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 tools and services needed to implement the workflows and take up the information needed. But also protocols and interfaces allowing to connect these tool, supporting the data flow between tools, 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

The Road-to-FAIR Strategy treats research data management as a distributed institutional responsibility rather than an activity that can be assigned exclusively to researchers or repositories. The information required to describe a dataset is created and maintained by different stakeholder groups, including researchers, technicians, administrative units, organizational management, libraries, data stewards, and repository operators. No single stakeholder normally possesses all information required for a complete and accurate description. The organization should therefore actively express expectations towards those stakeholder groups and document them through internal policies.

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

To coordinate these distributed responsibilities, the strategy defines workflows through which information is captured close to its point of origin, maintained by an appropriate authoritative source, and transferred to downstream data systems. This approach replaces the common practice of requesting all metadata from researchers only at the time of data publication. Instead, relevant information should flow continuously through institutional information interfaces connecting administrative systems, research processes, technical services, and data infrastructures. Its implementation requires both institutional support and technical services capable of exchanging, validating, aggregating, and enriching metadata.

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

In a fourth step, the Road-to-FAIR Strategy requires the definition, construction, and activation of a coordinated data ecosystem. This ecosystem comprises the services, tools, registries, interfaces, and technical infrastructures needed to support the stakeholder responsibilities and workflows defined in the preceding steps. Its purpose is to enable information to be captured at its point of origin, maintained by the responsible actors or authoritative systems, and transferred reliably to repositories and other downstream services. Relevant components may include persistent identifier registries, electronic laboratory and field notebooks, sample and instrument management systems, institutional information systems, vocabulary services, metadata editors, validation tools, interfaces, and aggregation services. These components must not operate as isolated applications, but as an interoperable environment in which metadata can be exchanged, validated, enriched, and reused across organizational and disciplinary boundaries. The data ecosystem therefore provides the technical foundation that translates assigned responsibilities and agreed workflows into sustainable operational practice.

The Road-to-FAIR Strategy complements the FAIR principles by translating their high-level objectives into a coordinated implementation framework. While FAIR defines the desired properties of research

data and metadata, it does not prescribe which standards, technologies, responsibilities, or processes should be used. The Road-to-FAIR Strategy addresses this gap by defining concrete building blocks, assigning responsibilities to relevant stakeholder groups, establishing workflows for the creation and maintenance of metadata, and specifying the technical services and interfaces required to support these processes.

A central element of the strategy is the development of shared community agreements. Persistent identifiers, metadata schemas, semantic vocabularies, exchange protocols, provenance models, licenses, and quality procedures must be selected and applied consistently across infrastructures. The strategy therefore treats FAIR implementation as a collaborative and incremental process rather than a binary state. Institutions can identify areas that are already comparatively mature, prioritize remaining gaps, and progressively improve interoperability, reusability, and machine actionability.

The strategy also recognizes that rich metadata alone are insufficient. Metadata must use harmonized semantics, explicit relationships, and interoperable formats so that information can be reliably interpreted, combined, and reused across organizational and disciplinary boundaries. At the same time, FAIRness is distinguished from unrestricted openness: data may remain access-controlled where legal, ethical, or contractual requirements apply, provided that access conditions and procedures are clearly documented.

Overall, the Road-to-FAIR Strategy establishes the socio-technical conditions needed to operationalize FAIR. It connects strategic goals with community agreements, institutional responsibilities, practical workflows, and an interoperable data ecosystem. In doing so, it provides a structured pathway from a general commitment to FAIR towards a progressively harmonized and reusable research data space.

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=1786010815

Last update: **2026/08/06 10:06**