Operationalising materials modelling workflows in industrial R&D – a benefits analysis

By Francesca L. Bleken[1], Michele Brigadoi[2], Alessandro Calvio[2], Vikki Cantrill[3], Jesper Friis[2], Gerhard Goldbeck[3*] and Otello Maria Roscioni[3].

[1] *SINTEF Industry, Strindvegen 4, 7034 Trondheim, Norway*

[2] *Department of Computer Science and Engineering, University of Bologna, Italy*

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

* Corresponding author: gerhard@goldbeck-consulting.com

Acknowledgement

This work has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 953167 (OpenModel).

Contents

Executive summary

For materials science and manufacturing, the use of semantic technologies is expected to be a key priority over the next 5–10 years as industries become increasingly reliant on data-driven decision-making, innovation, and sustainability practices. Indeed, semantic technologies are transforming materials science innovation by enabling efficient and effective integration of data and modelling in industrial research and development environments.

This article discusses the importance of materials modelling workflows within enterprises. It delves into the benefits, challenges and practicalities of implementing a semantic interoperability platform that integrates materials modelling more deeply into the enterprise, improves its efficiency and supports collaboration. The article highlights the work completed in this area as part of the European Union's Horizon 2020 project OpenModel.

Introduction

Efficient and effective materials innovation is a key priority to accelerate the development and manufacturing of advanced materials that deliver both safe and sustainable solutions to address the multitude of societal needs, from health to energy.

Digital technologies are seen as crucial to enhance efficiency (doing things faster) and effectiveness (doing the right thing)[1]. For example, physics-based modelling can provide data and insights leading to better research and development (R&D) decisions. Machine learning (ML) and artificial intelligence (AI) methods help to further speed up modelling and to work with large amounts of data.

Semantic technologies support the organisation of data and knowledge and are increasingly being used in combination with AI to gain targeted insights[2]. Hence, all three components (physics-based modelling, ML/AI and knowledge systems/semantic technologies) need to work together to enable data understanding and achieve the required breakthroughs in materials innovation. In particular, the complexities of materials science research and the large volumes of data generated highlight the importance and need for elegant data management solutions[3].

Indeed, the power of combining data organisation and modelling was showcased with the awards for both the Nobel Prize 2024 in Physics and Chemistry for work that uses ML with artificial neural networks and for the use of computational approaches to protein design and structure prediction, respectively. Similarly, knowledge management within materials research, manufacturing, and engineering is developing rapidly[4,5].

In this paper, we consider how to facilitate the use of materials modelling in industrial R&D by means of integrating its operations, including model and data management, into an overall knowledge management framework based on semantic technologies.

In the following sections, we will describe some approaches in materials modelling workflows in industrial settings and show how semantics and modelling can be used to better connect all parts of an organisation. We will then give an overview of the OpenModel project[6] and some of its applications, as well as analyse the benefits of using semantic technologies in materials science today alongside future perspectives.

Materials modelling in industrial settings

Since its inception and use by early adopters in the 1990s[7,8] computational chemistry and materials modelling have become well-established activities across a wide range of industries and applications. Increased reliance on these techniques is thanks to a combination of the much greater computational power, more efficient algorithms, better validation, and user-friendly software tools — including solutions that help create, manage, deploy and execute complex workflows — typically needed for industrial applications.

However, in many cases, the maturity of materials modelling in terms of its overall integration into R&D function and workflows remains a work-in-progress. Previously, we reported[9] some strategies and approaches for industry to engage in materials modelling and found that the maturity of modelling functions had not generally reached what could be considered a managed, or even optimized, level[9], and that the maturity level achieved strongly depended on an organisation's experience in the area.

The move to establish and put into operation complete materials modelling workflows within an industrial R&D environment is complex. It involves setting up and/or integrating computational tools, data management systems and automated processes. Overall, these changes aim to enhance and speed up the rate of design, discovery and development of materials. This approach can streamline modelling processes, making them quicker, more efficient and fit to solve the complex materials innovations challenges of the future.

Operationalising materials modelling workflows in industrial R&D involves easier access and use by non-experts, and better integration of modelling (input and output) data with other information systems. It requires the sharing of key 'control variables' within modelling operations across different R&D functions, so that information and data is provided in an understandable and usable way, i.e. as FAIR data[22,23].

Operationalising, typically by means of deploying web-apps, has also been referred to as democratisation, whereby modelling and simulation tools can be accessed by a broad range of users and not just domain experts. By being accessible to everyone, the barriers to using modelling tools are lowered and can be used by non-specialists to help drive innovation. However, in these approaches, the involvement of different types of users depends on the deployment provided to them. Users hence remain somewhat passive recipients and are not able to actively interact with integrated data systems.

Here, we will outline a more inter-connected, FAIR data, approach, which involves better conceptual interlinking of materials modelling with other R&D processes and innovation challenges. This integration of concepts, perspectives and understanding is achieved by using the Elementary Multiperspective Materials Ontology (EMMO)[10], a reference framework for concepts ranging from general systems and processes to materials science, modelling, characterisation and application details.

Enterprise modelling

To achieve our objective of better connecting materials modelling activities into other organisational processes, we first take a step back to enterprise modelling as a whole, to describe processes, information and resources in an organisation at different levels of granularity. Figure 1 outlines a simple, architectural model of processes in an enterprise by defining four levels of organisational processes in detail.

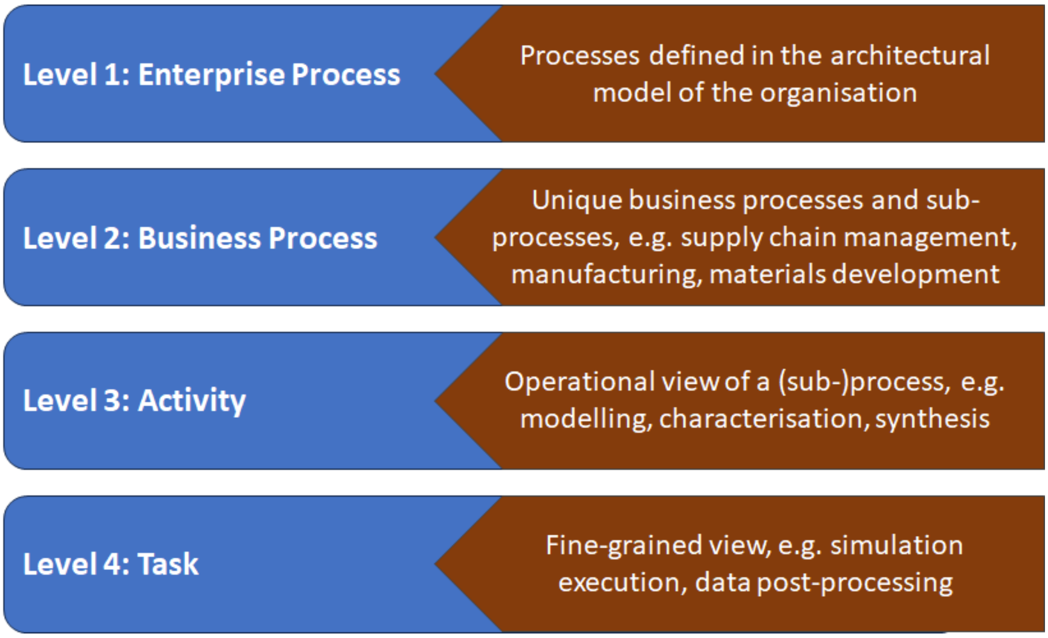

Figure 1. An enterprise model detailing the levels of processes and tasks within an organisation. Successful organisations have optimized processes within each of the levels and well-designed, efficient exchange of information between the levels.

Continuous improvement and optimization of processes and workflows requires an enterprise to work towards implementing an ontology or enterprise data model fed by an interconnected system of process models at every level. Ontologies connect materials and material systems to be investigated with information sources (data/information) that include materials simulations alongside the abovementioned organisational processes with the materials and material systems.

For example, at Level 4 (the task-level view for the process), a simulation expert uses dedicated simulation software and workflow tools. However, they operate on models that are managed as organisational assets (against certain parameters) that can be relevant and included at Level 3/2. For example, a Level 4 simulation uses data and produces data that support activities that take place at higher levels. In turn, the capabilities of the models and simulations are visible to and can be used by staff across different areas, thanks to workflow and data model interconnections.

An in-depth review of software and tools used at different levels is beyond the scope of this article. However, a general tool and standard for modelling, registering, sharing and even executing business processes is the Business Process Model and Notation (BPMN)[11]. The BPMN is open and used in a wide range of free and commercial software tools. It can be used as a graphical tool that can be executed, saved and shared in xml format.

BPMN has also been used to connect high-level modelling workflows to simulate workflow engines, effectively making links and connections between enterprise levels. In the Musicode project[12], for example, BPMN is connected to the MuPIF platform[13]. MuPIF is an open-source, modular, integration platform that combines complex multiphysics simulation workflows by combining existing simulation tools. The platform includes a data management system such that it can be used to develop digital twin representations of physical systems.

However, the challenge with these software systems and tools is the ability to simultaneously connect and manage all processes and related data throughout an organisation in an overarching data model or ontology. Progress in this direction has been made by connecting BPMN to the EMMO ontology[14].

Materials modelling processes

Materials modelling processes are placed at Level 3 and 4 within the enterprise modelling architecture. Organisations that aim to improve efficiency of related processes can nowadays choose from a wide range of software tools that support materials modelling workflows[15].

The state of the art in the field was reviewed by Schaarschmidt et al[16], in particular covering the following tools: AiiDA[17,18], FireWorks[19], KNIME[20], Pipeline Pilot[21], SimStack[22], PyIron[23] and MyQueue[24]. In the section on workflow documentation, the paper highlights the need for interoperability that can take the form of syntactic/format-based approaches that are most common, and semantic interoperability[25] that enables linking data entities by means of a reference data model or ontology. The paper notes that such semantic interoperability "remains an open challenge for the materials modelling domain due to the high heterogeneity, variety, complexity and dynamics of materials".

However, we believe that it is worthwhile to work to overcome challenges. The result will be the creation of more flexible workflow solutions. Furthermore, the data used and created will support better lateral interoperability within the simulation tasks level, and horizontal and cross-functional interoperability in terms of business levels and functions (see Figure 2).

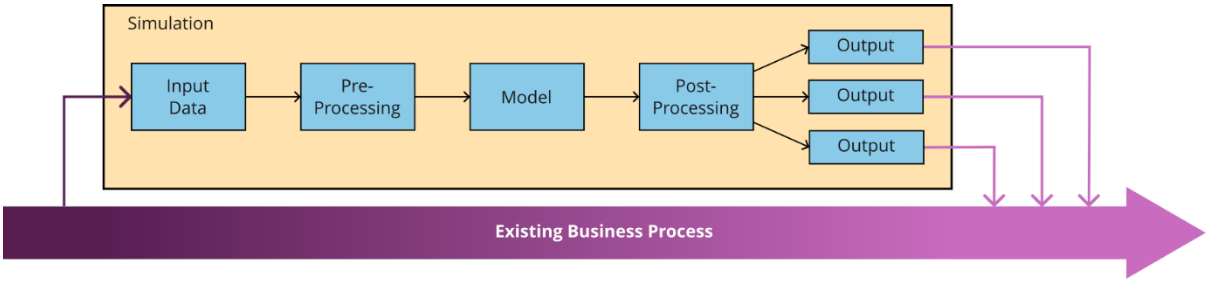

Figure 2. Simulation workflow including its different sub-processes transforming input data to outputs. The meaning of the input and outputs, as well as the meaning of the model should refer to terminologies/ontologies that make sure the simulation can be embedded into higher level business processes (such as "materials discovery" or "quality control").

Here, OpenModel is spearheading such an approach. It provides a semantic framework to describe and manage modelling workflows, built on the EMMO ontology, which ensures interoperability with all types of materials science concepts and data. In terms of specific workflow execution tools, OpenModel connects to the AIIDA workflow software. Details about materials simulations and their execution hence become more connected to the overall R&D challenges.

What is OpenModel?

OpenModel is a blueprint for semantic interoperability in materials modelling and provides a related technology stack, as shown at a high level in Figure 3.

In the semantic layer, OntoFlow supports the user to work at an abstracted level of workflow description and data annotation (see next section for more details). Such a concept has been discussed and implemented previously, e.g. in the abstract workflow description language[26]. In OpenModel, all relevant information, i.e. the concepts, their relations and instance data is handled by means of semantic web technologies and stored in a knowledgebase, implemented in a triplestore.

Separately, the execution of the model is described according to the workflow manager chosen in ExecFlow. An intermediate component (OntoConv) converts the semantic description into executable workflows operating in the syntactic layer. The semantic layer hence acts as a reference system that can, in principle, connect to different syntactic workflow solutions. Currently, in the OpenModel project AiiDA workflow solutions have been implemented. This flexible modular design opens for the integration in other workflow managers.

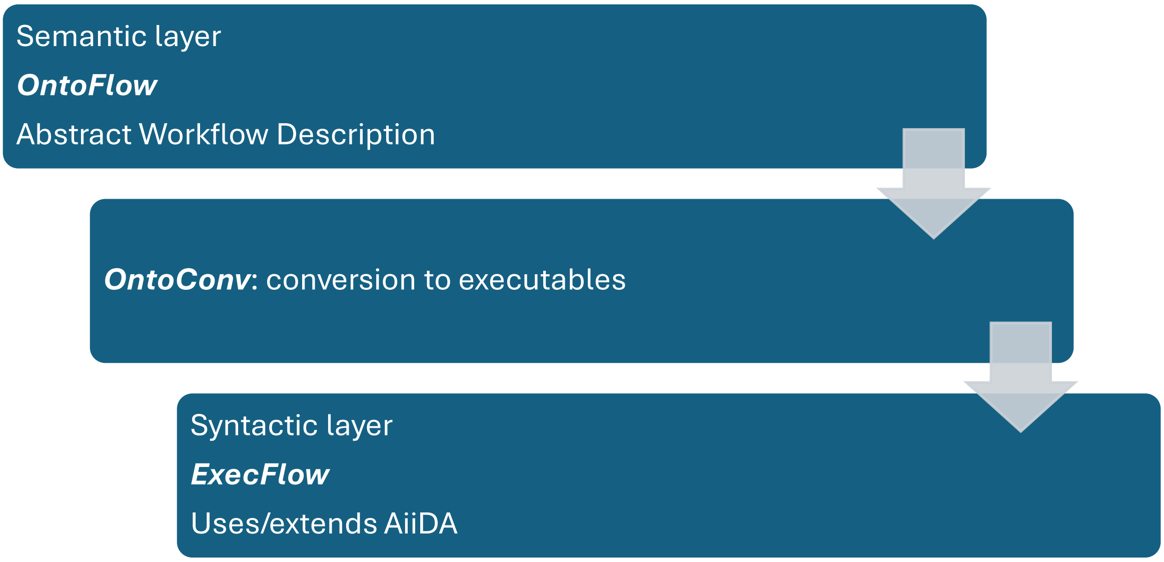

Figure 3. An outline of the technology stack used in OpenModel. The arrows represent data exchanged between the different layers.

The OpenModel stack consists of various open-source codes that store materials modelling information into a knowledge base to cover workflow building, semantic search, syntactic translation, workflow execution, and FAIR data[27,28] creation. In the following sections, we describe the OpenModel technology stack and its components in further detail.

OntoFlow

OntoFlow is a workflow designer and builder that uses semantic technology (triplestores and shared vocabularies and ontologies) to retrieve and build materials modelling workflows from documentation in a knowledge base. It provides interoperability between heterogeneous software tools by breaking down materials modelling workflows into a series of tasks, i.e. computational processes transforming input into output data.

OntoFlow discovers and recommends the most suitable set of models, tools, and operations necessary to achieve a specific desired outcome based on the characteristics of the final workflow. This can be achieved, for example, by only considering the available input datasets, and maximising or minimising a given set of key performance attributes.

At its core, the semantic layer of OpenModel is powered by two main components: a knowledge base, and decision system. The knowledge base, known as OntoKB — an output from the OntoTrans project[29] — uses a triplestore for storing and managing semantic information related to workflows, including the definition of models, modelling wrappers, tools, and datasets along with their attributes. Through its interaction with the triplestore SPARQL endpoint, OntoKB provides querying capabilities and finds possible workflows that satisfy the user query (Figure 4).

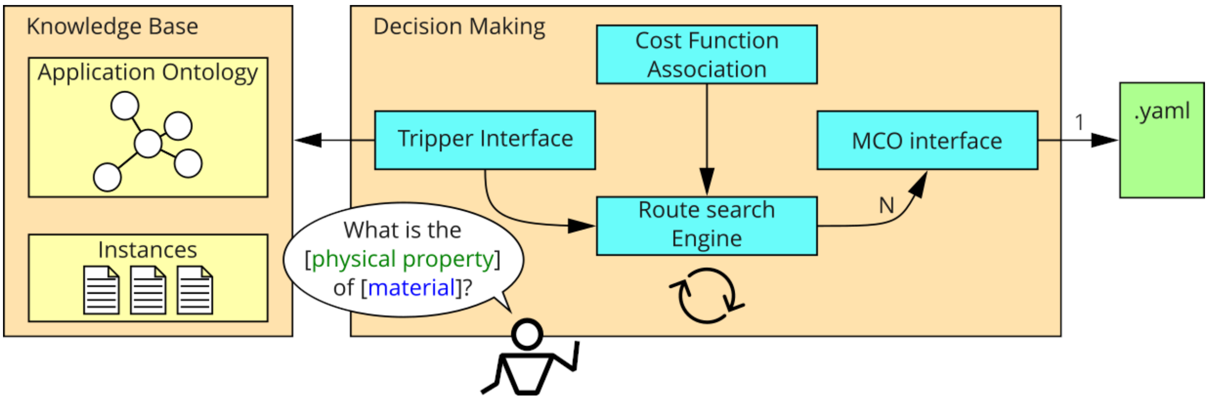

Figure 4. The semantic tool stack of OpenModel and how its parts interact. The route search algorithm begins with a user query that sets an objective, e.g. a given physical property, and a constraint, such as a particular material it refers to. Typically, several routes are found (N), which are ranked by the multi-criteria optimizer (MCO) and only one is executed by maximising some key performance attributes.

The decision-making component leverages the semantic knowledge contained within OntoKB to find the most suitable combination of models, tools, and operations necessary to achieve a specific desired outcome based on given inputs. Furthermore, it ranks all the available workflows by selecting the best one according to the key performance attributes and weights provided by the end user. The constraints may include performance indications, such as accuracy and computational time, or accessibility information on licensing. The optimal solution is obtained through several components, including an internal search engine, a cost tagger, and a multi-criteria optimization engine.

The interaction between the decision-making components and OntoKB is facilitated by an abstraction layer, called Tripper[30]. This layer enables the search engine to be agnostic with respect to the specific triplestore technology that can be changed with minimal effort.

Once the optimum workflow is identified, it is presented as a YAML file that contains the ontological classes describing computational processes (tasks) and datasets, and their topological relations. Proposed workflows with missing initial datasets can be filtered out.

Considering the modelling workflow representation in OntoFlow in more detail, each task in a workflow is connected to a wrapper, that is, a computer program that encapsulates and hides a subset of functions of another software by using a well-defined interface. The input and output data are semantically annotated using the ontology and documented at the numerical level within a common interoperability framework, as illustrated in Figure 5.

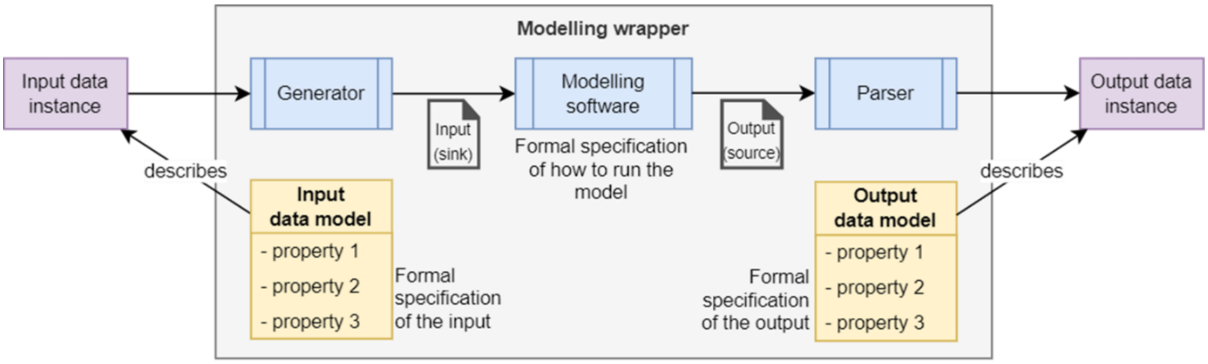

Figure 5. A model wrapper in OpenModel describes input and output data within the chosen interoperability framework, and how to execute the model within the chosen workflow manager.

Figure 6 shows how a workflow can be constructed based on the modelling wrappers. Each wrapper describes input and output semantically (so meaning is conveyed) and numerically (so numbers can be parsed and generated correctly). Models are therefore free to be combined according to the ontological description of the domain.

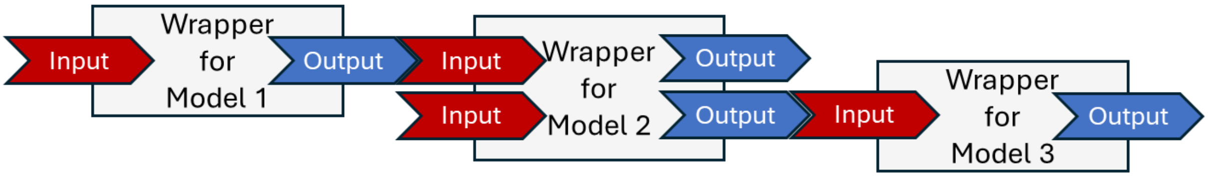

Figure 6. As a result of the documentation of input and output data for each model, models can be freely combined to a workflow based on the ontology.

Accordingly, a task only handles datasets exposing a subset of simulation parameters and links them to ontological concepts. In this way, a user can query for a way to compute a particular physical observable, and the system will suggest one or more workflows that connect the desired output to the relevant materials models that can predict the required observable based on available input datasets and registered modelling wrappers.

OntoConv

To be machine-executed, the workflow needs to be enhanced with additional information stored at the individual level that details the specific characteristics of models and data in terms of modelling wrappers, data source locations, and formats. This operation is performed by the OntoConv layer, which creates a declarative workchain that describes actual execution steps needed by the workflow manager. This description handles both actual execution of models and data transfer from one model to the next. OntoConv accesses the same knowledge base as OntoFlow but retrieves information at the individual dataset and model level. Once the workflow is converted in the declarative workchain syntax, it is passed to the workflow manager to be executed.

ExecFlow

ExecFlow is an interpreter designed to simplify the execution of complex computational workflows for materials modelling. ExecFlow operates on top of the workflow executor, AiiDA, and uses a declarative syntax to specify the execution of tasks within a workflow. This syntax is based on YAML and allows users to describe the tasks and data flow inside workflows without the need to create AiiDA plugins, which makes it easier to create modelling wrappers.

A key feature of ExecFlow is the ability to integrate with various software tools and executables through the use of modelling wrappers. Modelling wrappers encapsulate and freeze the use of a modelling software to a particular use case, e.g. by using only a subset of commands with a given syntax. The wrappers handle the inputs and outputs of a given software through OTEAPI pipelines[i,31], which converts the exchanged datasets to instances of ontological classes in the shared knowledge base.

ExecFlow has been designed to simplify the use of AiiDA and to facilitate the integration of modelling software without the burden of developing an AiiDA plugin. Along with the semantic description of wrappers and datasets, the ExecFlow component is a valuable contribution to the existing AiiDA ecosystem providing a potential way to ensure interoperability and explainability to the AiiDA materials modelling community.

Workflow documentation

Complex materials modelling workflows are semantically described as computational processes that transform input data into output datasets. The documentation framework is based on describing workflows in terms of the relationships between tasks and datasets.

A workflow is composed of multiple tasks, each representing a process that takes specific datasets as inputs and produces corresponding output datasets. The division of a workflow into tasks allows for a clear understanding of the data flow and dependencies between different processes. This modular approach facilitates the creation of complex workflows by combining individual tasks (that may be sub-workflows themselves) in a logical sequence and resulting in serial, parallel, or loop workflow topologies that are fully consistent with the mereological framework provided by the EMMO.

Each task is associated with input and output datasets, which are mapped to application ontologies providing a common vocabulary for data representation. This process is

[i] An OTEAPI pipeline is a framework for FAIR data documentation that has reusable strategies to process data from source to sink. It separates pipelines for documenting data sources (download, parse, and mapping filters) and consumers (mapping and serialization filters) to enable independent semantic availability. By mapping data model properties to ontological concepts, OTEAPI ensures semantic data integration and flow.

repeated for each task in the workflow, which enables the system to track the provenance of data throughout the entire workflow.

By decomposing complex simulation workflows into individual tasks and laying down the relationships between inputs, processes, and outputs, the OpenModel technology provides an efficient way to store the knowledge required to use advanced modelling techniques and thus helps non-expert users to access modelling tools. Furthermore, the use of semantic knowledge graphs provides interoperability between different software and allows the stored material models to be combined in new ways.

To validate the flexibility of this approach, six industrial cases were used. These cases test the OpenModel concept and demonstrate that it can integrate any type of model, whether electronic, atomistic, mesoscopic, continuum volume or data-based (see Figure 7).

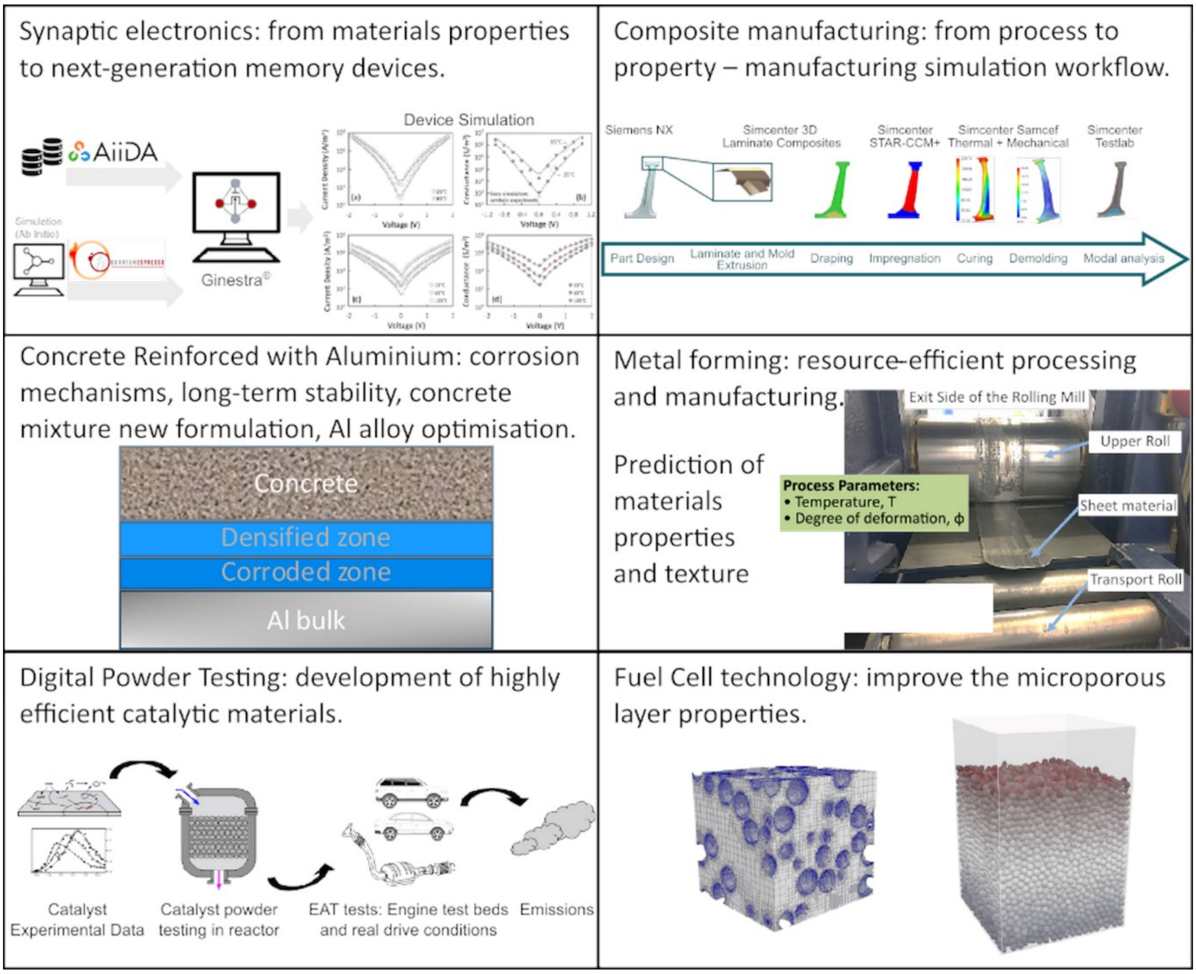

Figure 7. Schematic representation of the six industrial cases used to test the OpenModel technology stack.

The conceptualisation of a workflow in a knowledge graph allows us to visualise the physical models used to obtain a certain observable, along with the approximations made to represent the model investigated and the parameters used. This method also exposes these choices to scrutiny and ensures validation of the results obtained.

Figure 8 shows the complexity of a knowledge graph documenting a workflow that connects an electronic structure software (Quantum Espresso[32]) to a finite-element multi-physics software (Ginestra[33]) that uses a post-processing step to extract physical information from an output dataset and make it available as an input for another process. On top of documenting the workflow, the knowledge graph is enriched with semantic declarations, i.e. statements and relations that add meaning to each component of the workflow.

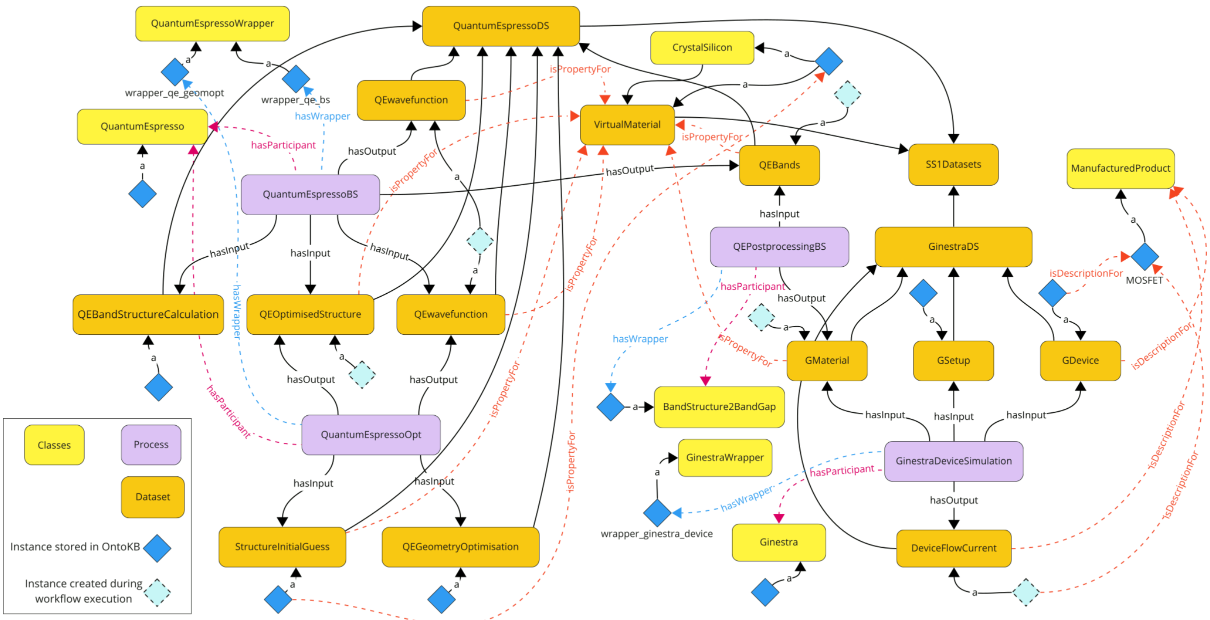

Figure 8. Diagram of the knowledge graph representing one of the six industrial cases used to demonstrate the OpenModel concept.

Semantic technologies in materials science

Within materials science, the application of semantic technologies has gained considerable attention and investment in recent years[34–37]. An overview of pertinent initiatives was undertaken by the Research Data Alliance as part of the working group entitled “Harmonised terminologies and schemas for FAIR data in materials science and related domains”[38].

Although the status of terminologies and ontologies in the field still shows considerable gaps[39], these are continuously being closed by projects developing a range of ontologies to interlink data associated with materials, processes, experiments, modelling and physical properties. Their use enables data integration, often of diverse data sources, and interoperability, which in turn means data can be more readily interpreted and used for knowledge discovery.

More understandable (and FAIR) data is of direct benefit to users at all organisational process levels, from task-level data users and business management alike. It opens up a broad range of data analytics and automated use tools. However, standardisation and harmonisation of data and ontologies can be challenging and requires continuous input once implemented to ensure quality assurance. It also unlocks potential opportunities to integrate with external data and enhance collaborations.

A strength of semantic technologies is that they can readily integrate numerous sources of data and align them into a single system. Any newly generated data and, moreover, new concepts can be readily added. Hence, the system is scalable and futureproof. Furthermore, a cycle of continuous and incremental improvements is entered into through the addition of new information. Bringing together different sources of information leads to enhanced insights and better decision making relating to ever more complex innovation requirements.

The application ontologies developed within the OpenModel project are based on the EMMO[10] top-level ontology. This approach was used because it provides a conceptualisation framework for materials science and related applications at both high and low levels, which facilitates accessibility to data and materials modelling solutions for all kinds of users.

Benefits

We argue that operationalising materials modelling workflows by integrating tasks, data and concepts horizontally and vertically in the organisation delivers a wide range of benefits.

Independent of use of semantics (beyond metadata), the benefits of workflow solutions were discussed by Schaarschmidt et al[16]. In particular, the paper highlights the following:

- Automation in various forms, which includes a reduced need for human intervention in data handling and analysis, reuse of data in subsequent steps, and automation in resource allocation for scalability and high-performance computing.
- Complexity reduction to facilitate passing validated workflows to non-expert users and expose only limited and well-controlled parameter choices to them.
- Provenance means registering workflows as persistent objects, log metadata and methods to track the origin of data and code, and ensure that a workflow can be automatically reproduced.
- Reliability and resilience by means of authenticating users, tracking errors, and recovering from failure. In this regard, data fidelity is key to workflow management systems being trusted and used. Area experts are needed to work within their domains of expertise to ensure the fidelity and reliability of a model before it can be used and shared.
- Rapid prototyping, productivity: reusing existing codes in a “drag-and-drop” fashion

The paper emphasises that "To leverage these benefits, workflow frameworks need to store not only the relevant data but also capture and store the associated metadata that describe in detail how the data was generated."

Time saving

The OpenModel industrial cases confirmed the major time saving due to workflow automation: Figure 9 shows an example of time saving in a fully automated workflow that is concentrated in the pre- and post-processing steps.

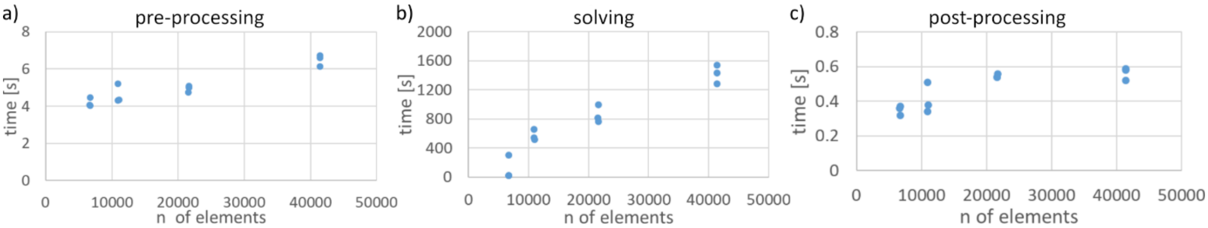

Figure 9. Wall time variation as a function of the number of elements for: a) pre-processing, b) solving, and c) post-processing in a fully automated workflow for the simulation of composites manufacturing.

Reduced cost

Once materials workflow management systems are in operation, R&D costs are notably reduced. In particular, experimental costs are lower because being able to more accurately predict a material's properties means fewer physical experiments are needed. The result is the more efficient use of material resources and significant associated time savings. A recent example is the virtual screening of non-biological peptide polymers by using high-throughput computational screening followed by experimental synthesis and characterisation[40]. In addition to screening, workflow management systems are open to AI-driven experiment planning (in Materials Acceleration Platforms, as demonstrated e.g. in the BIG-MAP project[41]), which effectively reduces the number of experiments needed to answer a given question.

Modelling and simulation governance

An efficient and effective use of modelling requires the careful management, maintenance and governance of the modelling solutions and simulations carried out. It includes aspects of model validation and verification, traceability, and making simulation workflows themselves findable, accessible, interoperable and re-usable, i.e. FAIR.

Workflow tools, as discussed by Schaarschmidt et al[16], make the creation of complex modelling workflows a lot easier, and contribute, in some ways, to interoperability. However, they can also lead to a proliferation in the number of workflows created and make findability and re-use harder. Semantic technologies can help to ensure better documentation, management and retrieval of modelling workflows.

Better and deeper data understanding

The quality of models is directly dependent on the quality of the data used to train it. Therefore, having access to larger datasets facilitates internal modelling and enables organizations to expand their knowledge base by incorporating external data sources. As a result, improved accuracy in modelling results leads to improved insights, which can significantly enhance collaborative efforts among stakeholders. Moreover, better insights ultimately contribute to the development of more sophisticated modelling outputs, fostering a continuous cycle of improvement in both model performance and organizational understanding.

In addition, there is also the possibility of gaining insight from the data itself that isn't immediately obvious to either the non-specialist or the specialist. Indeed, it is possible to identify gaps in knowledge to help explain the science that we are interested in. Data can effectively be discovered by the non-specialist, who lacks general domain knowledge, and by specialists, who may lack understanding of the broader issues.

Finally, workflow management systems afford a positive feedback loop: experimental data improves the models, which, in turn, leads to better experimental suggestions.

Sharing of models and workflows along the value chain

The ability to share models and workflows can be of significant benefit. Semantic technologies enable the sharing and controlled disclosure of model information. The capability to dynamically change what data is shared through workflow management systems can actively foster collaborations both within and outside an organisation or across disciplines. In this way, each user can filter or select results so that they only share the level of information and detail that they are comfortable with or are able to do, for instance, for confidentiality reasons. In the OpenModel platform, a user is able to share just the results of a model without sharing specific details about the process, assumptions or raw data.

Over the long term, the consistent mapping of existing and new data to ontologies or shared semantic models makes it easier for industries to manage and curate information exchange between departments along the value chain, thus further enhancing collaborative efforts.

Future outlook

Semantic technologies enable companies to develop more audacious strategic, innovation and growth goals because semantic technologies enable a holistic view of the enterprise knowledge. Similarly, companies are able to do so without compromising their ethical and sustainable values.

Even so, despite the clear benefits of semantic technologies, there are barriers to adopting these advanced systems, which can lead to hesitancy within organisations to make business-wide changes that require substantial time and expertise to implement. Industries may also implement semantic technologies to gain a competitive advantage through early technology adoption with the added benefit of ensuring they are not left behind as the sector moves forward.

Although workflow management systems offer myriad benefits, we should be clear to note that implementing semantic technologies is not to be the answer to all the problems that a company has. For relatively simple tasks and workflows, the use of a workflow management tool (e.g. OpenModel) simply adds to the overheads. Similarly, if a task is too specific, then workflow management systems don't make sense either. Indeed, used correctly, semantic technologies can enable augmented analyses, recommendations and modelling, including the use of natural language options; a level of depth and complexity (known as active metadata) that cannot be reached through IT or AI approaches alone.

Ultimately, interoperable and accessible data leads to better informed decision making within an organisation. Well managed data can also be readily shared and optimally exploited, which results in both cost and time savings. Yet, models and systems are only as good as the data — so input data needs to be clean and accurate — and checks and balances are needed from the outset.

Ontologies are a dependable way in which materials industries can provide a coordinated and controlled approach to data across an entire enterprise and involve domain experts dedicated to managing best practice, data quality, and data curation.

However, implementing semantic technologies requires a change in approach, not just in the way that data is used. Once companies have carefully selection an ontology or data system that works for them, then all users need to fully appreciate the value of the data they have and use and value its position as part of company-wide knowledge. Therefore, from the outset, a policy of well-managed change is needed, which can also help guard against resistance to change by staff who may be accustomed to relying on traditional methods to complete work.

Semantic technologies require a commitment to introducing new information management systems and provide the required skills training. Consequently, the long-term monetary and time savings are somewhat off-set in the short term because of the need for infrastructure investment and staff training/up-skilling.

Currently, the materials sector is mostly unregulated with respect to data management. However, as complex data and workflow management systems evolve, this situation may change, which could spur innovation through ontology development. It is anticipated that top-level ontologies will enable the development of sub-field branches (vocabularies, taxonomies and domain ontologies). So here, the EMMO ontology — being rigorous and able to describe all aspects of materials science — promises to be particularly valuable and could ensure organisations are ready and can adapt easily to future regulatory requirements.

Conclusions

Uptake of semantic technologies within industrial sectors is a strong growth area and one that is expected to continue to grow to be a multi-billion dollar sector over the next 5–10 years[42]. It is anticipated that many materials industries will adopt semantic technologies too because of the considerable benefits and the need to remain competitive within the sector.

Operationalising materials modelling workflows comes with short-term challenges, such as high initial costs, standardising data and implementation of updated data management systems and staff training. Nevertheless, when well aligned with company strategy in industrial R&D there are also significant advantages, the most pertinent of which are reduced costs, increased sustainability, more efficient and effective innovation and an associated reduction in time to market of new products.

References

1. Drucker, P. F. *The Essential Drucker: Selections from the Management Works of Peter F. Drucker*. (Butterworth-Heinemann, Oxford, 2001).

2. Kondinski, A., Bai, J., Mosbach, S., Akroyd, J. & Kraft, M. Knowledge Engineering in Chemistry: From Expert Systems to Agents of Creation. *Acc. Chem. Res.* **56**, 128–139 (2023).

3. Cantrill, V., de Baas, A., Roscioni, O. & Goldbeck, G. *Developing Ontologies in Materials Science: Semantic Materials Workshop Report - Semantic Data and Knowledge Management for Chemicals and Materials Industries*. https://zenodo.org/doi/10.5281/zenodo.11127892 (2024) doi:10.5281/ZENODO.11127892.

4. Friederich, P., Häse, F., Proppe, J. & Aspuru-Guzik, A. Machine-learned potentials for next-generation matter simulations. *Nat. Mater.* **20**, 750–761 (2021).

5. Fish, J., Wagner, G. J. & Keten, S. Mesoscopic and multiscale modelling in materials. *Nat. Mater.* **20**, 774–786 (2021).

6. The OpenModel EU project. *OpenModel* https://open-model.eu/.

7. Goldbeck, G. *The Economic Impact Of Molecular Modelling*. http://zenodo.org/record/44359 (2012) doi:10.5281/ZENODO.44359.

8. Scotti, L., Basoalto, H., Moffat, J. & Cogswell, D. Review of Material Modeling and Digitalization in Industry: Barriers and Perspectives. *Integr Mater Manuf Innov* **12**, 397–420 (2023).

9. Goldbeck, G. & Simperler, A. Strategies for industry to engage in materials modelling. (2019) doi:10.5281/ZENODO.3564454.

10. Del Nostro, P. *et al.* Elementary Multiperspective Material Ontology: Leveraging Perspectives via a Showcase of EMMO-Based Domain and Application Ontologies: in *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management* 135–142 (SCITEPRESS - Science and Technology Publications, Porto, Portugal, 2024). doi:10.5220/0012910200003838.

11. The Business Process Model and Notation Specification Version 2.0.2. https://www.omg.org/spec/BPMN/.

12. The Musicode EU project. http://musicode.eu/.

13. MuPIF platform is an open-source, modular, and object-oriented simulation platform designed to create complex, distributed, multiphysics simulation workflows. mupif (2024).

14. Feltus, C. *et al.* Towards Enhancing Open Innovation Efficiency: A Method for Ontological Integration of BPMN and EMMO. in *18th Conference on Computer Science and Intelligence Systems (FedCSIS), Warsaw, Poland,* 471–479 (2023). doi:10.15439/2023F8906.

15. Le Piane, F., Vozza, M., Baldoni, M. & Mercuri, F. Integrating high-performance computing, machine learning, data management workflows, and infrastructures for multiscale simulations and nanomaterials technologies. *Beilstein J. Nanotechnol.* **15**, 1498–1521 (2024).

16. Schaarschmidt, J. *et al.* Workflow Engineering in Materials Design within the BATTERY 2030+ Project. *Advanced Energy Materials* **12**, 2102638 (2022).

17. Pizzi, G., Cepellotti, A., Sabatini, R., Marzari, N. & Kozinsky, B. AiiDA: automated interactive infrastructure and database for computational science. *Computational Materials Science* **111**, 218–230 (2016).

18. Automated Interactive Infrastructure and Database for Computational Science — AiiDA. https://www.aiida.net/.

19. Jain, A. *et al.* FireWorks is a free, open-source code for defining, managing, and executing workflows. *Concurrency and Computation* **27**, 5037–5059 (2015).

20. KNIME provides visual workflows for complex data and AI work. https://www.knime.com/.

21. Pipeline Pilot operationalizes data-driven and AI-based workflows. *Dassault Systèmes* https://www.3ds.com/products/biovia/pipeline-pilot (2023).

22. SimStack – Computer-aided molecule design for advanced materials. https://www.simstack.de/ (2017).

23. PyIron develops building blocks for complex workflows in materials science. *pyiron* https://pyiron.org/.

24. Mortensen, J., Gjerding, M. & Thygesen, K. MyQueue: Task and workflow scheduling system. *JOSS* **5**, 1844 (2020).

25. Nyberg Åkerström, W. *et al.* Developing and implementing the semantic interoperability recommendations of the EOSC Interoperability Framework. (2024) doi:10.5281/ZENODO.10843882.

26. Qin, J. & Fahringer, T. Abstract Workflow Description Language. in *Scientific Workflows* 31–61 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012). doi:10.1007/978-3-642-30715-7_3.

27. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018 (2016).

28. Vogt, L. *et al.* FAIR 2.0: Extending the FAIR Guiding Principles to Address Semantic Interoperability. Preprint at https://doi.org/10.48550/ARXIV.2405.03345 (2024).

29. The OntoTrans EU project. https://ontotrans.eu/ (2020).

30. Tripper: A triplestore wrapper for Python. European Material Modelling Council (2024).

31. Adamovic, N. *et al.* Semantic Knowledge Management for Materials: the benefits of a FAIR data and model-based approach in industrial research and development. (2024) doi:10.5281/ZENODO.13304676.

32. Quantum espresso website. *Quantum Espresso* https://www.quantum-espresso.org/.

33. Ginestra simulation software platform website. https://www.appliedmaterials.com/eu/en/semiconductor/ginestra-software.html.

34. The European Materials Modelling Council website. *https://emmc.eu/* https://emmc.eu/.

35. Platform Material Digital Core Ontology website. https://materialdigital.github.io/core-ontology/index-de.html.

36. Nationale Forschungsdaten Infrastruktur (National Research Data Infrastructure) by the Deutsche Forschungsgemeinschaft (German Research Foundation). https://www.nfdi.de/?lang=en.

37. SEMANTICS-2024. The 20th International Conference on Semantic Systems. https://2024-eu.semantics.cc/.

38. Research Data Alliance Working Group, 'Harmonised terminologies and schemas for FAIR data in materials science and related domains'. *https://www.rd-alliance.org/* https://www.rd-alliance.org/groups/harmonised-terminologies-and-schemas-fair-data-materials-science-and-related-domains-wg/activity/.

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

40. Moyer, A. P. *et al.* Enumerative Discovery of Noncanonical Polypeptide Secondary Structures. *J. Am. Chem. Soc.* **146**, 25501–25512 (2024).

41. The BIG-MAP EU project. https://www.big-map.eu/.

42. Semantic Knowledge Graphing Market Size, Market Share, Application Analysis, Regional Outlook, Growth Trends, Key Players, Competitive Strategies and Forecasts, 2023 to 2031. https://www.researchandmarkets.com/reports/5867955/semantic-knowledge-graphing-market-size-market.