MAMBO: a lightweight ontology for multiscale materials and applications

Fabio Le Piane [a,b], Matteo Baldoni [a], Mauro Gaspari [b] and Francesco Mercuri [a,*]

[a] *DAIMON Lab, Istituto per lo Studio dei Materiali Nanostrutturati (ISMN), Consiglio Nazionale delle Ricerche (CNR), Italy*
E-mails: fabio.lepiane@ismn.cnr.it, matteo.baldoni@ismn.cnr.it, francesco.mercuri@cnr.it
[b] *Department of Computer Science, University of Bologna, Italy*
E-mail: mauro.gaspari@unibo.it

Abstract. Advancements of both computational and experimental tools have recently led to significant progress in the development of new advanced and functional materials, paralleled by a quick growth of the overall amount of data and information on materials. However, an effective unfolding of the potential of advanced and data-intensive methodologies requires systematic and efficient methods for the organization of knowledge in the context of materials research and development. Semantic technologies can support the structured and formal organization of knowledge, providing a platform for the integration and interoperability of data. In this work, we introduce the Materials and Molecules Basic Ontology (MAMBO), which aims at organizing knowledge in the field of computational and experimental workflows on molecular materials and related systems (nanomaterials, supramolecular systems, molecular aggregates, etc.). Linking recent efforts on ontologies for materials sciences in neighboring domains, MAMBO aims at filling gaps in current state-of-the-art knowledge modelling approaches for materials development and design targeting the intersection between the molecular scale and higher scale domains. With a focus on operational processes, lightweight, and modularity, MAMBO enables extensions to broader knowledge domains and integration of methodologies and workflows related to both computational and experimental tools. MAMBO is expected to advance the application of data-driven technologies to molecular materials, including predictive machine learning frameworks for materials design and discovery and automated platforms.

Keywords: Ontology, Materials Science, Nanomaterials, Molecular Materials, Knowledge Representation, Machine Learning

1. Introduction

The progress in a wide range of fields in science and technology has greatly benefited from the development of new tailored functional and advanced materials, addressing specific needs[1]. Accordingly, advancements in materials development and manufacturing are considered key sectors for innovation and socio-economical assets[2]. As well as for several strategic fields[3, 4], recent developments of data-driven technologies led to significant progress in research and innovation for materials[5–7]. Breakthrough results have also been achieved through the application of multiscale modelling and data-science approaches to materials science and engineering[8], also enabled by the advancements in high-performance and high-throughput computing (HPC/HTC), machine learning (ML) and artificial intelligence (AI) in general. The new trend of combining simulations with AI and ML-based techniques depicts a promising future for the discipline[9] where hybrid models, i.e. models seamlessly combining physical-based

*Corresponding author. E-mail: francesco.mercuri@cnr.it.

simulations with AI and ML, can support researchers in the translation of basic concepts and ideas into results and innovation value. Namely, ML and AI technologies can replace computing-intensive steps of simulation protocols, bridging gaps across multi-scale simulations, or can integrate information on materials at different scales to achieve a comprehensive description of complex systems[10].

State-of-the-art approaches for the design and development of novel materials are often based on a tight integration between computational and experimental tools. Computational techniques are currently able to tackle a manifold of different and complex scenarios[11], overcoming typical limitations of purely experimental tools. Besides, computational tools also give the possibility to investigate physical and chemical phenomena across a broad range of spatial and temporal scales, enabling the understanding of the interplay between materials properties at different levels[12]. From the experimental point of view, research workflows typically employ a variety of methodologies for gathering information on materials across the entire development process. However, the vast amount of data typically produced by both computational and experimental approaches is often unstructured and uncorrelated, due to the lack of shared operational strategies for the organization of knowledge and information on materials. In addition, the volume of data related to materials design and engineering is growing at a fast pace, leading to a strong need to organize and structure information and knowledge[13]. The efficient structuring of relevant data is at the basis of the progress for accelerating research on materials through automated and predictive platforms and AI. In this context, initiatives related to the FAIR (Findable, Accessible, Interoperable, Reusable) requirements are currently been pursued to push the development of materials towards a more integrated and structured use of data[14][15][16].

This strong need for organization of knowledge and information can be tackled by using ontologies, which have already shown their great potential in the field[17–19]. The term *ontology* is used to define a logical structure that represent concepts, categories, entities, properties and relations of a specific knowledge domain using formal definitions and naming conventions. Many academic disciplines uses ontology to deal with complexity and to organize raw data into usable information and knowledge. Ranging from the definition of a controlled common jargon to the description of concepts, taxonomies and relations, ontologies have a long history in supporting researchers to organize knowledge in specific domains, enabling faster communication and better design and management of experiments and activities, including complex workflows, processes and procedures.

Ontologies are one of the building blocks of semantic technologies supporting the formal encoding of semantic meanings alongside with data[20]. Semantic technologies enable knowledge sharing and discovering, helping to deal with heterogeneous data sources, to develop interoperable platforms, and to extract new information and knowledge from data, such as logical reasoning, which complements statistical learning[21, 22]. Moreover, by formally stating the inherent structure of knowledge within a specific subject area, encoding its properties and their mutual relations, new knowledge can potentially emerge, improving current processes and data collection procedures. Semantic platforms, backed by the cooperation among teams of researchers through ontologies and semantic technologies, can therefore potentially unify research and development efforts in the field of advanced materials and align existing research lines[23]. In addition, ontologies can provide platforms to enable interoperability across a broad range of technologies for materials development, integrating investigation and analysis methodologies at different levels of detail and at different scales within unified workflows[14, 24].

Recent work has been focused on the application of semantic technologies and methodologies to the development of materials and related applications[25–31], as in the cooperative effort leading to the Elementary Multiperspective Material Ontology (EMMO)[25]. Stemming from this seminal effort, many domain ontologies tailored for specific use cases have been developed[6, 26, 28]. Despite these notable endeavors, however, rational strategies for the organization of information and knowledge in several specific sub-fields of materials science and engineering are still missing. Indeed, previous research in the field has mostly been focused on either generic aspects related to materials and processing (for example, for bulk materials engineering and manufacturing) or materials properties defined at the atomistic level (individual molecules, ideal/periodic crystal structures, etc.)[25–31]. In this latter case, ontologies are mainly applied to organize information about structure/property relationships obtained from computational investigations at the electronic/atomistic level on materials models. Other recent work has targeted materials properties at the nanoscale, focusing mostly on the nanosafety and nanotoxicity domains[32]. Essentially, current available domain ontologies in materials science lack specific capabilities for modelling knowledge from the atomistic/molecular to the nano/micro scale, establishing a link up to the bulk/continuum scale with the support of top/middle-level ontologies such as EMMO. This is particularly the case of molecular materials, a specific class of

functional systems whose basic units are constituted of molecules. The physical and functional properties of molecular materials generally depend on both the structure of their individual molecular sub-units and on the molecular aggregation morphology in condensed phases. Accordingly, changes in the structure of individual molecules affect the properties of materials at larger (from nanoaggregation to the bulk phase) scales. In addition, the relationship between molecular structure and aggregation and morphology at different scales is ruled by fabrication and processing conditions, with a dramatic impact on the overall properties of materials[33–38]. Therefore, the comprehensive understanding of the interplay between the structure of molecular units and sub-units, processing, and aggregation morphology plays a crucial role in the development of molecular materials for strategic application fields, including organic electronics and optoelectronics (OLEDs, organic thin-film transistors)[35–37, 39], organic and hybrid photovoltaics (organic and hybrid solar cells),[34, 40, 41] and bioelectronics (neural and brain interfaces, devices based on biomaterials)[42–46]. The development of applications based on advanced functional materials, and in particular molecular materials, could therefore greatly benefit from efforts aimed at a more articulated structuring of knowledge in materials science.

The Materials and Molecules Basic Ontology (MAMBO) aims at filling this gap. MAMBO focuses on a specific domain related to materials science, including molecular materials, nanomaterials, supramolecular materials, molecular thin-films and other related systems. Also, MAMBO targets the improvement of the efficiency of data storage and retrieval infrastructures, merging information obtained from different, heterogeneous sources (for example, integrating computational and experimental tools) with seamless integration. As such, MAMBO is designed to provide an access point for a smoother integration between data-driven technologies and current computational and experimental workflows used to generate materials data, including ML-based techniques. The approach proposed aims at supporting the design and development of novel functional materials, which would strongly benefit from a unification of knowledge and information in the field.

The paper is structured as follows: in Section 2 the relationship between MAMBO and previous related work is discussed; Section 3 introduces MAMBO and its rationale; Section 4 presents the details of the development process and applied methodologies; in Section 5 the implementation of MAMBO is described, as well as the initial tests carried out to ensure the stability and expressiveness of MAMBO, while Section 6 shows how MAMBO can be exploited for research and development in materials science upon a case study in the context of simple workflows, providing an evaluation of the proposed approach; in Section 7 an outline of future perspectives and work is given; finally, Section 8 draws the conclusions.

2. Relations and integration with previous work

The development of ontologies for the organization of knowledge in fields related to materials design and applications has been the target of recent research efforts[16]. These efforts resulted in the development of both top/middle level ontologies, to create a general foundation for knowledge organization, and domain ontologies. These latter find application in the organization of knowledge related to specific workflows and procedures related to the development of advanced materials. In both cases, ontologies have proved to constitute a fundamental base for the future development of interoperable platforms and frameworks for materials. Most of the work in this context has been collected by the MatPortal initiative, which currently features 21 specialized ontologies[47]. The work leading to the development of MAMBO has largely been based on these efforts. Here, we briefly discuss the characteristics of existing ontologies for materials that are more related to our work.

EMMO

EMMO (the Elementary Multiperspective Material Ontology)[25] is the result of a multidisciplinary effort, aimed at the development of a standard representational ontology framework based on current materials modelling and characterization knowledge. Rather than starting from general upper level abstract concepts, as done by other ontologies, the development of EMMO has been based from the very bottom level of practical concepts, using the actual picture of the physical world coming from applied sciences, and in particular from physics and materials science. While being extremely useful as a starting point and as a common root, EMMO is (by design) general

and non-specific, and can be used as a top- and middle-level ontology. The specific role of EMMO is to focus and organize the knowledge of fundamental physics, chemistry and ontology (in the philosophical meaning) behind materials science and, more generally, applied science. The purpose of EMMO is to define all the basic building blocks that are necessary for organizing knowledge in the aforementioned scientific domains, while leaving to subsequent work the burden to actually define functional and enforceable categories and classes for practical applications. From this common parent, many different sub-ontologies have spawned, some of which we will discuss in the following. In the development of MAMBO, we mutuated from EMMO general design criteria and approaches, especially in the definition of relationships between classes, with the purpose of a possible use of MAMBO as a domain ontology of EMMO.

MDO

MDO (Materials Design Ontology)[26] defines concepts and relations to cover knowledge in the field of materials design and aims to define a knowledge representation that can unify experimental and computational results within a common framework. MDO is specifically designed for empowering information retrieval from databases collecting information in the computational materials science domain. Although potentially able to represent in detail knowledge in several sub-fields of the materials design domain, and especially computational tasks, MDO is principally focused on crystalline/periodic systems and single molecules. As such, MDO is less suited to represent and manage information about non-crystalline compounds and materials (e.g. molecular materials). MDO is structured upon Competency Questions[48], answered by specialists in materials science, and Use Cases[48], which define the aim and the foundations on which the ontology is actually designed. MDO is structured as a modular ontology, built around the core concepts of Structure, Provenance and Property. The term *Structure* defines the basic chemical structure of a specific material, while the term *Property* defines the actual chemico-physical properties of a material (both quantitative and qualitative). Namely, we reused these general concepts defined in MDO, reimplementing them in the context of the extension towards the target domain. As we will show later, we also linked in MAMBO some specific classes and attributes already defined in MDO.

DEB

DEB (the Device, Experimental scaffolds and medical Device ontology)[27] is an open resource for organizing information about biomaterials, their design, manufacture, and biological testing. It is developed using text analysis for identifying ontology terms from literature on biomaterials, systematically curated to represent the domain lexicon. DEB may be used for searching terms, performing annotations for machine learning applications, standardized meta-data indexing, and other cross-disciplinary data exploitation. Being an ontology for Biological materials, Devices and Experimental scaffolds, DEB is mainly linked to MAMBO from the point of view of foundations and approach. Moreover, the complexity and the very heterogeneous nature of the literature and operational approaches are very similar to those addressed by MAMBO. One of the interesting peculiarities of DEB consists in the semi-automatic selection of field names to be inserted in the ontology, as opposed to common approaches relying on domain experts. Moreover, the DEB terms set is dynamic, as new elements are added when the need for new terms emerges from new pieces of research. Finally, DEB is overtly oriented towards the integration of machine learning techniques with classical computational approaches. Namely, terms included in DEB have been selected via a bag-of-words approach, and the completeness of the selected terms set has been validated in two different phases: first, a curated selection of articles have been mined, verifying if all articles were easily findable with the terms present in the ontology; then, 70 domain experts have been asked to do the same with their articles and themes of interest, proposing new missing terms to be added to the ontology. The approach pursued in the development of DEB points therefore to self-updating ontologies, which can automatically be upgraded with newer terms and vocabularies leveraging some automation and data-mining techniques.

ChEBI

ChEBI (Chemical Entities of Biological Interest) [28] is an ontology, paired to a database, for molecular entities focused on "small" chemical compounds. ChEBI defines the term "molecular entity" as any "constitutionally or isotopically distinct atom, molecule, ion, ion pair, radical, radical ion, complex, conformer, etc., identifiable as a separately distinguishable entity". The molecular entities considered are either natural or synthetic potentially bioactive products. The concepts defined in ChEBI for the definition of the structural features of molecular systems are therefore particularly useful for the organization of knowledge in domains where individual molecules and their components are involved. We took inspiration from ChEBI's approach to the definition of molecular systems and sub-units to develop concepts related to individual molecules, particles and all the respective properties and hierarchies.

CHMO

CHMO, the CHemical Methods Ontology[49], describes methods used to collect data in chemical experiments, including characterization (mass spectrometry, electron microscopy), materials processing and isolation (ionisation, chromatography, electrophoresis) and to synthesise and fabricate materials (epitaxy, continuous vapour deposition). It also describes the instruments used in these experiments, such as mass spectrometers and chromatography columns. It is intended to be complementary to the Ontology for Biomedical Investigations (OBI)[50]. CHMO is a very mature ontology for experimental procedures. We reused sections of CHMO for complementing the general structure of MAMBO with classes and relationships related to experimental materials development.

MSEO

Another important ontology for materials science is MSEO: the Materials Science and Engineering Ontology[51]. It uses the Common Core Ontologies stack[52] and strive to provide material scientists the ability to represent their experiments and related data. MSEO aims at supporting researchers with semantic data management tools, with both human- and machine-readable capabilities, and that can be easily integrated into other scientific domains.

Other related ontologies and projects

MAMBO also relates to other existing projects, platforms, material databases:

- The Virtual Materials Marketplace Project (VIMMP)[29] provides an easily accessible, user-friendly hub to access all tangible and intangible components, such as information, knowledge, services and tools to support the efficient decision making, uptake and effective use of materials modelling by a wide range of manufacturing end-users, thereby facilitating an accelerated speed of development and market deployment of new materials. VIMMP proved to be an useful resource for re-using concepts, structures and relations.
- The Open Databases Integration for Materials Design (OPTIMADE)[53] is a consortium which aims to make materials databases interoperable by developing tools as for example a specification for a common REST API[54].
- NOMAD (NOvel MAterials Discovery) [30, 31] is a database that creates, collects, stores, and cleanses computational materials science data, computed by the most important materials-science codes available today. Furthermore, the projects develops tools for mining these data for finding structure, correlations, and novel information that could not be discovered by analysing smaller data sets.

3. The MAMBO approach and rationale

The MAMBO approach stems from a manifold of practical needs in domains related to the development and application of advanced materials, and in particular of molecular materials. One of the most relevant issues in the field concerns the need for a structured and standardised way to design and implement R&D activities, considering

both materials data, information and workflows. Targeting practical applications for common research-oriented activities, the MAMBO development approach is focused on the definition of a specific sets of concepts, relations, and tools, to support R&D on materials. Indeed, materials science requires intrinsically multidisciplinary efforts involving researches with different backgrounds, including physics, chemistry, engineering, computer science. In this multidisciplinary context, terminologies for the definition of research objects and procedures are unevenly scattered. However, additional concerns other then terminological ones may arise. For example, the definition of enabling conditions for a given experiment, in terms of required initial input data and conformed outputs, may be needed. Furthermore, research activities may require several interrelated steps to be completed. These steps can be modelled decomposing complex tasks into subtasks.

To deal with aspects related to the automation and optimization of processes and procedures, we decided to apply concepts mutuated from the Problem-Solving Method (PSM) approach throughout the development of MAMBO. PSMs are useful tools to implement task-based frameworks in Knowledge Engineering (KE) and can therefore link the representation of knowledge and information to operational tasks[55]. Accordingly, our approach was guided by the realization of a framework where tools are modelled as PSMs exploiting a domain ontology, which provides terms and concepts for the specification of tasks, data and workflows[20, 56].

Essentially, the the three basic ingredients of PSMs are the following:

- Competence: the description of the input and output behavior related to a given task, together with a description of what the PSM can achieve.
- Operational specification: the description of the reasoning process, which links the required knowledge to the specified competence.
- Requirements/assumptions: the description of the domain knowledge needed by the PSM to achieve the competence. Simply put, *requirements* and *assumptions* describe the pre-requisites needed by the inference steps described by the operational specification for the application of the PSM to achieve a target.

Therefore, PSMs can be used to define and accomplish tasks by applying domain knowledge. This means that the external context of a PSM is made of two pieces: the task to be realized and the domain knowledge that is needed. The execution of tasks can be defined in terms of methods, which are also related to pre- and post-conditions. Complex operational workflows can be defined in PSMs as a pre-constituted or an incremental (*on-the-fly*) composition of methods[20, 56]. This implies that the definition of an ontology can support the organization of knowledge required to apply PSM methods to a specific domain and, at the same time, the definition of PSMs requirements can be used as the starting point for shaping an ontology related to the domain of knowledge. Following this approach, MAMBO targets a "lightweight ontology"[57], specifically oriented to solve practical use-cases and to organize the required knowledge, rather than developing an all-inclusive ontology for materials science and/or molecular materials. This approach gives us the opportunity to focus on the real use-cases emerging from research activities of the considered domain. From the practical point of view, we can use the categories and relations defined in MAMBO to shape methods connected to individual tasks, into which complex problems are decomposed. In turn, we can further integrate in MAMBO concepts and relationships emerging from the analysis of new applications and problems, thus tailoring the ontology towards the solution of real-life tasks.

The development of MAMBO requires therefore the definition of typical tasks (goals and methods) and sub-tasks, methods and pre/post-conditions related to use-case scenarios within the domain considered. The instances of the pre-conditions will be related to ontology terms. Furthermore, relationships between concepts will be formalized for specific cases. A thorough analysis of the application scenarios and use-cases is therefore a fundamental prerequisite for the development of MAMBO.

3.1. Aims and applications

MAMBO targets typical situations emerging in the development of molecular materials and akin systems. The design of MAMBO was therefore guided from the analysis of use cases in the context of experimental and computational workflows, as it will be shown below. The generic development workflows and procedures considered can imply very complex tasks, involving knowledge, processes, information and data at different levels. This situation

is particularly relevant in multi-scale experimental and computational investigations[58–60], which are of crucial relevance for the development of molecular materials[61]. An ontology for representing knowledge within this domain should therefore be able to cover aspects related to the properties of materials at different spatial scales and to the properties of processes at different degrees of definition and granularity. In this context, workflows define and implements the steps that lead from a specific problem to the solution in terms of computational and/or experimental tasks. For example, MAMBO aims at representing the knowledge required, in terms of processes and data structures, to compute the properties of molecular aggregates starting from information on individual molecules only. Here, the initial information (for example, the structure of an individual molecule) is linked to the resulting output (for example, structural properties of a molecular aggregate in the condensed phase) through a complex sequence of individual tasks (i.e. retrieving the structure of a single molecule file from a data source, finding the optimal molecular structure upon a given interaction potential, building a model of a molecular aggregate according to a given simulation protocol, performing a simulation to model the morphology of the aggregate, etc.). Ontologies can supporting the design and implementation of complex workflows, assisting the organization of knowledge and the structuring of relevant data. This step enables the realization of high-throughput and automation strategies for the generation of structured materials data, compliance with the FAIR requirements and the implementation of efficient data-driven models. These latter include AI and ML methods for property screening and prediction, design of materials and process optimization across a broad range of scales. In addition, the semantic interoperability provided by the MAMBO approach will give researchers the possibility to implement platforms for integrating data and knowledge originating from both simulations and experiments.

4. Ontology development

4.1. Preliminary steps and use cases

First of all, we analysed the current literature in the field of ontology development in the materials science and development domain. This analysis led to the links with the ontologies mentioned above, while also highlighting the requirements of a specialized domain ontology in the field targeted by MAMBO. After this analysis step, we initially adopted a bottom-up approach in the development of the ontology, implementing strategies to identify possible real use-case scenarios where the use of MAMBO can support the organisation of knowledge. To this end, we identified a work group of about 10 domain experts, with competences in computational and experimental aspects of materials research and development, supported by knowledge engineers. We first asked the experts to describe typical operations, processes, objectives and goals related to their research activities, daily routine and tools used, gathering a set of relevant tasks and objectives. The answers provided by the group of experts can be qualitatively summarized in terms of the following use cases:

1. Retrieval and generation of structured information about molecular materials and related systems.
2. Definition and implementation of new, complex workflows for the development of molecular materials and related systems.

Accordingly, one of the main objectives of MAMBO consists in supporting interoperability in the implementation of workflows within the specific domain considered and for the organization of structured data. MAMBO should thus be able to support the following use case scenarios:

- Representing knowledge on integrated modelling and characterization workflows for advanced materials, processes and related technologies.
- Providing a standard representation of materials modelling and characterization workflows.
- Providing interoperability across modelling and characterization tools (for example, modelling software tools, characterization workflows, etc.).
- Providing a basis for the development of tools with search/query capabilities in the field of materials modelling and characterization.
- Connecting with existing knowledge in the materials science domain.

4.2. Competency questions

In agreement with other common ontology development schemes[48], the initial MAMBO draft was built also on the basis of competency questions (CQs)[48]. CQs are usually expressed in natural language and aim at assisting ontology engineers in defining the scope of the knowledge considered[48]. Essentially, the ontology should contain all the relevant knowledge needed to answer the CQs considered. Supported by the domain expert team, we identified an initial set of typical questions for which the information and organization in MAMBO should provide answers, including the following:

- What is the chemical formula of the molecules constituting the basic units of the material considered?
- Is the material made of a single molecular component or is it a blend?
- Is the material homogeneous at the molecular scale?
- What are the possible compositions of materials with computed properties (e.g. computed density) falling within a given range?
- What are the material properties and their values that are produced by a given materials simulation?
- What are the molecular structures of the interface between two given materials for which a computed property (e.g. computed density) falls within a given range?
- What is the method used for calculating a given material property?
- Which are the parameters used in a materials simulation and their respective value ranges?
- What is the method used for evaluating experimentally a given material property?
- For a calculated material property, which software produced the simulation results?
- What are the input and output structures of a materials simulation?
- What are the input and output structures of a materials characterization?
- Who are the authors of the simulation for a computed property?
- Who are the authors of the measurement for an experimentally-measured property?

Together with the UCs, from CQs we also gathered a first list of relevant words, which has been iteratively updated during the whole process.

4.3. Sketching MAMBO

In this phase we embraced an "hybrid" approach (bottom-up and top-down) to better represent the different nature of concepts involved in the development of the MAMBO ontology. First, we built an initial set of qualitative relationships among the emerged terms. Then, we defined the ontology classes, also refining their mutual relationships. As stated above, we pursued an essentially modular approach, based on a set of core concepts and relationships, and subsequently considered their extension to further specialize the structuring of the knowledge within the domain.

4.3.1. MAMBO main concepts and core structure

The entities constituting the core of MAMBO are strongly related to the tasks and use cases cited above: the main concept in MAMBO is that of **material**. This is linked to the concepts of **structure** and **property**, which are in turn linked to the concepts of **experiment** and **simulation**. This structuring reflects the main recurring terms emerging from the use cases targeted by MAMBO within the specific domain considered. Each of these qualitative concepts is strongly connected to at least one of the tasks related to the domain targeted by MAMBO. As we will show below, the actual implementation of MAMBO will strongly be based upon the concepts and relationships defined at this stage. A more detailed definition of this first set of concepts is provided below:

- **Material**: the most general concept related to a material, which defines a portion of matter with some specific attributes (kind, quality, etc.)[25]. Despite very generic in principle, the range of the specific materials covered by MAMBO is narrowed by the set of attributes considered. The **Material** concept is related to the following two concepts: **Structure** and **Property**.
- **Structure**: term representing all the concepts and relations connected to the structural characteristics of a **Material**. Sub-hierarchies of the **Structure** concept will specify knowledge related to typical material components within the targeted domain (atoms, molecules, etc.).

- **Property**: concept representing all the chemical and physical properties of a material (mechanical, electro-magnetic, etc.). The **Property** hierarchy is conceptually straightforward: new useful categories representing properties emerging from related and/or neighboring domains can easily be added to the ontology.
- **Simulation**: concept representing all the information related to a computational workflow leading to a specific result, from the scientific motivations down to the actual software and parameters. As it will be shown below, the hierarchies connected to the **Simulation** concept are, together with **Experiment**, the most complex of the entire ontology, as they should be able to represent a broad range of scenarios related to different simulation tools and protocols.
- **Experiment**: analogously to **Simulation**, this term represents all the processes and procedures needed to perform an empirical experiment. Together with **Simulation**, the **Experiment** concept represents the knowledge required to define the procedures providing the value of a **Property**, the characteristics of a **Structure** and/or their mutual link. Consequently, **Experiment** and **Simulation** can be viewed as different categories of methods providing different representations of given physical phenomena.

These concepts are deeply interconnected, as a consequence of the strong mutual relationship among them occurring within typical activities related to materials investigation. Accordingly, the core structure of MAMBO was designed by associating the main terms of interest described above to ontology classes (`Material`, `Simulation`, `Experiment`, `Structure` and `Property`, respectively). The link between classes was defined through relationships, which also emerged from the analysis of computational and experimental workflows addressing structure and property of materials. A sketch of the basic MAMBO core classes and relationships is depicted in Fig. 1.
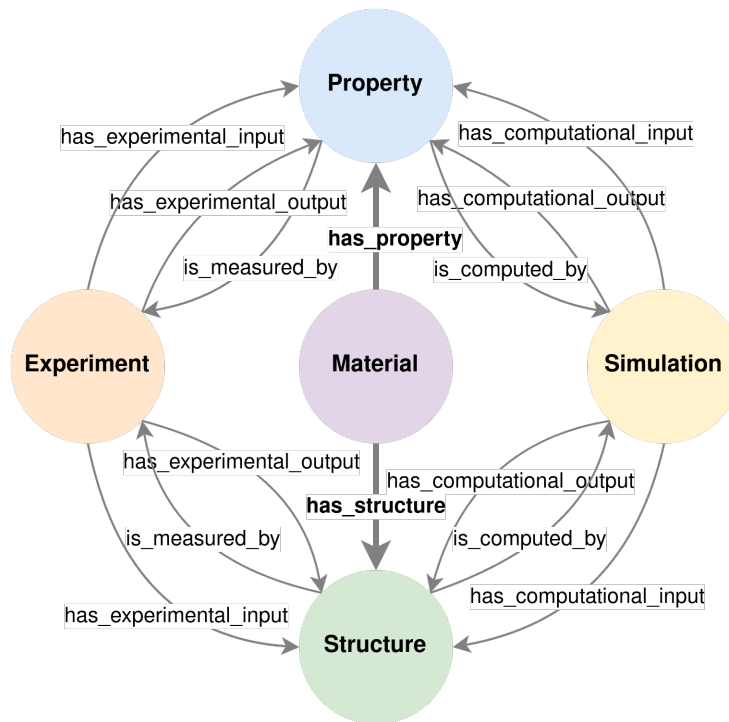

Fig. 1. MAMBO main core classes and relationships: the ontology revolves around the concepts of `Material`, `Simulation` and `Experiment`. An object (`Material`) is represented by its structural features (`Structure`) and properties (`Property`), while computational (`Simulation`) and experimental (`Experiment`) workflows are connected through a common interface to `Property` and to `Structure`.

4.4. Follow-up questionnaires

After the first sketching of MAMBO main structure and relationships, we ran through another round of interviews with domain experts in the materials development field. We asked computational and experimental domain experts to fill a questionnaire aimed at validating the effectiveness of MAMBO in representing typical tasks and workflows related to the materials development process. The interview was based on a set of very generic questions about protocols, requirements and use of information and data for research on advanced materials. This process helped us assessing the main technical requirements of experimental and computational workflows in the molecular materials domain, thus evaluating the general applicability of the core structure of MAMBO in real-life scenarios. It is worth noting that in this step we checked the consistency between the MAMBO core concepts and the materials development process within the domain considered. Moreover, the answers to the questionnaires provide valuable hints to sketch the deeper hierarchies of MAMBO, as it will be shown below. The questionnaire has been designed in order to offer experts the possibility to answer at different levels of detail. The questions posed to experts are the following[1]:

- Briefly describe your fields of study and interest.
- Which are the fundamental components of your work? (molecules, molecular aggregates)
- How do you usually structure your research process?
- Which of the aforementioned steps is usually the most difficult and time consuming?
- Which role do data have in your research activities?
- How big is the volume of data you usually have to manage? (Some megabytes, some gigabytes, many gigabytes, terabytes, other).
- Are data automatically collected? (yes, no, partially).
- If data are automatically collected only partially, explain which are the automated and manual steps, respectively, of the data acquisition process.
- Which is the rate of acquisition of data? How often are they collected? How long is this process?
- Redact a list of terms which are needed to describe your research activities.
- If a search engine for materials science existed, which kind of queries would you like to be able to do?
- Fully explain a particularly relevant experiment, from the design phase to the data collection phase.

We found that the vast majority of the terms and concepts emerged during this process can be mapped onto the core structure of MAMBO. We also found that concepts used for the description of specific computational and experimental workflows and protocols require a deeper level of specialization for the ontology. The specialized hierarchies of MAMBO were initially implemented for a set of specific cases, as those emerged from the interviews with experts. To this end, extensions of concepts and relationships shown in Fig. 1 were considered. However, the modular structure of MAMBO enables in principle extension towards a broader set of specialized topics.

4.5. Applicability and updates

In agreement with the design criteria and use cases outlined above, MAMBO is expected to be used by interdisciplinary teams of computational and experimental materials scientists and data scientists to develop and implement workflows for materials design and development, including ML/DL approaches. Moreover, MAMBO is expected to be used by interdisciplinary teams of materials scientists to collect and generate structured materials data and/or querying databases. Initial updates and maintenance will be carried out by the computational and data science teams involved in the development and use of MAMBO, supporting new versions on the basis of the needs related to modelling and experimental improvements and breakthroughs. The modularity provided by the approach pursued in the development of MAMBO is expected to activate a continuous integration process, expanding the scope of the ontology towards neighboring domains and involving a wider community in a bottom-up process.

---

[1]Closed questions have the possible answers reported inside parentheses

5. Implementation

The structure of MAMBO discussed above was implemented in OWL[62]. The implementation of MAMBO was assisted by Protégé[63], a mature, feature-full and cross-platform software for the implementation of ontologies. The current OWL implementation of MAMBO is publicly accessible on GitHub[2].

We initially implemented the core concepts and relationships shown in Fig. 1, gradually expanding MAMBO to cover topics related to specific case studies. The definition of MAMBO core classes and relationships follows closely the concepts outlined in section 4.3.1. A more formal definition of MAMBO core classes is provided below.

- `Material`: is the main class of MAMBO and is related to the `Property` and `Structure` classes via the `has_property` and `has_structure` relationships, respectively. An instance of this class represents the abstraction of a given material, and will be related to potentially many instances of `Structure` and `Property`, which in turn can be linked to different instances of the `Simulation` or `Experiment` classes.
- `Structure`: an instance of `Structure` represents the *specific* structural characteristics of a *specific* occurrence of a `Material`, being it a real world physical material or the object of a computational workflow. The `Structure` class is linked to the `Simulation` class via the `is_computed_by` relation and to the `Experiment` class via the `is_measured_by` relation.
- `Property`: this class defines the general properties of materials, either computed, measured experimentally or both. Same as for `Structure`, an instance of the `Property` class represents the specific property related to a specific case. Moreover, the `Property` class is linked to `Simulation` and `Experiment` via the relationships `is_computed_by` and `is_measured_by`, respectively, thus in analogy with the `Structure` class.
- `Experiment`: this class is linked to `Structure` and `Property` through two corresponding relationships, namely `has_experimental_input` and `has_experimental_output`, respectively. These relationships describe the `Structure` or a `Property` connected to a specific experiment in terms of input and/or the ability to produce a result.
- `Simulation`: this class essentially mirrors the `Experiment` class on the side of computational workflows. The `Simulation` class is linked to `Structure` and `Property` via `has_input` and `has_output` relations.

5.1. Deeper hierarchies, relationships and attributes

The core concepts described above and the respective ontology classes as sketched in Fig. 1 can be further structured into deeper hierarchies and other related concepts. We provide here an overview of the structuring of the main MAMBO classes in the initial implementation. A comprehensive description of all MAMBO classes and relationships is provided in the MAMBO implementation files. As stated above, the `Structure` class aims at defining the structural attributes of materials, with an initial focus on molecular systems and aggregates. Accordingly, the class features general attributes pertaining to materials structures (e.g., `spacegroup`, `lattice` and `composition`). However, the hierarchy of relations in this class is strongly based on the intrinsic characteristic of materials and chemical entities, which are in general composed by many different entities at different scales, ranging from elementary particles and atoms up to molecular aggregates and blends of different materials. Following the strongly operative nature of MAMBO, the entities representing the constituting units and sub-units of materials have been conceptualized on the same level. These entities are represented by the `StructuralEntity` class, which is linked to the `Structure` class via the `has_structural_entity` relation. The `StructuralEntity` class is further structured into sub-classes (namely `Atom`, `Particle`, `StructuralUnit`, `MolecularSystem`). These sub-classes inherit all the parameters of the `StructuralEntity` parent class, can share some common attributes (e.g., `formula`) and can feature specific attributes (e.g., `atomic_number`, `symbol`). Moreover, the `Structure` class can be structured in more detail through direct subclasses which specialize the concepts of the structure considered, for example the `MolecularAggregate` and the `Crystal` classes, respectively. These

---

[2]https://github.com/daimoners/MAMBO

subclasses represent the macro typologies of materials entities that we intend to represent and use. More categories can be easily added to the `Structure` class to extend the scope of MAMBO. To make a practical example of the hierarchies within the `Structure` class, let us consider a periodic model of a molecular aggregate composed of two different molecular species to be used in computational simulations. The structure of the system considered (the aggregate) is an instance of the `Structure` class. Each individual molecule constituting the aggregate is an instance of the `StructuralEntity` class. These entities are further defined in terms of the `MolecularSystem` subclass. In addition, we defined classes that will be used to link MAMBO to other ontologies for materials science. For example, the `Crystal` class was inspired by MDO, which, unlike MAMBO, is more specialized on crystalline structures. These links empower a more diffuse re-use of terms, relations and patterns and future cross-linking of ontologies for materials. Moreover, we also decided to inherit other classes from other related ontologies, aiming at a further extension of the scope of MAMBO. For example, we inherited from MDO the concept and class of `Coordinates`, used to describe the spatial placement of a `StructuralEntity`. However, we decided to define the ontology in more detail, using more specialized classes to represent different kinds of coordinates. Currently, we implemented for example the `CartesianCoordinates` class (imported from MDO) and the `PolarCoordinates` class. This extension emerges from the technical needs of the workflows and related PSM tasks considered to draft the ontology. However, their definition as sub-classes of the `Coordinates` class imported from MDO should enable applicability also in the context of MDO and othe related ontologies. A sketch of the `Structure` class is given in figure 2.
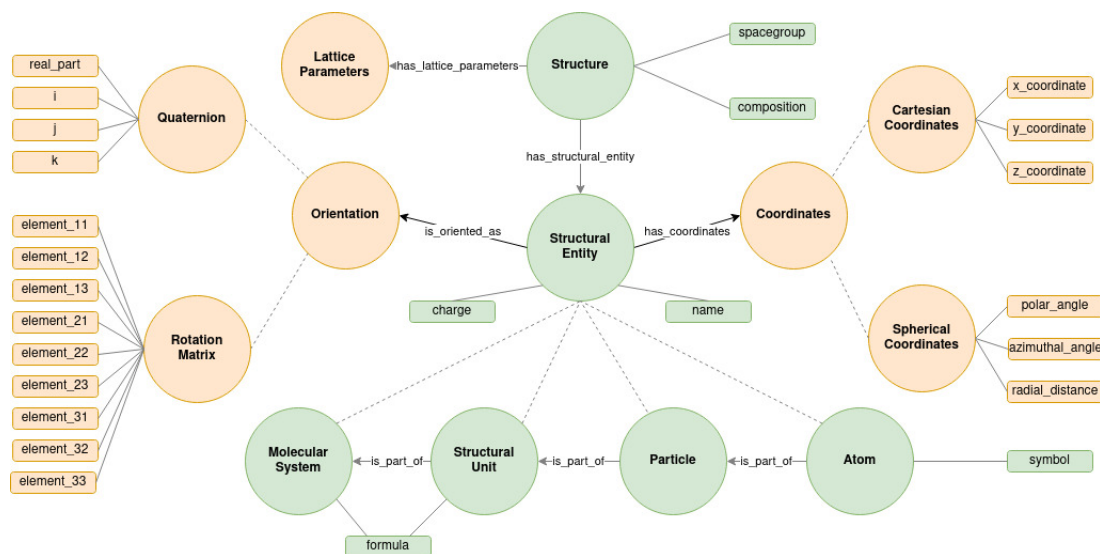

Fig. 2. Draft scheme of the `Structure` class. The main concepts and relationships used in the `Structure` class emerge from the analysis of actual workflows in typical problem solving tasks involving molecular materials. Terms and relationships are connected to both computational and experimental techniques and methods. Circles represent classes; squares represent some examples of attributes for the respective classes; dashed lines represent subclass relationships; arrows correspond to defined relations between classes, with the name of the relation written on top. For the sake of clarity, we omitted `Structure` direct subclasses, for example `MolecularAggregate` and `Crystal`. Green circles represent the classes directly related to structural entities, i.e. those directly related to `Structure` itself and `StructuralEntity`; orange circles represent the auxiliary classes which encode specific characteristics or spatial features, like all kinds of `Coordinates` types or `Orientation` subclasses, information related to the lattice (`LatticeParameters`) and so on.

The `Simulation` and `Experiment` classes aim at defining the more general set of workflows and procedures underlying a computational or experimental investigation, respectively, on materials. Although methodologies, technologies and quantitative characteristics (for example, time scales) are quite often very different from each other for these two classes, experimental and simulation workflows share several structures and patterns. For example, instances of both classes can be labelled according to a unique identifier to be retrievable (namely, an `ID` attribute).

Notes and information about results of simulations and experiments can be stored in a text form, which can be defined by a `log` attribute. The specific computational or experimental methods related to a simulation or an experiment are defined by the `ComputationalMethod` and `ExperimentalMethod` class, respectively, which include a collection of specific parameters. The modular approach pursued enables a swift definition of sub-methods and related parameters in terms of class/sub-class relationships. For example, `MolecularDynamics` is a subclass of the `Algorithm` class and `DFT` is a subclass of `QuantumMechanics`, which in turn is a subclass of the `InteractionPotential` class, related to `ComputationalMethod`. A sketch of the `Simulation` class is shown in Fig. 3.

An analogous structuring can be applied to the `Experiment` class, defining the properties and attributes of specific experimental methods and tools. The `Experiment` class is associated to the `ExperimentalMethod` class (linked through the `use_experimental_method` relation) that represents all the possible considered experimental methodologies. The sub-classes of `ExperimentalMethod` specialize the knowledge on the experimental methods used in the development of materials in more detail. In the initial implementation of MAMBO, we decided to import the subclasses defining specific experimental methods from the CHMO ontology[49]. The CHMO ontology already features a broad range of experimental methods of interest in the development and characterization of advanced materials. Other methods of interest can possibly be defined in MAMBO through the `ExperimentalMethod` class.

The `Property` class has general attributes related to values and units, which are used to gather quantitative information about materials properties. The `Property` class is structured in terms of sub-classes to specialize the knowledge about different kinds of materials properties. The hierarchical structuring of the `Property` class in terms of sub-classes is straightforward. This structuring reflects the representation of different chemico-physical properties of materials at a finer resolution, and can include a broad range of materials properties, as done in previous work[25]. In the initial implementation of MAMBO, we implemented explicitly a basic set of materials properties. However, further properties can possibly be included to extend the scope of MAMBO to broader application domains.

6. Exploiting MAMBO for Research and Development of Materials

The applicability of MAMBO in the organization of knowledge in the target domain was assessed by analysing simple typical workflows related to R&D for materials and in particular molecular materials. In the following, we focus on simulation workflows for investigations on molecular materials. The analysis of simulation workflows, in particular, allows us to define technical requirements and tune accordingly the expressiveness of MAMBO in addressing the specific knowledge involved in the description of materials at different scales (from particles to aggregates). Following the PSM approach, a general workflow connecting initial information and conditions (prerequisites) and final output (post-requisites) is decomposed into tasks and sub-tasks. The definition of tasks and subtasks and the domain knowledge is organized in terms of the structure provided by MAMBO. Let us first consider a simulation workflow for the evaluation of the chemico-physical properties of a molecular aggregate made of identical molecules by molecular dynamics (MD). While simple, this workflow exhibits the main features of more complex simulations. The consistent representation of this workflow within MAMBO can therefore be instructive of the approach pursued and gives possible hints of the ability to formalize more complex cases. This macro-task can be decomposed into several interconnected computational sub-tasks, which involve different operations on structured data. From the practical point of view, the overall workflow is generally realized by applying specialized simulation software, which implements specific computational methods, operating on structured input files and producing output files as results. Other operations may require the manipulation of files and data structures. In the case of the considered workflow, we need for example input files containing information about the structure of the considered molecule. This information is further processed by specialized software, implementing computational methods, which provide an output in terms of molecular properties. The considered methods can include for example structure manipulation tools (simulation box builders, etc.) and MD specific algorithms for equilibrating molecular aggregates in different conditions[64, 65]. The workflow produces a structured information containing for example a snapshot of the structure of the simulated aggregate in the considered conditions and/or derived properties (for
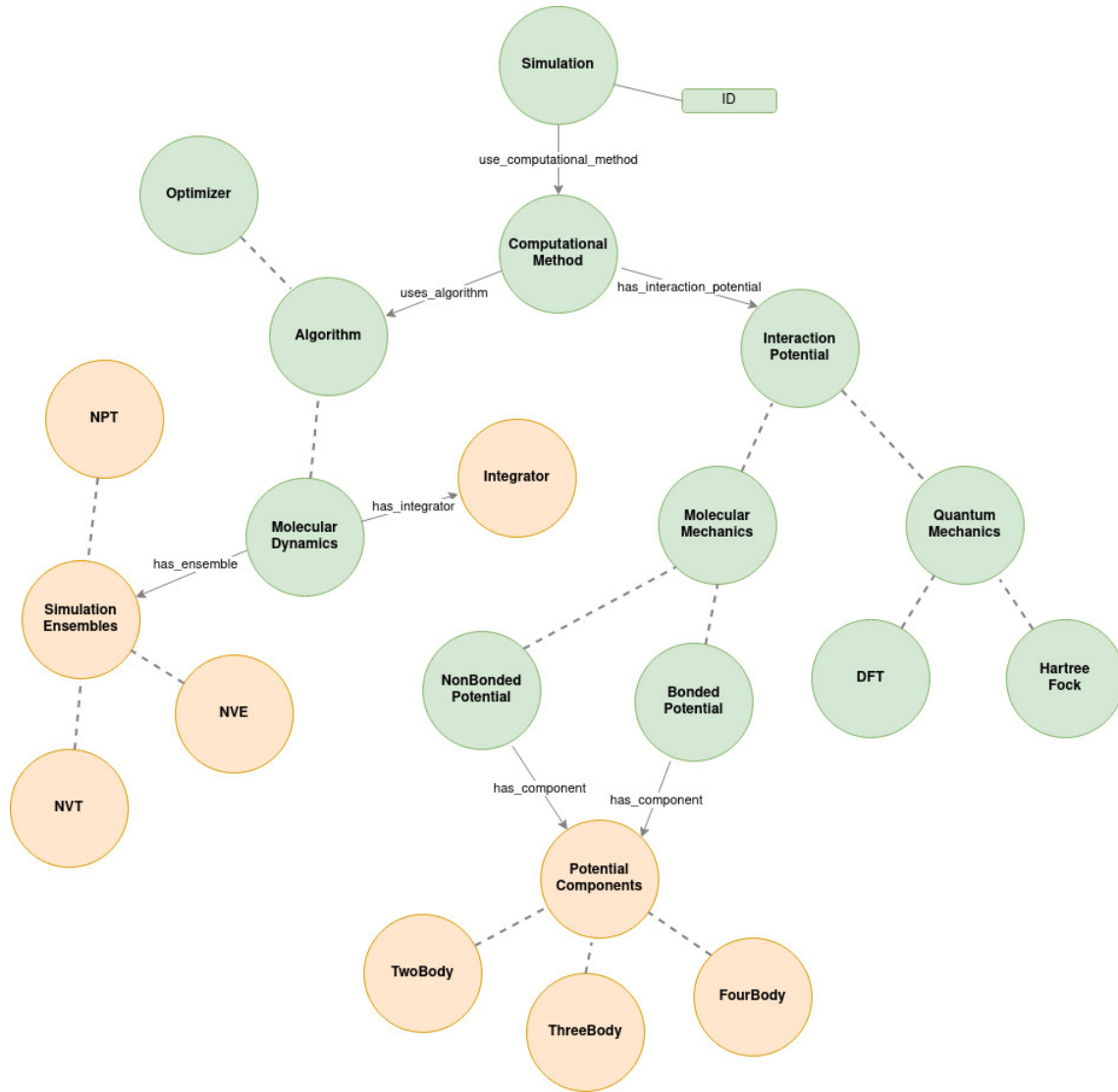
Fig. 3. Scheme of part of the `Simulation` class. The `ComputationalMethod` class gathers the different computational methods and their parameters and is related to the `Algorithm` class (the information about the specific algorithm used for the simulation) and to the `InteractionPotential` class (the information about the interaction potential used in simulations). These classes and their sub-classes are in green. Supplementing classes that collect specific information of algorithms or interaction potentials (e.g., `SimulationEnsemble` and its subclasses) are in orange. Arrows corresponds to defined relations between classes, with the name of the relation written on top. For example, `Simulation` and `ComputationalMethods` are connected by a relationship called `use_computational_method`. Dashed connections corresponds to the subclass relationships; for example, `Optimizer` is a subclass of `Algorithm`.

example, the computed equilibrium density of the aggregate in $kg\ m^{-3}$). A sketch of this workflow is shown in Fig. 4.

The decomposition of the workflow as sketched in Fig. 4 highlights the parallelism between the involved knowledge and instances of MAMBO classes. For example we can identify the following:

- The initial information about the molecular system considered is an instance of the `Structure` class, which is linked to the `Material` class via the `has_structure` relationship. In particular, the information pertains to the `MolecularSystem` subclass.
- More detailed knowledge on the molecular system considered can be structured in terms of instances of the
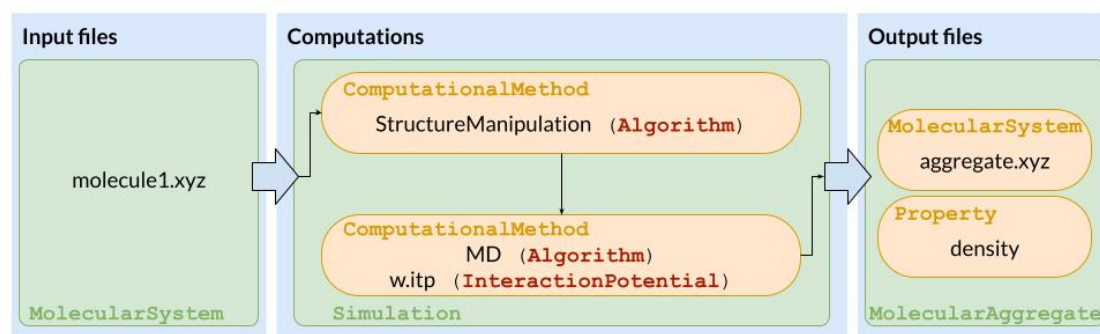
Fig. 4. A visual description of the workflow discussed in Sec. 6. The first block contains the input files, which are representable as `MolecularSystem` instances (see also Fig. 5) as individuals; the second block consists of all the files and software needed to perform the actual simulation (see Fig. 6 for more details); finally, the third block represents the output obtained from the simulation, with information about the structure of the molecular aggregate and the resulting computed density.

`Atom` class, which contain information about individual atoms of the molecule. In turn, the position of individual atoms corresponds to instances of the `CartesianCoordinates` class.

– Information on the tools for the manipulation of data structure and on MD algorithms can be represented as instances of the `ComputationalMethod` class.
– In analogy with the input data, part of the information provided by the workflow can be represented as an instance of the `Structure` class. In particular, the simulated structure of the molecular aggregate is an instance of the `MolecularAggregate` class.
– The computed property of the molecular aggregate (for example, the computed density) is an instance of the `Property` class.

An example of the parallelism between the structural information on a molecule stored as a file and encoded in a standard format in the context of molecular simulations (xyz format) and corresponding attributes of MAMBO classes is shown in Fig. 5. A similar example for attributes of classes pertaining to the `ComputationalMethod` class is shown in Fig. 6.

Fig. 5. An excerpt of a real-world input file containing structural information about a molecule encoded in the standard xyz format. In particular, the file contains information on the Cartesian coordinates and symbols of all the atoms in the molecule and the total number of atoms. Some of the involved MAMBO instances and class attributes are highlighted in different colors. Black: `Structure` instance; blue: `MolecularSystem` instance; orange: `Atom` instance and attributes; red: `CartesianCoordinates` instance and attributes.

The link between the structure provided by MAMBO and the data defining a specific computational workflow can be provided by metadata and/or annotations, which can be implemented in a variety of standard formats[67].

The applicability of MAMBO in the definition of the workflow considered above and defined by exploiting PSM techniques (competences - input/output, operational specifications and requirements) shows the potential of
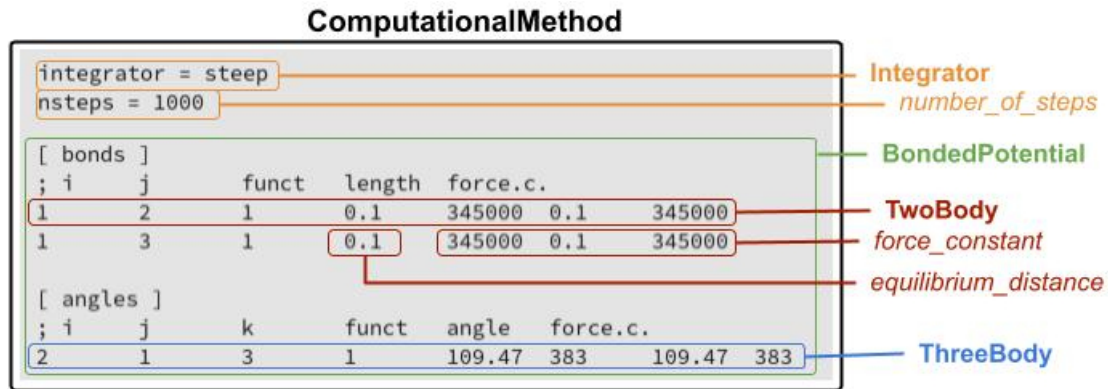
Fig. 6. An excerpt of a real-world configuration file containing information about a simulation. This example shows possible encoding in formats used by common software packages for MD simulations (here a syntax mutuated from the Gromacs[66] format is considered). In particular, the file contains information about the type of `Integrator`, the definition of the interaction potential used in MD simulations (for example, parameters for bonded potential terms, collected by an instance of `BondedPotential`). Involved MAMBO instances and class attributes are highlighted in different colors. Black: `ComputationalMethod` instance; green: `BondedPotential` instance; blue: `ThreeBody` instance; red: `TwoBody` instance and attributes; yellow: `Integrator` instance and attributes.

the proposed approach in the context of specific applications in the materials development pipeline. This approach can be easily extended to more complex systems and processes. The semantic interoperability ground provided by MAMBO in the materials science domain provides the basic components to represent complex workflows in terms of basic and reusable building blocks enabling high-throughput and automated data processing.

7. Future work

Work is currently in progress to design and realize platforms implementing the use cases targeted initially by MAMBO. In particular, we aim at applying MAMBO in data-intensive domains related to materials development, both for data retrieval and for storage. We plan to use MAMBO to organize and formally structure procedures for data management, processing and analysis, avoiding pitfalls both on a performance and on a logical point of view. The automation of data-related pipelines will enable fully automated data management and information retrieval and the univocal and standardized storage of information on materials. These activities will also be integrated into the more general effort of the materials science community to create common marketplaces and databases for materials development, implementing semantic technologies to define and store information using formally consistent axioms and definitions, thus overcoming current limitations.

Moreover, we plan to enforce MAMBO for developing formal methods for defining workflows using PSM-related languages and technologies. Our objective is to contribute to activities targeted to the support of researchers and users by providing a standardized and formal way to describe workflows in the materials research domain[68]. The interoperable definition of workflows, in turn, enables full automation and reproducibility. MAMBO classes and relation will be used to define and represent entities related to specific workflows. The semantic frameworks will be linked to data structures, software, procedures and protocols to define the operational aspects of the workflows and for implementing automated processes. The link between data and workflows will enable the realization of data-driven frameworks based on semantic technologies, such as decision support platforms and predictive ML tools.

Finally, future activities will focus on the extension of the scope of MAMBO to cover broader aspects of the knowledge involved in the development of materials. This work will be guided by the analysis of more complex workflows involved in specific applications, defined with the support of domain experts. For example, the knowledge related to the development of applications involving molecular materials (devices, sensors, etc.) can be considered. The increase in the degree of complexity of the workflows considered is expected to lead to the definition of a larger number of subclasses, specializing the knowledge in more detail. In this process, we also expect the possible revision

of some components of the overall MAMBO structure, and a tighter interaction and link with other related ontologies. While very promising in addressing the specific contexts considered, we can already envision some limitations of MAMBO and of the approach proposed, which highlight possible actions targeted to future improvements.

8. Conclusions

In this work, we discussed the main features of MAMBO, an ontology focused on representation and knowledge organisation in the molecular materials domain. Focusing on operative and data-driven approaches, MAMBO addresses interoperability between the experimental and computational sub-domains, thus potentially empowering data-intensive ML/DL applications. The structure of MAMBO is based on a broad range of concepts and relationships of common use in the field, including methods and approaches for the multiscale modelling of molecular materials.

The design of MAMBO was guided by the definition of computational and experimental workflows for materials development in the framework of PSM techniques. Here, ontologies provide a valuable semantic platform for defining the entities and objects emerging from the application of PSMs. The initial development of MAMBO was carried out according to a hybrid (top-down qualitative structuring of concepts and bottom-up formal specialization of knowledge) approach, starting from the identification of the possible use cases, the definition of competency questions and task definition. This phase led to a sketch of concepts, terms and relationships, followed by a formal implementation in OWL. Instantiation tests demonstrated the effectiveness and expressive power of MAMBO in the case of simple workflows in relation to the targeted domain. Here, MAMBO allowed us to re-construct and define formally the main components of a realistic computational workflow, representing processes, objectives and entities involved through MAMBO classes and relationships.

References

[1] Materials 2030 Manifesto. https://emmc.eu/wp-content/uploads/2022/02/advanced-materials-2030-manifesto.pdf.

[2] Key enabling technologies policy. https://ec.europa.eu/info/research-and-innovation/research-area/industrial-research-and-innovation/key-enabling-technologies_en.

[3] W. Li, Y. Liang and S. Wang, *Data Driven Smart Manufacturing Technologies and Applications*, Vol. 1, 1st edn, Springer, Cambridge, 2021, pp. 1–218. ISBN 978-3-030-66851-8. doi:https://doi.org/10.1007/978-3-030-66849-5.

[4] S.J. Qin, Survey on data-driven industrial process monitoring and diagnosis, *Annual Reviews in Control* **36** (2012), 220–234. doi:https://doi.org/10.1016/j.arcontrol.2012.09.004. https://www.sciencedirect.com/science/article/pii/S1367578812000399.

[5] L. Himanen, A. Geurts, A.S. Foster and P. Rinke, Data-Driven Materials Science: Status, Challenges, and Perspectives, *Advanced Science* **6** (2019). doi:10.1002/advs.201900808.

[6] H. Li, R. Armiento and P. Lambrix, A method for extending ontologies with application to the materials science domain, *Data Science Journal* **18** (2019), 1–21. doi:10.5334/dsj-2019-050.

[7] R. Pollice, G. dos Passos Gomes, M. Aldeghi, R.J. Hickman, M. Krenn, C. Lavigne, M. Lindner-D'Addario, A. Nigam, C.T. Ser, Z. Yao and A. Aspuru-Guzik, Data-Driven Strategies for Accelerated Materials Design, *Accounts of Chemical Research* **54** (2021), 849–860. doi:10.1021/acs.accounts.0c00785.

[8] A. Agrawal and A. Choudhary, Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science, *APL Materials* **4** (2016), 1–10. doi:10.1063/1.4946894.

[9] F.L. Piane, M. Baldoni and F. Mercuri, Predicting the properties of molecular materials: Multiscale simulation workflows meet machine learning, *arXiv* (2020), 1–14. https://arxiv.org/abs/2007.14832.

[10] H. Bhatia, F.D. Natale, J.Y. Moon, X. Zhang, J.R. Chavez, F. Aydin, C. Stanley, T. Oppelstrup, C. Neale, S.K. Schumacher, D.H. Ahn, S. Herbein, T.S. Carpenter, S. Gnanakaran, P.T. Bremer, J.N. Glosli, F.C. Lightstone and H.I. Ingolfsson, Generalizable Coordination of Large MultiscaleWorkflows: Challenges and Learnings at Scale, *International Conference for High Performance Computing, Networking, Storage and Analysis, SC* (2021). ISBN 9781450384421. doi:10.1145/3458817.3476210.

[11] L. Rosso and A.F. de Baas, *What makes a material function? Let me compute the ways... (Short Version)*, 2017, p. 264. ISBN 9789279265976. http://ec.europa.eu/research/industrial_technologies/pdf/modelling-brochure_en.pdf.

[12] M.J. Buehler, Materials by Design—A Perspective From Atoms to Structures, *MRS bulletin / Materials Research Society* **38** (2013), 169. doi:10.1557/MRS.2013.26. /pmc/articles/PMC3806500//pmc/articles/PMC3806500/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3806500/.

[13] J. Mavračić, C.J. Court, T. Isazawa, S.R. Elliott and J.M. Cole, ChemDataExtractor 2.0: Autopopulated Ontologies for Materials Science, *Journal of Chemical Information and Modeling* **61** (2021), 4280–4289. doi:https://doi.org/10.1021/acs.jcim.1c00446. https://pubs.acs.org/doi/full/10.1021/acs.jcim.1c00446.

[14] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. t Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J.V.D. Lei, E.V. Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao and B. Mons, Comment: The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* **3** (2016), 1–9. doi:10.1038/sdata.2016.18.

[15] H.S. Stein and J.M. Gregoire, Progress and prospects for accelerating materials science with automated and autonomous workflows, *Chemical Science* **10** (2019), 9640–9649. doi:10.1039/C9SC03766G. https://pubs.rsc.org/en/content/articlehtml/2019/sc/c9sc03766ghttps://pubs.rsc.org/en/content/articlelanding/2019/sc/c9sc03766g.

[16] X. Zhang, C. Zhao and X. Wang, A survey on knowledge representation in materials science and engineering: An ontological perspective, *Computers in Industry* **73** (2015), 8–22. doi:10.1016/J.COMPIND.2015.07.005.

[17] T. Ashino, Materials Ontology: an Infrastructure for Exchanging Materials Information and Knowledge, *Data Science Journal* **9** (2010), 54–61. doi:https://doi.org/10.2481/dsj.008-041. https://www.aaai.org/Papers/Symposia/Spring/2008/SS-08-05/SS08-05-003.pdf.

[18] K. Cheung, J. Drennan and J. Hunter, Towards an ontology for data-driven discovery of new materials, *AAAI Spring Symposium - Technical Report* **SS-08-05** (2008), 9–14. ISBN 9781577353614. https://www.aaai.org/Papers/Symposia/Spring/2008/SS-08-05/SS08-05-003.pdf.

[19] S.P. Voigt and S.R. Kalidindi, Materials graph ontology, *Materials Letters* **295** (2021), 129836. doi:10.1016/J.MATLET.2021.129836.

[20] T.R. Gruber, A translation approach to portable ontology specifications, *Knowledge Acquisition* **5** (1993), 199–220. doi:10.1006/KNAC.1993.1008.

[21] B. Bayerlein, T. Hanke, T. Muth, J. Riedel, M. Schilling, C. Schweizer, B. Skrotzki, A. Todor, B.M. Torres, J.F. Unger, C. Völker and J. Olbricht, A Perspective on Digital Knowledge Representation in Materials Science and Engineering, *Advanced Engineering Materials* **24** (2022). doi:10.1002/ADEM.202101176.

[22] S. Zhao and Q. Qian, Ontology based heterogeneous materials database integration and semantic query, *AIP Advances* **7** (2017), 105325. doi:10.1063/1.4999209.

[23] M.T. Horsch, S. Chiacchiera, Y. Bami, G.J. Schmitz, G. Mogni, G. Goldbeck and E. Ghedini, Reliable and interoperable computational molecular engineering: 2. Semantic interoperability based on the European Materials and Modelling Ontology, *arXiv* (2020), 1–25. https://arxiv.org/abs/2001.04175.

[24] M.T. Horsch, C. Niethammer, G. Boccardo, P. Carbone, S. Chiacchiera, M. Chiricotto, J.D. Elliott, V. Lobaskin, P. Neumann, P. Schiffels, M.A. Seaton, I.T. Todorov, J. Vrabec and W.L. Cavalcanti, Semantic Interoperability and Characterization of Data Provenance in Computational Molecular Engineering, *Journal of Chemical and Engineering Data* **65** (2020), 1313–1329. doi:10.1021/acs.jced.9b00739.

[25] E. Ghedini and G. Schmitz, EMMO the EUROPEAN MATERIALS MODELLING ONTOLOGY, *EMMC Workshop on Interoperability in Materials Modelling* (2017), 7–8. https://emmc.info/wp-content/uploads/2017/12/EMMC_IntOp2017-Cambridge_Ghedini_Bologna.pdf.

[26] H. Li, R. Armiento and P. Lambrix, An Ontology for the Materials Design Domain, in: *The Semantic Web – ISWC 2020*, J.Z. Pan, V. Tamma, C. d'Amato, K. Janowicz, B. Fu, A. Polleres, O. Seneviratne and L. Kagal, eds, Springer International Publishing, Cham, 2020, pp. 212–227. ISBN 978-3-030-62466-8. doi:10.1007/978-3-030-62466-8_14.

[27] O. Hakimi, J.L. Gelpi, M. Krallinger, F. Curi, D. Repchevsky and M.P. Ginebra, The Devices, Experimental Scaffolds, and Biomaterials Ontology (DEB): A Tool for Mapping, Annotation, and Analysis of Biomaterials Data, *Advanced Functional Materials* **30** (2020), 1909910. doi:10.1002/ADFM.201909910. https://onlinelibrary.wiley.com/doi/full/10.1002/adfm.201909910https://onlinelibrary.wiley.com/doi/abs/10.1002/adfm.201909910https://onlinelibrary.wiley.com/doi/10.1002/adfm.201909910.

[28] K. Degtyarenko, P.D. matos, M. Ennis, J. Hastings, M. Zbinden, A. Mcnaught, R. Alcántara, M. Darsow, M. Guedj and M. Ashburner, ChEBI: A database and ontology for chemical entities of biological interest, *Nucleic Acids Research* **36** (2008), 344–350. doi:10.1093/nar/gkm791.

[29] M.T. Horsch, S. Chiacchiera, M.A. Seaton, I.T. Todorov, K. Šindelka, M. Lísal, B. Andreon, E.B. Kaiser, G. Mogni, G. Goldbeck, R. Kunze, G. Summer, A. Fiseni, H. Brüning, P. Schiffels and W.L. Cavalcanti, Ontologies for the Virtual Materials Marketplace, *KI - Kunstliche Intelligenz* **34** (2020), 423–428. ISBN 0123456789. doi:10.1007/s13218-020-00648-9.

[30] C. Draxl and M. Scheffler, NOMAD: The FAIR concept for big data-driven materials science, *MRS Bulletin* **43** (2018), 676–682. doi:10.1557/mrs.2018.208.

[31] The Novel Materials Discovery (NOMAD), 2021. https://www.nomad-coe.eu.

[32] D.G. Thomas, R.V. Pappu and N.A. Baker, NanoParticle Ontology for cancer nanotechnology research, *Journal of Biomedical Informatics* **44** (2011), 59–74. doi:10.1016/J.JBI.2010.03.001.

[33] Y. Diao, L. Shaw, Z. Bao and S.C.B. Mannsfeld, Morphology control strategies for solution-processed organic semiconductor thin films (2014). doi:10.1039/c4ee00688g. www.rsc.org/ees.

[34] R. Fitzner, E. Mena-Osteritz, A. Mishra, G. Schulz, E. Reinold, M. Weil, C. Körner, H. Ziehlke, C. Elschner, K. Leo, M. Riede, M. Pfeiffer, C. Uhrich and P. Bäuerle, Correlation of π-conjugated oligomer structure with film morphology and organic solar cell performance, *Journal of the American Chemical Society* **134** (2012), 11064–11067. doi:https://doi.org/10.1021/ja302320c. https://pubs.acs.org/doi/abs/10.1021/ja302320c.

[35] A.O.F. Jones, B. Chattopadhyay, Y.H. Geerts, R. Resel, A.O.F. Jones, R. Resel, B. Chattopadhyay and H. Geerts, Substrate-Induced and Thin-Film Phases: Polymorphism of Organic Materials on Surfaces, *Advanced Functional Materials* **26** (2016), 2233–2255. doi:10.1002/ADFM.201503169. https://onlinelibrary.wiley.com/doi/full/10.1002/adfm.201503169https://onlinelibrary.wiley.com/doi/abs/10.1002/adfm.201503169https://onlinelibrary.wiley.com/doi/10.1002/adfm.201503169.

[36] A.A. Virkar, S. Mannsfeld, Z. Bao and N. Stingelin, Organic Semiconductor Growth and Morphology Considerations for Organic Thin-Film Transistors, *Advanced Materials* **22** (2010), 3857–3875. doi:10.1002/ADMA.200903193. https://onlinelibrary.wiley.com/doi/full/10.1002/adma.200903193https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.200903193https://onlinelibrary.wiley.com/doi/10.1002/adma.200903193.

[37] P. Kumar, K.N. Shivananda, W. Zajaczkowski, W. Pisula, Y. Eichen and N. Tessler, The Relation Between Molecular Packing or Morphology and Chemical Structure or Processing Conditions: the Effect on Electronic Properties, *Advanced Functional Materials* **24** (2014), 2530–2536. doi:10.1002/ADFM.201303571. https://onlinelibrary.wiley.com/doi/full/10.1002/adfm.201303571https://onlinelibrary.wiley.com/doi/abs/10.1002/adfm.201303571https://onlinelibrary.wiley.com/doi/10.1002/adfm.201303571.

[38] A. Lorenzoni, M. Muccini and F. Mercuri, A Computational Predictive Approach for Controlling the Morphology of Functional Molecular Aggregates on Substrates, *Advanced Theory and Simulations* **2** (2019), 1900156. doi:10.1002/ADTS.201900156. https://onlinelibrary.wiley.com/doi/full/10.1002/adts.201900156https://onlinelibrary.wiley.com/doi/abs/10.1002/adts.201900156https://onlinelibrary.wiley.com/doi/10.1002/adts.201900156.

[39] A. Lorenzoni, M. Muccini and F. Mercuri, Correlation between gate-dielectric morphology at the nanoscale and charge transport properties in organic field-effect transistors, *RSC Advances* **5** (2015), 11797–11805. doi:10.1039/C4RA10954F. https://pubs.rsc.org/en/content/articlehtml/2015/ra/c4ra10954fhttps://pubs.rsc.org/en/content/articlelanding/2015/ra/c4ra10954f.

[40] D. Zhang, B. Fan, L. Ying, N. Li, C.J. Brabec, F. Huang, Y. Cao and Y. Lei, Recent progress in thick-film organic photovoltaic devices: Materials, devices, and processing, *SusMat* **1** (2021), 4–23. doi:10.1002/SUS2.10. https://onlinelibrary.wiley.com/doi/full/10.1002/sus2.10https://onlinelibrary.wiley.com/doi/abs/10.1002/sus2.10https://onlinelibrary.wiley.com/doi/10.1002/sus2.10.

[41] M. Seri, F. Mercuri, G. Ruani, Y. Feng, M. Li, Z.X. Xu and M. Muccini, Toward Real Setting Applications of Organic and Perovskite Solar Cells: A Comparative Review, *Energy Technology* **9** (2021), 2000901. doi:10.1002/ENTE.202000901. https://onlinelibrary.wiley.com/doi/full/10.1002/ente.202000901https://onlinelibrary.wiley.com/doi/abs/10.1002/ente.202000901https://onlinelibrary.wiley.com/doi/10.1002/ente.202000901.

[42] J. Tropp and J. Rivnay, Design of biodegradable and biocompatible conjugated polymers for bioelectronics, *Journal of Materials Chemistry C* **9** (2021), 13543–13556. doi:10.1039/D1TC03600A. https://pubs.rsc.org/en/content/articlehtml/2021/tc/d1tc03600ahttps://pubs.rsc.org/en/content/articlelanding/2021/tc/d1tc03600a.

[43] S. Pradhan, A.K. Brooks and V.K. Yadavalli, Nature-derived materials for the fabrication of functional biodevices, *Materials today. Bio* **7** (2020). doi:10.1016/J.MTBIO.2020.100065. https://pubmed.ncbi.nlm.nih.gov/32613186/.

[44] D. Ohayon and S. Inal, Organic Bioelectronics: From Functional Materials to Next-Generation Devices and Power Sources, *Advanced Materials* **32** (2020), 2001439. doi:10.1002/ADMA.202001439. https://onlinelibrary.wiley.com/doi/full/10.1002/adma.202001439https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.202001439https://onlinelibrary.wiley.com/doi/10.1002/adma.202001439.

[45] B. Wang, Y. Sun, H. Ding, X. Zhao, L. Zhang, J. Bai and K. Liu, Bioelectronics-Related 2D Materials Beyond Graphene: Fundamentals, Properties, and Applications, *Advanced Functional Materials* **30** (2020), 2003732. doi:10.1002/ADFM.202003732. https://onlinelibrary.wiley.com/doi/full/10.1002/adfm.202003732https://onlinelibrary.wiley.com/doi/abs/10.1002/adfm.202003732https://onlinelibrary.wiley.com/doi/10.1002/adfm.202003732.

[46] A.I. Borrachero-Conejo, E. Saracino, M. Natali, F. Prescimone, S. Karges, S. Bonetti, G.P. Nicchia, F. Formaggio, M. Caprini, R. Zamboni, F. Mercuri, S. Toffanin, M. Muccini and V. Benfenati, Electrical Stimulation by an Organic Transistor Architecture Induces Calcium Signaling in Nonexcitable Brain Cells, *Advanced Healthcare Materials* **8** (2019), 1801139. doi:10.1002/ADHM.201801139. https://onlinelibrary.wiley.com/doi/full/10.1002/adhm.201801139https://onlinelibrary.wiley.com/doi/abs/10.1002/adhm.201801139https://onlinelibrary.wiley.com/doi/10.1002/adhm.201801139.

[47] Materials Open Laboratory MatPortal. https://matportal.org/.

[48] N. Noy and D. Mcguinness, Ontology Development 101: A Guide to Creating Your First Ontology, *Knowledge Systems Laboratory* **32** (2001).

[49] Chemical Methods Ontology, Ontology Lookup Service, EMBL-EBI. https://www.ebi.ac.uk/ols/ontologies/chmo.

[50] A. Bandrowski, R. Brinkman, M. Brochhausen, M.H. Brush, B. Bug, M.C. Chibucos, K. Clancy, M. Courtot, D. Derom, M. Dumontier, L. Fan, J. Fostel, G. Fragoso, F. Gibson, A. Gonzalez-Beltran, M.A. Haendel, Y. He, M. Heiskanen, T. Hernandez-Boussard, M. Jensen, Y. Lin, A.L. Lister, P. Lord, J. Malone, E. Manduchi, M. McGee, N. Morrison, J.A. Overton, H. Parkinson, B. Peters, P. Rocca-Serra, A. Ruttenberg, S.A. Sansone, R.H. Scheuermann, D. Schober, B. Smith, L.N. Soldatova, C.J. Stoeckert, C.F. Taylor, C. Torniai, J.A. Turner, R. Vita, P.L. Whetzel and J. Zheng, The Ontology for Biomedical Investigations, *PLOS ONE* **11** (2016), e0154556. doi:10.1371/JOURNAL.PONE.0154556. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0154556.

[51] T. Hanke, Material Science and Engineering Ontology, Materials Open Laboratory MatPortal. https://matportal.org/ontologies/MSEO.

[52] N. Otte, D. Kiritsi, M. Mohd Ali, R. Yang, B. Zhang, R. Rudnicki, R. Rai and B. Smith, An ontological approach to representing the product life cycle, *Applied Ontology* **14** (2019), 1–19. doi:10.3233/AO-190210.

[53] C. Andersen, R. Armiento, E. Blokhin, G. Conduit, S. Dwaraknath, M.L. Evans, A. Fekete, A. Gopakumar, S. Gražulis, V. Hegde, M. Horton, A. Merkys, F. Mohamed, A. Morris, C. Oses, G. Pizzi, T. Purcell, G.-M. Rignanese, M. Scheffler, M. Scheidgen, L. Talirz, C. Toher, M. Uhrin, D. Winston and C. Wolverton, The OPTIMADE Specification (2021). doi:10.5281/ZENODO.4251947. https://zenodo.org/record/4251947.

[54] C.W. Andersen, R. Armiento, E. Blokhin, G.J. Conduit, S. Dwaraknath, M.L. Evans, Ádám Fekete, A. Gopakumar, S. Gražulis, A. Merkys, F. Mohamed, C. Oses, G. Pizzi, G.M. Rignanese, M. Scheidgen, L. Talirz, C. Toher, D. Winston, R. Aversa, K. Choudhary, P. Colinet, S. Curtarolo, D.D. Stefano, C. Draxl, S. Er, M. Esters, M. Fornari, M. Giantomassi, M. Govoni, G. Hautier, V. Hegde, M.K. Horton, P. Huck, G. Huhs, J. Hummelshøj, A. Kariryaa, B. Kozinsky, S. Kumbhar, M. Liu, N. Marzari, A.J. Morris, A.A. Mostofi, K.A. Persson, G. Petretto, T. Purcell, F. Ricci, F. Rose, M. Scheffler, D. Speckhard, M. Uhrin, A. Vaitkus, P. Villars, D. Waroquiers, C. Wolverton, M. Wu and X. Yang, OPTIMADE, an API for exchanging materials data, *Scientific Data 2021 8:1* **8** (2021), 1–10. doi:10.1038/s41597-021-00974-z. https://www.nature.com/articles/s41597-021-00974-z.

[55] R. Studer, V.R. Benjamins and D. Fensel, Knowledge engineering: Principles and methods, *Data & Knowledge Engineering* **25** (1998), 161–197. doi:10.1016/S0169-023X(97)00056-6.

[56] R. Mizoguchi, J. Vanwelkenhuysen and M. Ikeda, (PDF) Task ontology for reuse of problem solving knowledge, 1995. https://www.researchgate.net/publication/243763523_Task_ontology_for_reuse_of_problem_solving_knowledge.

[57] V.R. Benjamins and A.G. Pérez, Knowledge-System Technology: Ontologies and Problem-Solving Methods (1999). http://www.swi.psy.uva.nl/.

[58] N. Kovachki, B. Liu, X. Sun, H. Zhou, K. Bhattacharya, M. Ortiz and A. Stuart, Multiscale modeling of materials: Computing, data science, uncertainty and goal-oriented optimization, *Mechanics of Materials* **165** (2022), 104156. doi:10.1016/J.MECHMAT.2021.104156.

[59] J. Fish, G.J. Wagner and S. Keten, Mesoscopic and multiscale modelling in materials, *Nature materials* **20** (2021), 774–786. doi:10.1038/S41563-020-00913-0. https://pubmed.ncbi.nlm.nih.gov/34045697/.

[60] D.L. Barreiro, J. Yeo, A. Tarakanova, F.J. Martin-Martinez and M.J. Buehler, Multiscale Modeling of Silk and Silk-Based Biomaterials-A Review, *Macromolecular bioscience* **19** (2019). doi:10.1002/MABI.201800253. https://pubmed.ncbi.nlm.nih.gov/30375164/.

[61] A. Mondal, L. Paterson, J. Cho, K.-H. Lin, B. van der Zee, G.-J.A.H. Wetzelaer, A. Stankevych, A. Vakhnin, J.-J. Kim, A. Kadashchuk, P.W.M. Blom, F. May and D. Andrienko, Molecular library of OLED host materials—Evaluating the multiscale simulation workflow, *Chemical Physics Reviews* **2** (2021), 031304. doi:10.1063/5.0049513.

[62] G. Antoniou and F. van Harmelen, Web Ontology Language: OWL, *Handbook on Ontologies* (2004), 67–92. doi:10.1007/978-3-540-24750-0_4.

[63] M.A. Musen and the Protégé Team, The Protégé Project: A Look Back and a Look Forward, *AI matters* **1** (2015), 4. doi:10.1145/2757001.2757003. /pmc/articles/PMC4883684/https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4883684/.

[64] M. Baldoni, A. Lorenzoni, A. Pecchia and F. Mercuri, Spatial and orientational dependence of electron transfer parameters in aggregates of iridium-containing host materials for OLEDs: Coupling constrained density functional theory with molecular dynamics, *Physical Chemistry Chemical Physics* **20** (2018), 28393–28399. doi:10.1039/c8cp04618b. https://pubs.rsc.org/en/content/articlehtml/2018/cp/c8cp04618b.

[65] A. Lorenzoni, A.M. Conte, A. Pecchia and F. Mercuri, Nanoscale morphology and electronic coupling at the interface between indium tin oxide and organic molecular materials, *Nanoscale* **10** (2018), 9376–9385. doi:10.1039/C8NR02341G. https://pubs.rsc.org/en/content/articlehtml/2018/nr/c8nr02341ghttps://pubs.rsc.org/en/content/articlelanding/2018/nr/c8nr02341g.

[66] H.J.C. Berendsen, D. van der Spoel and R. van Drunen, GROMACS: A message-passing parallel molecular dynamics implementation, *Computer Physics Communications* **91** (1995), 43–56. doi:10.1016/0010-4655(95)00042-E.

[67] T. Lv, P. Yan and W. He, Survey on JSON Data Modelling, *Journal of Physics: Conference Series* **1069** (2018), 012101. doi:10.1088/1742-6596/1069/1/012101.

[68] S.P. Huber, S. Zoupanos, M. Uhrin, L. Talirz, L. Kahle, R. Häuselmann, D. Gresch, T. Müller, A.V. Yakutovich, C.W. Andersen, F.F. Ramirez, C.S. Adorf, F. Gargiulo, S. Kumbhar, E. Passaro, C. Johnston, A. Merkys, A. Cepellotti, N. Mounet, N. Marzari, B. Kozinsky and G. Pizzi, AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance, *Scientific Data 2020 7:1* **7** (2020), 1–18. doi:10.1038/s41597-020-00638-4. https://www.nature.com/articles/s41597-020-00638-4.

[69] T. Tudorache, C. Nyulas, N.F. Noy and M.A. Musen, WebProtégé: A collaborative ontology editor and knowledge acquisition tool for the web, *Semantic Web* **4** (2013), 89–99. doi:10.3233/SW-2012-0057.

[70] J.B. Lamy, Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies, *Artificial Intelligence in Medicine* **80** (2017), 11–28. doi:10.1016/j.artmed.2017.07.002.

[71] W.O.W. Group, OWL 2 Web Ontology Language Document Overview, *OWL 2 Web Ontology Language* (2012), 1–7. http://www.w3.org/TR/owl2-overview/.

[72] D. Andrienko, *Multiscale Concepts in Simulations of Organic Semiconductors*, Springer International Publishing, 2020, pp. 1431–1442. ISBN 978-3-319-44677-6. doi:10.1007/978-3-319-44677-6_9.

[73] W. E, B. Engquist, X. Li, W. Ren and E. Vanden-Eijnden, Heterogeneous multiscale methods: A review, *Communications in Computational Physics* **2** (2007), 367–450.

[74] VIMMP Project Consortium Virtual Materials Marketplace, 2020. https://www.vimmp.eu.

[75] F. Farazi, N.B. Krdzavac, J. Akroyd, S. Mosbach, A. Menon, D. Nurkowski and M. Kraft, Linking reaction mechanisms and quantum chemistry: An ontological approach, *Computers and Chemical Engineering* **137** (2020), 814492. doi:10.1016/j.compchemeng.2020.106813.

[76] A. Menon, N.B. Krdzavac and M. Kraft, From database to knowledge graph — using data in chemistry, *Current Opinion in Chemical Engineering* **26** (2019), 33–37. doi:10.1016/j.coche.2019.08.004.

[77] G. Gygli and J. Pleiss, Simulation Foundry: Automated and F.A.I.R. Molecular Modeling, *Journal of chemical information and modeling* **60** (2020), 1922–1927. doi:10.1021/acs.jcim.0c00018.

[78] P. Ceravolo and E. Damiani, Introduction to ontology engineering, *Semantic Knowledge Management: An Ontology-Based Framework* (2008), 25–50. ISBN 9781605660349. doi:10.4018/978-1-60566-034-9.ch002.

[79] D. Kotkov, J. Veijalainen and S. Wang, How does serendipity affect diversity in recommender systems? A serendipity-oriented greedy algorithm, *Computing* **102** (2020), 393–411. doi:10.1007/s00607-018-0687-5.

[80] D. Fensel, E. Motta, F. van Harmelen, V.R. Benjamins, M. Crubezy, S. Decker, M. Gaspari, N.A.N. Available, R. Groenboom, W. Grosso, M. Musen, E. Plaza, G. Schreiber, R. Studer and B. Wielinga, The Unified Problem-Solving Method Development Language UPML, *Knowledge and Information Systems* **5** (2003), 83–131. doi:10.1007/s10115-002-0074-5.

[81] S. Bechhofer, Knowledge Acquisition (KA) From Knowing to Representation (2016), 1–17.

[82] J. Friis, B.T. Løvfall, E. Ghedini, G. Goldbeck, G. Schmitz and A. Hashibon, Documentation on Materials Modelling Ontology in UML (2020), 1–90. https://emmc.info/emmo-info/.

[83] M. Picklum and M. Beetz, MATCALO: Knowledge-enabled machine learning in materials science, *Computational Materials Science* **163** (2019), 50–62. doi:10.1016/j.commatsci.2019.03.005.

[84] H. Li, R. Armiento and P. Lambrix, Encyclopedia of Big Data Technologies, *Encyclopedia of Big Data Technologies* **117** (2020), 135502. ISBN 9783319639628. doi:10.1007/978-3-319-63962-8.

[85] M.-o. Lenz, L.M. Ghiringhelli, C. Baldauf and M. Sche, Ontologies in Computational Materials Science, 144. doi:10.17172/NOMAD/2017.03.15-1.

[86] N. Krdzavac, S. Mosbach, D. Nurkowski, P. Buerger, J. Akroyd, J. Martin, A. Menon and M. Kraft, An Ontology and Semantic Web Service for Quantum Chemistry Calculations, *Journal of Chemical Information and Modeling* **59** (2019), 3154–3165. doi:10.1021/acs.jcim.9b00227.

[87] S. Mosbach, A. Menon, F. Farazi, N. Krdzavac, X. Zhou, J. Akroyd and M. Kraft, Multiscale Cross-Domain Thermochemical Knowledge-Graph, *Journal of Chemical Information and Modeling* **60** (2020), 6155–6166. doi:10.1021/acs.jcim.0c01145.

[88] K. Hitzler, Web Ontology Language (OWL) Web Ontology Language (OWL) Encyclopedia of Social Network Analysis and Mining (2014). https://corescholar.libraries.wright.edu/cse.https://corescholar.libraries.wright.edu/cse/184.

[89] T.F. Hagelien, H.A. Preisig, J. Friis, P. Klein and N. Konchakova, A Practical Approach to Ontology-Based Data Modelling for Semantic Interoperability, *14th WCCM-ECCOMAS Congress 2020* **2100 - Other** (2021). doi:10.23967/WCCM-ECCOMAS.2020.035. https://www.scipedia.com/public/Hagelien_et_al_2021a.

[90] C.G. Crabtree, J. Roberto, J. Sarrao and L.A.N.L.R. LeSar, Computational Materials Science and Chemistry: Accelerating Discovery and Innovation through Simulation-Based Engineering and Science.