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.

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/