Contribution ID: **24** Type: **Talk**

A prototype platform for mapping heterogeneous metadata of the three domains health, environment and earth observation: the MetaMap³ project

Wednesday 11 October 2023 10:10 (20 minutes)

The environment plays an increasingly important role for human health and efficient linkage with environmental and earth observation data is crucial to quantify human exposures. Currently, there are no harmonized metadata standards for automatic mapping. This project aims to facilitate the linkage of data of different research fields by generating and enriching interoperable and machine-readable metadata for exemplary data of our three domains Health (HMGU), Earth & Environment (UFZ), and Aeronautics, Space & Transport (DLR) and by mapping these metadata so that they can be jointly queried, searched and integrated into HMC.
After reviewing several standards, strategies and tools, we developed an approach to align our metadata to a common structure and format. We identified spatial and time coverage as the main mapping criteria. For the environmental metadata and the epidemiological metadata that have a spatial component (study centers, recruitment districts) we converged to the standard ISO 19115 and to the eXtensible Markup Language (XML).
Additionally for the health domain, we reviewed several metadata standards for health data but currently available standards were set up mostly for medical or clinical data. Thus, they only fit our epidemiological cohort data in a very limited way. Nevertheless, we started to standardize and enrich the metadata of the LISA birth cohort study by supplementing more than 15,000 variables with Maelstrom categories and subcategories. Of these, approximately 900 entries include an ICD-10 code.
Moreover, we compiled a decision matrix to guide the selection of a suitable application to upload and store the harmonized metadata. Building on this conceptual work, we identified a catalog (GeoNetwork) to store a subset of our cross-domain metadata and set up a test instance on a HMGU server and uploaded the first metadata.
We are currently populating the mapping platform with further metadata and testing the full functionality of the tool, especially the filtering and search tools options of the application to enable the intended mapping. By the end of the project, we plan to release the platform to HMC and other researchers working in thematically related fields.

Please assign your contribution to one of the following topics

Data interoperability through harmonised metadata and interoperable semantics

Please specify ”other” (stakeholder)

In addition please add keywords.

Health, environment, earth observation, metadata mapping, metadata catalog

Please assign yourself (presenting author) to one of the stakeholders.

Researchers

Primary authors: Dr DALLAVALLE, Marco (Helmholtz Zentrum München GmbH - German Research Center for Environmental Health, Institute of Epidemiology, Neuherberg, Germany and Chair of Epidemiology, IBE, Faculty of Medicine, LMU Munich); GEY, Ronny (Helmholtz Centre for Environmental Research –UFZ, Research Data Management (RDM), Smart models and Monitoring, Leipzig); STAAB, Jeroen (German Aerospace Center (DLR),

German Remote Sensing Data Center, Geo-Risks and Civil Security, Oberpfaffenhofen, Weßling, Germany and Geography Department, Humboldt-University Berlin, Berlin, Germany); Dr STANDL, Marie (Helmholtz Zentrum München GmbH - German Research Center for Environmental Health, Institute of Epidemiology, Neuherberg, Germany); Dr BUMBERGER, Jan (Helmholtz Centre for Environmental Research –UFZ, Research Data Management (RDM), Smart models and Monitoring, Leipzig); Dr TAUBENBÖCK, Hannes (German Aerospace Center (DLR), German Remote Sensing Data Center, Geo-Risks and Civil Security, Oberpfaffenhofen, Weßling, Germany and Institute for Geography and Geology, Julius-Maximilians-Universität Würzburg, Würzburg, Germany); Dr WOLF, Kathrin (Helmholtz Zentrum München GmbH - German Research Center for Environmental Health, Institute of Epidemiology, Neuherberg, Germany)

Presenter: Dr WOLF, Kathrin (Helmholtz Zentrum München GmbH - German Research Center for Environmental Health, Institute of Epidemiology, Neuherberg, Germany)

Session Classification: Parallel Track 2

Track Classification: Facilitating connectivity of research data: Data interoperability through harmonised metadata and interoperable semantics