Development of recommendations for the implementation of semantic artifacts in HMC Earth and Environment

Dorothee Kottmeier[1] // Emanuel Söding[2] // Andrea Pörsch[3] // Stanislav Malinovschii[2] // Yousef Razeghi[4] // Sören Lorenz[2]
0000-0002-4263-4234 0000-0002-4467-642X 0000-0003-4502-6223 0009-0002-9792-6768 0000-0002-0007-630X 0000-0001-8577-6614

[1] Helmholtz Centre for Polar and Marine Research - AWI
[2] GEOMAR Helmholtz Centre for Ocean Research Kiel
[3] Helmholtz Centre Potsdam GFZ
[4] Helmholtz Centre for Environmental Research – UFZ

Introduction

The HMC Earth and Environment Hub is dedicated to establishing a framework for the semantic interoperability between the involved research data infrastructures in the field. Standardizing and semantically annotating metadata and harmonizing existing semantic resources are crucial for bridging the gap across diverse and complex data sets. Our goal is to work with the community to develop recommendations for implementing semantic concepts that meet the community's needs.

Establishment of the "Metadata – Semantics" working group

- Regular meetings with interested members from the Helmholtz community.
  - Work package 1: Instruments & methods
  - Work package 2: Observable properties
- Detailed analysis of challenges and common development of solutions and recommendations
- Communication through E-mail and the HMC community portal.
- Documentation of results in the HMC E&E Wiki.

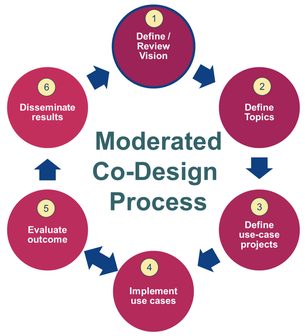
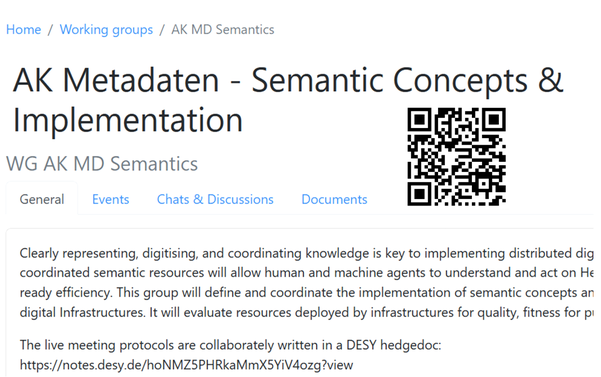


Fig. 1: Working procedure of the working group



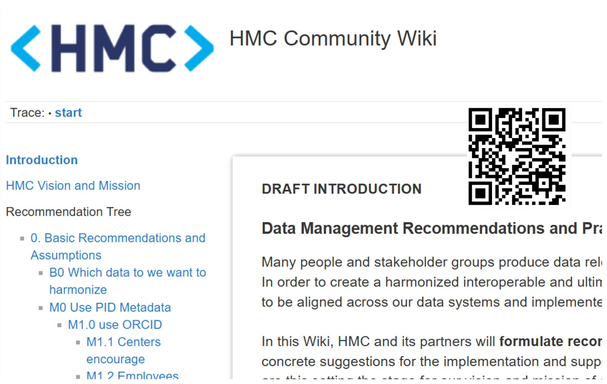


Fig. 2: Screenshots of HMC Community portal and the HMC Community Wiki

Identified challenges

- Despite the wide range of controlled vocabularies (CVs) in use, there is no consensus on the most suitable options. Many CVs are inadequate to meet all requirements.
- More support is needed to evaluate existing semantic resources in terms of their "FAIRness," sustainability, quality, relevance to stakeholders, utility, and user-friendliness etc.
- Metadata typically consists of diverse components, necessitating a detailed examination of which elements should be included in data descriptions and annotations (cf. I-ADOPT framework).
- Metadata classification is often ambiguous and prone to errors.
- The integration of data across the existing data infrastructures in the field is hindered by insufficient transparency in the use of CVs. User guidance—such as in interfaces or metadata naming (e.g., predefined syntax)—is still partially lacking.

Recommended solutions

1. A list of recommended vocabularies and a guideline that provides advice on the evaluation of existing semantic resources should be created.
2. Metadata should be decomposed into components using a consistent approach, with each component mapped individually. A common syntax should guide naming conventions, and fair controlled vocabularies should be employed for semantic linking whenever possible.
3. If multiple external vocabularies can be used for mapping metadata, it should be clearly defined which vocabulary should be used with priority. Even if specific needs aren't fully met, the focus should be on establishing a common framework.
4. Internal vocabularies should be published and contextualized through mapping to external fair vocabularies. If needed, the research community should collaborate to develop a common vocabulary.

Conclusions

- The "Metadata – Semantics" working group unites Helmholtz Earth and Environment community members to tackle semantic interoperability challenges.
- Prioritizing metadata standardization and fair controlled vocabularies is key to enhancing semantic linking
- Ongoing collaboration on a common vocabulary and clear metadata guidelines is vital for better data integration

Further Information and Contact:
HMC Earth and Environment
E-mail:
hmc-hub-ee@geomar.de
www.helmholtz-metadaten.de