

# **(Perceived) barriers to data sharing and reuse: Results from a discipline- specific needs analysis**

Agata Bochynska, University of Oslo

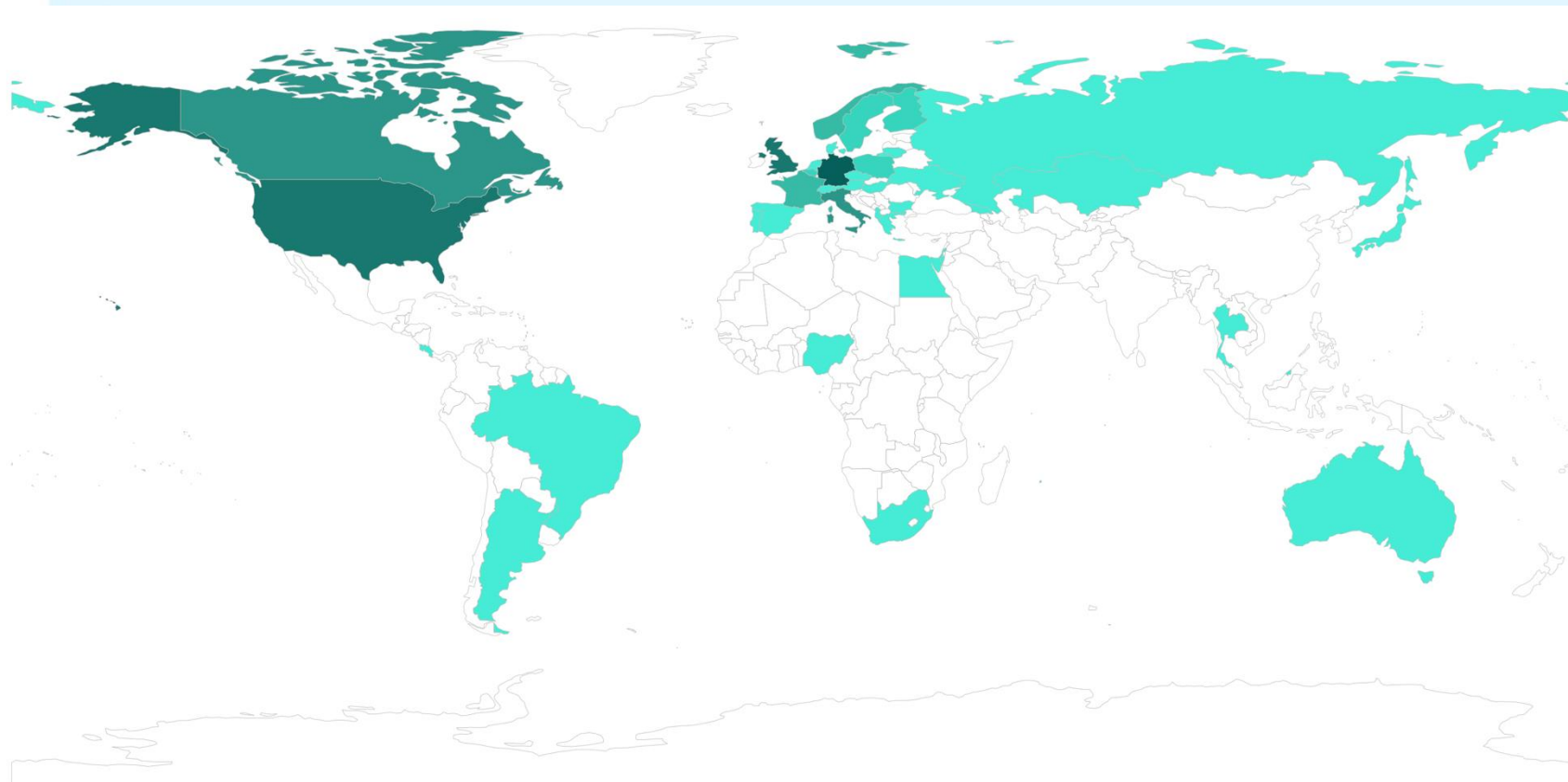
Lindsay Ferrara, NTNU, Norwegian University of  
Science and Technology



- An important facet of Open Science is data sharing and reuse.
- However, uptake of such practices faces challenges, especially in the human and social sciences → also in our field of linguistics (e.g. Berez-Kroeker et al., 2017; Gawne et al., 2017; Digital Science et al., 2022; Bochynska et al. 2023; Logan et al., 2025; Sendra et al., 2025).
- The Linguistics Data Interest Group (LDIG) within the Research Data Alliance set out to explore these challenges, among other topics, with a needs analysis survey.

- Questions related to three major themes:
  - knowledge and perceptions towards open science,
  - research data management attitudes and practices,
    - challenges to data archiving and reuse,
  - knowledge and experiences with infrastructures and initiatives supporting open science and recommended research data management practices
- Demographic questions about the respondents were also included.
- Survey was carried out between late 2023 and early 2025.

## Map of respondents

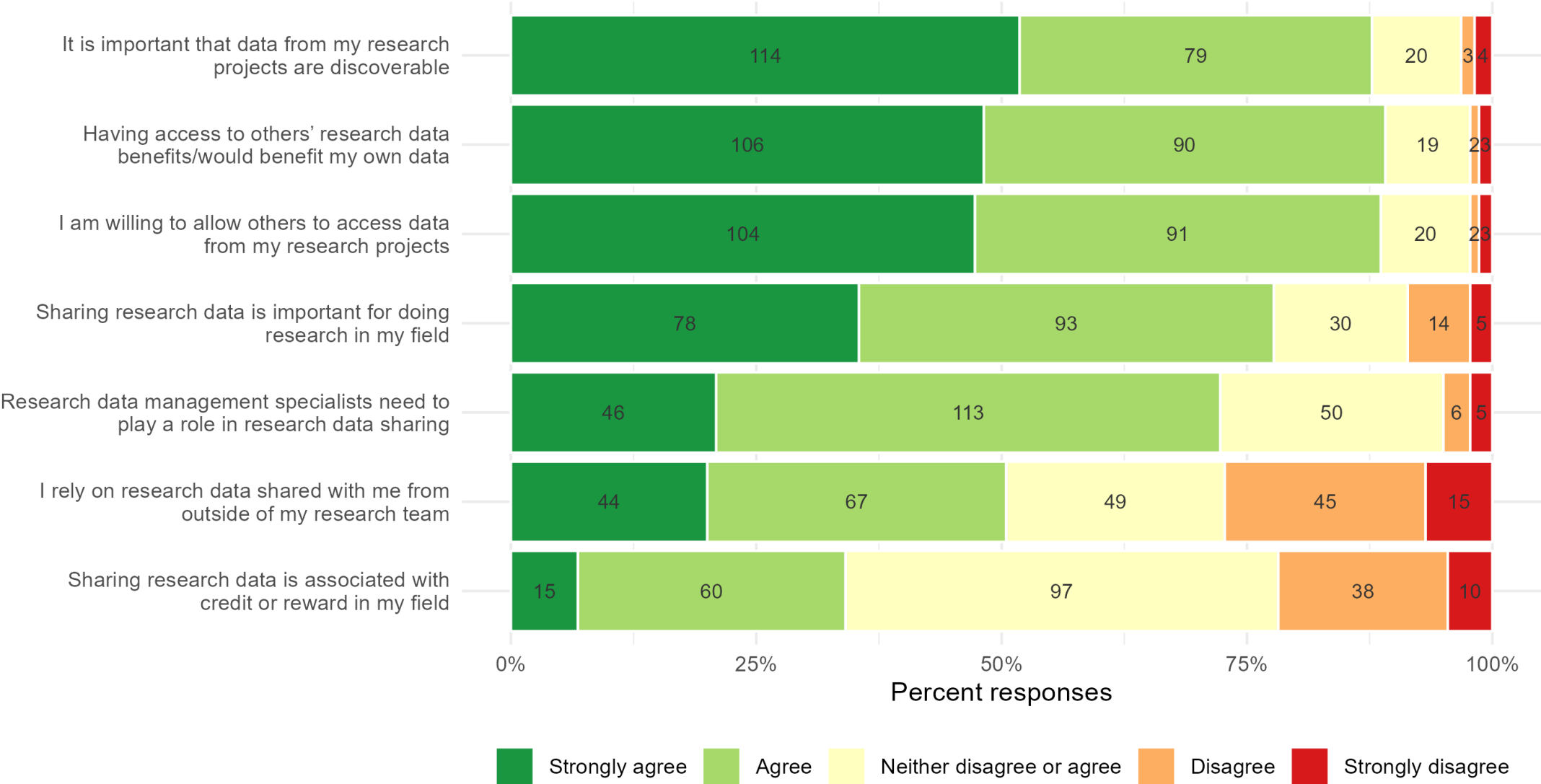


## Needs analysis survey: some respondent characteristics

- **Career stage:** early and mid-career researchers
- **Linguistic subfields:** applied linguistics, corpus linguistics, phonology, psycholinguistics, linguistic typology, language learning, lexicology, morphology...
- **Data types:** annotation data, audio/video data, indigenous data, spectrograms, treebanks, grammaticality judgements, eye tracking data
- **Data formats:** text, tabular, audio, video, programming, structured attribute-value'
- **Size of data:** less than 1 GB – more than one TB // less than 10 files – more than 1,000 files

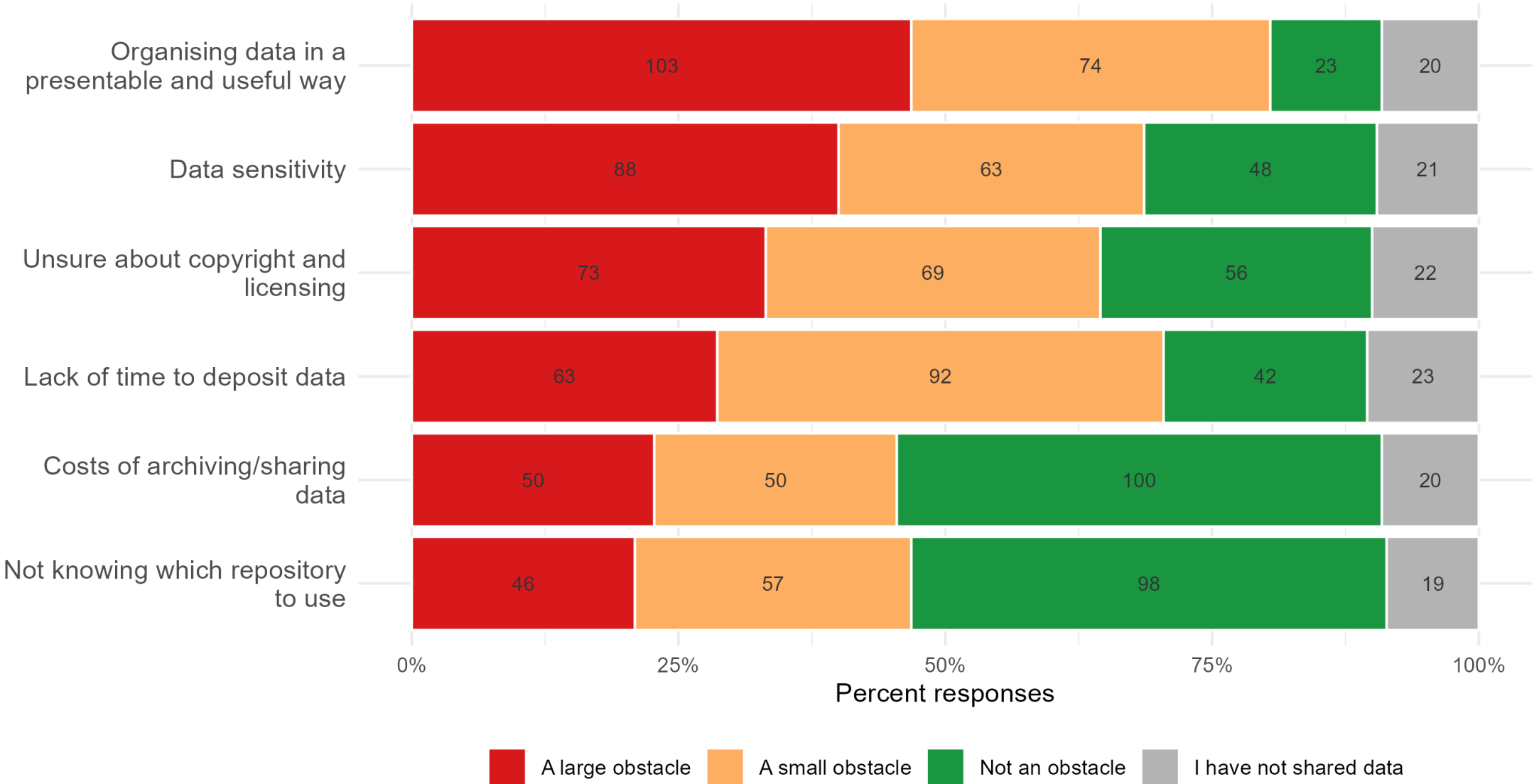
# Data sharing and re-use is important to linguists

How much do you agree or disagree with the following statements:



# Challenges to archiving and sharing

Do you face obstacles in sharing data that you use or create during your research?



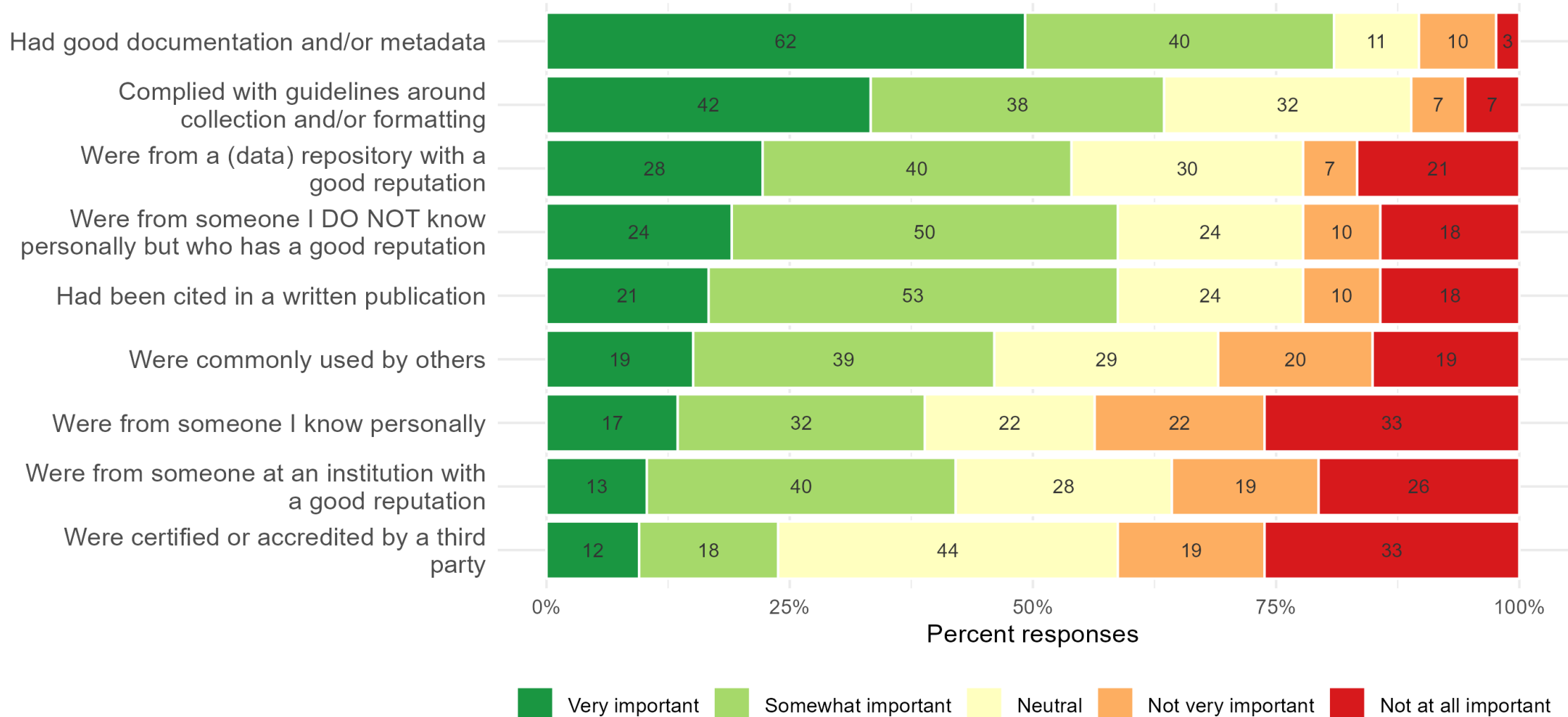
Additional challenges to archiving data (comments from respondents):

- Protectionist collaborators and colleagues
- GDPR
- IRB/Ethics boards who do not understand why data should be shared
- Constantly changing infrastructure



# (Challenges to) data re-use

How important were the following factors when deciding to use others' research data/code? That the data ...



## How to mitigate these challenges: some possible recommendations

- More support for **data management, organization, licensing** – the steps before the actual archiving
- Hands-on **training** on data **organization** and **documentation**
- Training (e.g., mini-lectures/courses) on data **licensing**
- Working with **IRB/ethics/data protection evaluation committees** (nationally, locally) to align messaging around legal/ethical protections and Open Science.

# Thank you



Linguistics Data  
Interest Group (RDA)

<https://www.rd-alliance.org/groups/linguistics-data-ig/activity/>



[eden-fidelis webpage](#)



[/company/fidelis-eosc/](#)



[@fidelis-eosc](#)

- Berez-Kroeker, A. L., Gawne, L., Kelly, B. F., & Heston, T. (2017). A survey of current reproducibility practices in linguistics journals, 2003-2012. <https://sites.google.com/a/hawaii.edu/data-citation/survey>
- Bochynska, A., Keeble, L., Halfacre, C., Casillas, J. V., Champagne, I.-A., Chen, K., Röthlisberger, M., Buchanan, E. M., & Roettger, T. B. (2023). Reproducible research practices and transparency across linguistics. *Glossa psycholinguistics*, 2(1). <https://doi.org/10.5070/G6011239>
- Digital Science, Goodey, G., Hahnel, M., Zhou, Y., Jiang, L., Chandramouliswaran, I., Hafez, A., Paine, T., Gregurick, S., Simango, S., Palma Peña, J. M., Murray, H., Cannon, M., Grant, R., McKellar, K., & Day, L. (2022). The State of Open Data 2022 (Version 5). Digital Science. <https://doi.org/10.6084/m9.figshare.21276984.v5>
- Gawne, L., Kelly, B. F., Berez-Kroeker, A. L., & Heston, T. (2017). Putting practice into words: The state of data and methods transparency in grammatical descriptions. *Language Documentation & Conservation*, 11, 157-189. <http://hdl.handle.net/10125/24731>
- Logan, J. A. R., Hayward, A. L., Swanz, L. E., & Ceviren, A. B. (2025). Education researchers' beliefs and barriers towards data sharing. *Quality & Quantity*, 59(5), 4061-4075. <https://doi.org/10.1007/s11135-025-02188-6>
- 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*, ahead-of-print. <https://doi.org/10.1108/ajim-12-2024-0959>