HMC Hub Earth and Environment

Our Vision

The vision of the HMC Hub Earth and Environment is to create a Helmholtz data space unifying Earth and Environmental Centers and infrastructures, powering a new wave of large-scale, globally oriented, data-driven research.

The Helmholtz FAIR Data Space (HFDS) is a “decentralized infrastructure for trustworthy data sharing and exchange in data ecosystems based on commonly agreed principles“ (Nagel L., Lycklama D., 2021).

Mission

The hub's mission is to federate the (meta)data systems Earth and Environment Centers and infrastructures throughout the Helmholtz Association, continuously aligning Helmholtz capacities to global norms and developments.

Basic Goals and Principles

Basic recommendations and assumptions to develop and activate a FAIR Dataspace for the Helmholtz Association

Description

[Status: Under development, Date: 2025-06-19, Version: 001]

Motivation for this Recommendation:

In order to achieve interoperability of datasets among various data infrastructures (DIS) within the Helmholtz Association,

- a common and agreed procedure to refer to common information is needed
- common semantic concepts need to be implemented
- common data exchange containers and formats need to be defined and implemented
- common interfaces and data exchange protocols need to be agreed upon

These following recommendations aim to describe procedures to achieve the above-mentioned goals, and identify activities and the relevant stakeholder groups to implement these. This implies that the identified stakeholders assume certain responsibilities in order to make these recommendations work.

As these goals require certain decisions to be made, we describe the decision processes from high-level to lower levels, resulting in agreed procedures. We then describe the conduct of these procedures with respect to the involved stakeholder groups, their responsibilities and the actions to be taken.

The activities may be prioritized for implementation in a specific order or for a particular stakeholder group, either because they are simple to set up, or because they are prerequisites for other recommended activities. These decisions have to be made on a case by case basis, in order to move forward establishing the envisioned data space.

Comment DK: *I think in this paragraph needs some modifications to put equal weight on all aspects of FAIR. The above version particularly addresses interoperability in its formulation. I also think that there is too much redundancy in the motivation paragraph with the recommendation paragraph, therefore I would here really only focus on the motivation. The suggested measures can come in the paragraph below. I suggest to write the following:*

The Helmholtz Association is committed to making its data available in accordance with the FAIR principles — ensuring that data are Findable, Accessible, Interoperable, and Reusable. This commitment is outlined in the “Recommendations for Policies of the Helmholtz Centers on Research Data Management” [1].

To achieve FAIR data across the various data infrastructures (DIS) within the Helmholtz Association, a number of coordinated efforts are required. The following recommendations aim to describe procedures to achieve the above-mentioned goals, and identify activities and the relevant stakeholder groups to implement these. This implies that the identified stakeholders will take on specific responsibilities to ensure the success of the FAIR-aligned approach.

Since achieving these goals requires deliberate decision-making, the document outlines decision processes from high-level strategy to operational detail, resulting in agreed procedures. It further describes how these procedures are to be implemented, specifying the roles, responsibilities, and actions required of each stakeholder group.

Activities may be prioritized for implementation based on factors such as ease of deployment or their role as prerequisites for other efforts. These decisions should be made case by case to effectively progress toward the envisioned Helmholtz data space.

Comment AP: *I also see the redundancy like DK and prefer a drawing of the building blocks. Here is my suggestion:*

The Helmholtz Association is determined to make their data available according to the FAIR principles, thus making it findable, accessible, interoperable and reusable. This is expressed in the “Recommendations for Policies of the Helmholtz Centers on Research Data Management” [1].

To achieve FAIR data across the various data infrastructures (DIS) within the Helmholtz Association, a number of coordinated efforts are required. The first step is to examine the existing infrastructures and their interfaces in order to identify the best practices to use as a model. With the help of coordination processes in various committees, which have agreed on different individual goals

(building blocks - see figure …), the detailed implementation and recommendation is agreed for each agreed building block and documented on the corresponding website.

The resulting recommendations are always recorded with the same questions and harmonized structure. There are therefore two perspectives to consider: on the one hand, the discussion and agreement on content, and on the other, how the administrative and technical implementation should take place. Each individual step required for implementation must be examined. The perspectives and roles of the various parties involved are taken into account in order to consider all influences in detail. Also the decision-making processes that lead to established procedures are outlined from the lower to the upper levels and vice versa.

This provides a foundation for optimizing the procedures concerning the individuals involved, their responsibilities, and the necessary actions. In order for the FAIR approach to be successful, it is essential that identified stakeholders take on specific responsibilities.

Basic FAIR Implementation Recommendations

Implement the **FAIR building blocks**:

- Common use of **PID metadata** to harmonize common metadata (M)
- Common use of **standardized interfaces** in repositories and data products to harvest and exchange data (I)
- Common use of **semantic resources** (S), e.g. vocabularies, ontologies
- Use of **data containers**, e.g. FDOs or data crates, to achieve machine actionability (C)

Other important topics where harmonize procedures are important are:

- Well defined **licenses** (L)
- **Quality** assessment and control (Q)
- **Provenance** tracking (P)
- Stakeholder **roles and responsibilities** (R)
- **Valuing RDM engagement** (V) by treating published data similar to other publications including DOIs and author references, citable, with licenses and credits to the authors. This includes accountability for publishing credit systems and others. Data citations improve the data set author’s score.

A very important precondition for the implementation of these recommendations is the **presence of skilled and highly trained data management personnel**, such as data stewards, data curators, and developers for data management and processing tools.

Comment/Suggestion DK: *I would suggest to rewrite this paragraph in a way that it sounds more like a recommendation and is less redundant with the motivation paragraph. I therefore merged and extended the bullet sections from both paragraphs. My suggestion is to write the following*:

It is recommended that coordinated efforts across the Helmholtz Association be undertaken to achieve the above-mentioned goals, with the implementation of a common set of **FAIR building blocks**. These building blocks are operational measures required to support the findability, accessibility, interoperability, and reusability of research data. They include:

- Common and agreed procedures to refer to shared entities and metadata values, particularly through the use of **metadata of persistent identifiers** (**PID**s), to reduce redundancy and improve consistency in referencing key entities such as people, organizations, instruments, and datasets (**rename to PID**?).
- Harmonized use of **semantic resources**, including controlled vocabularies and mapping tables, to standardize metadata element names and their meanings, reduce ambiguity, and ensure consistent interpretation across disciplines and systems (**S**).
- Agreements on **mandatory and optional metadata elements**, covering both domain-independent and domain-specific needs, to support accurate description, discovery, exchange, and reuse of data across infrastructures, user communities, and tools (**M**?).
- Common metadata or **exchange formats**, such as DataCite, DCAT, or ISO 19115, to ensure consistent structuring and exchange of metadata across systems (**E**)
- Common data **exchange containers**, such as FAIR Digital Objects or DataCrates, to enable machine-actionable reuse and portability (**C**).
- Common **interfaces**, to provide standardized, open mechanisms for data and metadata access and harvesting (**I**).

Other important topics where harmonization procedures are important are:

- Inclusion of **provenance information**, capturing data origin, transformation steps, and responsible agents, to enable assessment of data reliability and reproducibility (**P**).
- Clear **license information**, using standardized, machine-readable licenses, to define conditions of use and promote legal clarity (**L**).
- Documented procedures for assessing and communicating **data quality**, including uncertainty, validation, completeness, and versioning, to ensure data are fit for purpose (**Q**).
- Clear definition of **stakeholder roles** and responsibilities, including who is accountable for metadata provision, data stewardship, infrastructure maintenance, and policy implementation (**R**).
- **Valuating research data management** (RDM) engagement, through citation of datasets with DOIs, inclusion of author contributions, and formal acknowledgment of data curation efforts (**V**).

These elements form the structural foundation for the detailed recommendations presented in this wiki.

A key precondition for the implementation of these recommendations is the availability of skilled personnel, such as data stewards, data curators, and developers, who support data management and the technical and semantic infrastructure required to implement FAIR.

References

[1] Empfehlungen für Richtlinien der Helmholtz-Zentren zum Umgang mit Forschungsdaten, 2019, Helmholtz Open Science, https://os.helmholtz.de/open-research-data/forschungsdaten-policies/