

SEMANTIC INTEROPERABILITY BASED ON THE EUROPEAN MATERIALS AND MODELLING ONTOLOGY AND ITS ONTOLOGICAL PARADIGM: MEREOSEMIOTICS

M. T. Horsch,¹ S. Chiacchiera,² B. Schembera,¹ M. A. Seaton,² and I. T. Todorov²

¹ High Performance Computing Center Stuttgart (HLRS)
Nobelstr. 19, 70569 Stuttgart, Germany
{martin.horsch, bjoern.schembera}@hlrs.de

² UK Research and Innovation, STFC Daresbury Laboratory
Keckwick Ln, Daresbury, Cheshire WA4 4AD, United Kingdom
{silvia.chiacchiera, michael.seaton, ilian.todorov}@stfc.ac.uk

Key words: Research data infrastructure, semantic interoperability, top-level ontologies

Summary. The European Materials and Modelling Ontology (EMMO) has recently been advanced in the computational molecular engineering and multiscale modelling communities as a top-level ontology, aiming to support semantic interoperability and data integration solutions, *e.g.*, for research data infrastructures. The present work explores how top-level ontologies that are based on the same paradigm – the same set of fundamental postulates – as the EMMO can be applied to models of physical systems and their use in computational engineering practice. This paradigm, which combines mereology (in its extension as mereotopology) and semiotics (following Peirce’s approach), is here referred to as mereosemiotics. Multiple conceivable ways of implementing mereosemiotics are compared, and the design space consisting of the possible types of top-level ontologies following this paradigm is characterized.

1 INTRODUCTION

Semantic interoperability is the ability of multiple parties to exchange information with a well-defined, mutually agreed meaning. This may be codes that are coupled or linked to each other in a workflow or, *e.g.*, human or virtual agents that ingest, search, or extract information as users of a research data infrastructure. Interoperability is called semantic whenever it is not simply based on a common file format – which would be syntactic interoperability – but on a more abstract formalization of the meaning of data and metadata items, *i.e.*, on the semantics of a common language; on this basis, any serialization of the communicated content can be viably used as long as its semantics has been defined clearly in terms of a mapping from the employed syntax to the representation of meaning in terms of that common language. No FAIR data management, and hence no research data infrastructure that adheres to the FAIR principles (findability, accessibility, interoperability, and reusability), *cf.* Bicarregui [1], can be implemented without such a common language [2, 3]. More generally, any effort in simulation and data technology that addresses communities as a whole, including many teams of developers, cannot succeed without facilitating interoperability between the contributions of these teams, be it in terms of software, data, or anything else that needs to be integrated into a joint framework. For applications where it is

inevitable to connect a variety of components to each other reliably, a single file format often emerges as a de facto standard, delivering syntactic interoperability; in all other cases, even when there are only two formats between which a conversion becomes necessary, it is unavoidable to map them to each other in a way that is either implicitly or explicitly grounded in shared semantics. Making this explicit requires more work, but delivers solutions that are more robust and easy to extend to a greater number of parties. Whatever provides a definition of such an underlying mutually agreed semantic space is called a semantic asset or also, equivalently, a semantic artefact. The two most common approaches to realizing this are hierarchical metadata standards, most typically based on XML/XSD technology, where one element acts as a container for others, and semantic-web technology where resources (*i.e.*, data and metadata items) are connected to each other in a non-hierarchical way, yielding a knowledge graph; in the latter case, an RDF schema or an OWL ontology based on description logic (DL) can be employed as a semantic asset, while RDF triples are used to assert propositions concerning the data [4, 5].

Semantic assets are further distinguished by the level of abstraction at which they operate: At one end, they can be very generic and remote from any concrete application, as it is the case for top-level ontologies (also known as foundational or upper ontologies), while other standards are domain-specific. An ontology is a top-level ontology if it claims to formalize all that exists and all ways in which objects can be related to each other. Both technically and by the logic of the approach itself, it is by no means necessary to use a top-level ontology in order to realize semantic interoperability. Instead, top-level ontologies play a role that is analogous to that of widespread libraries and APIs in software engineering; they can make coding efforts less redundant by providing a reliable generic basis for dedicated developments that are more specific, and wherever they are employed by multiple components, it becomes easier to integrate them into a single coherent solution. However, the use cases that drive metadata standardization and hence the development of semantic assets almost always reside within a comparably narrow domain of knowledge; accordingly, the semantic space to be characterized, *i.e.*, the expressive capacity of the agreed language in terms of what sort of propositions can be formulated, pertains to a specific field that constitutes only a very small fraction of all that which can be known or asserted in general.

The present line of work, in particular, is concerned with use cases from molecular and multiscale modelling and simulation as well as research data management concerning thermodynamic, mechanical, and kinetic properties that are of interest to chemical and process engineering. Gygli and Pleiss [6] argue that despite significant efforts at making molecular simulation data and metadata FAIR, it is still mainly due to the lack of a recognized “community wide standard” for semantic interoperability, or of an “MDML” (molecular dynamics markup language), that it is hard to communicate molecular models and simulation workflows “from one molecular dynamics package to another.” According to Edwards *et al.* [7], metadata can serve as a key contributor to diminish “science friction” and enable interoperability. Semantic heterogeneity, however, can contribute to “metadata friction” which Edwards *et al.* [7] define as the effort required to use metadata as a product (*e.g.*, the effort to tag data with metadata). In our sense, metadata friction would also occur when a metadata model or semantic asset lacks recognition as a community wide standard, and it would include the effort required to achieve a broad community agreement. Since the emergence of semantic technology, there have been many endeavours to develop metadata standards for modelling and simulation [8]. Incidentally, a molecular simulation markup language (MSML) had been in existence since 2014, *cf.* Grunzke *et al.* [9], and a markup-language based standard for engineering metadata (EngMeta) had been released by the University of Stuttgart [10, 11] – but as it is clear from the assessment by Gygli and Pleiss [6], it is broad community uptake that counts, not the development of

a semantic asset as such. Similarly, for the development of a molecular model database [12], substantial work had to be done to clarify the employed nomenclature; not because there was no clear understanding of the terminology, but precisely due to the absence of a broad community agreement. In this and many similar cases, including work done by a series of projects funded from the Horizon 2020 research and innovation programme such as VIMMP [13, 14, 15, 16] and EMMC-CSA [17, 18, 19], the *Review of Materials Modelling* by de Baas [20] has been playing the role of an underlying foundational element.

Ontologies targeted at simulation-based engineering include OntoCAPE [21, 22], PhysSys [23] and, more recently, the Physics-based Simulation Ontology (PSO) [24], the Ontology for Simulation, Modelling, and Optimization (OSMO) [13, 14, 15, 16], and the European Materials and Modelling Ontology (EMMO), a top-level ontology developed by Ghedini *et al.* [25], *cf.* Goldbeck *et al.* [26] and Francisco Morgado *et al.* [27], motivating the present discussion. Among the metadata standards mentioned above, the EMMO is unique in that it is a top-level ontology: It operates at the highest level of abstraction and hence constitutes a contribution to the philosophy of modelling and simulation in its own right. Moreover, the EMMO is also particularly innovative as a top-level ontology by following a novel approach that raises non-trivial issues and challenges when users attempt to apply it to concrete scenarios. Some of these issues will be examined here with respect to *mereosemiotics*, the paradigm of the EMMO, understood to consist in a series of postulates that underly not merely the EMMO itself, but its philosophical approach as such, of which there can be a variety of implementations that would all face similar challenges. As Francisco Morgado *et al.* [27] summarize, the EMMO is “the top and middle level ontology developed by the EMMC” and, at the same time, pursues domain-specific targets by aiming to become “the only materials science ontology able to explicitly capture all granularity levels of description of processes and materials.” Accordingly, in addition to its top-level part, the EMMO also contains modules that have the status of domain ontologies. These modules, however, are not the focus of the present discussion which concerns “metaontology” in the sense suggested by Berto and Plebani [28], *i.e.*, it deals with design choices that need to be made when a top-level ontology is developed. For a discussion of the related, but distinct problem of making appropriate use of the existing domain-specific metadata standards in materials modelling, the reader is referred to an article dedicated to that topic [29].

The present work therefore deals with the ontological task of characterizing what models and simulations *are* and how to describe this formally in a coherent way at an abstract level [8]. It is structured as follows: In Section 2, those aspects from the approach of the EMMO are isolated that are most fundamental and can thus be seen as constituting a philosophical or ontological paradigm, here called mereosemiotics. Section 3 introduces three challenges that any ontology implementing this paradigm needs to address when it is applied to scenarios from engineering data technology, and Section 4 discusses four design choices to be made top-level ontology developers. Section 5 concludes this article by characterizing the top-level ontology design space that is accessible to ontologies that implement mereosemiotics.

2 TOP-LEVEL ONTOLOGIES AND MEREOSEMIOTICS

Formal ontology (particularly, metaontology [28]) may appear rather philosophical and abstract, and rightly so, but it can nonetheless deliver actual technical support to successful FAIR data management based on ontology-driven semantic technology. As Guizzardi states, “the opposite of ontology is not non-ontology, but just bad ontology” [3]; a well-designed top-level ontology ensures that concepts and relations are integrated into a sound and coherent formalization. Unsurprisingly, a broad spectrum of

solutions for this problem are available. The reader is referred to a recent UKCIH survey [30] where the existing top-level ontologies are discussed in detail. Many of them precede the EMMO by a long time; *e.g.*, in 2003, Masolo *et al.* [31], conducted a systematic comparison of the BFO, DOLCE, and OCHRE, and at a 2006 “summit,” a *Joint Communiqué* [32] was issued wherein a group of developers representing the BFO, DOLCE, GUM, ISO 15926, ISO 18629, OpenCyc, and SUMO, agreed to henceforth act as “upper ontology custodians.” A comparison of how top-level ontologies deal with temporal phenomena, focusing on the BFO [33], DOLCE [34], and the GFO [35], was published by Galton [36]. With the EMMO, another stone has been added to this mosaic. Most of the pre-existing top-level ontologies distinguish between continuants and occurrents, following Johnson [37]; the EMMO, however, is based on a novel approach that combines 4D mereotopology and Peircean semiotics: Mereosemiotics.

The term *mereosemiotics* is employed here for a paradigm that is constituted by the following core tenets:

- Mereology, *i.e.*, one of the fundamental relations is stated to be (improper) spatiotemporal parthood P such that $x P y$ means “ x is (an improper) part of y .” Mereology can be extended to mereotopology by including an irreducible predicate for connectedness [38, 39, 40, 41].
- Semiotics following Peirce, by which representamina (signs) engage in a dialectical relationship with represented objects through a process the elementary steps of which are conceptualized as triads [42, 43, 44, 45]; representation R , where $s R o$ would mean “the representamen s represents the object o , its referent” always generates a third element when it occurs in practice. By semiosis, in particular, an interpretant s' is created as a new representamen for the same object.
- Nominalism, by which only concrete things exist; ontological commitment exclusively extends to individuals (*i.e.*, in other words, objects, instances, or concreta). Quantification cannot be applied to concepts (*i.e.*, in other words, universals, classes, or abstracta), only to individuals, and only individuals, not concepts, can instantiate a concept. This is in line with the OWL DL approach to formulating ontologies in a decidable fragment of first-order logic [5]; however, it contrasts with Peirce’s point of view, which is Platonist (or realist), *i.e.*, non-nominalist.
- Physicalist materialism; *i.e.*, only that exists which can be conceived of as being physically present either in actual reality or in a hypothetical scenario to which the same laws of physics apply.
- Spatiotemporal monism; all that exists are finite four-dimensional spacetime regions, *i.e.*, objects with a non-zero extension in three spatial dimensions and the fourth dimension of time [46]. Therein, monism means that all objects *are* spacetime, as opposed to a dualist approach where objects would be *in* spacetime so that there would actually be two kinds of entities: The container (spacetime) and the contained (object). Here, instead, an object is identified with its extension in spacetime. Similarly, continuant-occurrent dualism is absent from this paradigm: Processes and the objects that participate in them all enjoy equal status as 4D spatiotemporal regions.
- Semiotic monism; the postulates above already exclude mind-body dualism, since only physical objects exist and therefore mental processes are a kind of physical processes. Beyond this, the present approach also excludes sign-object dualism: There is no rigid distinction between things that can act as a representamen and things that can act as an referent. In Peirce’s work there are many cases where real-world objects, for which there are signs, act as signs themselves, particularly as an “index,” a kind of sign. Examples from Augustine, who also held this position, include the word *sign*, which is itself a sign and can therefore act as its own referent [47], and a situa-

tion where “smoke signifies fire” because we see smoke and know that there must be a fire [48]. This contrasts with Saussurean semiotics where the roles of the signifier and the signified (and the referent, where applicable) are usually understood to be taken by entities from disjoint categories.

Mereosemiotics combines a materialist approach that is well suitable for discussing materials and their properties with semiotics grounded in a complex sign-object interaction. In this way, intricate scenarios from modelling and simulation as well as their relation to experimental data can be captured. Physicalism suggests a view of operating with signs, and a view of thought in general, as being constituted by processes that are simultaneously logical, social, and material; much of Peirce’s work can be read in this way. To model something means to represent it. This point of view is, *e.g.*, advanced by Durán [49] who states that by analysing the way in which this representation occurs, “three kinds of models based on their representational capacity emerge, namely, phenomenological models, models of data, and theoretical models.” The Peircean approach can contribute such an analysis; in particular, it can provide a formalization of representation relations and cognitive processes to be employed by digital infrastructures that deal with multiscale modelling and simulation of physical systems, possibly in combination with experimental procedures and data. Thereby, computer simulations are semioses, *i.e.*, processes by which a representamen (namely, a model) is evaluated, producing another representamen (a different model or a computed property, *i.e.*, an interpretant), and data management needs to take the material preconditions of this process into account, since they constitute relevant metadata: Where was the simulation done, how were the input and the output stored and exchanged, *etc.*? Any research data infrastructure that aims at making models and simulation results FAIR must account for such metadata [50]. Similarly, metadata on the provenance of sensory data are crucial for integrating “field devices 4.0” into model-driven process control [51]. In view of this, such an approach seems to be well suited for its purpose.

3 PRACTICAL CHALLENGES

Mereosemiotics permits describing modelling and simulation as semiosis. In this way, central concepts from simulation-based engineering data technology can be given a fundamental function in terms of entities defined in a top-level ontology; further domain ontologies can then be aligned with the top-level ontology, yielding a data infrastructure where models, data, and services can be integrated in accordance with the principles of FAIR data stewardship [1, 2, 3, 50]. However, the core tenets of this ontological paradigm also raise issues that are not necessarily trivial to address. Three examples are given below:

3.1 Mereotopology and collectives

Mereotopology permits the distinction between spatiotemporal objects that are connected components (called “items” by the EMMO) as well as mereotopological collectives, called “collections” by the EMMO, that are composed of multiple connected components. As discussed by Masolo *et al.* [52], many kinds of collective-like entities occur frequently in practice (and hence will also appear on research data infrastructures) that do not straightforwardly align with this categorization. This includes *all Nobel laureates*, which can act as the referent of a sign, namely, the text “all Nobel laureates;” but some Nobel laureates met and shook hands during their lifetime, by which they are spatiotemporally connected and belong to the same “item,” whereas others did not. It is hard to relate the entity *all Nobel laureates* to the individual people systematically by a single membership relation based on mereotopology. On the

other hand, the “collection” concept is so generic that it also applies to entities that do not really function as collectives at all, since their members are very different from each other, such as *the spatiotemporal fusion (union) of Konrad Zuse and the coffee vending machine on the first floor of the HLRS building*.

3.2 Contingent and counterfactual phenomena

It is inherent in the nature of modelling and simulation that it does not limit itself to that which is actually present. Simulating something that is already right in front of our eyes is the least interesting use of a model; instead, simulations often concern scenarios that are *contingent* (“should we run the process at 200 or 400 kPa?”) or *counterfactual* (“what if the weather after Chernobyl had been different?”). Even practically *impossible* scenarios are relevant, since simulations permit considering idealized states and processes, inaccessible by experiment, by which theories and higher-order models can be validated or parameterized. A knowledge base supporting such use cases needs to be able to consider a simulation of a system where the proposition φ holds while simultaneously asserting that $\neg\varphi$ obtains in actual reality. Similarly, it needs to be able to consider φ as one eligible option during optimization and $\neg\varphi$ as another, without thereby asserting the contradiction that both φ and $\neg\varphi$ simultaneously hold in the real world.

Moreover, as Kaminski [53] points out, it is characteristic of technological innovation to induce states of *hypercontingency* where it is contingent whether φ is contingent, or where it is possible that φ is possible. In model-driven engineering design this is not the exception, but the rule: It may or may not be possible to actually build the nanosurface tailored to a specific purpose that was obtained as a theoretical optimum from a series of simulations. Obversely, if model checking for a discrete process model does not succeed at proving conclusively that a given requirement is satisfied under all circumstances, this does not necessarily mean that the requirement *can* be violated, it means that it *could*. Logically expressing this kind of hypercontingency needs to rely on modal pluralism [54], *e.g.*, by combining different types of possibility operators; in the present example, if φ is a violation of the requirement, $\diamond_1\varphi$ is the possibility of a violation, and $\diamond_0\diamond_1\varphi$ is the possibility of that possibility, where \diamond_0 and \diamond_1 are not the same operator.

3.3 Application to copying, input/output, and data transfer

Dealing with multiple copies of the same data or metadata item is one of the most basic functions of data technology. It is a prerequisite for any exploitation of semantic interoperability in practice. But under what conditions are multiple signs (*e.g.*, models or simulation results), or multiple semioses (*e.g.*, simulations) to be regarded as the same, as equivalent, as similar enough, or as manifestations of the same information content? How can a unification or subsumption of multiple entities under a shared identity be expressed when they occur in different spacetime regions, *e.g.*, on different computers? Peirce defines two different types of signs for this purpose; “the word ‘the’ will usually occur from fifteen to twenty-five times on a page. It is in all these occurrences one and the same word, the same legisign. Each single instance of it is a replica” [42]. The *legisign* is to be understood as a pattern or rule (or a law, hence the name) that needs to be applied to assess whether something is a *replica* of it.

4 VARIATIONS OF THE PARADIGM

On the basis of the core tenets of mereosemiotics, it is possible to take a variety of perspectives with regard to specific issues, some of which are explored below. In particular, this concerns 1) the relation between semiotics and physicalism; 2) the relation between triads and dyads in Peircean semiotics; 3) the role of objects from fictional or counterfactual scenarios and their participation in semiosis; 4) the equivalence of multiple instances or copies of the same symbol, data item, or simulation workflow.

4.1 Relation between semiotics and physicalism

Two variants will be considered:

- *Semiotics is fundamental* (element S_1); *i.e.*, it is not assumed that the relation of represented and representing entities during semiosis can be expressed straightforwardly in physical terms. This includes conceptualizations where representation/semiosis and spatiotemporal parthood are both fundamental on an equal footing, or where symbolic reasoning is considered on the basis of formal logic only, without specifying any precise connection to the characterization of spacetime.
- *Physical participation is fundamental* (element P_1); in this view (held by the EMMO), semiosis is a process, *i.e.*, a spatiotemporal region, in which the sign and the interpretant actually participate physically; arguably, this means that the representamen and the cognition in which it is involved are 4D regions that overlap spatiotemporally. In a *perception*, *e.g.*, a measurement or experiment, the represented object needs to be physically present as well. In the case of an *interpretation*, the physical participation of the object in the process is not required; the representation is carried over from the previous step, ultimately pointing back to a perception where the object was present.

Peirce's semiotics contains aspects of both elements: It is fundamentally non-nominalist insofar as it admits the existence of universals, and universals cannot participate in a physical process in the same way as individuals that are understood to be spatiotemporal regions. However, Peirce requires a "real causal connection" to obtain between the sign and the object; for this purpose, he distinguishes between semioses where the object is necessarily physically present and cases where the causal connection between the sign and the object may be indirect. The latter include references to hypothetical entities (permitted by Peirce) which by their nature cannot overlap with (and thereby participate in) a non-hypothetical process.

4.2 Relation between triads and dyads

Two types of realizations of the paradigm can be distinguished on the basis of treating the semiotic triad either as fundamentally irreducible or as non-elementary and reducible to a dyadic relation:

- *Irreducibility of the semiotic triad* (element I_2). Following Peirce, mere dyadic representation of the type *s R o* does not occur; representation can only be realized in combination with a third element. Hence, the relations that are elementary to semiosis do not connect the sign directly to the object but, instead, the sign to the semiosis process, the object to the semiosis process, *etc.*
- *Elementary dyadic representation* (element E_2). Obversely, the relation that is seen as elementary to semiosis is the dyadic one that directly connects the representamen and the referent, as it is the case with the relation "hasSign" in the EMMO.

4.3 Modal propositions

A rough classification might begin by differentiating ontologies that permit making statements on that which is possible, impossible, contingent, *etc.* (*cf.* Section 3.2), from ontologies that do not permit this:

- *Modal propositions can be expressed* (element M_3). Top-level ontologies that account for modality can further be distinguished in a variety of ways; this includes whether they implement Kripke’s notion of possible worlds [55] (or even “impossible worlds” [56]) or reject that interpretation [57], modal monism versus modal pluralism, and the distinction between contingentism, according to which the existence of an individual may be contingent, and necessitism, stating that everything (*i.e.*, all that exists) necessarily exists and that modal operators can only be applied to relations [58].
- *No modal propositions can be expressed* (element N_3). Mereosemiotics does not require the notions of possibility and necessity; a top-level ontology following this paradigm can restrict itself to describing how things are rather than how they could or must be. This is the case for the EMMO.

Peirce’s semiotics explicitly permits referring to hypothetical objects through signs that are factually present; an ontology that aims at capturing such scenarios therefore needs to be of the first type.

4.4 Relation between multiple replicas

Possible approaches for implementing the replica-legisign formalism, *cf.* Section 3.3, may include:

- *Unification by universals* (element U_4). Concreta can share a feature by partaking in the same abstractum. This is closest to Peirce’s realism (*i.e.*, Platonism). For an implementation that is compatible with nominalism, a variety of solutions can be conceived of; *e.g.*, sets or collectives may be defined that contain individuals that are similar in a certain respect, or an equivalence relation may be employed to state that its subject and object are replicas of the same item. What these approaches have in common is that it is *externally posited* that two items are similar: The strings “the” and “the” are the same word because the knowledge base states it. It is not possible for one interpreter to recognize them as the same and for another to believe them to be different.
- *Absence of unification* (element A_4). If whatever exists is spacetime, and exists as spacetime, different regions of spacetime by definition cannot be the same; accordingly, “the” and “the” are just different physical objects, one of which is printed more to the left, while the other is printed more to the right. It is *not posited* that they are similar. This corresponds to an ontology that strongly prioritizes mereology (or mereotopology) over semiotics, as it is the case for the EMMO.
- *Unification by semiosis* (element S_4). This solution proposes that assessing equivalence is a process of pattern matching, which is a semiosis. Thereby, the sign is a pattern, the object is the item that is matched against the pattern, and the interpretant is the outcome of the pattern matching process, which may be acceptance, rejection, or any other assessment of how the object matches the pattern. If two objects match the same pattern, they are equivalent in a certain respect; however, this is *not externally posited*, but subject to evaluation, where multiple interpreters may disagree.

5 CONCLUSION

By combining elements from Sections 4.1 to 4.4, a variety of strategies can be followed for the design of a top-level ontology and its application to research data infrastructures in the engineering sciences. The choices for each of the four points mentioned above can be made independently; no combinations of elements appear to be absolutely irreconcilable with each other. Therefore, the whole product space

$$\{P_1, S_1\} \times \{E_2, I_2\} \times \{M_3, N_3\} \times \{A_4, S_4, U_4\}$$

is accessible, providing a landscape of possible types of top-level ontologies within the paradigm of mereosemiotics. Peirce permits a reading that positions him comparably closely to the top-level ontology type $S_1I_2M_3U_4$, as far as such a claim may be upheld for any ontology that follows nominalism rather than Platonist realism. The EMMO opposes Peirce on each of the four issues. However, the main purpose here should not consist in an exegesis of Peirce’s body of work, since there may be very good reasons to deviate from his views; it remains to be explored by future work to what extent the different types from the accessible design space are suitable as a semantic foundation for research data infrastructures.

Acknowledgment. This article reports on a contribution to the Minisymposium on Ontology-Based Materials Modelling, Optimization, and Design Applied to Modelling Translation Services and Business Decision Support Systems (OBM-MODA-MTS-BDSS) at WCCM-ECCOMAS 2020, which was held in 2021. This work was supported by activities of the Innovation Centre for Process Data Technology (Inprodat e.V.), Kaiserslautern; the minisymposium was organized by N. A. Konchakova and P. Klein. The co-authors M.T.H. and B.S. acknowledge funding from the German Research Foundation (DFG) through the National Research Data Infrastructure for Catalysis-Related Sciences (NFDI4Cat) within the National Research Data Infrastructure (NFDI) programme of the Joint Science Conference (GWK).

REFERENCES

- [1] J. BICARREGUI: 2016, ‘Building and sustaining data infrastructures: Putting policy into practice’. Technical report. doi:10.6084/m9.figshare.4055538.v2.
- [2] S. COLLINS, F. GENEVA, N. HARROWER, S. HODSON, S. JONES, L. LAAKSONEN, D. MIET-CHEN, and R. PETRAUSKAITÉ: 2018, *Turning FAIR into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data*. EU Publications. doi:10.2777/54599.
- [3] G. GUIZZARDI: 2020, ‘Ontology, ontologies and the “I” of FAIR’. *Data Intel.* **2**, 181–191. doi:10.1162/dint_a.00040.
- [4] D. ALLEMANG and J. HENDLER: 2011, *Semantic Web for the Working Ontologist*. Morgan Kaufmann, 2nd edition. ISBN 978-0-12-385965-5.
- [5] F. BAADER, I. HORROCKS, C. LUTZ, and U. SATTLER: 2017, *An Introduction to Description Logic*. Cambridge Univ. Press. ISBN 978-0-521-87361-1.
- [6] G. GYGLI and J. PLEISS: 2020, ‘Simulation foundry: Automated and F.A.I.R. molecular modeling’. *J. Chem. Inf. Model.* **60**, 1922–1927. doi:10.1021/acs.jcim.0c00018.
- [7] P. N. EDWARDS, M. S. MAYERNIK, A. L. BATCHELLER, G. C. BOWKER, and C. L. BORGMAN: 2011, ‘Science friction: Data, metadata, and collaboration’. *Social Studies of Science* **41**, 667–690.

- [8] C. TURNITSA, J. J. PADILLA, and A. TOLK: 2010, ‘Ontology for modeling and simulation’. In: *Proc. 2010 WSC*. IEEE, pp. 643–651. doi:10.1109/WSC.2010.5679124.
- [9] R. GRUNZKE and OTHERS: 2014, ‘Standards-based metadata management for molecular simulations’. *Concur. Comput.: Pract. Exper.* **26**, 1744–1759. doi:10.1002/cpe.3116.
- [10] B. SCHEMBERA and D. IGLEZAKIS: 2019, ‘The genesis of EngMeta: A metadata model for research data in computational engineering’. In: *Metadata and Semantic Research*. Springer, pp. 127–132. doi:10.1007/978-3-030-14401-2_12.
- [11] B. SCHEMBERA and D. IGLEZAKIS: 2020, ‘EngMeta: Metadata for computational engineering’. *Int. J. Metadata Semant. Ontol.* **14**, 26–38. doi:10.1504/IJMSO.2020.107792.
- [12] S. STEPHAN, M. T. HORSCH, J. VRABEC, and H. HASSE: 2019, ‘MolMod – an open access database of force fields for molecular simulations of fluids’. *Mol. Sim.* **45**, 806–814. doi:10.1080/08927022.2019.1601191.
- [13] M. T. HORSCH and OTHERS: 2020, ‘Semantic interoperability and characterization of data provenance in computational molecular engineering’. *J. Chem. Eng. Data* **65**, 1313–1329. doi:10.1021/acs.jced.9b00739.
- [14] M. T. HORSCH and OTHERS: 2020, ‘Ontologies for the Virtual Materials Marketplace’. *Künstl. Intel.* **34**, 423–428. doi:10.1007/s13218-020-00648-9.
- [15] M. T. HORSCH, S. CHIACCHIERA, M. A. SEATON, I. T. TODOROV, B. SCHEMBERA, P. KLEIN, and N. A. KONCHAKOVA: 2020, ‘Pragmatic interoperability and translation of industrial engineering problems into modelling and simulation solutions’. In: *Proc. DAMDID/RCDL 2020*. Voronezh State Univ., pp. 50–53. ISBN 978-5-6045486-0-8.
- [16] M. T. HORSCH, S. CHIACCHIERA, W. L. CAVALCANTI, and B. SCHEMBERA: 2021, *Data Technology in Materials Modelling*. Springer. ISBN 978-3-03068596-6.
- [17] EMMC COORDINATION AND SUPPORT ACTION: 2017, ‘EMMC Translation Case Template’. <https://emmc.info/emmc-translation-case-template/>; date of access: 23rd January 2021.
- [18] CEN-CENELEC MANAGEMENT CENTRE: 2018, ‘Materials modelling: Terminology, classification and metadata’. CEN workshop agreement 17284:2018 E.
- [19] D. HRISTOVA-BOGAERDS and OTHERS: 2019, ‘EMMC Translators’ Guide’. Technical report. doi:10.5281/zenodo.3552260.
- [20] A. F. DE BAAS (ed.): 2017, *What Makes a Material Function? Let me Compute the Ways . . .*. EU Publications. ISBN 978-92-79-63185-6.
- [21] J. MORBACH, A. WIESNER, and W. MARQUARDT: 2008, ‘OntoCAPE 2.0: A (re-)usable ontology for computer-aided process engineering’. *Comp. Aid. Chem. Eng.* **25**, 991–996. doi:10.1016/S1570-7946(08)80171-X.
- [22] W. MARQUARDT, J. MORBACH, A. WIESNER, and A. YANG: 2010, *OntoCAPE: A Re-Usable Ontology for Chemical Process Engineering*. Springer. doi:10.1007/978-3-642-04655-1.
- [23] P. BORST, H. AKKERMANS, and J. TOP: 1997, ‘Engineering ontologies’. *Int. J. Hum.-Comp. St.* **46**, 365–406. doi:10.1006/ijhc.1996.0096.

- [24] H. CHEONG and A. BUTSCHER: 2019, ‘Physics-based simulation ontology: An ontology to support modelling and reuse of data for physics-based simulation’. *J. Eng. Design* **30**, 655–687. doi:10.1080/09544828.2019.1644301.
- [25] E. GHEDINI, G. GOLDBECK, J. FRIIS, A. HASHIBON, and G. SCHMITZ: 2020, ‘European Materials and Modelling Ontology: Version 1.0.0-beta’. <https://emmo-repo.github.io/versions/1.0.0-beta/emmo.html>; date of access: 10th February 2021.
- [26] G. GOLDBECK, E. GHEDINI, A. HASHIBON, G. J. SCHMITZ, and J. FRIIS: 2019, ‘A reference language and ontology for materials modelling and interoperability’. In: *Proc. NWC 2019. NAFEMS*, p. NWC_19_86.
- [27] J. FRANCISCO MORGADO, E. GHEDINI, G. GOLDBECK, A. HASHIBON, G. J. SCHMITZ, J. FRIIS, and A. DE BAAS: 2020, ‘Mechanical testing ontology for digital-twins: A roadmap based on EMMO’. In: *Proc. SeDiT 2020. CEUR-WS*, p. 3.
- [28] F. BERTO and M. PLEBANI: 2015, *Ontology and Metaontology*. Bloomsbury. ISBN 978-1-4411-9195-3.
- [29] M. T. HORSCH, J. FRANCISCO MORGADO, G. GOLDBECK, D. IGLEZAKIS, N. A. KONCHAKOVA, and B. SCHEMBERA: 2021, ‘Domain-specific metadata standardization in materials modelling’. In: *Proc. DORIC-MM 2021*. UKRI STFC. To appear.
- [30] C. PARTRIDGE, A. MITCHELL, A. COOK, D. LEAL, J. SULLIVAN, and M. WEST: 2020, ‘A survey of top-level ontologies: To inform the ontological choices for a foundation data model’. Technical report, UK Construction and Innovation Hub. doi:10.17863/CAM.58311.
- [31] C. MASOLO, S. BORGO, A. GANGEMI, N. GUARINO, and A. OLTRAMARI: 2003, ‘WonderWeb Deliverable D18: Ontology Library (final)’. Technical report, ISTC-TNR.
- [32] L. OBRST, P. CASSIDY, S. RAY, B. SMITH, D. SOERGEL, M. WEST, and P. YIM: 2006, ‘The 2006 Upper Ontology Summit Joint Communiqué’. *Appl. Ontol.* **1**, 203–211.
- [33] R. ARP, B. SMITH, and A. D. SPEAR: 2015, *Building Ontologies with Basic Formal Ontology*. MIT Press. ISBN 978-0-262-52781-1.
- [34] S. BORGO and C. MASOLO: 2010, ‘Ontological foundations of DOLCE’. In: *Theory and Applications of Ontology: Computer Applications*. Springer, pp. 279–295. doi:10.1007/978-90-481-8847-5_13.
- [35] H. HERRE: 2010, ‘General Formal Ontology (GFO): A foundational ontology for conceptual modelling’. In: *Theory and Applications of Ontology: Computer Applications*. Springer, pp. 297–345. doi:10.1007/978-90-481-8847-5_14.
- [36] A. GALTON: 2018, ‘The treatment of time in upper ontologies’. In: *Proc. FOIS 2018*. IOS, pp. 33–46. doi:10.3233/978-1-61499-910-2-33.
- [37] W. E. JOHNSON: 1924, *The Logical Foundations of Science*. Cambridge Univ. Press.
- [38] N. ASHER and L. VIEU: 1995, ‘Toward a geometry of common sense: A semantics and a complete axiomatization of mereotopology’. In: *Proc. IJCAI’95*. Morgan Kaufmann, pp. 846–852. ISBN 978-1-55860-363-9.

- [39] A. C. VARZI: 1996, 'Parts, wholes, and part-whole relations: The prospects of mereotopology'. *Data Knowl. Eng.* **20**, 259–286. doi:10.1016/S0169-023X(96)00017-1.
- [40] B. SMITH: 1996, 'Mereotopology: A theory of parts and boundaries'. *Data Knowl. Eng.* **20**, 287–303. doi:10.1016/S0169-023X(96)00015-8.
- [41] B. SMITH and A. C. VARZI: 2000, 'Fiat and bona fide boundaries'. *Philos. Phenomenol. Res.* **60**, 103–119. doi:10.2307/2653492.
- [42] C. S. PEIRCE: 1955, 'Logic as semiotic: The theory of signs'. In: *Philosophical Writings of Peirce*. Dover, pp. 98–119. ISBN 978-0-48620217-4.
- [43] C. S. PEIRCE: 1991, *Peirce on Signs*. Univ. North Carolina Press. ISBN 978-0-80784342-0.
- [44] J. ZEMAN: 1977, 'The esthetic sign in Peirce's semiotic'. *Semiotica* **19**, 241–258. doi:10.1515/semi.1977.19.3-4.241.
- [45] T. L. SHORT: 2007, *Peirce's Theory of Signs*. Cambridge Univ. Press. ISBN 978-0-521-84320-1.
- [46] S. L. WILLIAMS: 2010, 'Spatiotemporal monism and region theory: A non-standard account of the nature of spacetime'. M.A. thesis, UC Boulder.
- [47] A. AUGUSTINUS: 1998, *De magistro: Über den Lehrer*. Reclam. ISBN 978-3-15-002793-6.
- [48] A. AUGUSTINUS: 2002, *Die christliche Bildung*. Reclam. ISBN 978-3-15018165-2.
- [49] J. M. DURÁN: 2018, *Computer Simulations in Science and Engineering*. Springer. doi:10.1007/978-3-319-90882-3.
- [50] B. SCHEMBERA and J. M. DURÁN: 2019, 'Dark data as the new challenge for big data science and the introduction of the scientific data officer'. *Philos. Technol.* **33**, 93–115. doi:10.1007/s13347-019-00346-x.
- [51] M. MAIWALD: 2020, 'Integrated and networked systems and processes: A perspective for digital transformation in thermal process engineering'. *ChemEngineering* **4**, 15. doi:10.3390/chemengineering4010015.
- [52] C. MASOLO, L. VIEU, R. FERRARIO, S. BORGO, and D. PORRELLO: 2020, 'Pluralities, collectives, and composites'. In: *Proc. FOIS 2020*. IOS, pp. 186–200. doi:10.3233/FAIA200671.
- [53] A. KAMINSKI: 2015, 'Potenzialerwartungen von Biotechnologien: Ein Modell zur Analyse von Konflikten um neue Technologien'. *Filoz. Društvo* **26**, 519–539. doi:10.2298/FID1503519K.
- [54] K. FINE: 2002, 'The varieties of necessity'. In: *Conceivability and Possibility*. Clarendon, pp. 253–281. ISBN 978-0-19825090-6.
- [55] S. KRIPKE: 1980, *Naming and Necessity (2nd ed.)*. Harvard Univ. Press. ISBN 978-0-674-59846-1.
- [56] F. BERTO and M. JAGO: 2019, *Impossible Worlds*. Oxford Univ. Press. ISBN 978-0-19881279-1.
- [57] B. VETTER: 2011, 'Recent work: Modality without possible worlds'. *Analys. Rev.* **71**, 742–754. doi:10.1093/analys/anr077.
- [58] T. WILLIAMSON: 2015, *Modal Logic as Metaphysics*. Oxford Univ. Press. ISBN 978-0-19-870943-5.