

OpenSemanticLab: A Reference Implementation for Open Semantic Materials Research

Stier, S. (Speaker)^{1*}; Gold, L.¹; Popp, M.¹; Räder, A.¹; Triol, A.¹

¹Fraunhofer Institute for Silicate Research ISC, Würzburg

*simon.stier@isc.fraunhofer.de

In materials science, complex relationships exist between the properties of materials and their composition and processing. Therefore, **digital transformation and acceleration** in this domain represents a particularly **big challenge**. Although it is generally agreed that data must be linked by means of semantics and ontologies to form holistic data spaces, there is still a lack of suitable tools for integrating the necessary structures into the **everyday work of scientists**.

This challenge must be addressed with a broad-based strategy that closely links activities at all relevant levels, including automated lab infrastructure, machine-readable specification and documentation of scientific processes and the harmonization of generated data structures in accordance with international standards efforts to build common data spaces (c.p. **IDS, GAIA-X**).

With **OpenSemanticLab (OSL)**[1] we developed a reference implementation to fulfil this wide spectrum of requirements. Core of the resulting OpenSource solution architecture is the central web data platform that links **people (knowledge)**, **machines (data)** and **algorithms (AI)** equally by supporting both unstructured and structured content in an integrated and machine readable form (see Fig. 1).

OSL does not specify a fixed data model. Instead, data schemas based on **JSON-SCHEMA**[2] can be individually installed from an **open repository**[3], imported from an existing **ontology (OWL, SHACL)**, or created and extended by the user. Regardless of the way a schema is created and updated, both **object oriented** program code[4] and visual interfaces for searching, viewing and editing the corresponding **JSON-LD**[5] datasets are always generated automatically for the user. Thus, OSL makes it possible to apply semantic structures and ontologies in the background, including their mapping to **Large Language Models (LLM)**, without the user needing in-depth knowledge. Conversely, the user can also use the platform to provide pre-structured knowledge for the creation and extension of domain ontologies.

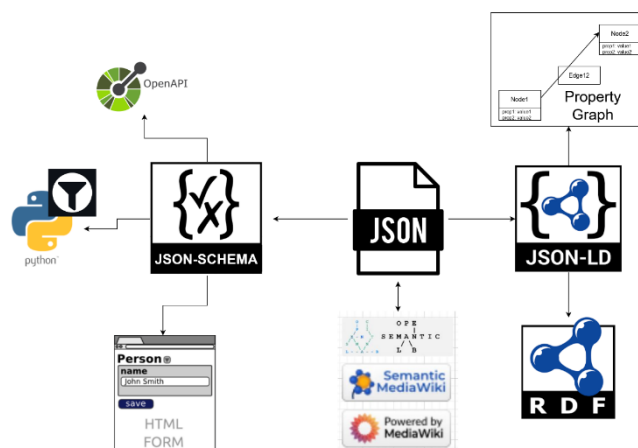


Figure 1. The OpenSemanticLab Software Stack utilizing Semantic MediaWiki as an interactive document store for both unstructured (wikitext) and unstructured data (JSON). JSON data is shaped by a JSON-SCHEMA graph, allowing visual/form editing, API and code generation. JSON-LD context embedded in the schemas allows a mapping to a build-in RDF-store and prepares for usage of property graphs in the future.

This talk will provide an overview about the features and their technical implementation, which can be followed on a **public demo instance**[6-7].

References

- [1] <https://github.com/OpenSemanticLab> / <https://doi.org/10.5281/zenodo.8290365>
- [2] <https://json-schema.org/>
- [3] <https://opensemantic.world>
- [4] <https://doi.org/10.5281/zenodo.8198098>
- [5] <https://json-ld.org/>
- [6] <https://demo.open-semantic-lab.org>
- [7] <https://onto-wiki.eu/wiki>