Developing Ontologies in Materials Science

By Vikki Cantrill[1], Anne de Baas[1], Otello Roscioni[2], and Gerhard Goldbeck[1*]

[1] *Goldbeck Consulting Ltd, St John's Innovation Centre, Cambridge CB4 0WS, UK*

* Corresponding author: gerhard@goldbeck-consulting.com

The OntoTrans Project received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 862136

Developing Ontologies in Materials Science

Collaboration promises a pathway for materials industries to harmonise and put meaning into data

Data problems

The complexities of materials science research alongside the ever-growing volumes of associated data highlights the importance and need for efficient data management and data exchange.[1,2]

In April 2024, Goldbeck Consulting Limited, in partnership with the European Union's Horizon 2020 project OntoTrans, hosted a one-day workshop to look into semantic data and knowledge management for chemicals and materials industries.

The workshop covered industrial challenges and requirements, research and ontology developments, and contemporary solutions to implementation. Panel discussions at the event focussed on the latest advances in semantic knowledge management and collaborative, industrial approaches for the future.

Discussing solutions

A number of domain-level ontologies within materials science have emerged as valuable tools to describe properties, processes and structures, and provide data integration, interoperability, and knowledge discovery. But standardisation and harmonisation of such ontologies — to make them interoperable across data systems and research fields — remains a challenge.

Barriers

Although advances in data management have evolved considerably in recent years, semantics and semantic technologies — that attach meaning to data — could unify data and offer augmented analyses, recommendations and modelling. This level of depth and complexity (known as active meta data) cannot be reached through IT or AI approaches.

Today, few materials science industries use completely integrated data systems. Yet, as the volume of data grows so does an enterprises' interest in managing it; but there are often limited or insufficient central resources allocated to do so, including staff with specialised competences and expertise.

Unlike the pharmaceutical sector, materials science has few regulatory requirements for data management that would encourage and push innovative, ontology development. The diverse nature of materials science means that, even though some industries use technologies and taxonomies developed within subfields, no over-arching ontology exists to connect them. Many research and industrial enterprises agree that a top-level ontology and tools to allow the development of sub-field branches (vocabularies, taxonomies and domain ontologies) are much needed.

Dedicated ontology

The Elementary Multiperspective Material Ontology (EMMO) provides a conceptualisation framework for materials science and related applications. This framework offers a standard approach to represent data and knowledge from chemistry details to products, as well as materials and manufacturing processes, properties and models. The EMMO can be regarded as a top-level ontology because it

The OntoTrans Project received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 862136

covers all concepts from low level to high level yet is accessible to users with limited or no philosophical background.

The EMMO has been substantially developed and exploited within the OntoTrans project to address a number of industrial challenges, and the know-how acquired will be carried forward into other projects both publicly and privately funded in the near future.

Clear benefits

Adopting semantic technologies, including ontologies and taxonomies, brings numerous benefits for both the data user and business strategists. Applied ontologies can integrate diverse data sources, so relevant data can be accessed quicker and explored to uncover new insights. This enables broader data analytics, even by non-specialists, and provides opportunities to integrate with external data, which supports efforts towards open data and innovation collaborations.

At the enterprise level, interoperable and accessible data leads to better informed decisions that create business value. And because data is retained and can be shared across all parts of the business, then the risks associated with development are minimised, which lead directly to time and cost savings.

Future success

Despite the clear benefits of semantic technologies, enterprises cite a number of barriers to adopting these advanced systems. Currently, there is some hesitancy within organisations to make business-wide changes, changes that require substantial time and expertise to implement.

In the past, when a problem with data and its analysis has arisen, new data platforms were adopted to solve the concern at a local level. The use semantic technologies means a transition away from working with data when needed, towards an attitude whereby people realise a fundamental approach to knowledge is needed. Furthermore, organisations and institutions need to be sure that they are making the right choices when constructing or adopting an ontology that works for them.

Here, the EMMO provides a reliable way forward for materials-based industries, with a coordinated, controlled approach from area experts with a dedicated interest in promoting best practice for development, quality, management, and curation of ontologies and knowledge graphs. Additional collaborations between experts in ontology development, industry, and semantic technology providers also help pave the way.

A strong, top-level ontology can reduce the variability of data models and can help show industry how data goes into a model and produces meaningful insights. The link between data, and its integration with both other data and modelling methods thus becomes clear. Similarly, the link between business, IT and contributions from data/knowledge specialists is highlighted. Once the needs of industry are addressed, what is at stake is clear, and the potential gains from both a business and profession perspective are understood, then the barriers to adopting semantic technologies are lowered.

As a next step, a Semantic Materials Consortium, a collaborative effort between EMMO developers and materials industries has been proposed.

1. Zhang, X., Zhao, C. & Wang, X. A survey on knowledge representation in materials science and engineering: An ontological perspective. *Comput. Ind.* **73**, 8–22 (2015).

2. De Baas, A. *et al.* Review and Alignment of Domain-Level Ontologies for Materials Science. *IEEE Access* **11**, 120372–120401 (2023).

The OntoTrans Project received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 862136