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

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

[1] Department of Computer Science, University of Potsdam, Potsdam, Germany

[2] Department of Psychology, University of Potsdam, Potsdam, Germany

DFG Deutsche Forschungsgemeinschaft

SFB 1294 Data Assimilation

Open Science

Pixabay.com, CC0

Open science is perceived positively across scientific disciplines

Stracke et al. (2020), https://doi.org/10.1007/978-981-15-4276-3_2

Open Science definition

Open science is defined as an inclusive construct that combines various movements and practices aiming to **make multilingual scientific knowledge openly available, accessible and reusable for everyone** [...]

UNESCO (2021), https://doi.org/10.54677/MNMH8546

But what does Open Science mean?

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

Therefore, considerations and sensitivity to the characteristics of the individual research fields and research cultures are crucial to implementing such practices

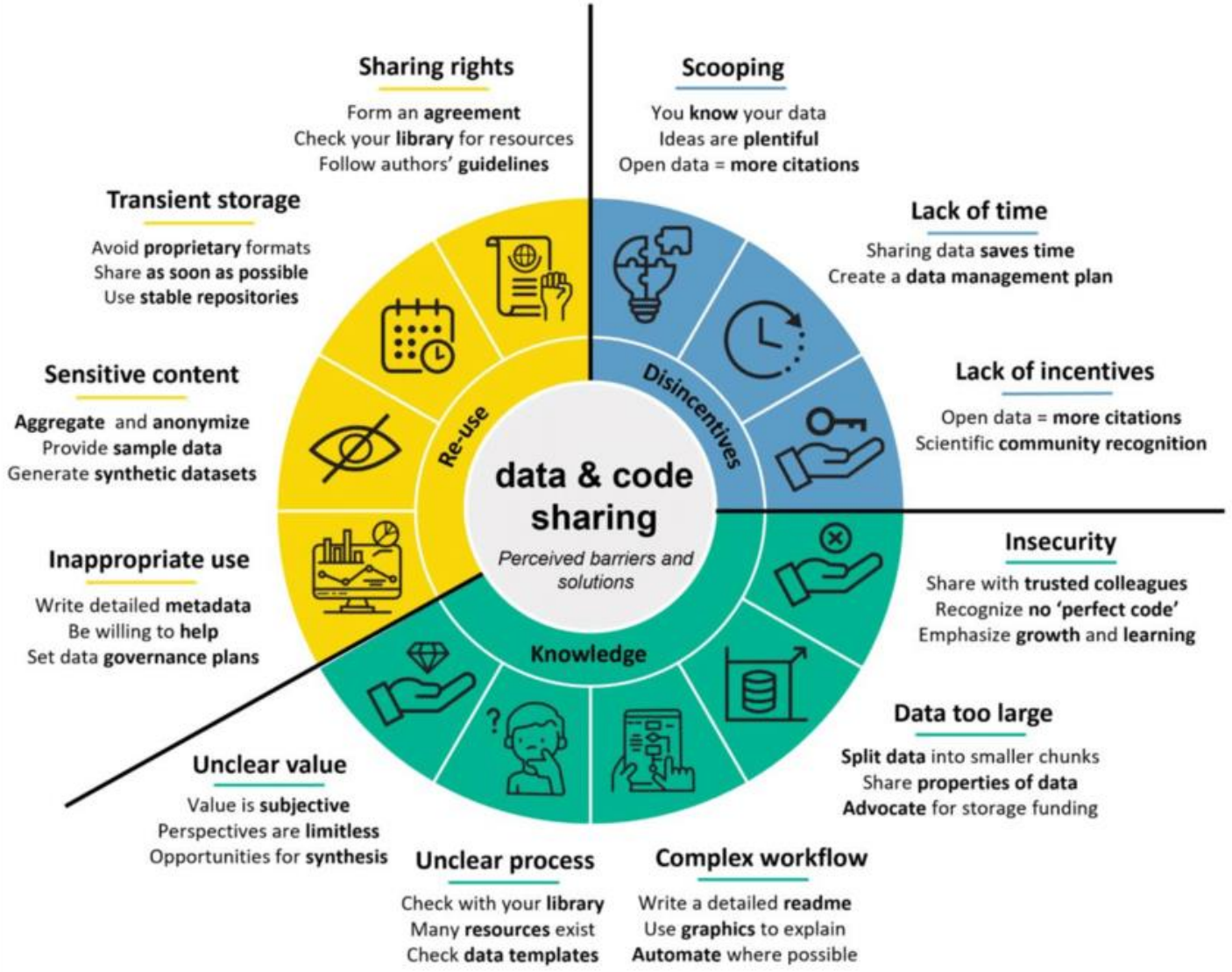
Banks et al. (2018), https://doi.org/10.1007/s10869-018-9547-8
Feger et al. (2020), https://doi.org/10.1145/3415212
Stracke et al. (2020), https://doi.org/10.1007/978-981-15-4276-3_2

Structural and Individual Challenges

- **Individual** challenges (challenges that discourage researchers from adopting open science practices)
  - Additional effort required to prepare and share data
  - Lack of formal training in data and software management and open science practices
  - Concerns over intellectual property and the potential misuse of shared data
  - Research culture (e.g., only share data upon specific requests rather than proactively)
  - Concerns about data and software quality
  - Navigating legal and licensing issues
- **Structural** Challenges (imposed by specific research cultures and institutional frameworks)
  - Technical hindrances (e.g., interoperability issues)
  - The reliance on closed and proprietary tools
  - Challenges associated with large data volumes
  - Lacking institutional support structures and regulatory frameworks
  - Strategic concerns complicating the willingness of providers (e.g., privacy, security, legal issues)
  - Misaligned career incentives, where the current academic reward system does not sufficiently recognize or reward open science efforts

Banks et al. (2018), https://doi.org/10.1007/s10869-018-9547-8; Beno et al. (2017), https://doi.org/10.29379/jedem.v9i2.465; González-Teruel et al. (2022), https://doi.org/10.3145/epi.2022.may.05; Feger et al. (2020), https://doi.org/10.1145/3415212; Houtkoop et al. (2018), https://doi.org/10.1177/2515245917751886; Pasek & Mayer (2019), https://doi.org/10.29173/istl12

Limitations to Data & Code Sharing



Gomes et al. (2022), https://doi.org/10.1098/rspb.2022.1113

How do researchers from different fields perceive the barriers to open science?

Objectives

Questionnaire:

- Based on the Gomes et al. barriers, we created a questionnaire to investigate the respective limitations
- Shared in CRC 1294 – „Data Assimilation“ and intra-faculty unit ’Cognitive Science’ at the University of Potsdam

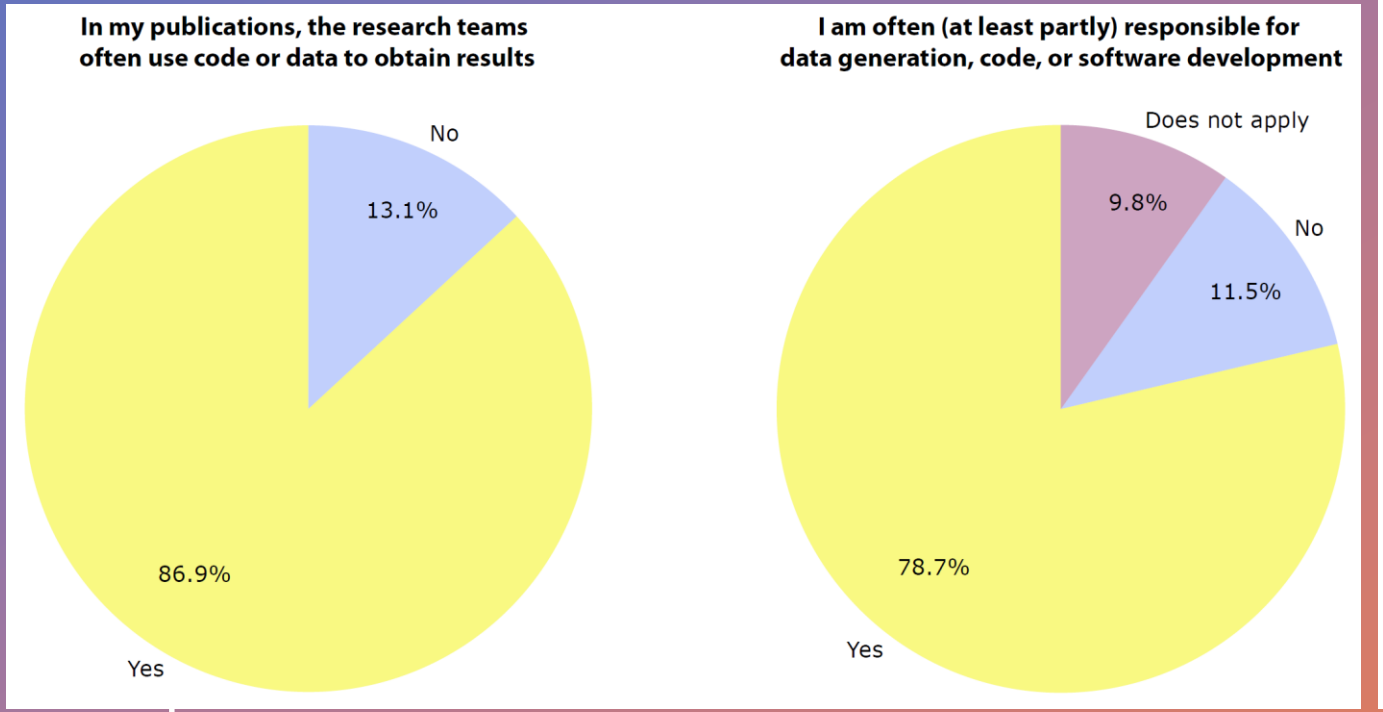
Objectives:

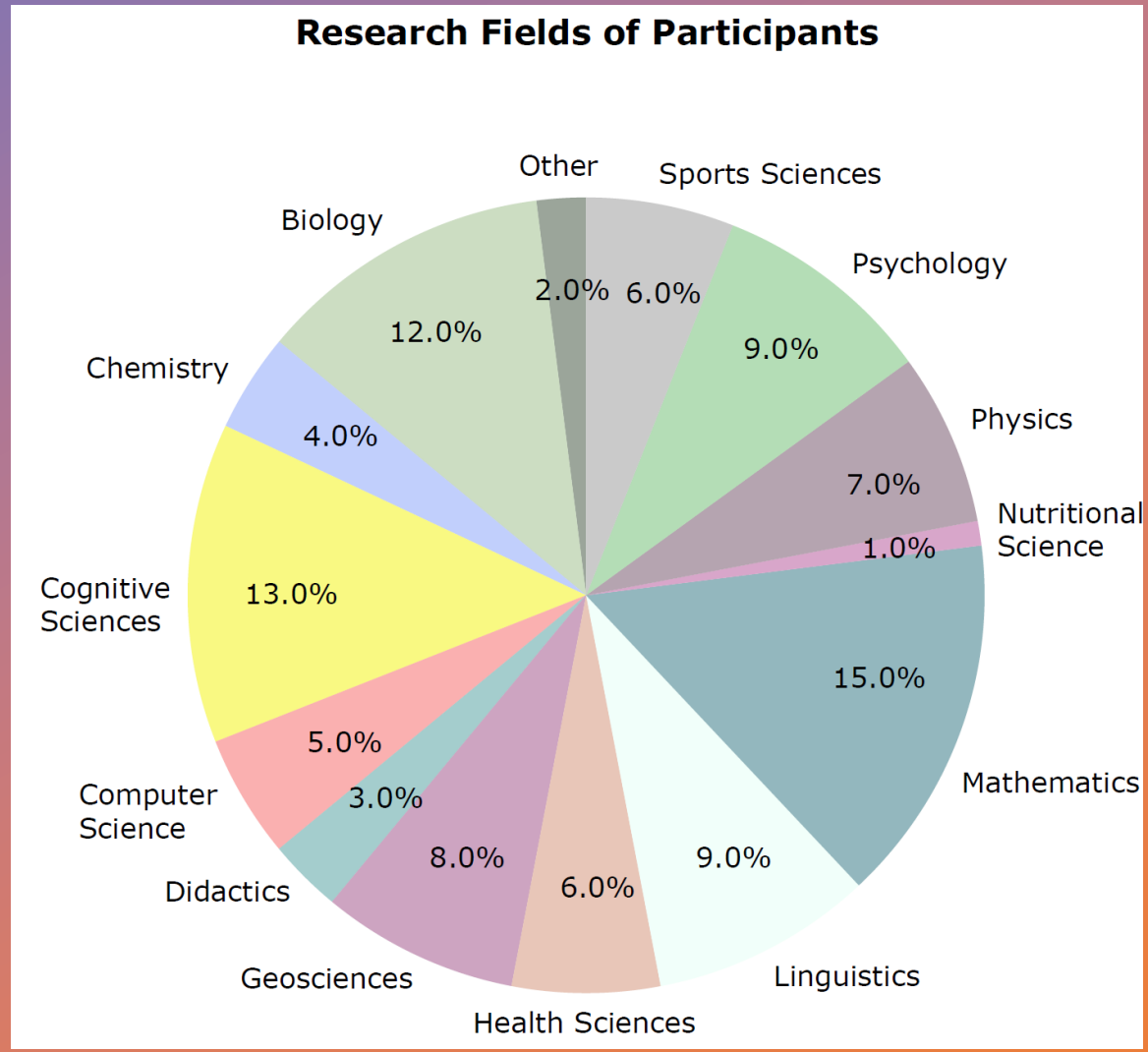
- How do researchers perceive the individual barriers?
- Which discipline-specific characteristics in perceived barriers and data/software use can we investigate?

Questionnaire

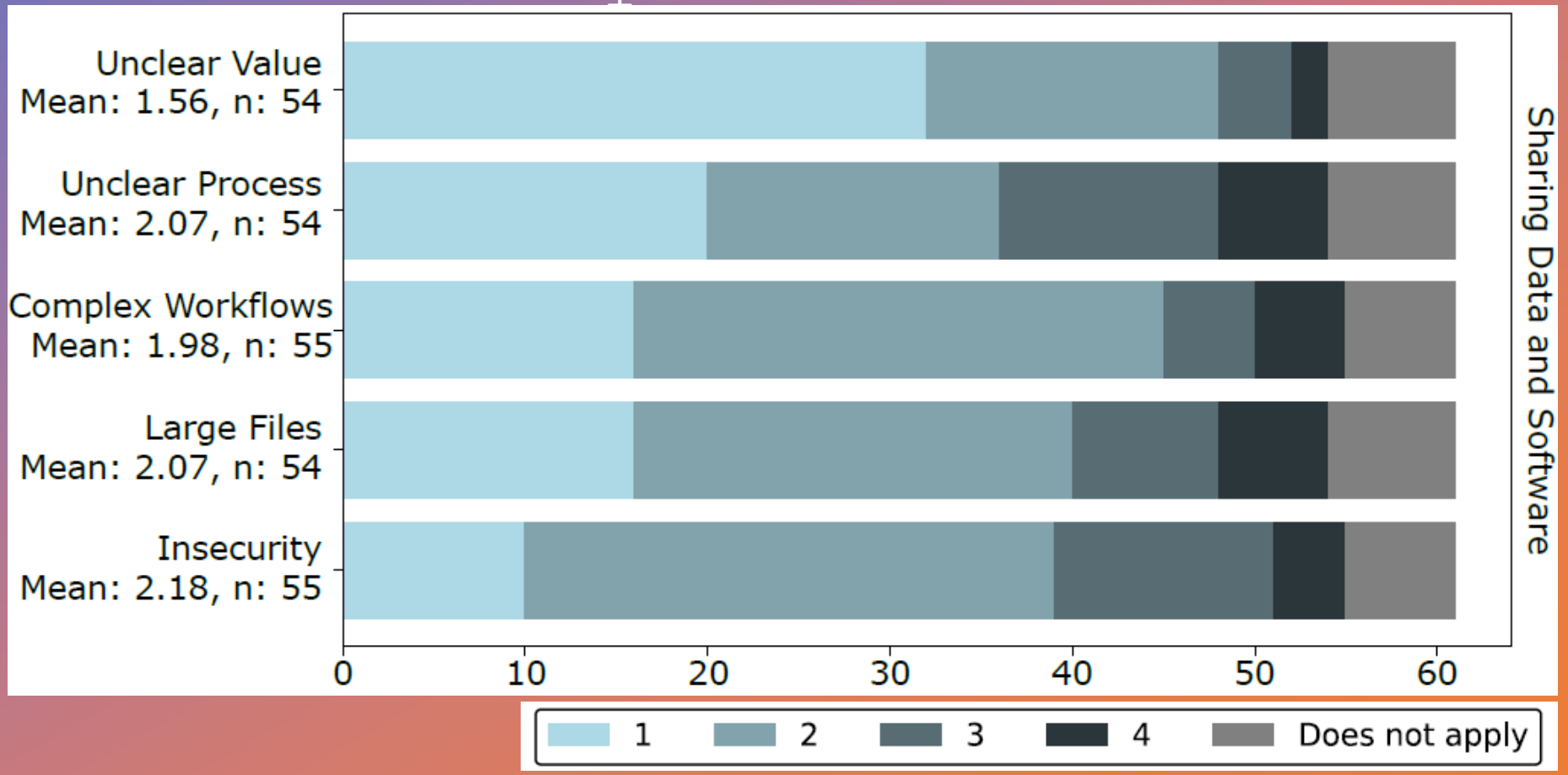
General - Questions	
Q1. In my publications, the research teams often use code or data to obtain results.	Yes, No, Not sure
Q2. In my publications, I am often (at least partly) responsible for data generation, code, or software development	Yes, No, Does not apply
Q3. I conduct research in the following 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

Results – General

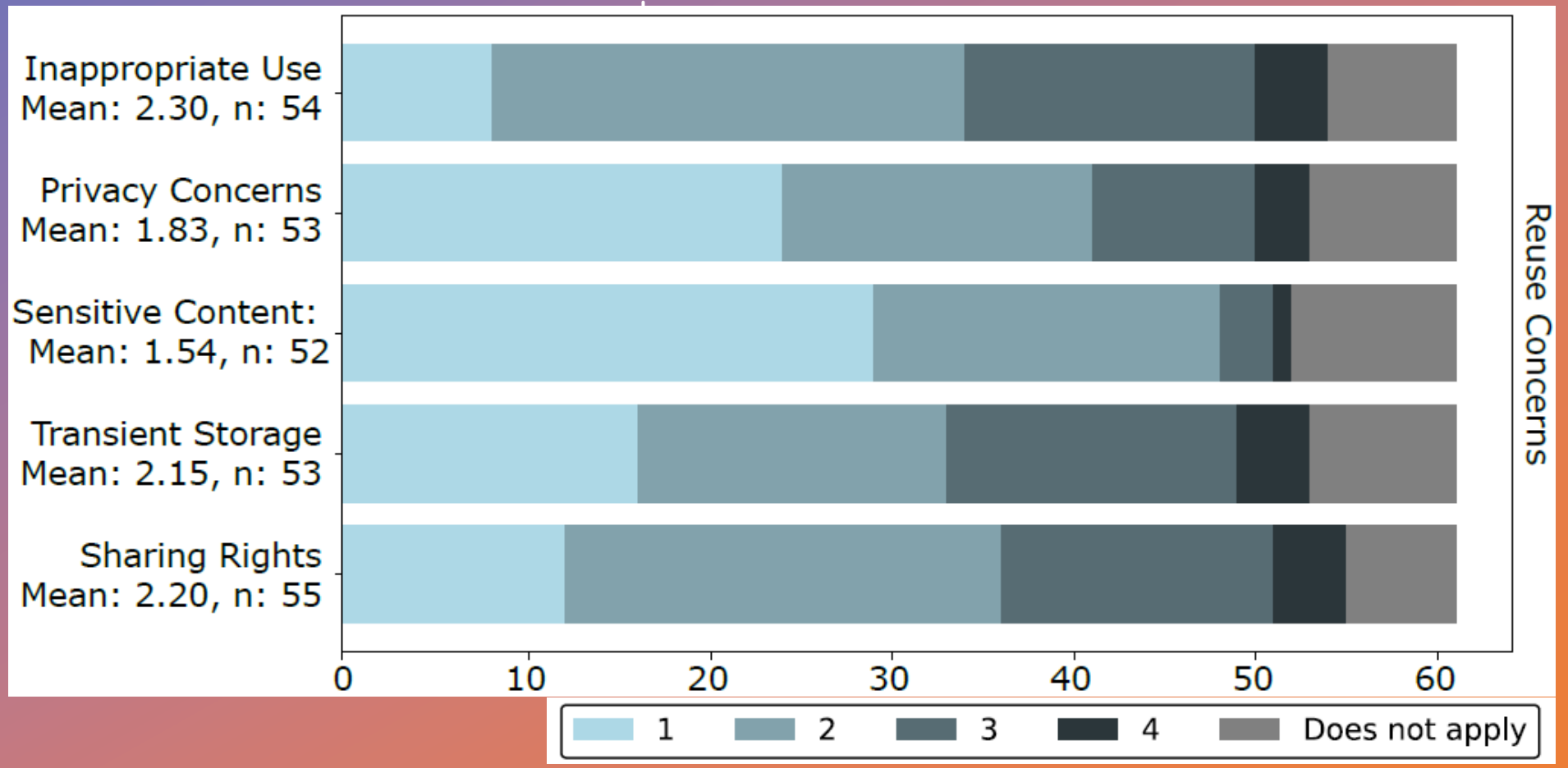




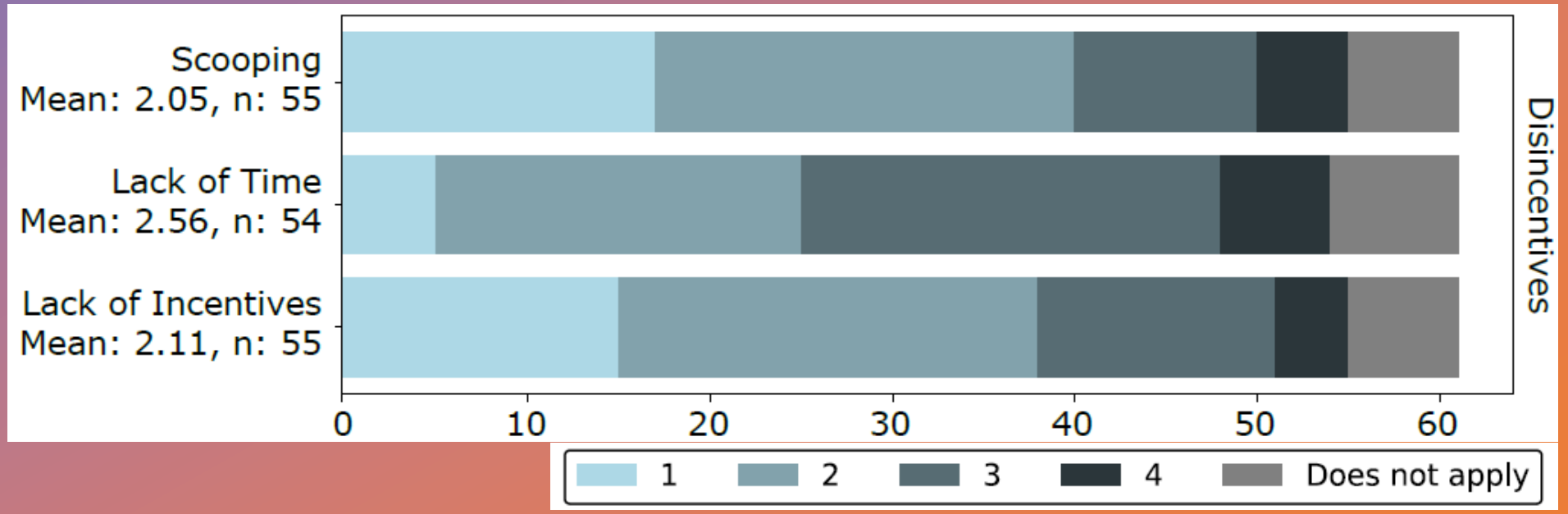
Results – Sharing Data and Software



Results – Reuse Concerns



Results – Disincentives



Training as a Solution for Perceived Barriers (Individual Barriers)

Insecurity, Complex Workflows, Lack of Time

- Training on research data and software management
- Software development with a focus on automation

Scooping, Unclear Value

- Highlight increased citations and collective benefits

Inappropriate Use, Sharing Rights, Privacy

- Training on software documentation and licensing
- Implementing restrictive mechanisms like synthetic datasets or access restrictions for sensitive data

Field-Specific Needs

Data-intensive fields (e.g., Geosciences, Biology, Linguistics, Physics)

- Focus on strategies for publishing and archiving large datasets

Fields handling sensitive data (e.g., Cognitive Sciences, Psychology, Didactics)

- Focus on anonymization techniques

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

Barriers Beyond Training

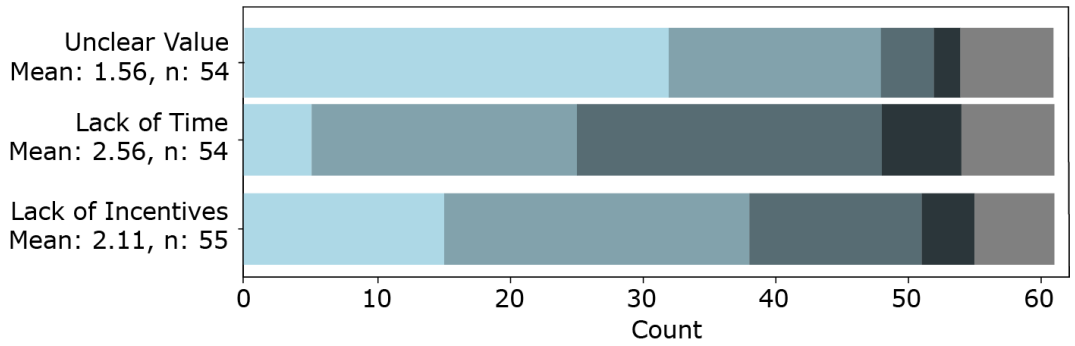
Certain barriers to sharing research data and software are rooted in cultural and infrastructure frameworks

Infrastructure:

- Fields like Geosciences, Biology, and Physics face challenges with Large Files and Transient Storage
- Solution: Provide access to discipline-specific repositories for publishing and archiving large datasets

Systemic Issues in Academia

- Barriers like 'Lack of Time' and 'Lack of Incentives' critical
- The perceived value, stated in the ’Unclear Value’ barrier, is not seen as a major hindrance
- Discrepancy between the perceived value of data/software sharing and the lack of incentives in academia
- Academic reward structures do not align with the importance of open science practices



Intermediate Conclusion

Addressing Barriers

- Targeted training can help overcome individual perceived barriers
- Limitations are part of a larger structural context

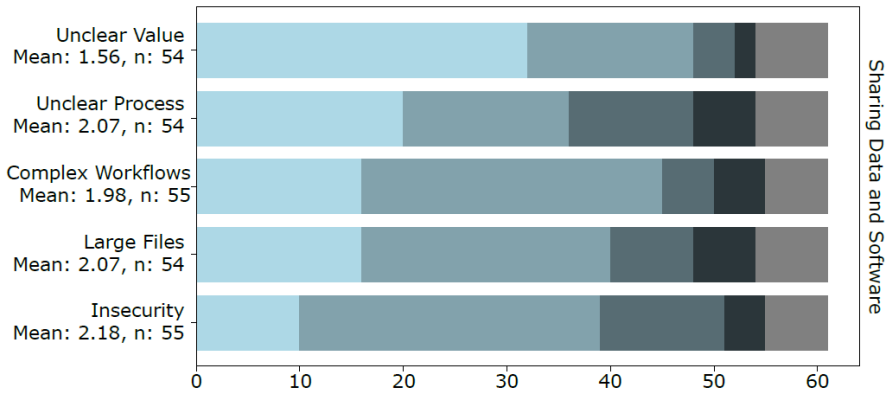
Need for Support

- Establish appropriate training opportunities and IT infrastructure
- Enable policies from funding agencies, journals, and research institutions
- Promote cultural change towards open science

Discipline-Specific Considerations

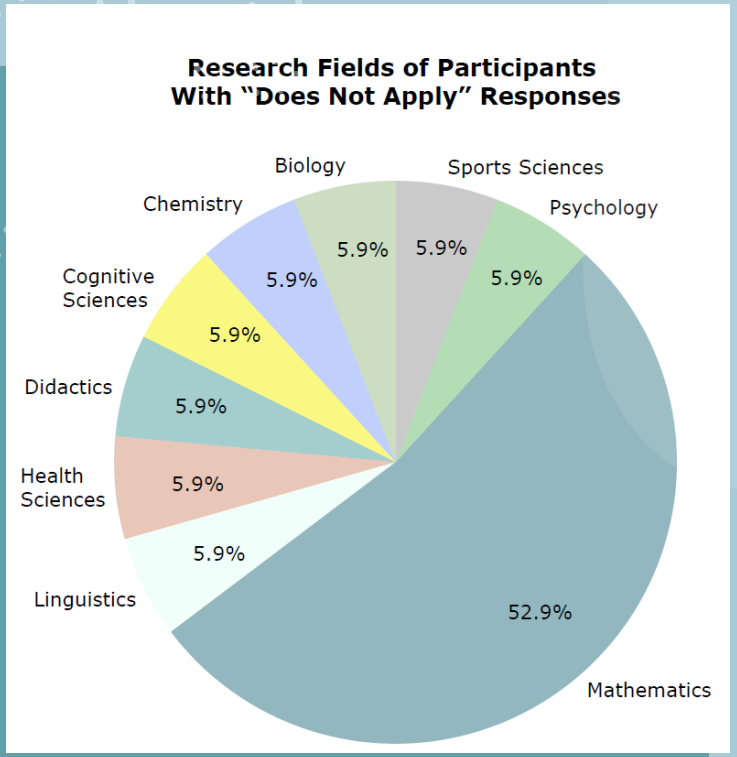
- Recommendations should be connected to discipline-specific open science demands
- Consider the role of data and software in the research process and associated research culture

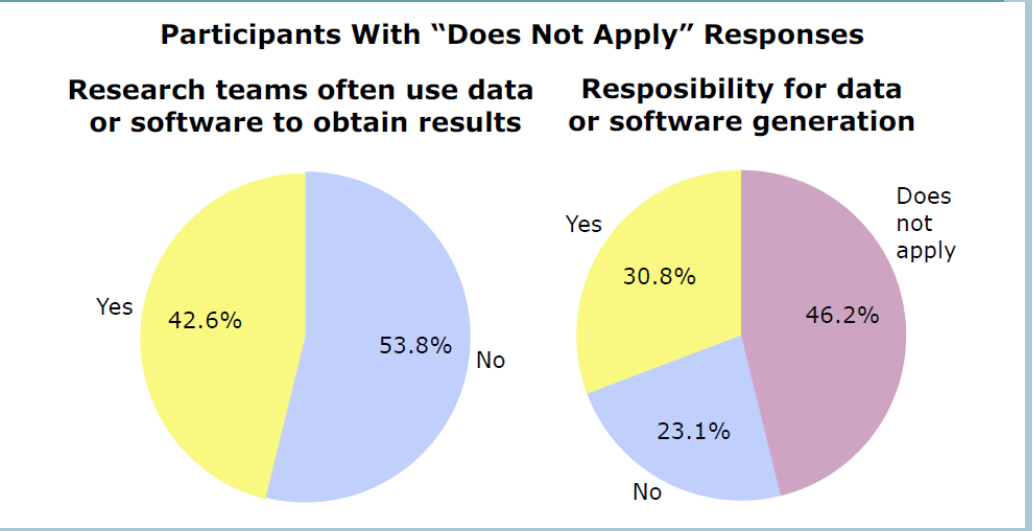
Participants selected „Does not Apply“ across all questions, which indicates that the definined limitations are irrelevant in some research contexts.



Research Culture: When Barriers Do Not Apply

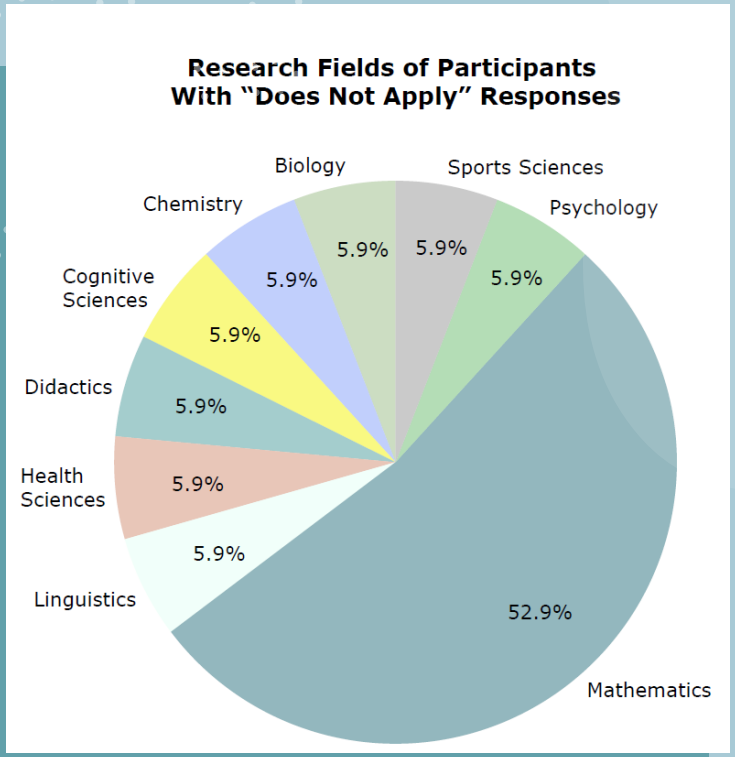
- A major proportion of participants who found at least one barrier inapplicable were conducting research in Mathematics

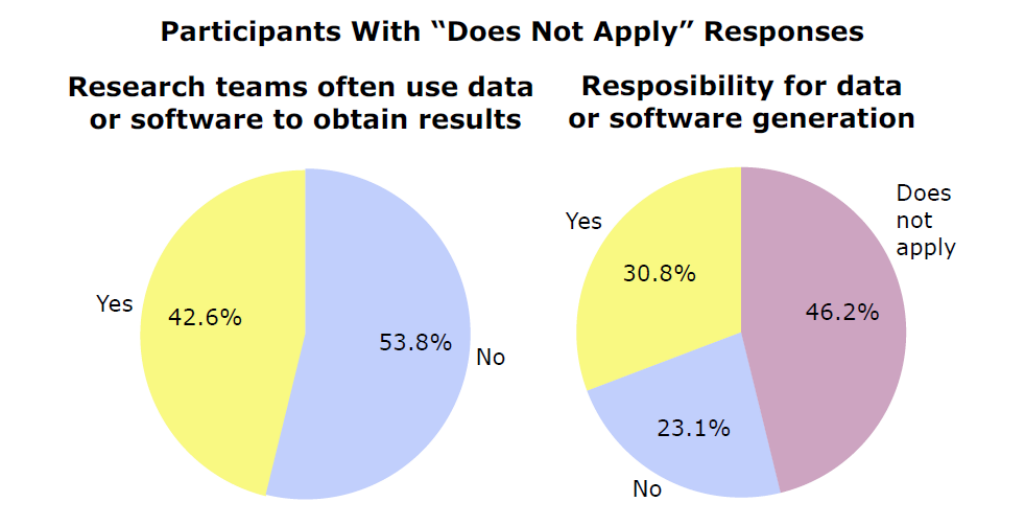




Research Culture: When Barriers Do Not Apply

- Research culture characterized by methodologies where open science practices are less dependent on data / software
- Research data and software management is not integral to research activities



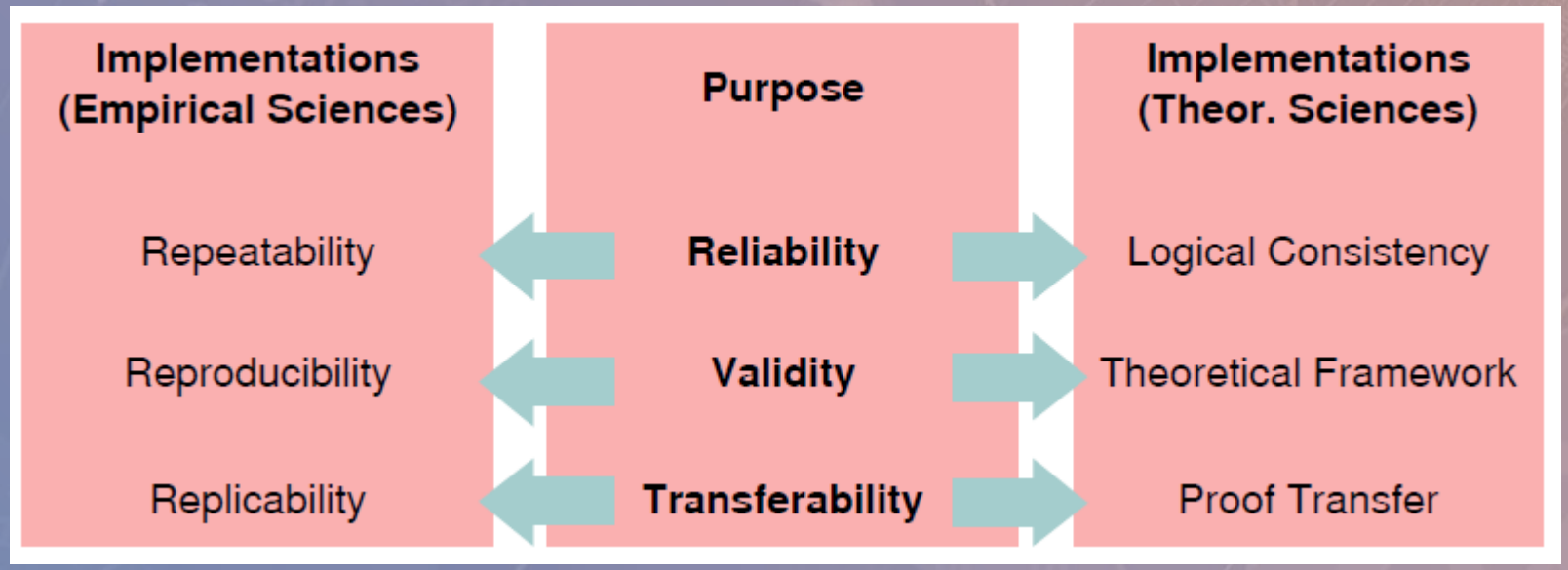


Conclusions

- Researchers recognize the value of Open Science but face significant barriers
- **Key Barriers** :
  - **Knowledge Barriers**: Lack of education in data and software management
  - **Infrastructure Barriers**: Insufficient digital infrastructure for storing and sharing data
  - **Cultural Barriers**: Lack of incentives and academic rewards for open science practices
- **Recommendations**:
  - **Education**: Enhance training in research data and software management
  - **Infrastructure Investments**: Research institutions must invest in digital infrastructure
  - **Recognition Systems**: Revise academic hiring and rewards to prioritize open science
  - **Research culture**: Investigate discipline-specific needs, especially in fields with a strong focus on theoretical research, where open science practices differ from empirical sciences.

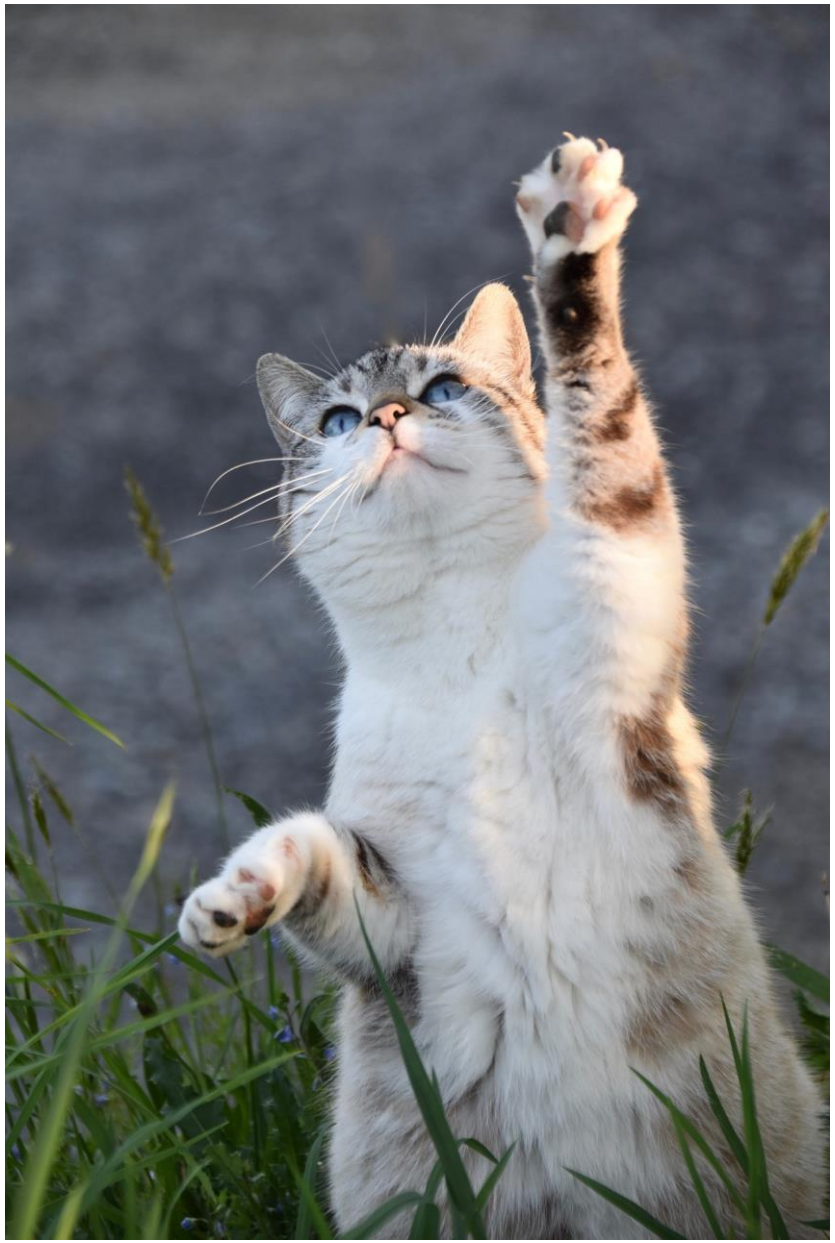
Transdisciplinary Discussions

Would prioritizing the purpose over the implementation enhance transdisciplinary discussions about Open Science?



Thankyou!

christian.riedel@uni-potsdam.de



Pixabay.com, CC0

This work was supported by DeutscheForschungsgemeinschaft (DFG) - Project-ID 318763901 - SFB1294, project Z03.