Molecular and Materials Basic Ontology: development and first steps

Fabio Le Piane[12], Matteo Baldoni[1], Mauro Gaspari[2], and
Francesco Mercuri[1]

[1] Consiglio Nazionale delle Ricerche (CNR), Istituto per lo Studio dei Materiali
Nanostrutturati (ISMN), Bologna, Italy
{fabio.lepiane, matteo.baldoni}@ismn.cnr.it, francesco.mercuri@cnr.it
[2] Alma Mater Studiorum, Università di Bologna, Bologna, Italy
mauro.gaspari@unibo.it

Abstract. Advanced materials and their applications have become a
key field of research, and it looks like this trend is not going to change
soon. For that reason, the need for systematic and efficient methods for
organizing knowledge in the field and conduct computational or exper-
imental investigations is stronger than ever. In this work, we present
a basic implementation of MAMBO - an ontology for molecular ma-
terials and their applications in real-life scenarios. The development of
MAMBO has been guided by the needs of the research community in-
volved in the development of novel materials with functional properties,
with particular attention to the nanoscale. MAMBO aims at extending
the current work in the field, while retaining a modular nature in order
to allow straightforward extension of concepts and relations to neighbor-
ing domains. Our work is expected to enable the systematic integration
of computational and experimental data in specific domains of inter-
est (nanomaterials, molecular materials, organic an polymeric materials,
supramolecular and bio-organic systems, etc.). Moreover, MAMBO is
developed with a strong focus on the applications of data-driven frame-
works for the design of novel materials with tailored characteristics.

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

1 Introduction

The progress of a wide range of fields in science and technology has greatly
benefited from the development of new tailored functional materials, address-
ing specific needs. For that reason, advancements in materials development and
manufacturing are considered key sectors for innovation and socio-economical
assets [1]. Moreover, the recent developments of data-driven technologies led to
significant progress in most strategic fields [2, 3], one of which is research and
innovation for materials [4, 5, 6]. Another piece of the puzzle is the amazing
progress made in multiscale modelling and data-science approaches [7, 8], and
the specific advancements in high-performance and high-throughput computing

(HPC/HTC) and artificial intelligence served as a solid base for the applications of derived techniques.

The actual state-of-the-art approach for the design and development of novel materials is based on tight integration between computational and experimental methods. Computational techniques are able to tackle a multitude of scenarios [9], while also giving the possibility to employ multi-scale techniques to link knowledge about materials spanning across a range of spatial and temporal scales. Moving to the experimental workflows, researchers often employ a variety of methodologies in order to gather information about materials during the entire development process. Both approaches share a trait: they are able to produce a large quantity of unstructured information, and because of that, the dimension of data related to materials science increased enormously, leading to a strong need to organize and structure such information. Initiatives related to FAIR (Findable, Accessible, Interoperable, Reusable) requirements will further push the development on functional molecular materials [10].

This strong need for organization can be fulfilled by ontologies, which are already showing their great potential in the field [11, 12]. The creation of prolific platforms for data sharing in materials science is bound to the cooperation of group of researchers motivated to realize semantic technologies able to unify all the efforts and research lines already existing [13].

Indeed, we are already witnessing a huge amount of work in this direction; a particularly relevant case is the European Materials Modelling Ontology (EMMO) [14]. Stemming from this seminal effort, many domain ontologies tailored for specific use cases were born [15, 5, 16]. However, for materials where aggregation properties at the molecular level are relevant, we can still face deficiencies in the development and application of structured knowledge.
MAMBO — the Materials And Molecules Basic Ontology - aims at filling this gap, focusing on a specific domain related to materials science, which include molecular materials, nanomaterials, supramolecular materials, molecular thin-films and other similar systems. Many strategic fields like organic electronics and optoelectronics (OLEDs, organic thin-film transistors), organic and hybrid photovoltaics (organic and perovskite solar cells), bioelectronics (neural and brain interfaces) and molecular biomaterials strongly depend on this kind of materials.

Also, MAMBO is intended to lead to efficient data storage and retrieval infrastructures, merging information obtained via computational or experimental method with seamless transition. It can also provide the basis for a easier integration between data-driven technologies and classical materials science workflows. For example, machine learning based techniques for the design and development of novel functional materials would strongly benefit from a unification of knowledge on molecular materials and their representations.

2 Related Work and Integration with Existing Ontologies

There are already different efforts in the field of ontologies for materials science domain focusing on different aspects and details. The already mentioned

EMMO constitutes a significant example of a general ontology for the whole domain of materials modelling [14], from which many others spawned focusing on specific use cases or operational applications. Two relevant examples are ChEBI (Chemical Entities of Biological Interest) [16] and MDO (Materials Design Ontology) [15]. In fact, despite a strong focus on specific use cases, concepts from these ontologies can be reused in domains, and many of the concepts we introduced in MAMBO are borrowed from ChEBI and MDO. In particular, MAMBO is linked to ChEBI via the concepts related to individual molecules, and we integrated the organization for crystals (usually inorganic) inside MAMBO, also using it as a first reference for our approach to molecular (organic) materials. However, it must be noted that a better integration of MAMBO with these ontologies is still a work in progress.

Moreover, even ontologies developed in other related domains (like digitalisation and virtualisation) can be related to MAMBO, like OSMO (ontology for simulation, modelling, and optimization), and ontologies developed within the European project VIMMP (Virtual Materials Marketplace Project) [13] also proved to be useful resources for re-using concepts, structures and relations.

Lastly, MAMBO also aims at connecting with pre-existing materials databases, like OPTIMADE and NOMAD [13, 17, 18].

3   Application Scenarios

MAMBO is tailored to the typical frameworks for the development of molecular materials and akin systems. In particular, we identified the following two main scenarios: i) retrieving structured information on molecular materials and ii) supporting the development of new, complex workflows for modelling systems based on molecular materials.

These can be complex tasks, where data can contain information about the basic entities that constitute parts of a target system (i.e. molecules, polymers, etc.). A good example is that of multi-scale modelling and characterization data on OLEDs, such as those discussed in [19, 20]. Another example use case for MAMBO could be the modelling of complex computational workflows for specific problems related to materials science. Moreover, MAMBO can help to organize the process of using data obtained by simulations in order to implement data-driven techniques in order to realize predictive models for tasks like property prediction, designing new materials and so on. This will also benefit from the semantic interoperability provided by MAMBO, which will give researchers the ability to integrate data between simulations and empirical experiments.

4   Development process, principles and methods

The whole development process started with meetings with domain experts, aiming to define possible applications. These meetings allowed us to define:

- A set of questions that MAMBO should answer (competency questions).

– A set of tasks that MAMBO should help to organize.
– A set of use cases.

Due to the peculiar nature of the typical development approaches pursued in the considered application area, we modelled the main concepts of the ontology associating them to specific problem solving methods (PSMs) [21]. PSMs gives the possibility to define operations able to fulfill specific requirements and to reach the goals of a specific task, decomposing it into simpler subtasks, and then defining pre- and post-conditions for each of them. Thanks to this approach, we were able to identify the indispensable terms needed to describe materials science, together with the connections that resides between different concepts stemming from such terms.

Thanks to these first steps, an initial representation of the concepts and relations was drafted using a "hybrid" approach (bottom-up and top-down) in order to better represent the different nature of concepts involved in the development of the MAMBO ontology. A tentative set of relationships among terms was initially built. Further details about the development process of MAMBO will be provided in a future work. We then realized a first representation of the main concepts and their respective relations, drawing from the terms identified in the previous step in order to better represent concepts from different scales and domains.

5 Realization of MAMBO

We then proceed to define the core concepts and their mutual relationships, respecting the design principles described previously. We strove to give to MAMBO a modular structure in order to make it as easy to extend as possible in order to cover new domains and use cases.

5.1 Core Concepts

The very core of MAMBO includes the most fundamental terms stemmed from the aforementioned process. The general structure emerging is the following:

– The central concept is that of `Material`, which identifies the actual object of investigation
– `Material`s are defined mainly by their `Structure`, which is the class containing the information about the structural characteristics of the material
– `Material`s have properties, which describes how they interact with the rest of the system/environment (and which are described in the `Material Properties` class)
– `Material Properties` and `Material Structures` can be the input and output of an experimental process (here `Measurement`) or a computational one (here `Calculation`)
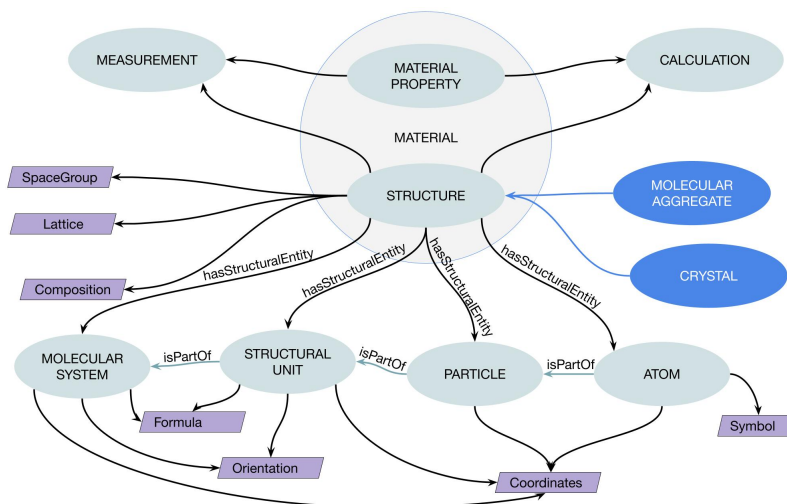
Fig. 1. Draft scheme of the `Structure` class and its relation to MAMBO's core. The main concepts and relationships used in the `Structure` class are related to the analysis of actual workflows emerging from typical problem solving tasks involving molecular materials. Terms and relationships are connected to both computational and experimental techniques and methods.

We then proceeded to defining the `Structure` class. The concepts and relationships identified at the time of this writing are shown in Fig. 1.

As already mentioned, the `Structure` class role is to contain the information regarding the structural characteristics (in 3D space and time) of an object. The main choice we made in this realm is to describe a structure as composed by one or many "structural entities" (like atoms, particles, functional groups, molecules and so on) having different features. Morever, we defined focused subclasses of the `Structure` class in order to represent more complex but fundamental systems like `Molecular Aggregate`s and `Crystal`s. Then we introduced features for the aforementioned structural entities, like `Coordinate`s (center of mass, cartesian coordinates etc), `Orientation` (Euler angles, quaternion and rotation matrix) and so on. Finally, the `Structure` class have properties related to the material in its integrity, like its periodicity. For the sake of clarity, only a subset of all these concepts and relations are shown in Fig. 1.

We then shifted to the other core concepts of MAMBO, namely `Property`, `Measurement` and `Calculation`, while also investigating their mutual relationships.

These three classes are strongly interconnected (and are also connected with the `Structure` class): a `Property` or a `Structure` could be the results of a experimental measurement or of a computational workflow, respectively, represented by `Measurement` and `Calculation`. These last two classes are intended to be as similar as possible, meaning that the will have similar organisation and symmet-

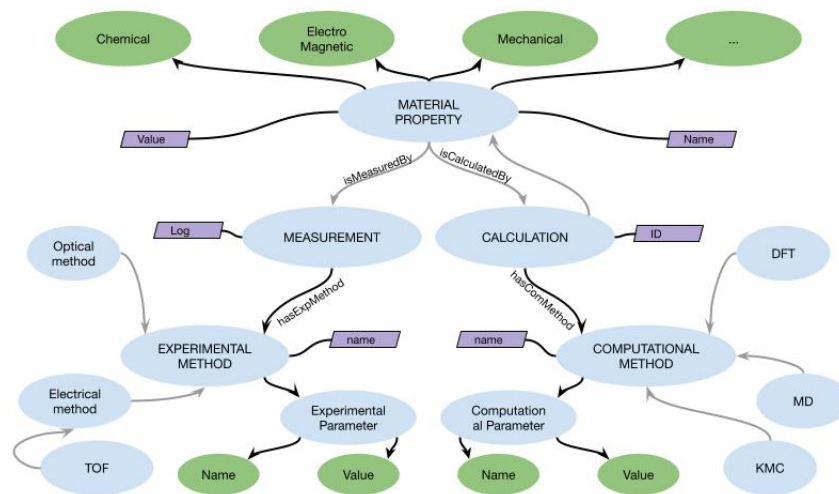**Fig. 2.** `Property` connections with `Measurement` and `Calculation` are developed in order to allow interoperability between experimental and computational workflows and data. `Measurement` and `Calculation` both have their respective "method" class, `Experimental Method` and `Computational Method`, respectively, which lead to the different experimental and computational methods, while gathering their parameters.

rical relations with the other classes. This design is part of our strategy to make computational and experimental workflows as interoperable as possible. At the same time, is important to be able to distinguish data and results coming from computational or experimental research, so we introduced both `Experimental Method` and `Computational Method` which are used to represent many different methodologies and their respective parameters.

This organization is shown in Fig. 2.

5.2 Formalization and Implementation Procedures

To implement MAMBO, we started by drawing the informal representation of a module, then trying to define the relations between the selected concepts, and finally identifying the main properties for each class. This also meant that we had to sketch the main hierarchies for classes, which have been identified using the hybrid approached already discussed. To this end, we used the OWL 2 language [22][3], using the RDF/XML syntax. At the time being, the MAMBO core is implemented with the corresponding relations, and also `Structure` and `Property` general structure have been implemented but relations with their nested subclasses and other related classes are still a work in progress.
We then conducted brief instantiation tests considering the case of a simulation of

---

[3] A draft version of the OWL implementation of MAMBO is available on GitHub at: https://github.com/daimoners/MAMBO

liposomes in water solution. The main entity analyzed is the liposome structure, which is actually a lipid bilayer with a specific shape. It is straightforward to say that the liposome is going to be the instance of `Molecular aggregate`, while the phospholipid which compose the liposome will be the instance of `Molecular System`. Going forward, we can classify the molecule of the phospholipid as a `Structural Unit`, having the related `Property`s like charge. One of its phosphate group is and instance of the `Particle` class and, finally, a phosphorus atom is easily assignable to the `Atom` class. It should also be noted that the water surrounding the liposome (and the water actually contained within the liposome cavity) should be considered as a second instance of `Structure`. We found that our reasoning was solid but slightly imperfect: for example, we found out that we needed the `Molecular Aggregate` and `Crystal` classes, and some of the original hierarchies have been pruned and modified and ended up being the one discussed in this paper.

6   Future Steps

MAMBO is still under active and intense development, in particular we need to keep working on instantiation and modellation of real-world workflows in order to see if the implemented architecture holds. While the core and the main concepts proved to be effective, a certain amount of work will be needed in order to give consistent and proper naming to the relations used in order for MAMBO to be more easily understandable for domain experts.

Then, our attention will shift to specialized domain like to formally organizing computational and experimental knowledge gained through research on molecular materials in a as-unified-as-possible fashion. Because of that, MAMBO needs to address a broad range of concepts and their respective relations in subjects like multiscale computational modelling and experimental characterization for many specific class of materials. It is fundamental to be able to easily and efficiently reuse more terminology coming from other ontologies while progressively add new ones for different use cases.

Finally, we would like to use MAMBO in order to design a database for molecular materials, giving researchers the power a semantic approach to realize complex and deep queries based on a flexible yet solid organization of knowledge of the field.

7   Conclusions

In this paper we introduced MAMBO, a new ontology for molecular materials research and design both in the realm of computational and experimental workflows, striving to make the two fully interoperable.

The project yarn for being able to model a wide spectrum of concepts and relationships used in the filed of molecular materials, including methods and approaches coming from disciplines like multiscale modelling. Giving a common

interface for data coming from empirical and computational workflows will enable a full integration of such data, which would prove to be a great added value both for the creation of a database containing pre-existing data and for the application of data-driven techniques, like machine learning, which will give researchers the possibility to gather new information (and then, new data) at a faster pace. Moreover, the development approach used during the development of MAMBO is meant to allow the extension of the semantic asset towards related fields in the domain of molecular materials, and the concepts and relationships defined within MAMBO can also be easily reused while developing other top-level ontologies.

Initial assessment and instantiation tests demonstrate how the structure of MAMBO holds and allows for great expressivity and representability in the specific field of molecular materials and nanostructures. The formal implementation is still a work in progress, in particular for extending the scope of classes while testing performance in the intended use cases and applications.

References

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

[2] Weidong Li, Yuchen Liang, and Sheng Wang, eds. *Data Driven Smart Manufacturing Technologies and Applications.* Chambridge: Springer, 2021. ISBN: 978-3-030-66851-8. DOI: https://doi.org/10.1007/978-3-030-66849-5.

[3] S Joe Qin. "Survey on data-driven industrial process monitoring and diagnosis". In: *Annual Reviews in Control* 36.2 (2012), pp. 220–234. ISSN: 1367-5788. DOI: https://doi.org/10.1016/j.arcontrol.2012.09.004. URL: https://www.sciencedirect.com/science/article/pii/S1367578812000399.

[4] Lauri Himanen et al. "Data-Driven Materials Science: Status, Challenges, and Perspectives". In: *Advanced Science* 6.21 (2019). ISSN: 21983844. DOI: 10.1002/advs.201900808.

[5] Huanyu Li, Rickard Armiento, and Patrick Lambrix. "A method for extending ontologies with application to the materials science domain". In: *Data Science Journal* 18.1 (2019), pp. 1–21. ISSN: 16831470. DOI: 10.5334/dsj-2019-050.

[6] Robert Pollice et al. "Data-Driven Strategies for Accelerated Materials Design". In: *Accounts of Chemical Research* 54.4 (2021), pp. 849–860. DOI: 10.1021/acs.accounts.0c00785. URL: https://doi.org/10.1021/acs.accounts.0c00785.

[7] Ankit Agrawal and Alok Choudhary. "Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science". In: *APL Materials* 4.5 (2016), pp. 1–10. ISSN: 2166532X. DOI: 10.1063/1.4946894. URL: http://dx.doi.org/10.1063/1.4946894.

[8]   Fabio Le Piane, Matteo Baldoni, and Francesco Mercuri. "Predicting the properties of molecular materials: Multiscale simulation workflows meet machine learning". In: *arXiv* (2020), pp. 1–14. URL: `https://arxiv.org/abs/2007.14832`.

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

[10]   Mark D. Wilkinson et al. "Comment: The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific Data* 3 (2016), pp. 1–9. ISSN: 20524463. DOI: `10.1038/sdata.2016.18`.

[11]   Toshihiro Ashino. "Materials Ontology: an Infrastructure for Exchanging Materials Information and Knowledge". In: *Data Science Journal* 9.July (2010), pp. 54–61.

[12]   Kwok Cheung, John Drennan, and Jane Hunter. "Towards an ontology for data-driven discovery of new materials". In: *AAAI Spring Symposium - Technical Report* SS-08-05 (2008), pp. 9–14.

[13]   Martin Thomas Horsch et al. "Ontologies for the Virtual Materials Marketplace". In: *KI - Kunstliche Intelligenz* 34.3 (2020), pp. 423–428. ISSN: 16101987. DOI: `10.1007/s13218-020-00648-9`. URL: `https://doi.org/10.1007/s13218-020-00648-9`.

[14]   Emanuele Ghedini and Georg Schmitz. "EMMO the EUROPEAN MATERIALS MODELLING ONTOLOGY". In: *EMMC Workshop on Interoperability in Materials Modelling* November (2017), pp. 7–8. URL: `https://emmc.info/wp-content/uploads/2017/12/EMMC_IntOp2017-Cambridge_Ghedini_Bologna.pdf`.

[15]   Huanyu Li, Rickard Armiento, and Patrick Lambrix. "An Ontology for the Materials Design Domain". In: *The Semantic Web – ISWC 2020*. Ed. by Jeff Z Pan et al. Vol. 12507 LNCS. Cham: Springer International Publishing, 2020, pp. 212–227. ISBN: 9783030624651. DOI: `10.1007/978-3-030-62466-8{\_}14`. URL: `http://dx.doi.org/10.1007/978-3-030-62466-8_14`.

[16]   Kirill Degtyarenko et al. "ChEBI: A database and ontology for chemical entities of biological interest". In: *Nucleic Acids Research* 36.SUPPL. 1 (2008), pp. 344–350. ISSN: 03051048. DOI: `10.1093/nar/gkm791`.

[17]   Claudia Draxl and Matthias Scheffler. "NOMAD: The FAIR concept for big data-driven materials science". In: *MRS Bulletin* 43.9 (Sept. 2018), pp. 676–682. ISSN: 08837694. DOI: `10.1557/mrs.2018.208`. URL: `http://link.springer.com/10.1557/mrs.2018.208`.

[18]   N. Konchakova D. Hoeche, T. Hagelien M. Zheludkevich, and J. Friis. "Ontology Assisted Modelling of Galvanic Corrosion of Magnesium". In: *WCCM-ECCOMAS2020* (2021). URL: `https://www.scipedia.com/public/Hoeche_et_al_2021a`.

[19]   Denis Andrienko. "Multiscale Concepts in Simulations of Organic Semiconductors". In: *Handbook of Materials Modeling: Methods: Theory and*

Modeling. Ed. by Wanda Andreoni and Sidney Yip. Cham: Springer International Publishing, 2020, pp. 1431–1442. ISBN: 978-3-319-44677-6. DOI: 10.1007/978-3-319-44677-6{_}39. URL: https://doi.org/10.1007/978-3-319-44677-6_39.

[20]   Matteo Baldoni et al. "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". In: *Physical Chemistry Chemical Physics* 20.45 (2018), pp. 28393–28399. ISSN: 14639076. DOI: 10.1039/c8cp04618b. URL: https://pubs.rsc.org/en/content/articlehtml/2018/cp/c8cp04618b.

[21]   Dieter Fensel et al. "The Unified Problem-Solving Method Development Language UPML". In: *Knowledge and Information Systems* 5.1 (2003), pp. 83–131. ISSN: 0219-1377. DOI: 10.1007/s10115-002-0074-5.

[22]   W3C OWL Working Group. "OWL 2 Web Ontology Language Document Overview". In: *OWL 2 Web Ontology Language* December (2012), pp. 1–7. URL: http://www.w3.org/TR/owl2-overview/.