

The HMC Information Portal for enhanced metadata collaboration in the Helmholtz FAIR data space

Lucas Kulla¹[\[https://orcid.org/0000-0002-2484-2742\]](https://orcid.org/0000-0002-2484-2742), Jens Bröder²[\[https://orcid.org/0000-0001-7939-226X\]](https://orcid.org/0000-0001-7939-226X), Constanze Curdt³[\[https://orcid.org/0000-0002-9606-9883\]](https://orcid.org/0000-0002-9606-9883), Markus Kubin⁴[\[https://orcid.org/0000-0002-2209-9385\]](https://orcid.org/0000-0002-2209-9385), Helen Kol-lai⁵[\[https://orcid.org/0000-0003-0214-1336\]](https://orcid.org/0000-0003-0214-1336), Christine Lemster⁶[\[https://orcid.org/0000-0001-7764-1517\]](https://orcid.org/0000-0001-7764-1517), Marco Nolden¹[\[https://orcid.org/0000-0001-9629-0564\]](https://orcid.org/0000-0001-9629-0564), Kai Schmieder⁷[\[https://orcid.org/0000-0002-1171-1428\]](https://orcid.org/0000-0002-1171-1428), Annika Strupp⁸[\[https://orcid.org/0000-0002-0070-4337\]](https://orcid.org/0000-0002-0070-4337), Karl-Uwe Stucky⁷[\[https://orcid.org/0000-0002-0065-0762\]](https://orcid.org/0000-0002-0065-0762), Emanuel Söding⁵[\[https://orcid.org/0000-0002-4467-642X\]](https://orcid.org/0000-0002-4467-642X), Konstantin Pascal Walter⁴[\[https://orcid.org/0009-0003-4693-0932\]](https://orcid.org/0009-0003-4693-0932) and Arndt Witold⁹[\[https://orcid.org/0000-0002-7713-9647\]](https://orcid.org/0000-0002-7713-9647),

¹ Division of Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany

²Forschungszentrum Jülich GmbH

³GEOMAR Helmholtz Centre for Ocean Research Kiel

⁴Helmholtz-Zentrum Berlin

⁵Helmholtz Centre for Environmental Research UFZ

⁶GEOMAR Helmholtz Centre for Ocean Research Kiel

⁷Karlsruhe Institute of Technology

⁸Forschungszentrum Jülich

⁹German Aerospace Center

The **Helmholtz Metadata Collaboration** (HMC) platform was launched in late 2019 to turn **FAIR** (Findable, Accessible, Interoperable, Reusable) research data into reality within the Helmholtz Association and beyond. The **Information Portal** was initiated to enable the structured **cartography of metadata** and FAIR landscape of Helmholtz, providing information for multi-level decision-making and creating a curated knowledge base for research data managers, scientists and other stakeholders.

Developed through a top-down approach, 18 categories, and associated metadata schemas were defined and aligned by an HMC taskforce. Data curation followed, with resources collected from different domains based on the aligned metadata schema. The Information Portal is a **web application** for capturing **FAIR data practices** across all Helmholtz domains, offering a unified user interface for collecting and exploring results.

Built using **state-of-the-art technologies**, including Python and Docker, the Information Portal leverages GitLab as a database. It offers a public / central read-only version for stakeholders and a personal instance for curation - synchronized to a GitLab repository. Git-based systems offer advantages, such as raw data accessibility, flexible data curation, easy synchronization, and customizable repositories.

The single-page web application is **user-friendly** and developed in multiple iterations for an intuitive and flexible interface. The Information Portal is crucial for creating a **sustainable, distributed, semantically** enriched Helmholtz data space, promoting seamless data sharing and reuse.

Keywords: Helmholtz Metadata Collaboration, Information Portal, FAIR

The Helmholtz Metadata Collaboration (HMC) platform was launched in late 2019. Its mission is to leverage the visibility and reusability of data across the Helmholtz Association and beyond and to turn FAIR (Findable, Accessible, Interoperable, Reusable) [2] research data into reality. HMC operates as a federated collaboration embedded in different Helmholtz research fields, providing a unique opportunity to translate global metadata concepts into practical implementations. HMCs vision is to create a sustainable, distributed, semantically enriched Helmholtz data space that scientists can use to seamlessly share and re-use data in new ways.

To achieve this goal, the development of the Information Portal was initiated. The Information Portal enables structured cartography of metadata and FAIR landscape of Helmholtz and beyond, providing information for multi-level decision-making for different stakeholders, and creating a curated knowledge base for research data managers and scientists.

Developed in a top-down approach, 18 categories and their associated metadata schemas were aligned and defined between the domains by a taskforce of HMC such as repositories, metadata standards and terminologies. Followed by the alignment, data collection and curation were the next steps. Therefore, the landscape of different domains within the Helmholtz collaboration was explored and data for the various resources were collected based on the defined and aligned metadata schema. The Information Portal is a web application to capture FAIR data practices and resources across all Helmholtz. It uses a unified user interface for capturing and exploring the results of the landscape design process in different areas. It is built using state-of-the-art technologies, with a Python and Docker-based system, leveraging the GitLab Backend of the Helmholtz Cloud [1] for the management of contributions, users, and quality control through CI pipelines. Currently, there are two versions available. The first is a public version, intended to be used as a knowledge base for different stakeholders within Helmholtz (<https://informationportal.helmholtz-metadaten.de>). This version features a read-only view, where users can explore the landscape and find useful information for their work. The second is a personal instance, which is intended to be hosted either on an internal server or on a personal device. It is based on the public version but includes all the tools necessary to curate resources. The edited, deleted, or added resources are then synchronized to a pre-defined GitLab repository. Reusing a Git-based system as a database has multiple advantages over a standard, e.g. Postgres database. The first is that the raw data is accessible by all users of GitLab even without access to the Information Portal. Secondly, curation may or may not be done through the Information Portal, as everyone has the option to add data to the GitLab repository in their preferred way. Thirdly, this allows easy synchronization between multiple private Information Portal instances. Lastly, it is not necessary to use the default GitLab repository. To change the destination of truth a single URL needs to be changed.

Creating a private instance is due to the use of docker-compose a straightforward process, regardless if its deployment is on a local machine or on a server. In addition, the single-page web application is built as user-friendly as possible. Developed in multiple iterations with multiple technological approaches, the user interface is as intuitive, flexible, and target oriented as possible.

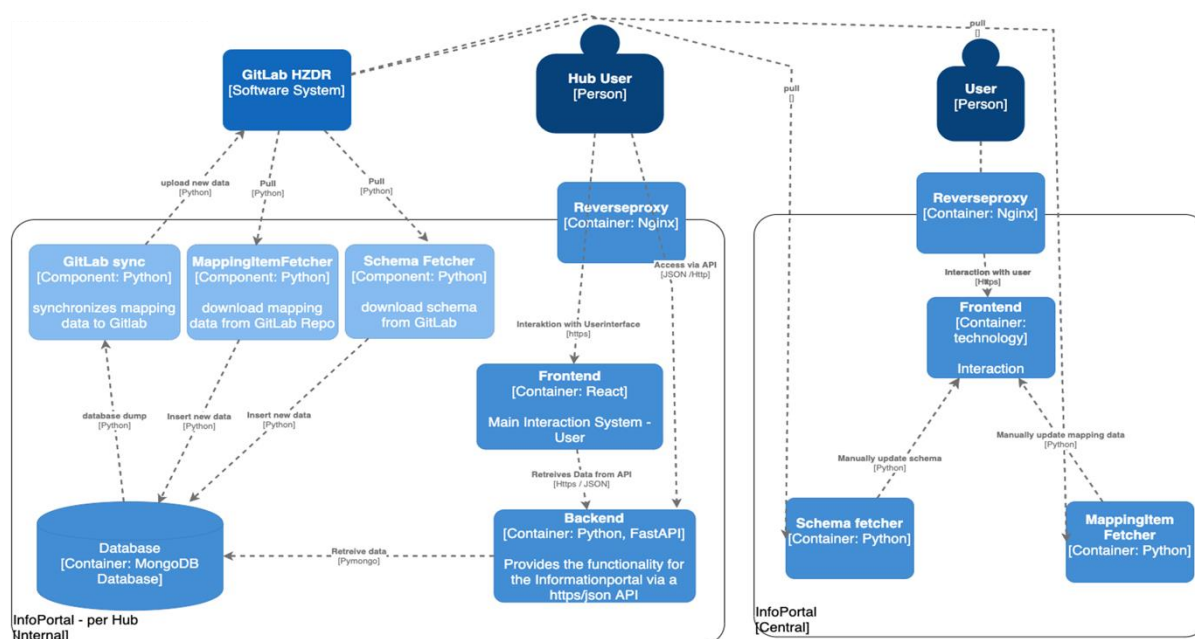


Figure 1. Architecture of the Information Portal

In conclusion, the development of the InformationPortal within HMC has been a crucial step in achieving the goal of creating a sustainable, distributed, semantically enriched Helmholtz data space that scientists can use to seamlessly share and reuse data in new ways. Through a top-down approach and successful collaboration across different units, the Information Portal has been developed utilizing state-of-the-art technologies, providing a centralized resource for metadata and FAIR landscape mapping.

The future outlook of the Information Portal includes several improvements. First, establishing a public contribution and quality assurance process. Secondly, semantic enrichment in the context of Linked Data. The goal is to further utilize the use of data in a way that data is not only seen as independent entries but being seen as a network of entries that are connected to each other.

In conclusion, HMC's pioneering efforts not only empower researchers to unlock the full potential of their and others' data but also pave the way for unprecedented scientific discoveries and collaborations, ultimately driving innovation and shaping the future of research in the Helmholtz Association and beyond.

Author contributions

In the following paragraph, author contributions are summarized following the CRediT taxonomy.

Name	Concep- tual- iza- tion	Data cura- tion	In- vesti- ga- tion	Meth- odol- ogy	Pro- ject ad- min- istra- tion	Soft- ware	Su- per- vi- sion	Writ- ing – origi- nal draft	Writ- ing – re- view & ed- iting
Lucas Kulla	X	X	X	X	X	X		X	
Constanze Curdt	X	X	X	X	X				
Markus Kubin	X	X	X	X	X				X
Helen Kollai	X	X	X	X	X				
Christine Lemster	X	X	X	X	X				
Marco Nolden					X		X		X
Oonagh Mannix					X		X		X
Kai Schmieder	X	X	X	X					
Annika Strupp	X	X	X	X					
Karl-Uwe Stucky	X	X	X	X					X
Emanuel Söding	X	X	X	X	X		X		
Konstantin Pascal Walter	X	X	X	X					
Arndt Witold	X		X		X		X		

Competing interests

The authors declare that they have no competing interests.

Funding

This publication was supported within the hub Health at the German Cancer Research Center by the Helmholtz Metadata Collaboration (HMC), an incubator-platform of the Helmholtz Association within the framework of the Information and Data Science strategic initiative.

References

1. DESY, D. E.-S. (2023). *HIFIS*. Deutsches Elektronen-Synchrotron DESY. Retrieved 26.04.2023 from <https://hifis.net>
2. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., & Bourne, P. E. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9.