Innovationszentrum für Prozessdatentechnik e.V.

Technical report no. 2021–D

original version

OSMO: Ontology for Simulation, Modelling, and Optimization

Date: 9th July 2021

Authors: Horsch, M. T.; Toti, D.; Chiacchiera, S.; Seaton, M. A.; Goldbeck, G.; Todorov, I. T.

Dissemination:

- Submitted to FOIS 2021, Ontology Showcase track

Funding information:

- DFG, project NFDI4Cat, project no. 441926934
- Horizon 2020, project VIMMP, grant agreement 760907

Accessibility:

- doi: 10.5281/zenodo.5084393
- https://zenodo.org/communities/inprodat/

OSMO: Ontology for Simulation, Modelling, and Optimization

Martin T. Horsch[1,2], Daniele Toti[3], Silvia Chiacchiera[2], Michael A. Seaton[2], Gerhard Goldbeck[4] and Iilian T. Todorov[2]

[1] *High Performance Computing Center Stuttgart, Nobelstr. 19, 70569 Stuttgart, Germany*
[2] *UK Research and Innovation, STFC Daresbury Laboratory, Keckwick Ln, Daresbury WA4 4AD, UK*
[3] *Catholic University of the Sacred Heart, via della Garzetta 48, 25133, Brescia, Italy*
[4] *Goldbeck Consulting Ltd., St. John's Innovation Centre, Cowley Rd, Cambridge CB4 0WS, UK*

Abstract

This work describes the ontology OSMO, *i.e.*, an ontologization and extension of MODA, a workflow metadata standard that constitutes a mandatory requirement within a number of European calls and projects in the context of materials modelling. OSMO was developed within the Horizon 2020 project VIMMP (Virtual Materials Marketplace) and is part of a larger effort in ontology engineering driven by the European Materials Modelling Council, with the European Materials and Modelling Ontology (EMMO) as its core. As such, OSMO provides connections and alignments with other related domain ontologies in computational engineering, including the EMMO itself. This work summarizes the domain, purpose, and design choices underlying OSMO, commenting on the implementation of OSMO and its applications.

Keywords

applied ontology, simulation workflows, research data provenance, process data technology, MODA (Model Data), OSMO, European Materials and Modelling Ontology (EMMO)

1. Introduction

The Ontology for Simulation, Modelling, and Optimization (OSMO) is a metadata standard for modelling and simulation workflows in computational engineering [1]. Comparable semantic artefacts from related lines of development include EngMeta [2, 3], MODA [4] (see below), MVWR [5], and PIMS-II [5, 6, 7]; research data provenance in general, not tailored to materials modelling, is addressed by widespread ontologies such as PROV-O [8] and the BPMN 2.0 ontology [9]. OSMO is part of the system of marketplace-level domain ontologies developed by the Virtual Materials Marketplace (VIMMP) project for use in the logical layer of the digital marketplace platform [3, 10], supporting interoperability between internal components and with external digital platforms. The platform interoperability aspects rely on the line of work toward domain-ontology

FOIS 2021 (Ontology Showcase track), Bolzano–Bozen, 13th–16th September 2021
EMAIL: martin.horsch@hlrs.de (M. T. Horsch); daniele.toti@unicatt.it (D. Toti)
ORCID: 0000-0002-9464-6739 (M. T. Horsch); 0000-0002-9668-6961 (D. Toti); 0000-0003-0422-7870 (S. Chiacchiera); 0000-0002-4708-573X (M. A. Seaton); 0000-0002-4181-2852 (G. Goldbeck); 0000-0001-7275-1784 (I. T. Todorov)

CEUR Workshop Proceedings (CEUR-WS.org)

development advanced by the European Materials Modelling Council (EMMC ASBL), and accordingly, the European Materials and Modelling Ontology (EMMO) is employed as a foundational ontology to facilitate crosswalks [3, 11, 12, 13]. OSMO comprises a complete ontologization of MODA ("Model Data"), a pre-existing metadata standard based on tables with plain-text content intended for human-to-human communication, *cf.* CEN workshop agreement 17284:2018 [4]. MODA is a template for the standardized description of materials models; it is endorsed by EMMC ASBL along with the EMMO, and the uptake of MODA has been mandatory in a series of calls from Horizon 2020 and continues to be an explicit requirement in the Horizon Europe work programme [14, pp. 173, 177]. Accordingly, a substantial corpus of MODA workflow descriptions has been compiled from Horizon projects; a representative selection of such workflows is included in an Annex of the Review of Materials Modelling (RoMM) [15] with OSMO as the ontology version of MODA, these human-readable metadata become machine-processable and can be integrated into ontology-based semantic technology frameworks.

2. Ontology design

The original purpose of RoMM [15] and MODA [4] was to guide users toward a complete high-level documentation of materials models, starting from use cases and including all relevant computational details. In this regard, it was meant to provide all necessary aspects for describing, reproducing, curating, and interfacing with other models, guiding modellers step-by-step throughout their modelling workflows. However, the semantic expressivity of the graph notation from MODA is rather limited, making workflow descriptions too ambiguous for common applications (*e.g.*, technical-level reproducibility). The connections between steps, or sections, of a workflow lack clearly characterized semantics altogether; connections are uniformly shown as blue arrows in the MODA graph notation, by which different ways of coupling, linking, user interaction, and data transfer cannot be distinguished. Moreover, it does not immediately facilitate an integration of resources distributed over the semantic web: Descriptors in MODA are essentially plain-text labels; they do not canonically contain any resolvable or non-resolvable identifiers for RDF semantic-web resources or FAIR digital objects. To integrate MODA metadata into knowledge bases that rely on ontology-driven semantic technology, it was necessary to develop an ontology version of MODA; specifically, this requirement from the VIMMP project (and, similarly, the MarketPlace project with which VIMMP aims at complete platform interoperability[1]) motivated the development of OSMO. Therefore, the main motivations behind the design of OSMO design were, on the one hand, to make MODA-based annotations machine-processable and, on the other hand, to address the limitations of MODA mentioned above. These objectives were attained by constructing a domain ontology and providing an alignment and interconnection with other related ontologies, including the EMMO, creating crosswalks to a variety of domain ontologies from other projects [12, 13]. Accordingly, OSMO integrates the representation of application cases

[1] Virtual Materials Marketplace (VIMMP): H2020 grant agreement no. 760907, https://vimmp.eu/; MarketPlace: H2020 grant agreement no. 760173, https://the-marketplace-project.eu/.

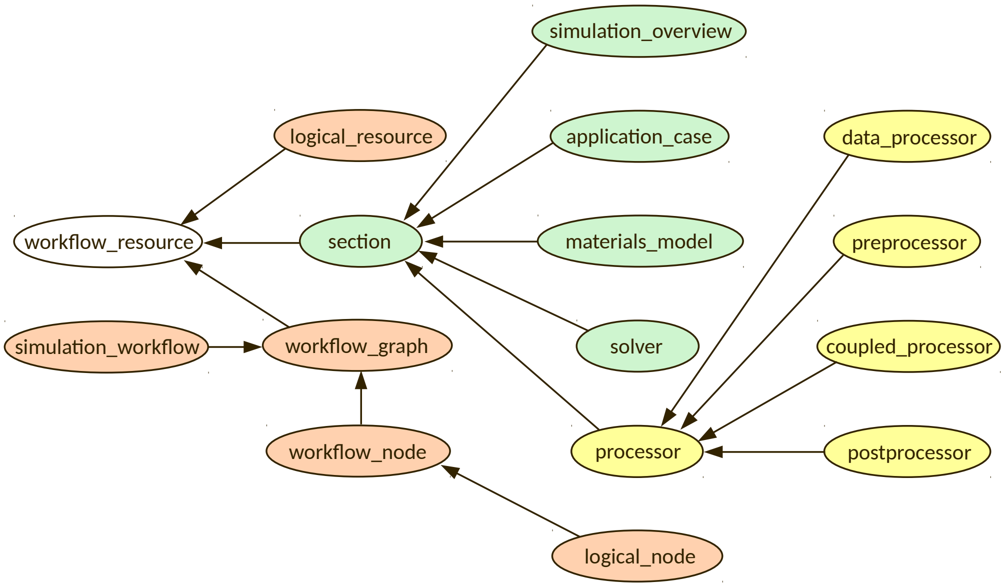

Figure 1: Fragment of the taxonomy for the workflow_resource branch of OSMO; arrows indicate subsumption, *i.e.*, rdfs:subClassOf.

(including MODA use cases), solvers, models, and processors from MODA; these are enriched with additional semantic information given by the structuring of these elements into a class hierarchy, as well as by including elements for logical data transfer and explicit dependencies among the workflow sections [1], *cf.* Fig. 1.

3. Datasets and applications

As a metadata standard for modelling and simulation workflows, application scenarios for OSMO fall into two categories: First, it can be used to specify workflows or parts of a workflow in advance at multiple stages, *e.g.*, to support project planning (Horizon proposals often contain MODA workflows) and simulation deployment by open simulation platforms (OSPs) [1, 5]; specifically, the description can restrict itself to the application case (*e.g.*, a translation case [16, 17] or a MODA use case [4]), by which a problem description can be submitted to a materials modelling marketplace where it can be addressed by "translators," *i.e.*, experts in academia-to-business knowledge transfer [10, 16, 17]. Second, simulation results can be annotated with OSMO metadata to characterize their provenance, which is crucial to making research data FAIR by facilitating their reproducibility and providing *epistemic grounding* to substantiate the extent to which the results can be trusted [6, 18]; this aspect of using MODA and OSMO is most relevant to research data infrastructures where discipline-specific data with a heterogeneous provenance (*e.g.*, grounded in data-driven, model-driven and experimental scientific practices) need to be integrated while keeping track of their epistemic status [19, 20].

Accordingly, OSMO is mainly used to provide metadata on plans or requirements, on the one hand (for workflows to be carried out in the future, or alternatively for purposes or reporting, in hindsight), and metadata meant to support workflow orchestraction or to document the provenance of simulation results at a more technical level, on the other hand. In the first case, the annotation is naturally comparably abstract, since it aims at supporting project management or matchmaking during translation where it is necessary to communicate information in a more aggregated way. It is here that the MODA templates work best; RoMM [15, pp. 79–374] includes a series of examples from 36 projects. The ontologization provided by OSMO, including the additional features mentioned above, supports technical-level documentation, facilitating the exchange of workflow metadata with components such as OSPs [1, 21]. To develop platform interoperability for the *molecular model database* (MolMod DB) of the Boltzmann-Zuse Society for Computational Molecular Engineering (BZS), *cf.* Stephan *et al.* [22], OSMO was used to specify rigid intermolecular pair potentials for low-molecular fluids [23]. Industrially relevant application cases from VIMMP include multi-phase flow simulation with OpenFOAM [24, 25] and molecular dynamics simulation of wetting and adhesion [26].

4. External ontologies and crosswalks

The VIMMP system of ontologies is connected to about 20 external semantic artefacts [3, 10]. These links are not specific to OSMO itself which is mainly connected to the other VIMMP marketplace-level domain ontologies directly; the relation between OSMO and the Materials Modelling Translation Ontology (MMTO), which is particularly close, is discussed in previous work [16]. The projects associated with EMMC ASBL, including VIMMP, jointly mainly rely on the EMMO [11] for semantic interoperability between platforms and to align their respective domain ontologies [12, 13]. However, MODA and the EMMO are not very closely related in their foundational approach and require a major effort to obtain a viable alignment; the main challenge in situating OSMO in relation to other semantic artefacts therefore consisted in facilitating crosswalks from MODA to the EMMO. Two crosswalk strategies have been developed for this purpose. It depends on the circumstances and requirements what connection is more suitable for a concrete task; however, they can also be applied jointly, complementing each other.

First, mediated by the PIMS-II mid-level ontology for cognitive processes [5, 6, 7], sections are mapped to steps of a cognitive action and described using sign-object-interpretant triads, following the approach to semiotics introduced by Peirce, a formalism that is also employed by the EMMO [5, 11, 27]. This alignment scheme, discussed in detail elsewhere [5], preserves the structure of a MODA/OSMO workflow correctly; fine-grained information content may be lost, since PIMS-II is comparably close to the EMMO itself and operates at a substantially higher degree of abstraction than MODA/OSMO. The second crosswalk employs multiple elements for intermediate transformation steps, including VIPRS (VIMMP primitives), the mid-level ontology module of the VIMMP semantic interoperability architecture [3], the European Virtual Marketplace Ontology (EVMPO), which reflects agreements between the VIMMP and MarketPlace project consortia at an early stage of development, and the EMMO-VIMMP Integration (EVI) module. Therein, OSMO concepts are mapped to EMMO concepts by means of EVMPO and EVI, whereas OSMO relations are mapped to *mereosemiotic chain relations* [3, 6], transforming the structure of the knowledge graph; for an example, *cf.* previous work [23]. This connection makes less use of foundational elements of the EMMO; however, it retains greater taxonomical detail, and the detailed content of the MODA tables is better preserved if this route is followed. As a major mechanism of the second crosswalk, OSMO concepts including logical_term, section, section_aspect, and workflow_resource are mapped to the EMMO concept Symbolic, and where one entity can be shown to act as a representamen for another, it is mapped to a Sign, and hasSign from the EMMO is employed for the representation relation; in this way, a simulation_overview (the OSMO rendering of the MODA cover sheet) is understood to act as a representamen for a simulation. All relations from OSMO are subsumed under EMMO relations (or chains of EMMO relations), and all concepts from OSMO are subsumed under EMMO concepts.

5. Technical details and availability

OSMO is specified using the RDF, RDFS and OWL formalisms,[2] and it has been deployed and tested on a Stardog[3] triplestore. Supporting tools for development and testing have been developed in Java version 1.8+. The OWL axiomatization of the ontology in TTL format is available at http://www.molmod.info/semantics/osmo.ttl.

Acknowledgments

This work was funded from the European Union's Horizon 2020 research and innovation programme through grant agreement no. 760907, Virtual Materials Marketplace; it was facilitated by activities of the Innovation Centre for Process Data Technology (Inprodat e.V.). The co-author M.T.H. also acknowledges funding from DFG project no. 441926934, NFDI4Cat, within the NFDI programme of the German Joint Science Conference (GWK).

References

[1] M. T. Horsch, et al., J. Chem. Eng. Data 65 (3) (2020) 1313–1329, doi:10.1021/acs.jced.9b00739.

[2] B. Schembera, D. Iglezakis, Int. J. Metadata Semant. Ontol. 14 (1) (2020) 26–38, doi:10.1504/IJMSO.2020.107792.

[3] M. T. Horsch, S. Chiacchiera, W. L. Cavalcanti, B. Schembera, Data Technology in Materials Modelling, Springer, ISBN 978-3-03068596-6, 2021.

[4] CEN-CENELEC Management Centre, workshop agreement 17284:2018, CEN, 2018.

[5] H. A. Preisig, P. Klein, N. Konchakova, M. T. Horsch, in: DAMDID 2021 (submitted), 2021.

[2]Resource Description Framework (RDF): http://www.w3.org/RDF/; RDF Schema (RDFS): http://www.w3.org/TR/rdf-schema/; Web Ontology Language (OWL): https://www.w3.org/OWL/.

[3]Stardog Enterprise Knowledge Graph: https://www.stardog.com/platform/.

[6] M. T. Horsch, in: FOUST 2021 (submitted), preprint doi:10.5281/zenodo.4849611, 2021.
[7] M. T. Horsch, Mereosemiotics: Five scenarios, technical report 2021–B, Inprodat e.V., doi:10.5281/zenodo.4679522, 2021.
[8] L. Moreau, P. Groth, J. Cheney, T. Lebo, S. Moles, J. Web Semant. 35 (4) (2015) 235–257, doi:10.1016/j.websem.2015.04.001.
[9] M. Rospocher, C. Ghidini, L. Serafini, in: P. Garbacz, O. Kutz (Eds.), Proc. FOIS 2014, IOS, ISBN 978-1-61499-437-4, 133–146, 2014.
[10] M. T. Horsch, et al., Künstl. Intell. 34 (3) (2020) 423–428, doi:10.1007/s13218-020-00648-9.
[11] EMMC ASBL, European Materials and Modelling Ontology (EMMO), https://github.com/emmo-repo/EMMO, 2021.
[12] J. Francisco Morgado, E. Ghedini, G. Goldbeck, A. Hashibon, G. J. Schmitz, J. Friis, A. de Baas, in: Proc. SeDiT 2020, CEUR-WS, 2020.
[13] D. Höche, N. Konchakova, M. Zheludkevich, T. Hagelien, J. Friis, in: Proc. WCCM-ECCOMAS 2020, Scipedia, doi:10.23967/wccm-eccomas.2020.263, 2021.
[14] European Commission, Horizon Europe work programme. 7. Digital, industry, and space, decision C(2021)4200, EC, 2021.
[15] A. F. de Baas (Ed.), What Makes a Material Function? Let Me Compute the Ways, EU Publications Office, Luxembourg, ISBN 978-92-79-63185-6, 2017.
[16] M. T. Horsch, S. Chiacchiera, M. A. Seaton, I. T. Todorov, B. Schembera, P. Klein, N. A. Konchakova, in: Proc. DAMDID 2020, Springer, ISBN 978-3-030-81199-0, doi:10.1007/978-3-030-81200-3_4, 2021.
[17] M. Pezzotta, J. Friis, G. J. Schmitz, N. Konchakova, D. Höche, D. Hristova-Bogaerds, P. Asinari, L. Bergamasco, G. Goldbeck, Report on translation case studies describing the gained experience, Technical report, EMMC, doi:10.5281/zenodo.4457849, 2021.
[18] M. Williams, Philos. Phenomenol. Res. 60 (3) (2000) 607–612, doi:10.2307/2653818.
[19] S. Herres-Pawlis, O. Koepler, C. Steinbeck, Angew. Chem. Int. Ed. 58 (32) (2019) 10766–10768, doi:10.1002/anie.201907260.
[20] C. Wulf, et al., ChemCatChem, doi:10.1002/cctc.202001974, 2021.
[21] S. P. Huber, et al., Sci. Data 7 (2020) 300, doi:10.1038/s41597-020-00638-4.
[22] S. Stephan, M. T. Horsch, J. Vrabec, H. Hasse, Mol. Sim. 45 (10) (2019) 806–814, doi:10.1080/08927022.2019.1601191.
[23] M. T. Horsch, et al., in: Proc. ISWC 2020 Demos and Industry Tracks, CEUR-WS, 134–137, 2020.
[24] A. Marcato, G. Boccardo, D. Marchisio, Chem. Eng. J. 417 (2021) 128936, doi:10.1016/j.cej.2021.128936.
[25] G. Tronci, A. Buffo, M. Vanni, D. Marchisio, Comp. Fluids, doi:10.1016/j.compfluid.2021.105026, 2021.
[26] M. Chiricotto, G. Giunta, H. A. Karimi-Varzaneh, P. Carbone, Soft Mater. 18 (2–3) (2020) 140–149, doi:10.1080/1539445X.2019.1701497.
[27] M. T. Horsch, S. Chiacchiera, B. Schembera, M. A. Seaton, I. T. Todorov, in: Proc. WCCM-ECCOMAS 2020, Scipedia, doi:10.23967/wccm-eccomas.2020.297, 2021.