Development of a Robust Framework for Seamless Semantic Interoperability in Earth and Environmental Research

Dorothee Kottmeier[1] // Andrea Pörsch[2] // Yousef Razeghi[3] // Stanislav Malinovskiy[4]// Sören Lorenz[4]//Emanuel Söding[4]

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

Introduction

The HMC Earth and Environment Hub is dedicated to establishing a robust framework for seamless semantic interoperability 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. Increasing semantic interoperability makes it possible to connect data sets from different repositories using common metainformation and thus support a harmonized common FAIR data space.

Current activities

1. Metadata Prioritization

We assess elements essential for meaningful data connections. Our intention is to determine which metadata should be prioritized for successful semantic linking through annotation.
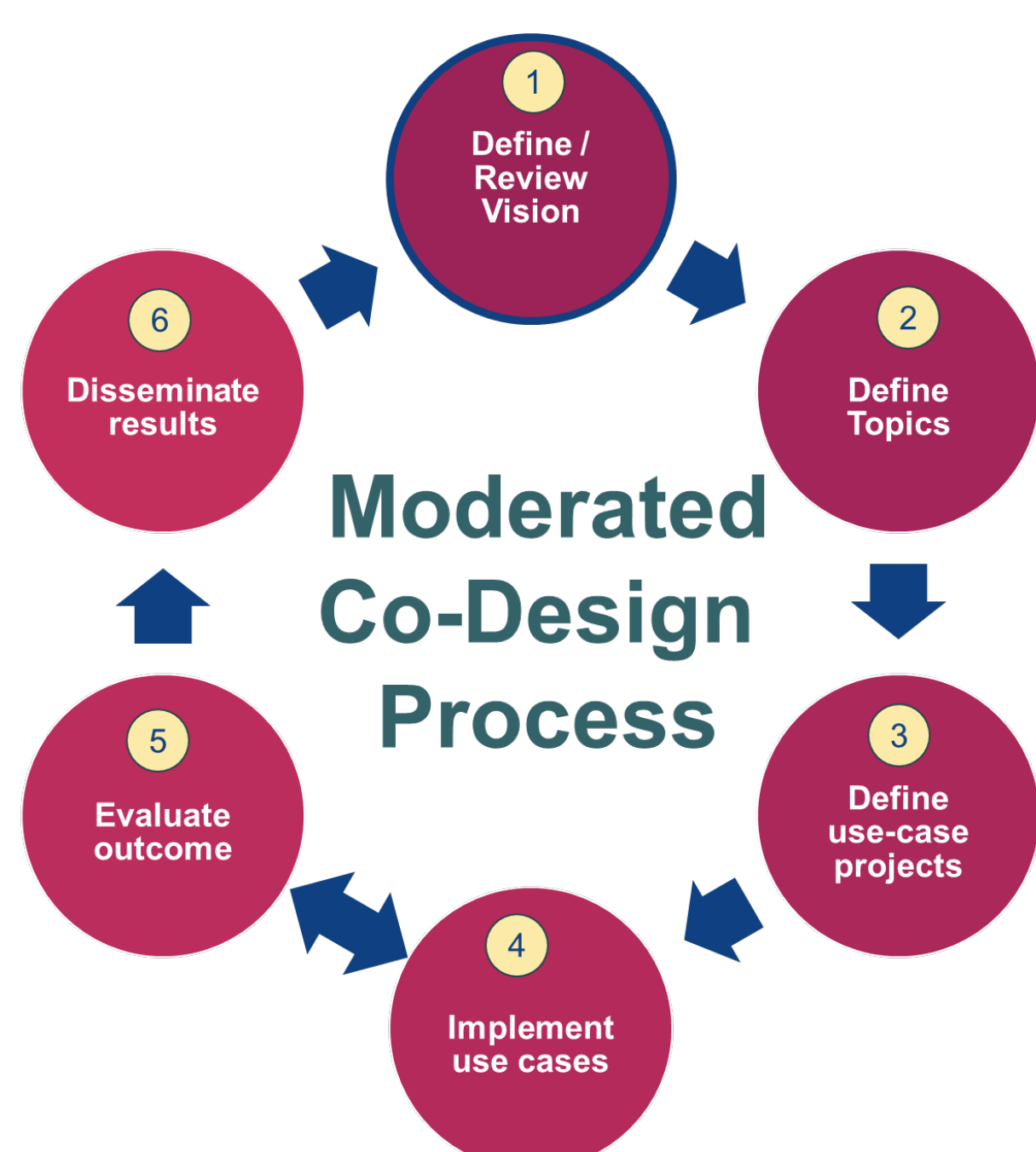
2. Recommendations

We evaluate existing vocabularies and services and define sustainability and recommendation criteria for their use. We suggest crucial implementations for semantic interoperability.

3. Networking

Building a unified data space requires consensus at national and international levels. We collaborate with networks, aligning processes, and adapting services to community needs

4. Community collaboration

Facilitating community collaboration is vital for harmonization. We moderate processes like agreeing on metadata schema and vocabulary

Moderated Co-Design Process

1 Define / Review Vision
2 Define Topics
3 Define use-case projects
4 Implement use cases
5 Evaluate outcome
6 Disseminate results

Status quo and demands of the community

To find out about the state of our infrastructures, we compiled a list of all **controlled vocabularies (CV) currently used in Helmholtz data infrastructures (DIS):**

Parameters/variables/ measured quantities	CF Metadata Convention, Observations Data Model (ODM2, version 2), American metorological society, World meteorological organization, envthes, SeaDataNet Parameter Discovery Vocabulary, Climate and Forecast Standard NamesMarine Strategy Framework Directive indicators 2010/477/EU, Global Change Master Directory Science Keyword variables, BODC parameter semantic model parameter entity names, Marisaurus Thesaurus, BODC parameter semantic model chemical substances, EMODnet Chemistry aggregated parameter names, Chemical Entities of Biological Interest (ChEBI), The enviromental ontology (ENVO), Phenotype and Trait Ontology (PATO), World Register of marine Species (WoRMs), ITIS, SeaDataNet Parameter Discovery Vocabulary, BODC Parameter Usage Vocabulary, MEDATLAS Parameter Usage Vocabulary, and more
Units	BODC-approved data storage units, qudt
Instruments/Devices/ platforms/methods	SeaDataNet device categories (L05), SeaVoX Device Catalogue (L22), American metorological society, Observations Data Model (ODM2, version 2), coastMap method list
Licenses & data access & restriction	Software Pacakge Data Exchange, SeaDataNet data access mechanisms, SeaDataNet Data Access Restriction Policies
Keywords & further sample/data descriptions	Global Change Master Directory (GCMD) Keywords, gemet, ISO Topic Categories, Software Pacakge Data Exchange
Topics	ISO 19115 Topic Category
Sphere names	BODC parameter semantic model sphere names
Vessels & platforms	ICES Platform Codes, SeaVoX Platform Categories (L06),
Countries, locations etc.	International Standards Organisation countries, SeaDataNet Ports Gazetteer (C38)
Coordinates & data	SeaDataNet geographic co-ordinate reference frames, SeaDataNet depth measurement reference planes
Format (data files)	SeaDataNet data transport formats
Organization name	Research Organization Registry

Conclusions:

- Despite a wide range of vocabularies utilized, consensus on the most suitable ones is lacking. Existing CVs are partly insufficient to fulfill all demands
- Measurement variables and other metadata encompass diverse components, necessitating a more detailed examination of which components they should consist of (cf. i-adopt framework)
- Coordination is essential for selecting controlled vocabularies and determining mapping strategies. This process should consider both internationally recognized controlled vocabularies and the specific needs of the community
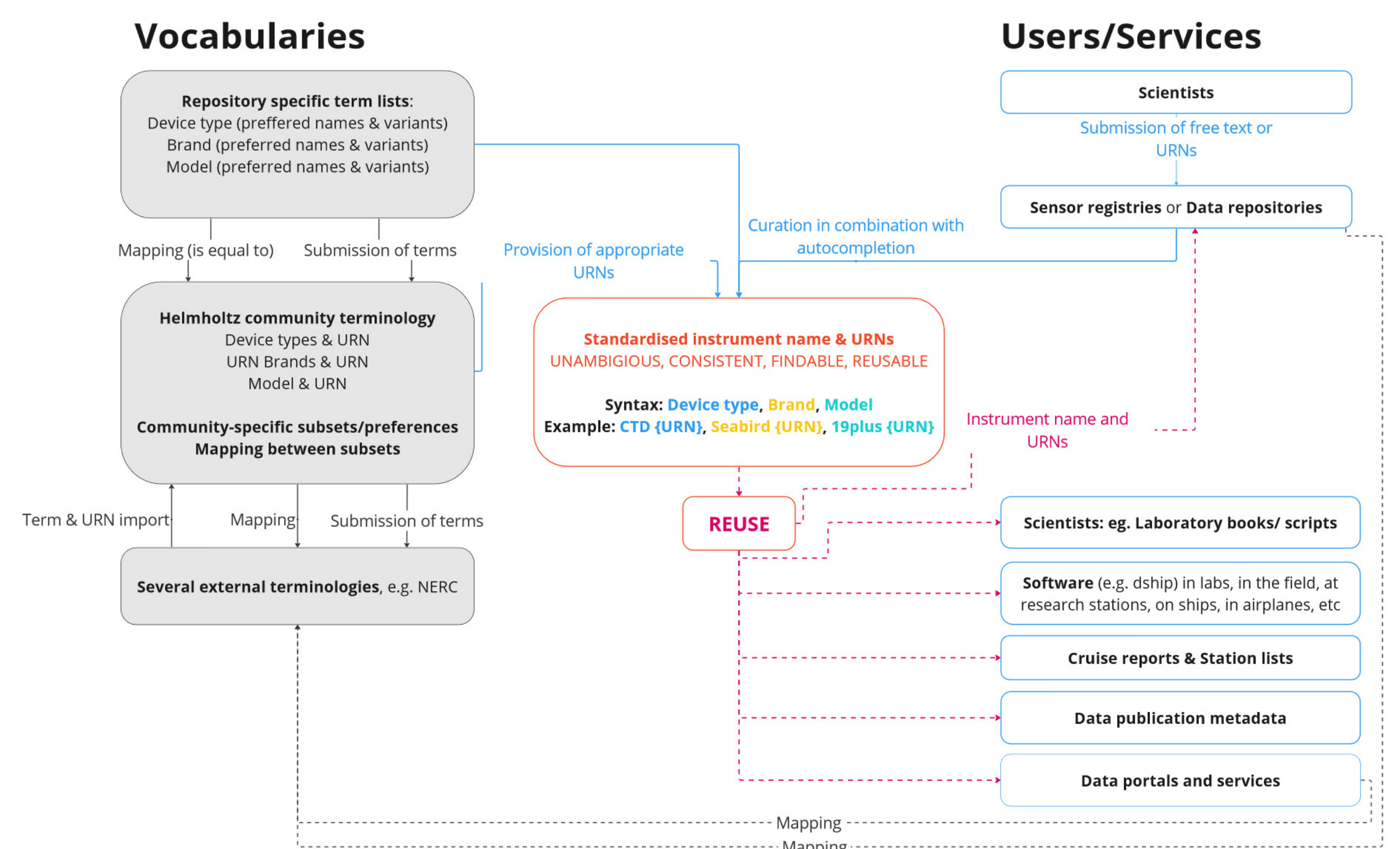
Proposal for a workflow for the use and maintenance of vocabularies within the Helmholtz Earth & Environment community using the example of devices. Objective: Regardless of the service used, the same syntax or combination of URNs for specifying types, brand, and model should be used within a community

Together we can make FAIR happen!
Contact us for more info at
hmc-hub-ee@geomar.de