A data-model driven approach for semantic interoperability in scientific software

Thomas F. Hagelien[1], Jesper Friis [1] and Petter Rønningen[1]

[1] SINTEF, NO-7465 Trondheim, Thomas.F.Hagelien@sintef.no, Jesper.Friis@sintef.no, Petter.Ronningen@sintef.no,
https://www.sintef.no/en/

Key Words: *Data modelling, interoperability, technology adoption, scientific software.*

ABSTRACT

The European Material Modelling Ontology (EMMO[1]) formalises knowledge representation of materials, modelling and characterization. The emerging technology for enabling semantic interoperability, has not been widely adopted in the industry yet. Feedback from data- and application providers indicate that industrial onboarding is a challenge. In industrial applications, the structure and knowledge about existing data is often well known., and metadata and/or data schemas can be provided with little effort. However, the ability to map this knowledge to an ontology and/or develop a new domain ontology requires special expertise. To construct domain ontologies and/or map information to ontological concepts (i.e., EMMO concepts) requires training on semantic technologies and an understanding of the fundamental idea behind the structuring of the ontology. Here we present a data model that represents the physical perspective i.e., provides a data-representation close to the physical data source. The data model can be included in a representation of the logical perspective i.e., the relationships between multiple data models in a software system. Furthermore, the data model can be enhanced further by mapping its properties and attributes to ontological concepts that describes the information from a conceptual perspective. Importantly, the mapping can be performed in retrospect, by someone other than the original provider of the data model. Furthermore, the syntactical data model representation can be tailored to fit the most technical needs, with formats such as JSON, YAML, XML etc. The data model can also be realized as RDF-triples that maps to concepts in representational frameworks such as EMMO, supporting semantic representation (mapping). The RDF-triple representation allows for the later adoption of the mapping.

We show that it is possible to accelerate the onboarding process by defining knowledge-base- and database agnostic data-models that can be retrospectively mapped to ontological concepts and specific data points.

REFERENCES

[1] Ghedini, E., Goldbeck, G., Friis, J., Hashibon, A. and Schmitz, G.J. European Materials & Modelling Ontology, https://github.com/emmo-repo/EMMO.