Road-to-FAIR-Strategy

While the vision of HMC is to create a FAIR data space, it does not automatically make it obvious how this vision could be implemented. As the FAIR principles are a rather rough guideline, outlining the four major goals we strive to achieve for our data, in fact, a deeper implementation strategy is required, which defines who does what, in which order tasks are taken, and which measures precede others. However, there are so many aspects in data management, FAIR, AI readiness and so on, that it is difficult to keep track of issues, put measures and goals into context, while keeping the FAIR principles in sight. It is therefore important and helpful to formulate a narrative that can be easily understood by everyone involved and that acts as a guide, against which implementation measures can be prioritized and evaluated. This narrative, putting everything into context, proportion, and prioritizing tasks, is our Road-to-FAIR strategy (RTF).

The RTF strategy aims to

1. define and prioritize our **goals**, reflected by the **Road-to-FAIR building blocks**, ensuring interoperability and re-usability of data from diverse origins.
2. define **core concepts**, like **stakeholder groups**, playing **roles** in the RDM processes. It explains how and where we envision stakeholders to **assume responsibility** for their role in RDM.
3. define **workflows**, that can be followed by the stakeholders, resulting in coordinated activities and **cooperation** on their way to establish the harmonized data space, as the FAIR metadata space can not be achieved individually.
4. define aspects of the technical **data ecosystem**, like protocols, interfaces, and services, which support the data flow within institutions, and across organizations, and which allow stakeholders to implement and activate the measures outlined as building blocks.

The RTF strategy thus sets the beginning and end of all our activities. It starts with defining RDM as a community task, orchestrated by the RDM concepts and personnel, where many people, often beyond their current awareness, play a role in documenting science, creating knowledge, and ultimately allowing for next-level science in the form of AI and ML applications. It finally suggests ways how to make this knowledge accessible and usable. The strategy supports the various types of information in where they are created, along and across the documentation process, towards their intermediate location in the repositories, ready for aggregation and enrichment processes, on their way to serve as fuel for system understanding, scientific models and simulations, and ultimately decision making for a better future of mankind served by the harmonized FAIR metadata space.

1. The beginning - the FAIR building blocks

The FAIR building blocks define the core and the some immediate goals on our road to FAIR. The building blocks are aligned along the FAIR princpiles, as we ask ourselves: where do we stand in the implementation of FAIR? In Helmholtz we assume, that a significant part of the FAIR principles are already solved and implemented. These parts comprise Findability and Accesability for all data, that has been deposited in well managed institutional or disciplinary repositories. This information is relatively easy to find, e.g. through meta-databases like re3data, fairsharing or OpenAire tools. The metadata and often the data itself can typically be accessed, as Helmholtz follows an open science policy, and strives to publish its data whereever possible. The big challenges within our organization are the Interoperability and the Reusability of our data. All repositories, projects, research programs, disciplines, networks follow own rules, standards, and procedures in describing their data. This makes

it nearly impossible to aggregate information from more than one repository, or even often within repositories. As a consequence data reuse on very heterogeneous datasets is very hard to conduct. In fact projects, who try to model on heterogeneous data sets and aggreagate information, spend by far most of their time, to clean data, harmonize semantic expressions, and enrich metadata with missing information, in order to compile useful, high-quality data sets.

2. The Core concepts: Who is responsible for what or defining stakeholder groups and their roles in our organisation

A complete and comprehensive description of a dataset is not trivial. It requires in-depth knowledge about techniques how to describe the different actors, organizations, measured parameters, instruments and methods used to create the dataset and many other aspects needed to describe the circumstances under which a dataset or sample was produced or acquired. Experience shows, that not one person can provide all this information herself. We therefore suggest to adopt a community approach in collecting information related to the description of datasets, and take a look, who in our organisation assumes which role and what data the people assuming this role are possible administrating. We also take a look at the organisation itself,

3. Activating the community: Defining workflows and standard procedures to capture cor information and pass it on within the system

4. Making it possible: Enabling the community by defining and providing the technical **data ecosystem** supporting the data

How to achieve a FAIR state of our data ecosystems What does FAIR mean in that respect?

Data is documented in the most complete way, allowing F, A, I and R to be achieved.

What are the problems in the current practices?

How do we move forward, what are the priorities?

A brief description of the roadmap.

From:
https://earth-and-environment.helmholtz-metadaten.de/wiki/ - **HMC Earth and Environment Community Wiki**

Permanent link:
https://earth-and-environment.helmholtz-metadaten.de/wiki/doku.php?id=wiki:1_road-to-fair-strategy:start&rev=1785767018

Last update: **2026/08/03 14:23**