

# HMC: Towards a FAIR Data Space for Helmholtz Earth and Environment Data Infrastructures

Emanuel Söding<sup>1</sup> // Martin Weinelt<sup>1</sup> // Andrea Pörsch<sup>2</sup> // Pier Luigi Buttigieg<sup>3</sup> // Sören Lorenz<sup>1</sup>

<sup>1</sup> GEOMAR Helmholtz Centre for Ocean Research  
<sup>2</sup> Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences  
<sup>3</sup> Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research (AWI)

## 1 HMC Goal: Establish a common FAIR data space

What is the FAIR data space?

The 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). Within Helmholtz we may alter details of the concept to our specific needs.

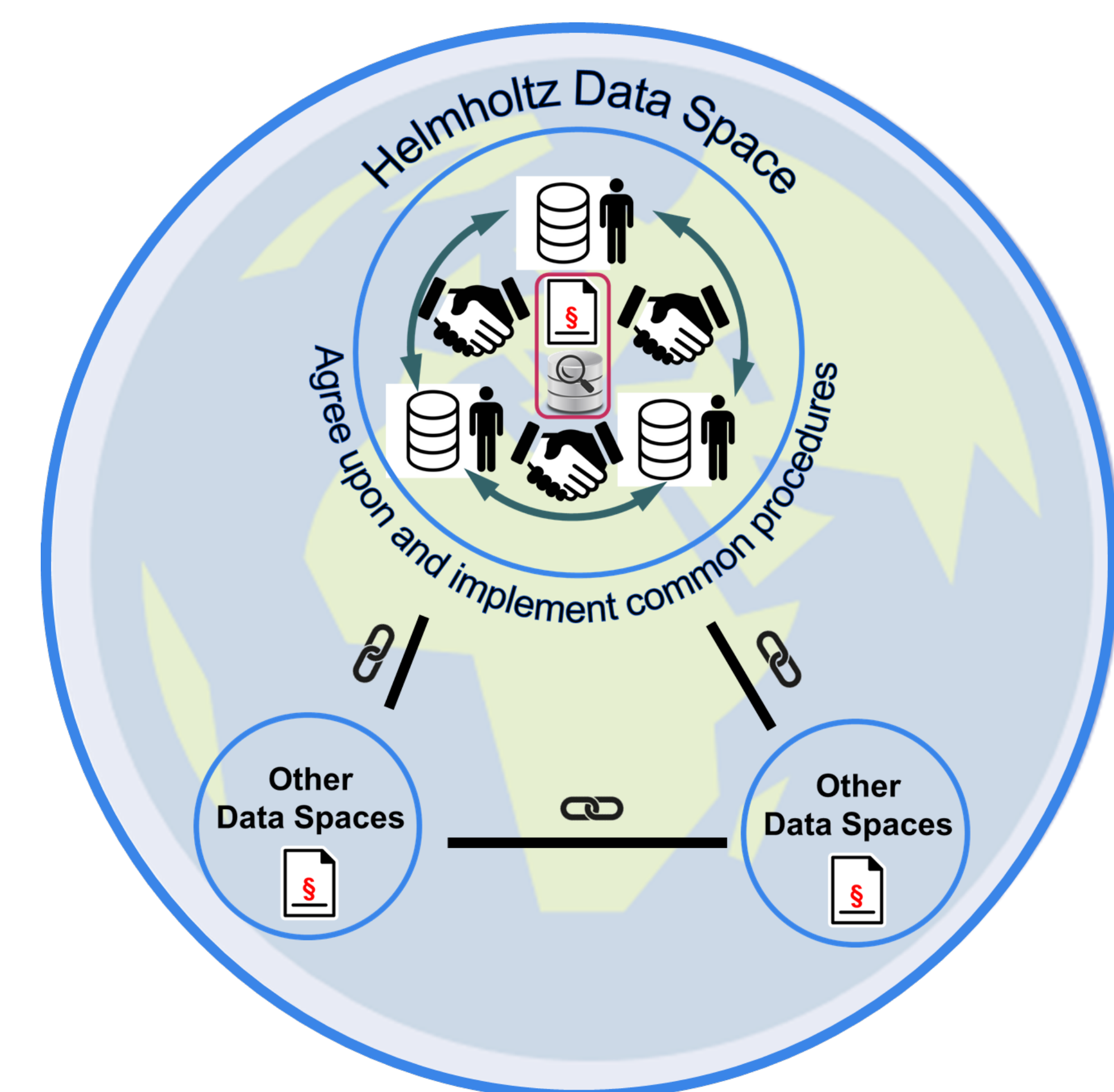
What are the components of the FAIR data space?

The data space consists of:

- common **agreements** (see 4) to implement the FAIR building blocks (see 2), leading to internal interoperability of data.
- a **data integration system**, which will act as a data broker between data infrastructures, providing internal and external integration and data access opportunities.

Is this a Helmholtz-only data space?

No. This is an **open development**. Everyone may join, and anyone who follows the same principles agreed among the partners, is automatically part of this data space.



## 2 Concept: Building Blocks of the FAIR Data Space

What do we need to agree upon?

To build an interoperable FAIR data space, agreements on the necessary building blocks must be reached. These building blocks are:

**PID** -metadata, to create references for redundant information

Common metadata information needs to be kept in agreed **PID systems**, e.g. DOI, ROR, ORCID, IGSN, DataCite, DataTypes (e.g. <http://typeregistry.org>) or others.

**Semantic** -concepts and knowledge graphs to link information.

Consistent semantic concepts, e.g. vocabularies or ontologies, should be used in data infrastructures. They should harmonize with already existing, **community-agreed, semantic systems** to form a **global knowledge graph**.

**Container** - machine-actionable digital objects (FDO)

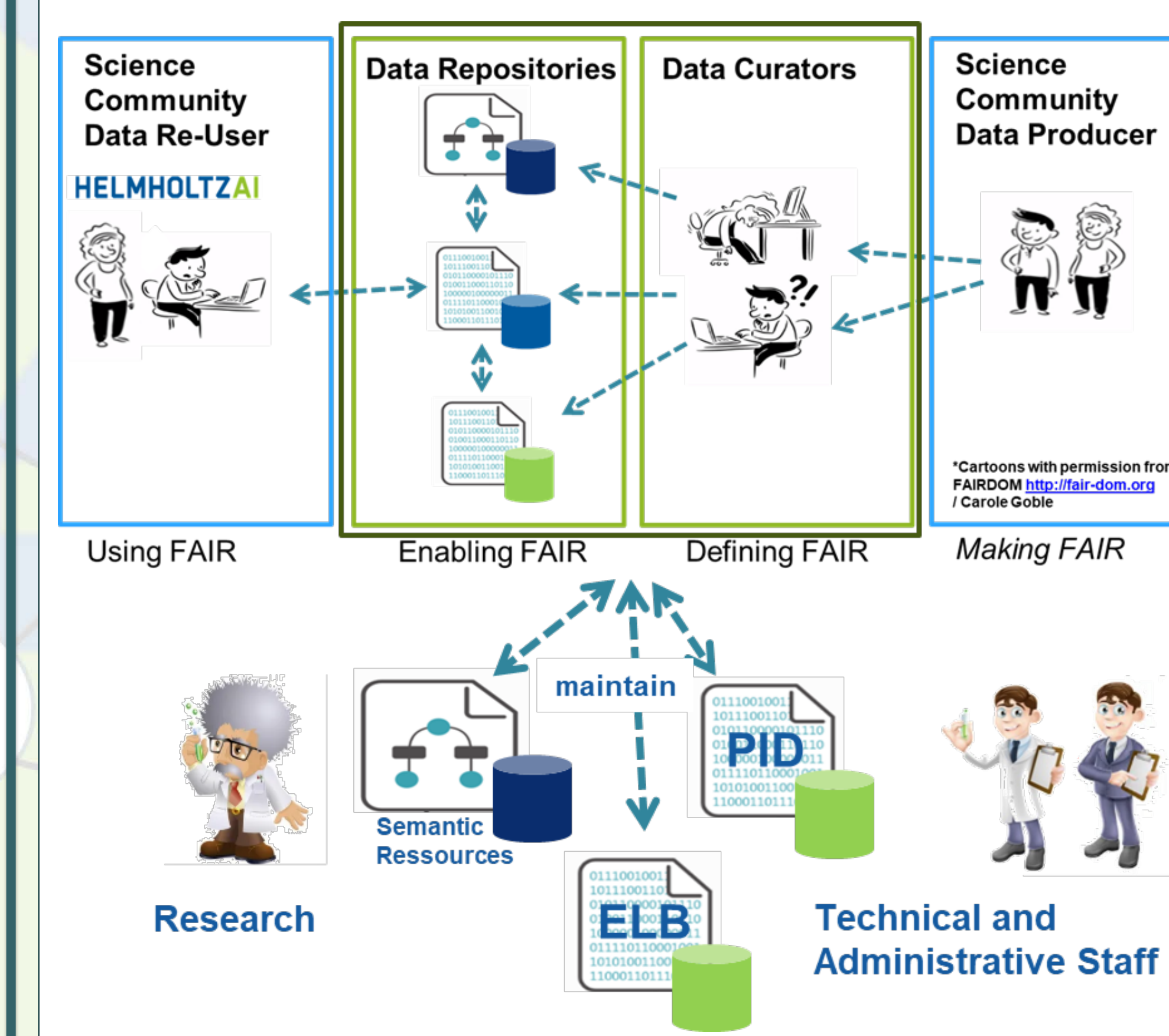
Metadata must be packaged in data containers e.g. **FAIR Digital Objects**. This allows the metadata to be **interpreted by machines** and will greatly enhance possibility to automatically merge datasets.

**Interfaces** - well-documented standardized interfaces and APIs

Standardized interfaces and interoperable APIs will allow **queries across multiple repositories**, independent of the domain. It will allow to construct a knowledge graph and **well formed common search index** of all data in the dataspace.

## 3 Who does what? Roles and Responsibilities

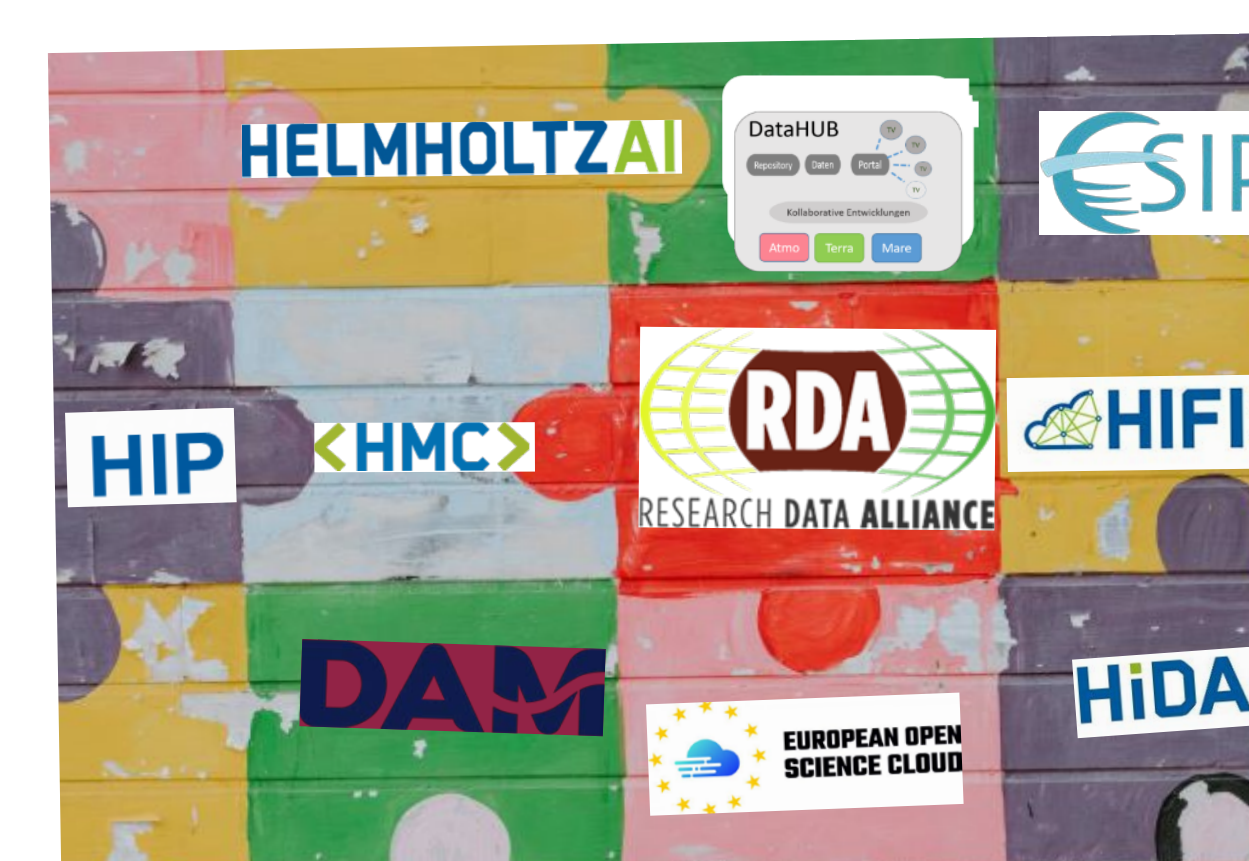
The actors implementing the data space



Responsibilities implementing the blocks

Concept	Lead community	Partner communities
PID metadata	Data curators, DIS	researchers
Semantic concepts	Data providing and re-using researchers	Data curators, DIS
Data Containers / Types	Data curators	researchers
Interfaces and APIs	DIS	Data re-users

Don't forget the wider data community

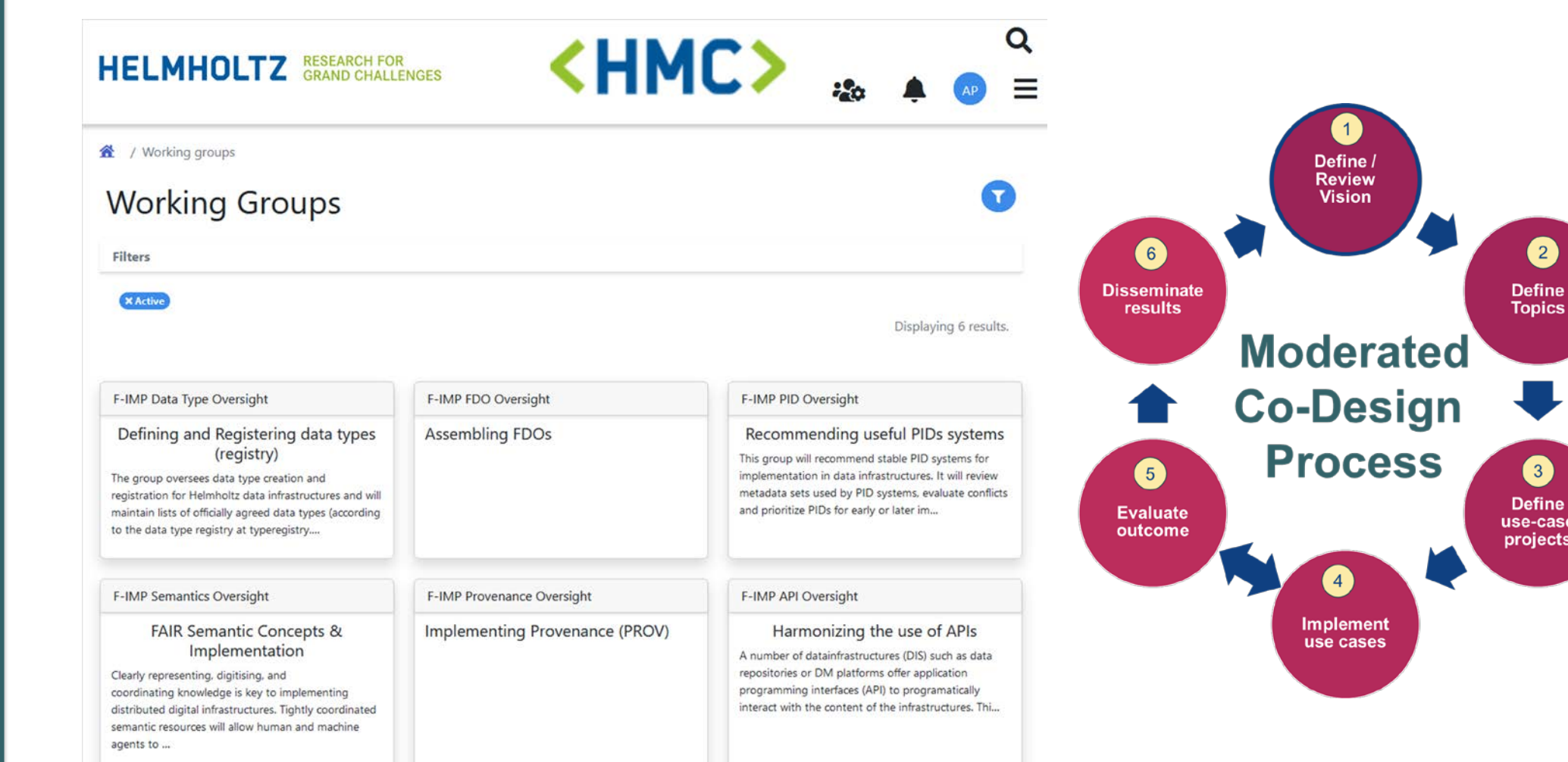
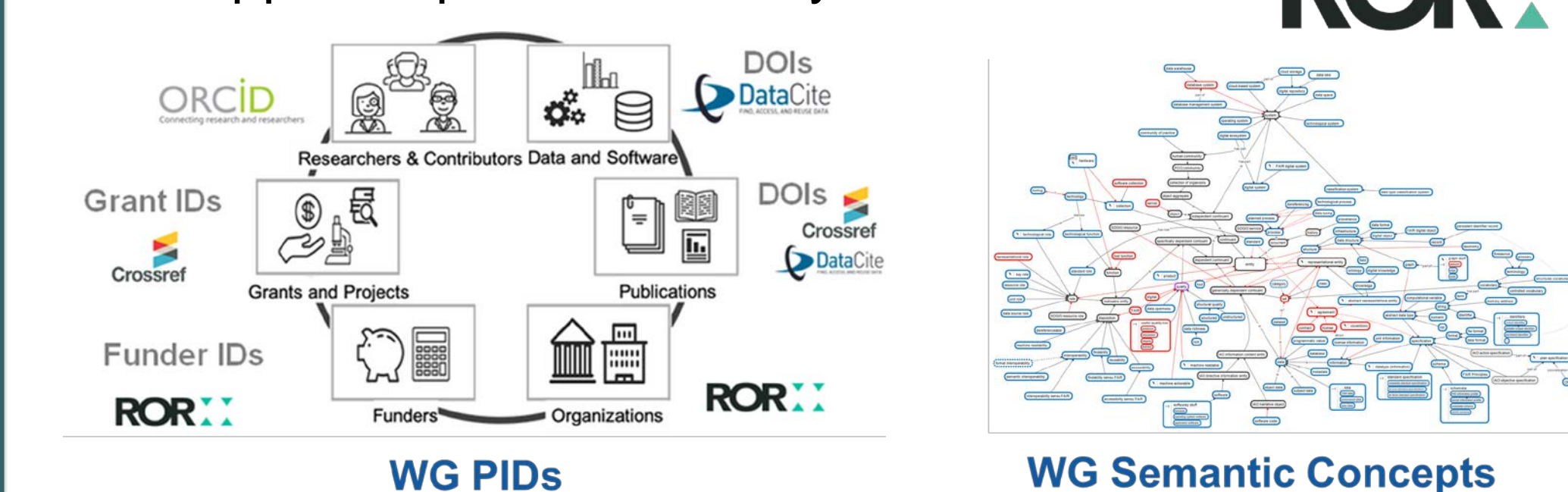


## 4 What's next? Implementing the Data Space!

**Commonly agreed principles** will be discussed and evaluated in a moderated Co-design process on the **community platform**:

Example: Harmonize PID metadata:

1. Evaluate the stability of PID systems.
2. decide and agree:
  - a. who registers
  - b. who maintains
  - c. how implemented
3. document agreement
4. support implementation by other DIS.



Further Information and Contact:



**HMC Earth and Environment**  
E-mail: [hmc-hub-ee@geomar.de](mailto:hmc-hub-ee@geomar.de)  
[www.helmholtz-metadaten.de](http://www.helmholtz-metadaten.de)