Nanosafety data made interoperable using semantic modeling and linked-data knowledge graphs

Ammar Ammar[1,*], Egon Willighagen[1]

[1]Department of Bioinformatics - BiGCaT, NUTRIM, Maastricht University, the Netherlands

Abstract

Achieving data interoperability is a critical challenge in the increasingly complex landscape of the nanosafety field, where ensuring the safe use of nanomaterials is of great importance. The unique properties of nanomaterials, stemming from their size and structure, necessitate comprehensive and standardized data to evaluate potential risks and hazards. One significant challenge lies in the diversity of experimental approaches, measurement techniques and exchange formats employed in nanosafety research [1]. Fortunately, semantic modeling coupled with linked-data knowledge graphs emerges as a powerful solution. Semantic modeling involves structuring data in a way that adds meaning and context to the information, facilitating better harmonization and standardization. Linked-data knowledge graphs take this a step further by establishing relationships between diverse datasets and their metadata, creating a web of interconnected information. That allows for better understanding and seamless data integration and exchange across different domains and applications. Moreover, the semantic approach inherently complies with the FAIR principles (Findable, Accessible, Interoperable and Reusable) [2] and covers several of its sub-principles. Thus, making the data more accessible and reusable for the community.

The semantic model presented in this work adopts several standardized ontologies to describe both the datasets and their metadata. For metadata representation, DCAT [3] and VoID [4] ontologies were used. Moreover, more specialized ontologies were used to represent nanosafety data, namely, NPO [5] for nanomaterial entities and BAO [6] and eNanoMapper [7] ontologies to represent the bioassays, experimental conditions and measured outputs. The model captures two types of assays, toxicity assays and gene expression assays allowing to reveal insights from gene expression signatures at concentrations where a nanomaterial is deemed toxic.

This approach utilizes the RDF Mapping Language (RML) [8] and its extension (YARRML) [9] to represent the semantic model as a set of reusable mapping rules to convert related datasets into a knowledge graph. Then, the knowledge graph can be explored using SPARQL query language. For example, finding relations between up/down-regulated genes and toxicity levels of a nanomaterial in a specific cell line. Another example, enriching the nanomaterials with information about the key events in adverse outcome pathways where they take effect. This can be done using federated queries against external sources like AOP-Wiki [10] and ENM-MIE [10] knowledge graphs. Furthermore, this knowledge representation allows morphing the data model into another one that adheres to different ontologies or vocabularies [11]. For example, using SPARQL, the nanomaterial data can be remapped to a Schema.org compliant model which then can be used to annotate relevant web pages with semantic metadata.

In summary, making data interoperable through semantic modeling and linked-data knowledge graphs is essential for advancing our understanding of nanomaterials' safety profiles. This approach not only enhances data FAIRness but also promotes seamless integration and exchange of information, fostering a more interconnected ecosystem for diverse applications.

Keywords

Semantic Modeling, Knowledge Graph, Linked Data, Nanosafety Data, FAIR, Interoperability,

SWAT4HCLS, February 26–29, 2024, Leiden, The Netherlands

a.ammar@maastrichtuniversity.nl (A. Ammar); egon.willighagen@maastrichtuniversity.nl (E. Willighagen)
0000-0002-8399-8990 (A. Ammar); 0000-0001-7542-0286 (E. Willighagen)

CEUR Workshop Proceedings (CEUR-WS.org)