

Research Data Management in Helmholtz – Findings from Data Professionals Survey 2024

Status quo and opportunities for enhancement

Silke Christine Gerlich¹  <https://orcid.org/0000-0003-3043-5657>, Markus Kubin^{2*}  <https://orcid.org/0000-0002-2209-9385>,
Lucas Kulla^{3*}  <https://orcid.org/0000-0002-2484-2742>, Christine Lemster⁴  <https://orcid.org/0000-0001-7764-1517>, Andreas
Schmidt⁵  <https://orcid.org/0000-0002-9911-5881>, Jan Schweikert⁵  <https://orcid.org/0000-0003-4774-2717>, Sangeetha
Shankar^{6*#}  <https://orcid.org/0000-0003-0387-7740>, Karl-Uwe Stucky⁵  <https://orcid.org/0000-0002-0065-0762>

** These authors contributed equally to this work*

Presenter and corresponding author

¹ Institute Materials Data Science and Informatics (IAS-9), Forschungszentrum Jülich

² Helmholtz-Zentrum Berlin für Materialien und Energie

³ German Cancer Research Center Heidelberg

⁴ GEOMAR Helmholtz Centre for Ocean Research Kiel

⁵ Karlsruhe Institute of Technology

⁶ Institute of Transportation Systems, German Aerospace Center (DLR)

Abstract

The Helmholtz Metadata Collaboration (HMC) is a platform that supports researchers and data professionals in making their research data findable, accessible, interoperable, and reusable (FAIR [1]) through collaborative development of tools, recommendations and other services. HMC also monitors FAIR (meta)data management practices in the Helmholtz Association through community surveys. In 2021, the first survey targeted at researchers in the association revealed the status of data management and data publication practices among researchers. The results can be found in [2].

This submission focuses on the second survey conducted by HMC, the Data Professionals Survey 2024. The survey was targeted at all employees of Helmholtz who work on research data management (RDM), henceforth mentioned as “data professionals”. This target group includes not only data stewards, librarians and data managers, but also researchers and technicians who are actively involved in the management of the research data generated in their department or Helmholtz center. The goals of the survey were to identify the data professionals within Helmholtz and understand their RDM practices, the application of FAIR principles, alignment of their work to data policies of their center as well as the challenges they face and the needs they have in the area of RDM.

The questionnaire consisted of 17 questions that explored various parameters addressing the above-mentioned goals. The survey was implemented using an open-source web-based tool for online surveys called LimeSurvey [3]. The survey was anonymous and participation was voluntary. It was disseminated at different Helmholtz centers over a six-month period between April and October 2024. Data analysis involved the usage of an open-source framework called HIFIS-surveyval [4] as well as Llama 3.3 [5] for identifying common

categories among free-text answers. Python scripts were created for data cleaning and analysis as well as for generating an anonymized dataset for publication. The survey was answered by 156 data professionals working at different Helmholtz centers.

The survey results reveal that the largest group involved in RDM are researchers with close to 39% of the respondents. They spend about 33% of their working time on RDM tasks. On the other hand, 31% of the respondents are explicitly employed as data professionals and they spend close to 80% of their working time on RDM tasks. The respondents were found to be operating at various organizational levels ranging from division level to overarching level. However, only one-fourth of the respondents reported having had a formal training for their RDM tasks. Their reported RDM practices were found to adhere more to findability (61%) and accessibility (49%) aspects of FAIR than to interoperability (39%) and reusability (41%). Furthermore, just over 75% of the respondents mentioned that they are aware of the guidelines and policies of their center and 70% of them also reported at least partially following those guidelines.

Over 90% of the respondents mentioned they face at least one challenge in their daily RDM work and over 85% of the respondents expressed interest in at least one of the service formats. These results help HMC in aligning the services to address the needs reported by the data professionals, thus contributing to the improvement of data management practices in Helmholtz. A complete report on the survey responses as well as a cleaned dataset will be published soon.

Keywords: Data Professionals, Helmholtz Association, Research Data Management, Research Data Practices, FAIR

Resources

—Repository with LimeSurvey questionnaire and analysis scripts:
<https://codebase.helmholtz.cloud/hmc/hmc-public/surveys/hmc-data-professionals-survey-2024>

Author contributions

S.C.G., M.K., L.K., C.L., J.S., and S.S. conceptualized the study. M.K., L.K., S.S. and K.U.S. curated the data. M.K., L.K. and S.S. formally analyzed the data. M.K., L.K., and S.S. conducted the investigation. S.C.G., M.K., L.K., C.L., J.S., and S.S. developed the methodology. L.K. and S.S. worked on the software. M.K., L.K., A.S. and S.S. validated the data. M.K., L.K., and S.S. visualized the data. M.K., L.K. and S.S. wrote the original draft of this report. M.K., C.L., S.S., and K.U.S. reviewed and edited the report. M.K. and L.K. administrated the project.

Competing interests

The authors declare that they have no competing interests.

Funding

The activities presented in the current work were funded by the Helmholtz Metadata Collaboration (HMC).

Acknowledgment

We thank the Helmholtz Metadata Collaboration and the Helmholtz Association of German Research Centers for supporting this project. Many thanks to all individuals in the Helmholtz Association who helped distribute the survey questionnaire and also to those taking the time for answering the survey questionnaire. We thank the survey team of the Helmholtz Open Science Office for strategic discussions, and the HIFIS team for providing the HIFIS-surveyval framework used for analyzing the survey data.

References

- [1] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
- [2] Arndt, W., Gerlich, S. C., Hofmann, V., Kubin, M., Kulla, L., Lemster, C., Mannix, O., Rink, K., Nolden, M., Schweikert, J., Shankar, S., Söding, E., Steinmeier, L. und Süß, W. and Helmholtz Metadata Collaboration (HMC) Working Group "Taskforce Survey" (2022) [*A survey on research data management practices among researchers in the Helmholtz Association.*](#), ed. by Lorenz, S., Finke, A., Langenbach, C., Maier-Hein, K., Sandfeld, S. und Stotzka, R.. HMC Report, 1 . HMC-Office, GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany, 39 pp. DOI [10.3289/HMC_publ_05](https://doi.org/10.3289/HMC_publ_05).
- [3] "LimeSurvey — Free Online Survey Tool". <https://www.limesurvey.org/> (last accessed on 07.04.2025)
- [4] HIFIS Software Services, "hifis-surveyval · PyPI". <https://pypi.org/project/hifis-surveyval/> (last accessed on 07.04.2025)
- [5] "Llama3.3". <https://ollama.com/library/llama3.3> (last accessed on 07.04.2025)