



Contribution ID: 122

Type: POSTER&PITCH

Connecting information across repositories – a keyword-based approach

Monday 4 November 2024 15:00 (1 hour)

Knowledge Graphs help to connect and organize information from different sources and entities. They can be used to apply advanced search and filtering techniques on very large datasets and reveal connections and dependencies across the data. To be useful, however, they require highly uniform and harmonized data sets. So far, most knowledge graphs on scientific data have used bibliographic data to build a network of information. These data are of limited use for scientific purposes because they contain little scientifically relevant information. In order to enhance the scientific usability in the Helmholtz research area Earth and Environment, we aim to identify seven parameters in data sets and build a knowledge graph from it:

Measuring Instrument (type, manufacturer, model)

Methodology

Measured Attribute (e.g. sulfur content)

Measured Parameter (e.g. MS Spectrum)

Measured Unit (e.g. velocity)

Measured Object / Medium (e.g. rock sample)

Sample ID (e.g. as IGSN)

DataCite, ISO191XX and schema.org are among the most common standards currently implemented by repositories, to retrieve and exchange metadata. However, most of the mentioned parameters are not yet well documented in the common metadata standards used to export data from repositories. Repositories thus apply very different approaches to include this information within their metadata.

In this poster we discuss our approaches, challenges and successes to harvest this information from several repositories from the Helmholtz Earth and Environment research field. We also discuss the potential to create knowledge graphs from this data, and how the quality of these graphs can be improved. Finally, we present some statistics on the harvested data and make suggestions on how the data can be improved.

Please specify "other"

In addition, please add 3 to 5 keywords.

Knowledge graph, metadata harmonization, PID, Metadata schema, semantics

Please specify "other"

For whom will your contribution be of most interest?

Researchers

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

Data professionals and stewards

Primary authors: PÖRSCH, Andrea (HMC Hub EE at GFZ); KOTTMEIER, Dorothee (HMC E&E @PAN-GAEA/AWI); SÖDING, Emanuel (GEOMAR); MALINOVSCHII, Stanislav (HMC); LORENZ, Sören; RAZEGHI, Yousef

Presenters: SÖDING, Emanuel (GEOMAR); MALINOVSCHII, Stanislav (HMC)

Session Classification: Poster Session B

Track Classification: Connecting research data: 5. Technical solutions for findable and machine-readable metadata