HELMHOLTZ METADATA COLLABORATION

Data management practices among Helmholtz's research communities – A survey on the status quo and on community-specific demands.

Silke Gerlich[1] , Volker Hofmann[1] , Markus Kubin[2] , Lucas Kulla[3] , Christine Lemster[4] , Oonagh Mannix[2] , Katharina Rink[3] , Marco Nolden[3] , Jan Schweikert[5], Sangeetha Shankar[6] , Emanuel Söding[4] , Leon Steinmeier[7] , Wolfgang Suess[5]

[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
[7] Helmholtz Institute Freiberg

Abstract

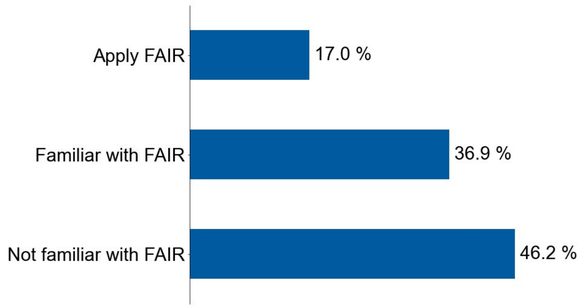
- In 2021, HMC conducted a community survey to understand the *status quo*, gaps, and needs of research data management practices across the Helmholtz Association.
- 631 completed survey replies were obtained from researchers in all Helmholtz centers from all six Helmholtz research fields.
- An official report with a Helmholtz-wide analysis of the survey data and a data publication will be available soon [1].

Approach

- The survey was designed to dynamically **adapt to the respondents' expertise** across subject topics.
- Survey consisted of **49** (sub-) questions.
- Data analysis was carried out with the **HIFIS-surveyval** python framework [2].

Personal Background

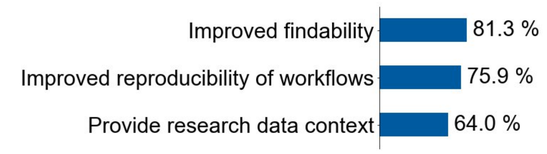
- Most responses from **research associates** (36%) and **principal investigators** (23%).
- Most responses from participants who have been working in research more than 10 years (49%).



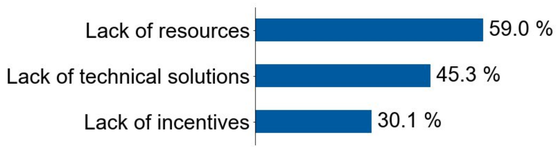
"How familiar are you with the FAIR data guidelines?" *(All hubs, except Hub Information; Single choice; n = 599; relative amounts refer to n.)*

Research Data Management

- Only 17% of respondents store their data on external servers and repositories after the end of a research project. 83% of respondents store their research data in internal servers.
- **35% document their research data in a structured way**, out of which 62% use internationally used **schemas and standards**.
- 67% work on purely or mostly self-generated data.
- 22% work on purely or mostly reused data.



Top 3 motivations to document work in a structured way *(All hubs)*

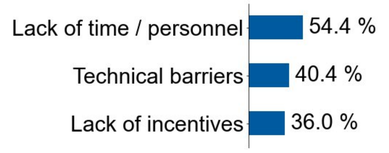


Top 3 difficulties encountered in collecting metadata *(All hubs)*

Data Publication

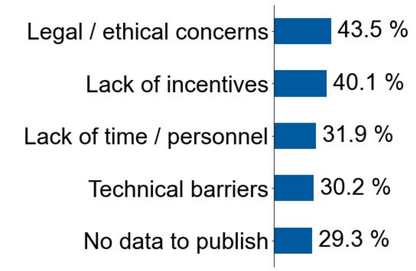
- **60% have published at least some of their research data.**
- Respondents who publish some of their data said they **published data as a supplement to journal publications** (64%) and **repositories** (52%) or data journals (10%).

Top 3 motivations to publish research data *(All hubs)*



Top 3 obstacles encountered in publishing research data *(All hubs)*

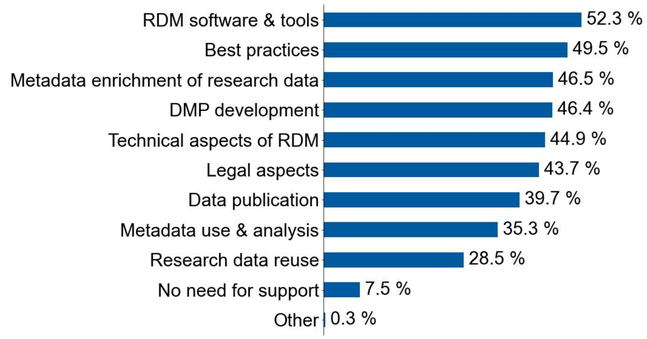
- **40% have never published their data.**



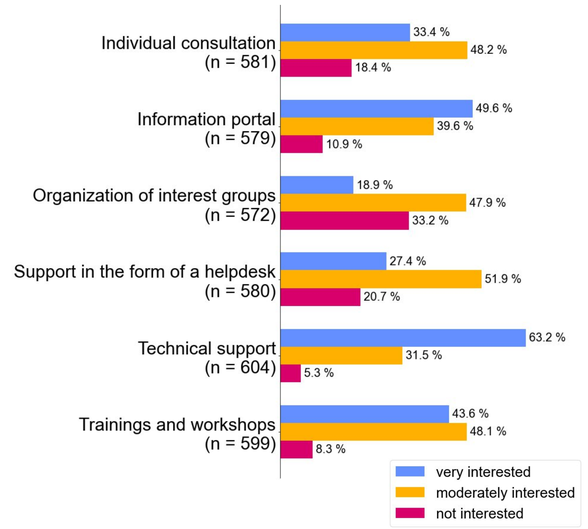
Top 5 obstacles that discouraged participants from publishing data *(All hubs)*

Services

- The **community** has expressed a **need for support** in RDM software & tools, best practices, etc.



"In which areas of research data management do you perceive a need for supporting services?"*. (All hubs; Multiple-choice; n = 604; relative amounts refer to n.)*



"Please rate your interest in the following service formats"*. (All hubs; Multiple-choice; One non-mandatory interest rating per service format n = 612; relative amounts refer to n.)*

References

[1] HMC Community Survey 2021 (data publication) doi: https://doi.org/10.7802/2433
[2] HIFIS-Surveyval python framework https://gitlab.hzdr.de/hifis/overall/surveys/hifis-surveyval