **Inappropriate Use:** Concerns over misinterpretation or misuse of data or software when shared broadly
- **Sensitive Content:** Potential harm from publishing data or software due to privacy concerns or concerns about vulnerable subjects or environments
- **Transient Storage:** Uncertainties about the long-term availability of repositories
- **Sharing Rights:** Complexities in data and software ownership that might indicate a loss of rights

Disincentives contain the worries and perceived lack of rewards associated with sharing data and software:

- **Scooping:** Fears that others publish results using shared data before the original researchers
- **Lack of Time:** Perceived substantial time investment required to prepare data and software for sharing
- **Lack of Incentives:** Insufficient career incentives for sharing data and software, despite potential long-term benefits

Based on the barriers by [Go22a], we created a questionnaire (see Tab. 1) to investigate the respective limitations, and the participants' usage of research data and software, as well as their generation. We also asked participants to specify their research fields, with the option to select multiple disciplines, and divided the 'sensitive content' barrier described by [Go22a] into privacy-related concerns and concerns about vulnerable subjects or environments. Furthermore, the 'Does not apply' response was available for data and software generation questions and individual perceived barriers to accommodate differing perspectives on research data and software. The survey was conducted from June 2023 to April 2024.

General - Questions	
Q1. In my publications, the research teams often use code or data to obtain results.	Yes, No, Not sure
Q2. In my publications, I am often (at least partly) responsible for data generation, code, or software development .	Yes, No, Does not apply
Q3. I conduct research in the following scientific field(s) (multiple selections possible):	Biology, Chemistry, Cognitive Sciences, Computer Science, Didactics, Geosciences, Health Sciences, Linguistics, Mathematics, Nutritional Science, Physics, Psychology, Sports Sciences, Other
Barriers - Questions	
1-Strongly disagree, 2-Disagree, 3-Agree, 4-Strongly agree	
Sharing Data and Software	
Q4-Unclear Value: I do not see a significant benefit in sharing data or code associated with my publications.	1, 2, 3, 4, Does not apply
Q5-Unclear Process: I am still determining where and how to upload my data or code for sharing.	1, 2, 3, 4, Does not apply
Q6-Complex Workflows: My workflows are usually very complex and can not be shared and applied easily.	1, 2, 3, 4, Does not apply
Q7-Large Files: My datasets are often too large to be shared.	1, 2, 3, 4, Does not apply
Q8-Insecurity: The quality of my generated data or code needs to improve to be appropriately reused by others.	1, 2, 3, 4, Does not apply
Reuse Concerns	
Q9.-Inappropriate Use: The data or code I generate could be misinterpreted or misused.	1, 2, 3, 4, Does not apply
Q10-Privacy Concerns: The data or code I generate contains information that would raise privacy concerns when shared.	1, 2, 3, 4, Does not apply
Q11-Sensitive Content: The data or code I generate includes content that may not be in the best interest of science or society when shared.	1, 2, 3, 4, Does not apply
Q12-Transient Storage: There need to be more appropriate long-term publication platforms to publish my data or code	1, 2, 3, 4, Does not apply
Q13-Sharing Rights: I am concerned that the generated data or code ownership is not in my hands anymore when shared.	1, 2, 3, 4, Does not apply
Disincentives	
Q14-Scooping: Sharing the data or code limits my ability to generate further publications from the investigation.	1, 2, 3, 4, Does not apply
Q15-Lack of Time: The commitment to preparing and publishing data or code takes too much of my time.	1, 2, 3, 4, Does not apply
Q16-Lack of Incentives: Sharing my data or code does not benefit my academic career.	1, 2, 3, 4, Does not apply

Tab. 1: Questionnaire on Perceived Barriers to Open Science Practices

3 Results

A total of 61 researchers from various disciplines participated in the survey. Participants were given the option to select multiple scientific fields. Despite some asymmetries, the scientific fields in question are represented with considerable uniformity. Fig. 1 shows that Mathematics (15%), Cognitive Sciences (13%), and Biology (12%) are the most active fields for participants. In contrast, Nutritional Science, various unspecified fields collectively categorized as 'Other,' and Didactics had the least amount of interaction with only 1%, 2%, and 3%, respectively.

The results of our survey indicate that there is a clear trend towards the integration of data and software into research methodologies among the participants (see Fig. 2). Most respondents (86.9%) reported that their research teams frequently rely on data or code to achieve research outcomes, and the majority of respondents (78.7%) is involved in data generation or software development. However, there is a smaller yet considerable proportion of participants (13.1%) who reported that their teams do not regularly use data or software to obtain results. Furthermore, 11.5% of respondents are not involved in generating data or software, and 9.8% considered these actions not applicable to their research. These figures suggest that certain research areas may use different experimental or theoretical methods or that some segments of the academic community might be less engaged with data- and software-driven approaches.

The survey results for each perceived barrier are presented in Fig. 3, using a raincloud plot to show the distribution of answers. Participants were asked to rate each barrier on a scale of 1 (strongly disagree) to 4 (strongly agree), with an option for 'Does not apply.' Mean scores were based only on responses ranging from 1-4, excluding 'Does not apply' answers. Therefore, the number of responses used to calculate the mean is lower than the total number of 61 participants. However, all identified barriers were acknowledged as actual limitations by at least some researchers, as indicated by responses of 3 (agree) or 4 (strongly agree) for every question. The intensity with which these barriers are perceived varies, suggesting that some barriers are viewed as more limiting than others within the contexts of the participants' research activities.

The participants expressed that the most significant concern related to data and software sharing is the insecurity of publishing poor-quality data and software. This concern is likely because many researchers have self-taught data management and software development skills, which can lead to unstructured and unsystematic approaches in research and software data management [Fe20]. Additionally, the data and software sharing process remains unclear to many, and the substantial size of the data and software to be published also poses strong concerns. Since the largest groups of participants who 'agreed' or 'strongly agreed' that large files pose a barrier conduct research in Geosciences, Biology, Linguistics, and Physics (See Tab. 2), this issue is likely most pronounced in data-intensive disciplines where large datasets are common. Some participants also perceive complex workflows that require multiple procedures and software tools to reproduce scientific results as a barrier.

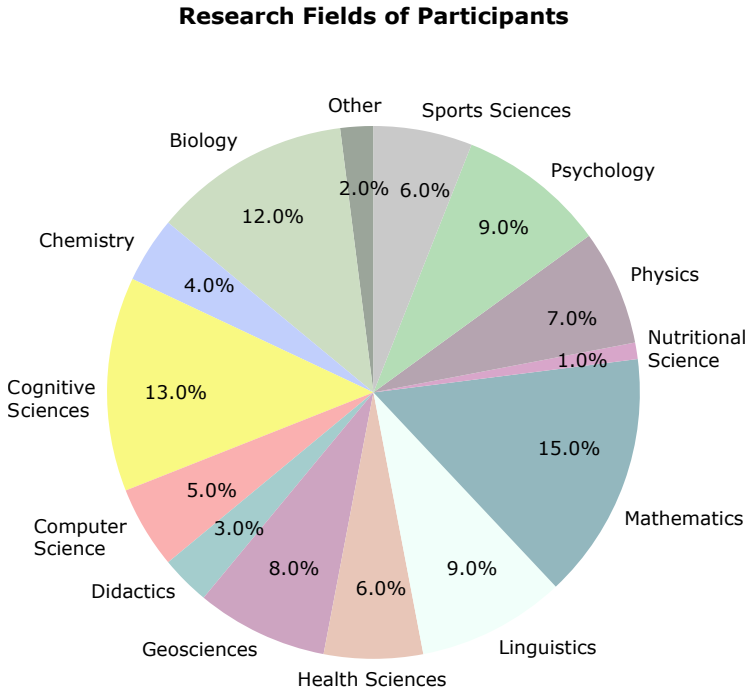

Fig. 1: Distribution of Research Fields Among Survey Participants

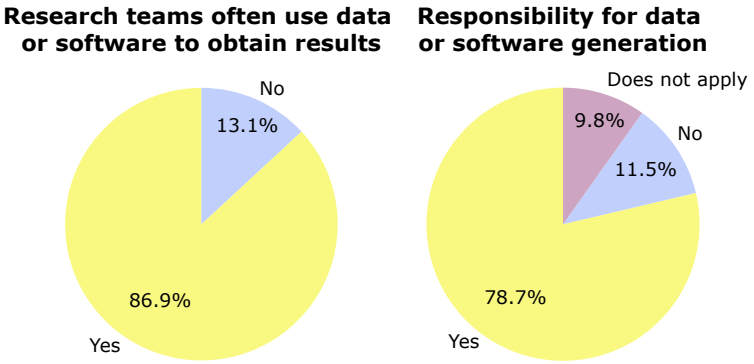

Fig. 2: Involvement in Data and Software Usage and Development Among Participants

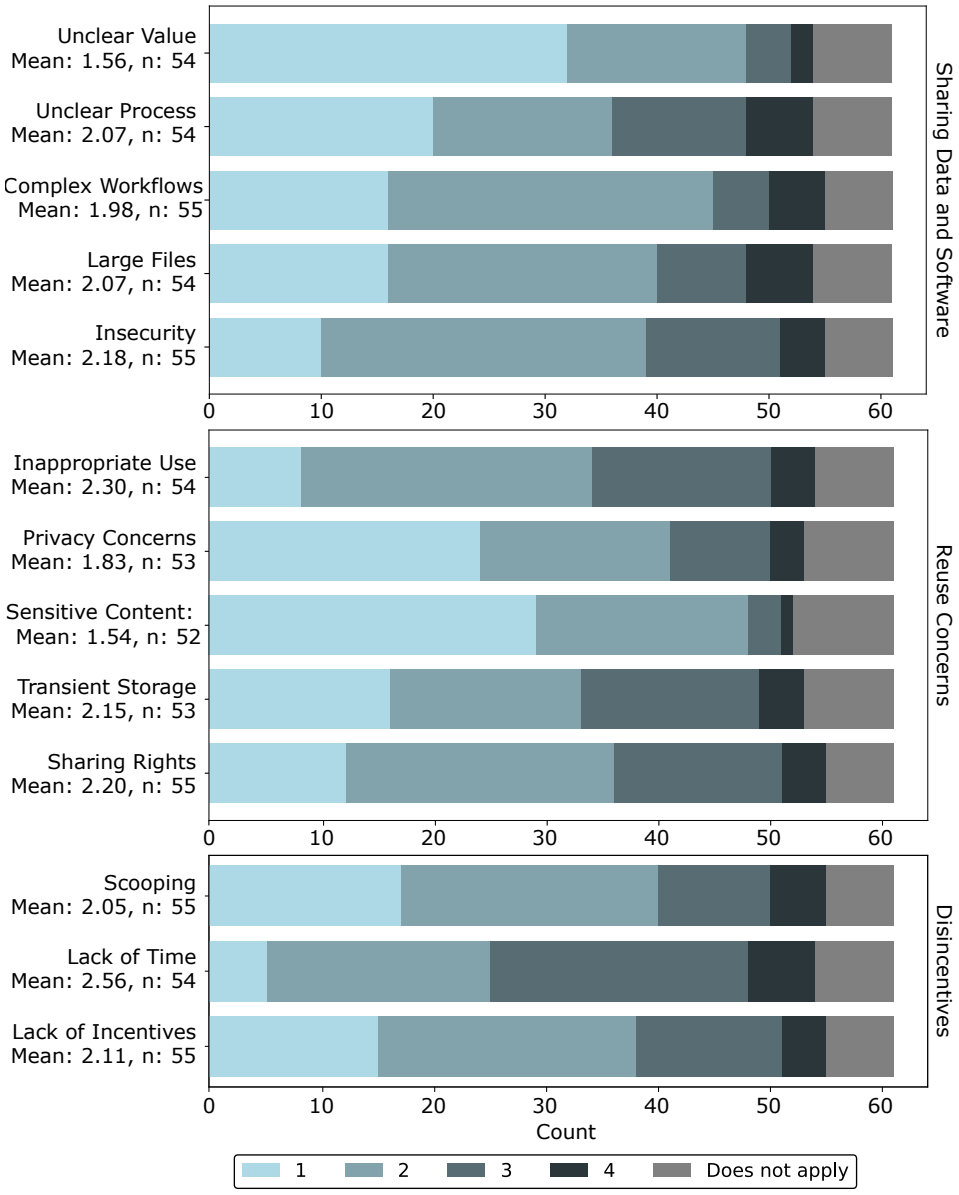

Fig. 3: Perceived Barriers Among Survey Participants

However, the majority of participants recognize the additional value of openly sharing data and software, with only a few considering the lack of additional value as a deterrent to sharing their work.

Among the reuse concerns considered in the survey, the fear of inappropriate use, where researchers are concerned about the misinterpretation or misuse of published data or software, emerged as the most significant barrier. Concerns about sharing rights followed this, indicating worries about losing control over data or software when sharing. Additionally, a lack of appropriate long-term publication platforms poses a strong barrier for many participants, particularly those in Geosciences, Biology, and Physics (See Tab. 2), linking to the barriers associated with managing large files. Privacy concerns were not as common, but they still pose challenges for some researchers, particularly those in Cognitive Sciences, Psychology, and Didactics, suggesting that the relevance of this issue varied by research field, particularly those working with personal data. Lastly, sensitive content, potentially threatening vulnerable subjects or environments, was rated as the least of the concerns. Nonetheless, it remains a relevant issue for some researchers.

Disincentives present significant barriers to the participants. A notable concern among many is the risk of scooping, where researchers fear that others might use their publicly shared data or software to publish results before they do. Additionally, the lack of time to adequately prepare data and software for publication is the most significant of all barriers reported by the survey participants. This issue is accompanied by a lack of incentives, where many participants feel that their efforts toward open science are not sufficiently recognized and rewarded in academia.

Our survey reveals that participants selected the option 'Does not apply' across all barrier questions, indicating that the definitions were irrelevant in some research contexts. Fig. 4 provides a breakdown of the research fields of participants who found at least one barrier inapplicable, hinting to diverse research cultures. A major proportion (52.9%) of those who selected 'Does not apply' for any barrier were conducting research in Mathematics, while the remaining fields (Linguistics, Health Sciences, Didactics, Cognitive Sciences, Chemistry, Biology, Sports Sciences, and Psychology) have equal representation. In contrast, all participants who conduct research in Geosciences, Computer Sciences, Physics, Nutritional Sciences, and Other fields found the defined barriers applicable to their field. We found that participants who considered the barriers irrelevant had less integration of data and software in their research methodologies than those who did not. Fig. 5 shows that 53.8% of this group does not rely on data or software for research outcomes, 23.1% were not involved in data or software generation, and 46.2% considered such responsibilities inapplicable. This data suggests a distinct research culture, particularly prevalent among mathematicians, characterized by methodologies where open science practices are less dependent on data and software or a perception where research data and software management are not integral to their research activities exists.

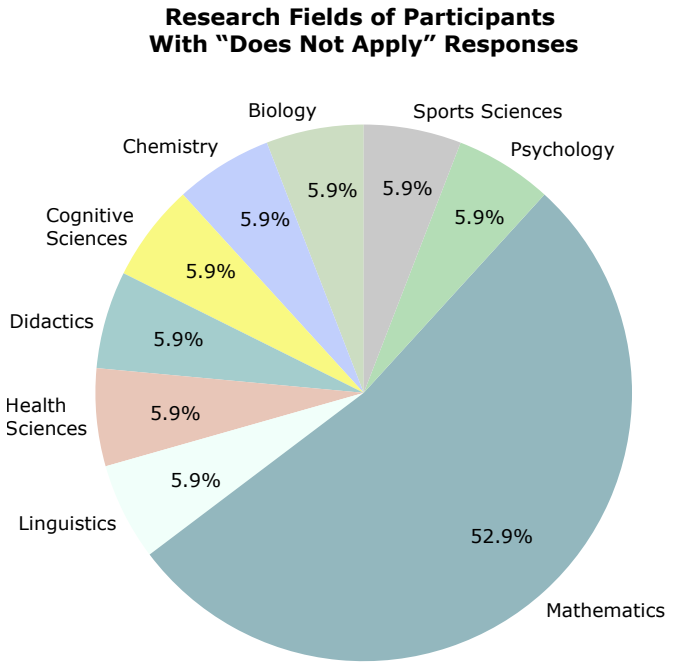

Fig. 4: Distribution of Research Fields Among Participants Who Reported That Specific Barriers Were Irrelevant to Their Research

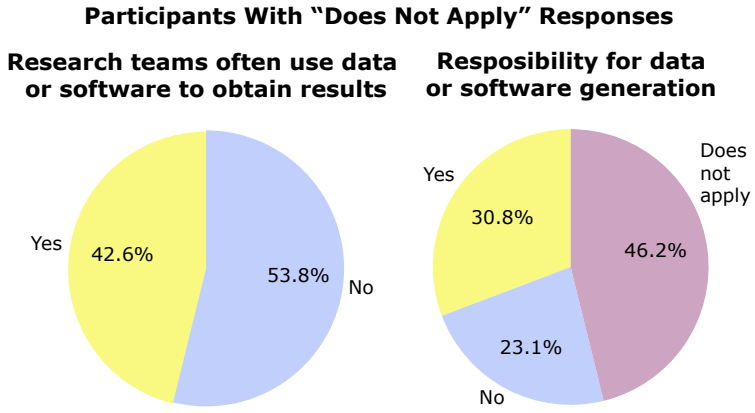

Fig. 5: Involvement in Data and Software Usage and Development Among Participants Who Reported That Specific Barriers Were Irrelevant to Their Research

Research Field	Large Files	Transient Storage	Privacy
Biology	2	4	-
Chemistry	1	2	1
Cognitive Sciences	-	3	4
Computer Science	1	1	-
Didactics	-	-	3
Geosciences	3	5	1
Health Sciences	1	-	2
Linguistics	2	3	1
Mathematics	-	2	2
Nutritional Science	-	-	-
Physics	2	4	1
Psychology	1	2	3
Sports Sciences	1	3	2
Other	-	-	1

Tab. 2: Frequency of Agreement and Strong Agreement (3 and 4) on Barriers Related to 'Large Files', 'Transient Storage', and 'Privacy' Among Researchers

4 Recommendations

Our survey shows patterns in perceived barriers and research culture related to open science. However, the relatively small sample size and the method of voluntary participation may have resulted in a bias toward researchers who are already interested in open science. Therefore, the results may not be applicable across a broader research landscape with statistical significance as some research cultures are potentially over or underrepresented. Despite these limitations, the survey allows for deriving recommendations to improve open science practices.

Essentially, all addressed barriers were identified as limitations to open science practices. While targeted training can help overcome many perceived barriers on an individual level, the perceived limitations are embedded in a larger structural context. Therefore, it is also necessary to establish appropriate training opportunities and IT infrastructure, and to enable corresponding policies by funding agencies, journals, and research institutions to further promote a cultural change towards open science. This should be done within the context of discipline-specific open science demands, associated with research culture and the role of data and software in the research process.

Many of the individually perceived barriers among the 'Data and Software Sharing' and 'Reuse Concerns' categories can be addressed through training. For example, education on adequate research data and software management, as well as software development, with a focus on automating processes, can be essential in addressing issues related to 'Insecurity,' 'Complex Workflows,' and 'Lack of Time.' Similarly, researchers who consider 'Unclear Process' a barrier can be educated on the publication process and how to choose a repository.

To address concerns about 'Scooping' and 'Unclear Value,' education programs can highlight the benefits of open science, such as increased citations and the collective advantages of making research data and software available. Barriers related to 'Inappropriate Use,' 'Sharing Rights,' and 'Privacy' can be addressed through training sessions that focus on creating appropriate software documentation, data and software licensing, and implementing restrictive mechanisms that still align with open science principles, such as creating synthetic datasets or setting access restrictions on sensitive data. However, the survey results show that adjusting these training programs to the specific needs of different research fields is crucial. For instance, data-intensive groups such as those in Geosciences, Biology, Linguistics, and Physics might focus more on strategies for publishing and archiving large datasets, while fields dealing with sensitive personal data, such as Cognitive Sciences, Psychology, or Didactics, might require more training on anonymization techniques. Beyond these knowledge-based competencies, such programs should also facilitate the internalization of open science values, ensuring that training impacts more than just the cognitive aspects of knowledge transfer.

Still, certain barriers to sharing research data and software cannot be resolved through training alone, as they are embedded within the academic community's cultural and infrastructure frameworks. For example, adequate research data and software management require appropriate infrastructure, such as collaborative work systems, backup and storage, as well as archiving solutions. This requirement is important in data-intensive fields, such as Geosciences, Biology, and Physics, which encounter barriers related to 'Large Files' and 'Transient Storage.' To meet the needs of these disciplines, discipline-specific repositories that can publish and archive large datasets must be accessible to researchers. Furthermore, the perceived barriers concerning 'Lack of Time' and 'Lack of Incentives' emphasize the need to better acknowledge and reward the creation of data and software as scientific achievements. This is particularly relevant because the perceived value of data and software sharing, stated in the 'Unclear Value' barrier, is not seen as a major hindrance, but the absence of incentives within academia is. This discrepancy highlights a systemic issue where the academic reward structure does not align with the perceived importance of open science practices among researchers.

To overcome such structural and cultural barriers, research institutions, funding agencies, and academic journals must promote this process. For example, academic journals can help by requiring data and software to be included with submitted research articles. Funding agencies can encourage data and software sharing by requiring data management plans in grant applications and by providing financial support for the development of data storage, publication, archiving, and management tools. Universities and research institutions can contribute to removing barriers through additional infrastructure investments. They can also support by revising their tenure and promotion criteria to prioritize and reward the creation and sharing of data and software. Furthermore, scientific repositories are encouraged to offer data and software publication solutions that are adapted to meet the needs of the respective research communities, particularly in terms of storage.

These recommendations are particularly relevant in fields where data and software are applied in the research process. However, our study revealed that this is not always the case, especially for researchers in mathematics. This discrepancy may indicate either a need for better integration of open science practices concerning research data and software or a need to expand the definitions that describe open science mechanisms to apply to mathematics. In any case, further investigation and discussions are required both across the respective research communities.

5 Conclusions

This study explored the perceived barriers to integrating open science practices, particularly data and software sharing, by members of the Faculty of Science and the intra-faculty unit 'Cognitive Science' at the University of Potsdam and the CRC 1294 - 'Data Assimilation.' Our investigation reveals that despite general recognition of the value of open science practices, significant limitations due to individual knowledge barriers, a lack of infrastructure, and cultural constraints within the academic system persist. The study shows that all investigated barriers can be perceived as limitations to implementing open science practices by researchers and that the intensity of a barrier can depend on the research field. The study suggests that researchers perceive limitations due to knowledge barriers in sharing data and software, reuse concerns, and a lack of incentives. This highlights the need for additional education in research data and software management, as well as policies by academic journals, funding agencies, and research institutions to support the integration of open science principles. For example, academic journals can contribute by making it mandatory to include data and software during the publication process. Funding agencies are advised to include data management plans for grant requirements. They should also provide dedicated funding for data sharing and archiving tools. Universities and research institutions must invest in digital infrastructure to support storing, managing, sharing, and archiving research data and software, particularly for large datasets. Furthermore, they should consider revising academic hiring and recognition systems to better reward open science practices. Particularly notable is the contrast in perceptions of barriers related to research data and software, particularly among participants who conduct research in mathematics, where the role of open science, research data, and software appears to be different and should be investigated more closely in the future. This observation implies that measures to remove barriers to integrating open science practices in research processes must consider discipline-specific demands that are associated with research culture and the integration of data and software into the respective research methodologies.

Acknowledgements

This work was supported by Deutsche Forschungsgemeinschaft (DFG) - Project-ID 318763901 - SFB1294, project Z03. The data and code are accessible via https://doi.org/10.5281/zenodo.12572928.

References

[Ba18] Banks, G. C. et al.: Answers to 18 Questions About Open Science Practices. Journal of Business and Psychology 34 (3), pp. 257–270, ISSN: 1573-353X, DOI: 10.1007/s10869-018-9547-8, URL: http://dx.doi.org/10.1007/s10869-018-9547-8, 2018.

[Be17] Beno, M. et al.: Perception of Key Barriers in Using and Publishing Open Data. JeDEM - eJournal of eDemocracy and Open Government 9 (2), pp. 134–165, ISSN: 2075-9517, DOI: 10.29379/jedem.v9i2.465, URL: http://dx.doi.org/10.29379/jedem.v9i2.465, 2017.

[Fe20] Feger, S. S. et al.: "Yes, I comply!": Motivations and Practices around Research Data Management and Reuse across Scientific Fields. Proceedings of the ACM on Human-Computer Interaction 4 (CSCW2), pp. 1–26, ISSN: 2573-0142, DOI: 10.1145/3415212, URL: http://dx.doi.org/10.1145/3415212, 2020.

[Go22a] Gomes, D. G. E. et al.: 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), ISSN: 1471-2954, DOI: 10.1098/rspb.2022.1113, URL: http://dx.doi.org/10.1098/rspb.2022.1113, 2022.

[Go22b] González-Teruel, A. et al.: Drivers and barriers in the transition to open science: the perspective of stakeholders in the Spanish scientific community. El Profesional de la información, ISSN: 1386-6710, DOI: 10.3145/epi.2022.may.05, URL: http://dx.doi.org/10.3145/epi.2022.may.05, 2022.

[Gr19] Grahe, J. E. et al.: Open Science Promotes Diverse, Just, and Sustainable Research and Educational Outcomes. Psychology Learning & Teaching 19 (1), pp. 5–20, ISSN: 1475-7257, DOI: 10.1177/1475725719869164, URL: http://dx.doi.org/10.1177/1475725719869164, 2019.

[Ho18] Houtkoop, B. L. et al.: Data Sharing in Psychology: A Survey on Barriers and Preconditions. Advances in Methods and Practices in Psychological Science 1 (1), pp. 70–85, ISSN: 2515-2467, DOI: 10.1177/2515245917751886, URL: http://dx.doi.org/10.1177/2515245917751886, 2018.

[Le14] Lenhardt, W. C. et al.: Data Management Lifecycle and Software Lifecycle Management in the Context of Conducting Science. Journal of Open Research Software 2 (1), e15, ISSN: 2049-9647, DOI: 10.5334/jors.ax, URL: http://dx.doi.org/10.5334/JORS.AX, 2014.

[Pe13] Peters, M. A.: Open Science, Philosophy and Peer Review. Educational Philosophy and Theory 46 (3), pp. 215–219, ISSN: 1469-5812, DOI: 10.1080/00131857.2013.781296, URL: http://dx.doi.org/10.1080/00131857.2013.781296, 2013.

[PM19] Pasek, J. E.; Mayer, J.: Education Needs in Research Data Management for Science-Based Disciplines: Self-Assessment Surveys of Graduate Students and Faculty at Two Public Universities. Issues in Science and Technology Librarianship (92), ISSN: 1092-1206, DOI: 10.29173/istl12, URL: http://dx.doi.org/10.29173/istl12, 2019.

[St20] Stracke, C. M.: Open Science and Radical Solutions for Diversity, Equity and Quality in Research: A Literature Review of Different Research Schools, Philosophies and Frameworks and Their Potential Impact on Science and Education. In: Lecture Notes in Educational Technology. Springer Singapore, pp. 17–37, ISBN: 9789811542763, DOI: 10.1007/978-981-15-4276-3_2, URL: http://dx.doi.org/10.1007/978-981-15-4276-3_2, 2020.

[Te21] Tedersoo, L. et al.: Data sharing practices and data availability upon request differ across scientific disciplines. Scientific Data 8 (1), ISSN: 2052-4463, DOI: 10.1038/s41597-021-00981-0, URL: http://dx.doi.org/10.1038/s41597-021-00981-0, 2021.