Linked Data Schema Repositories for Interoperable Data Spaces

Stier, S.[1]; Gold, L.[1]; Popp, M.[1]; Räder, A.[1]; Triol, A.[1]

In this talk we discuss the improved interoperability and enlarged tool landscape of data spaces by stronger coupling between semantic web, general web technologies, and programming languages and give insights into a pilot data space for battery research.

In a material science context, metadata models need to cover all relevant levels, including data acquisition, processing, interpretation, publication and documentation and should be subject to harmonization within a data space community. For this purpose, both IDS and GAIA-X propose [1][2] RDF and it's JSON-LD serialization as a common data representation, and consequently RDFS, OWL, and SHACL for data schemas and validation. While the expressiveness of RDF is theoretically unlimited and thus fulfills all requirements, there are practical limitation and barriers in its practical and scaled application.

We therefore propose the closer integration between RDF/SHACL and object orientated programming by a modular hierarchical self-description of data assets by JSON-LD annotated JSON-SCHEMAs in combination with direct coupling with object-oriented programming through a linked data schema repository. In addition, a search index, visual renderings and auto-generated edit forms provide a human-centered interface. A published and maintained python package [3] allows the add-hock generation [4] of data classes from data schemas and vice versa, enabling conformance by design of programmatically generated and used data through an abstract knowledge graph interface. This includes not only pure data classes, but also mapped controller classes that provide corresponding functions/methods with a strong link to data space services/apps via automatically generated APIs.

By providing a JSON-SCHEMA for OWL classes the documentation, visualization and graphical editing of large and linked ontologies also becomes an important use case [5] for both their development and governance. Thereby OWL classes and data schemas are regarded as complementary and linked concept combining their strength in formal logic on the one hand and closeness to concrete implementations on the other hand. With respect to the existing reference architectures this approach integrates the hierarchical concept of GAIA-X with the graph structure of IDS and provides a user and programmer friendly toolset to extend their application to detail/domain levels beyond the core metadata models.[6][7]

[1] https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Position-Paper-GAIA-X-and-IDS.pdf

[2] https://library.oapen.org/handle/20.500.12657/57901

[3] https://github.com/OpenSemanticLab/osw-python

[4] https://opensemantic.world/wiki/Item:OSW659a81662ff44af1b2b6febeee7c3a25

[5] https://onto-wiki.eu/wiki

[6] https://gaia-x.gitlab.io/gaia-x-community/gaia-x-self-descriptions/core/core.html

[7] https://github.com/International-Data-Spaces-Association/InformationModel