Ontology and semantic net based technology applied to smart corrosion protection modelling & simulation

Peter Klein [a,*], Heinz A. Preisig [b], Martin Thomas Horsch [c,d] and Natalia Konchakova [e]
[a] *Department Optimization, Fraunhofer ITWM, Kaiserslautern, Germany*
E-mail: peter.klein@itwm.fraunhofer.de
[b] *Department of Chemical Engineering, Norwegian University of Science and Technology, Trondheim, Norway*
E-mail: heinz.preisig@chemeng.ntnu.no
[c] *STFC Daresbury Laboratory, UK Research and Innovation, Daresbury, United Kingdom*
E-mail: martin.horsch@stfc.ac.uk
[d] *School of Psychology and Computer Science, University of Central Lancashire, Preston, United Kingdom*
E-mail: mthorsch@uclan.ac.uk
[e] *Institute of Surface Science, Helmholtz-Zentrum Hereon, Geesthacht, Germany*
E-mail: natalia.konchakova@hereon.de

Abstract. Ontology-based integrated materials modelling for an active protective coating system design is presented and applied to a practical example. For this purpose, an ontological methodology implemented using the Process Modelling suite (ProMo) is developed to be used with an open simulation platform (OSP), *i.e.*, a workflow management and orchestration framework that can be integrated into digital infrastructures. The target infrastructures, which are under development in various Horizon 2020 projects include modelling marketplaces, open innovation platforms, and open translation environments, among others. Semantic interoperability for the communication between the involved digital infrastructures relies on ontologies that are aligned with Elementary Multiperspective Material Ontology (EMMO) at the top level; in particular, the Ontology for Simulation, Modelling, and Optimization (OSMO) is employed as an ontologization of the Modelling Data metadata schema (MODA) in combination with the Physicalistic Interpretation of Modelling and Simulation Interoperability Infrastructure (PIMS-II) mid-level ontology. The challenge of addressing semantic heterogeneity is addressed by working toward crosswalks between domain-specific and mid-level ontologies for industrially relevant problems, where knowledge graph transformation is evaluated as a candidate solution for a future implementation strategy. The involved semantic artefacts are platform-agnostic, and their EMMO compliance allows for a specification of executable modelling and simulation workflows on multiple EMMO-compliant OSPs. We demonstrate the presented approach on an industrially relevant example for developing active corrosion protection of metallic surfaces.

Keywords: ontology alignment, process data technology, graph transformation, active protective coating, molecular modelling

1. Introduction

Europe has been dedicating substantial efforts in its Framework Programs FP7 and Horizon 2020 to the development of Open Simulation Platforms (OSPs), which consist of model orchestration tools for the construction of

*Corresponding author. E-mail: peter.klein@itwm.fraunhofer.de.

materials modelling workflows and materials modelling and simulation tool repositories.[1] Aiming to make materials modelling widely available, particularly to industrial users, these OSPs generally enrich their simulation backbones by including additional services on top of their OSP cores. This goal is achieved by interoperating with external platforms, usually also from Horizon 2020 projects, such as business decision support systems, materials modelling marketplaces, open translation environments, and open innovation platforms [2]. Technologies developed in such projects advance dedicated domain-specific solutions for digital twins [3] and interoperability with disciplinary research data infrastructures [4] and within EOSC (European Open Science Cloud [2, 5]). One central idea is to ensure that data from modelling and simulation become and remain FAIR[2] [11, 12], avoiding the generation of *dark data* that cannot be validly reused [13–15]. The harmonisation of these developments is grounded in ongoing work on the Elementary Multiperspective Material Ontology[3] (EMMO) top-level ontology [16–18], the MODA (Modelling Data) metadata schema [19] for documenting simulation workflows, and OSMO (Ontology for Simulation, Modelling, and Optimization [20–23]), the ontology version of MODA; work has also progressed toward an analogous approach to provenance metadata standardisation for experimental research workflows based on CHADA (Characterization Data [24]). Similarly, the provenance of chemical compounds from a network of chemical reactions has been characterised ontologically [25]. To support standardised documentation of research workflows at large, combining modelling and simulation, synthesis, and characterisation, among other many potential steps, PIMS-II has been developed as a mid-level ontology for cognitive processes [26, 27]. All these efforts, at various levels of maturity and rigour, support aspects of *translation* [21, 28–30], which at its core deals with the design of simulation workflows that are fit for industrial challenges.

Beside translating industrial problems to modelling workflows, translators typically need to indirect implement EMMO-interoperable semantic technology related to the required industrial and simulation workflows. In this paper, we argue that humans providing these translation services are unnecessarily expected to be ontology experts. In our opinion, this is an overburdening of the translator role that needs to be (semi-) automatised. In this paper, we present such an approach. In previous work, the authors constructed a set of ontology-based tools with the idea to provide a usable abstraction on top of EMMO and MODA supporting interoperability between OSP cores [31]; in these and similar scenarios, a human in the loop, referred to as a Scientific Data Officer (SDO), *cf.* Schembera and Durán [32], can greatly improve the outcome from the ontological tools developed. Here, these tools are applied in a somewhat different arena, namely for the construction of simulation workflows for industrially relevant applications and the potential to implement automation of simulation workflow construction based on the developed ontological tools. It turns out that this challenge is tractable, but a human inspection and modification of the outcome is still advantageous for constructing optimal modelling workflows and for providing an insightful documentation. However, even if no such human support is available and the involved researchers are unable to provide and communicate metadata by themselves without any dedicated support, all is not lost – automated metadata extraction can go a long way to reduce the amount of dark data [14, 33]. The present work discusses the release of a corrosion inhibitor in an active protective coating as a practical demonstrator, thereby demonstrating the viability of our ontology-based modelling approach.

The remainder of this article is structured as follows: In Section 2, the use case of developing *active corrosion protection* is introduced from a phenomenological and industrial point of view. We introduce the suggested MODA documentation for the used case and discuss some of the weaknesses of the MODA approach in this example. Along the lines of this use case, a brief introduction into OSMO, an ontologised version of MODA, and of a ProMo based modelling language for modelling workflows as cognitive processes represented in PIMS-II follow. In Section 3, a crosswalk from the MODA/OSMO to PIMS-II/EMMO is described based on a knowledge graph transformation system. This transformation is rule-based and can therefore be automatised – it can substantially alter the shape of the knowledge graph, which is necessary for a pair of metadata standards that fundamentally differ in their design, as is the case for MODA and the EMMO. Section 4 discusses the potential for dedicated human support at ensuring metadata compliance between different modelling workflow realisations on OSPs. The conclusion, Section

[1] This article is based on a FOMI 2021 workshop manuscript [1] that was presented at the *2021 Bolzano Summer of Knowledge*.
[2] Findable, accessible, interoperable, and reusable [5–10].
[3] Previously known as the European Materials and Modelling Ontology, with the same abbreviation.

5, discusses the results of this work from the perspective of the primary users of the presented approach, traditionally called "translators" in the context of the European materials modelling ecosystem.

2. Use case and simulation workflow: from MODA to OSMO/ProMo

2.1. Inhibitor leaching in active protective coating

It is well known that aggressive environmental exposure and corrosion-induced damage often limit engineering structures' service life. Functional coatings are the most effective and efficient way to protect bridges, off-shore equipment, cars, trains, buildings, ships, aircraft, and daily consumables. The main role of the coating in degradation and damage protection is to provide a dense barrier against corrosive species. However, defects appear in the protective coatings during exploitation of the coated structures opening direct access for aggressive agents to the metallic surface. Including anti-corrosive inhibitor agglomerates, so-called nano-containers, into coating polymer matrix provides "smart" corrosion protection when the coating (barrier) is damaged (cracks) by its ability to release a corrosion inhibitor (leaching) which accumulates in the cracks [34]. The inhibitor, released from the nano-containers, forms an active layer on the substrate surface and thus prevents the corrosion development on the metallic substrates [35].

Corrosion inhibitor release (leaching) is the main active protective mechanism of corrosion inhibiting primers used in the field of protective coatings [36]. These primers are loaded with sparingly soluble pigments to provide a reservoir of corrosion inhibiting ions released through ion exchange as depicted in Fig. 1 (top image).

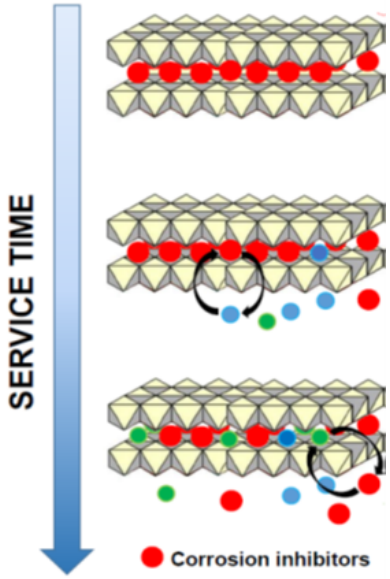

Fig. 1. Cartoon of atomistic ion exchange mechanism in pigments mobilizing inhibitor molecules (red). The inhibitor release is accompanied by trapping corrosives like chlorides (green) or water (blue). ©Smallmatek Lda, Portugal

In case of a defect, the coating will be exposed to the environment and absorb moisture, which triggers the mobilisation of the corrosion inhibitor in the coating matrix (see Fig. 1, middle image). The mobilised molecules diffuse through the coating matrix and leach into the damage region (crack), where they agglomerate on top of the metal substrate. The formed inhibitor film prevents the substrate surface from corrosion. The molecular structure of the corrosion inhibitor, the inhibitor leaching rate, and the critical concentration in the damage region (crack) are

the key factors influencing the immediate and the long-term corrosion prevention by this active protective coating process.

The development of new coating formulations is a complex process, usually performed by enormous time and resources consuming experimental efforts. An effective way to accelerate novel coating development is by supplementing experimental efforts with molecular and multiscale modelling and simulation to obtain quantitatively reliable *surface process-morphology-property relationships* (OMEB, from an abbreviation of the German technical term [37]).

The approach, as summarised above, is one of the core scientific ideas of the ongoing H2020 EU project VIPCOAT [38] which is funded under call DT-NMBP-11-2020, "Open Innovation Platform for Materials Modelling." The main scientific objectives of the project are to establish an Open Innovation Platform for the development of inhibiting active protective coatings and corresponding accelerator tests for assessing their in-service durability, as well as to promote the development of a green technology for active protective coatings based on materials modelling and optimisation.

One of the most effective classes of inhibitor agglomerates is Layered Double Hydroxides (LDH) nanocontainers [39]. If LDH nano pigments are present in the corrosive media, a significant corrosion rate reduction is observed. LDHs are anion-exchange systems consisting of stacks of positively charged, mixed-metal hydroxide layers, intercalated by layers of anionic species and solvent molecules, see Fig. 1. These environmentally friendly nanostructures have already demonstrated their ability to control the release of active inhibition species under certain environmental conditions. Their functioning is twofold: not only to release the species that impart active protection but also to trap corrosive ionic agents (for example, Cl^-) [40].

2.2. MODA description of an inhibitor performance simulation workflow

The example workflow described here closely resembles the development process of a new coating. The primer of the coating is assumed to be predefined, for example as specified by the end customer. The usual implemented experimental trial-and-error development process starts by selecting the inhibiting molecule known to form LDH structures. To this end, a large number of experiments is done. In each experiment, an inhibiting molecule is dissolved in water and brought in contact with the metallic substrate and the performance is experimentally evaluated. In the next step, the LDH structures are generated experimentally in a self-assembly process using candidate inhibitor molecules. The LDH nano-containers are then mixed with the primer to form a liquid coating formulation which is brushed on the metallic surface, where the liquid coating cures and results in a solid film that contains a microstructured distribution of the nano-containers. The function of the coating loaded by the nano-containers is then assessd at a system level by introducing damage (a small crack) into the coating film and monitoring the inhibition performance under various lab conditions emulating in-service-life. In particular these experiments are very time consuming.

Next, we partially describe the flow of four models using the MODA standard [19] supporting the experimental development. We note that the experimental steps of formulating the liquid coating, the application on a metallic surface and the curing of the coat are **not supported by modelling** in VIPCOAT and are therefore not part of the modelling workflow since no predictive modelling approaches are available.

Any MODA description of materials modelling, which proceeds in a top-down manner, starts with an overview section. In this use case, the overview is shown in Fig. 2.

The individual models in the MODA workflow are identified along the following lines of reasoning:

- **Model 1**: The first set of experiments provide information on the inhibition performance of dissolved molecules. The data gathered are used to train a machine learning model (ML), which predicts the inhibitor performance of molecules that are not necessarily experimentally characterised at the metallic substrate/surface. This approach, as applied in VIPCOAT, is described in Ref. [41].
- **Model 2**: The coating applied to a metal substrate consists of a polymer matrix enriched with inhibitor nanocontainers. The mobilisation of inhibiting molecules from a single nano-container is modelled using molecular modelling, resulting in a descriptor role describing the inhibitor mobilisation. The result of the simulation is a nano-scale model that describes the ion exchange mechanism depicted in Fig. 1.

MODA for active corrosion protection in project VIPCOAT

1	USER CASE	Active corrosion protection by leaching of corrosion inhibitor molecules from a coating through a water film (condensation of water after aircraft landing) to the metallic surface in a coating defect (micro-crack).	
2	CHAIN OF MODELS	MODEL 1	Machine Learning model for inhibitor efficiency prediction as a function of inhibitor concentration in water close to metal surface.
		MODEL 2	Mobilization of inhibitor molecules from pigments/capsules.
		MODEL 3	Mesoscopic CFD model simulating the transport and leaching of mobilized inhibitor molecules from coating matrix in defect volume under wet conditions.
		MODEL 4	Macroscopic CFD system model representing coating as pseudo-phase described in an effective medium approximation for prediction of corrosion protection efficiency.
		DATA MINING METHODOLOGY	Model 1: Supervised Learning.
3	PUBLICATION PEER-REVIEWING THE DATA	Data retrieved from scientific network and owned by industrial partners of VIPCOAT consortium.	
4	ACCESS CONDITIONS	At this point in time: confidential, FAIR access conditions under negotiation.	
5	WORKFLOW AND ITS RATIONALE	Model 1 is used to identify candidate inhibitors based on their inhibition efficiency. Model 2 is used to extract an inhibitor mobilization model for pigments used in Model 3 as input to simulate the transport and leaching of the coating and to extract pseudo-phase transport coefficients for Model 4 which simulates inhibition (system) performance in use.	

Fig. 2. Sketch of the MODA overview section for a simulation workflow of inhibitor selection, inhibitor leaching and inhibitor performance at a system level. Note: MODA's "user case" corresponds to "use case".

- **Model 3**:The leaching within the microstructure is modelled using mesoscopic Computational Fluid Dynamics (CFD). The mobilisation of the inhibiting molecules from the nano-containers in the coating matrix uses the molecular descriptor roles from the previous step as a source term. The microstructure of the coating is fully resolved. The post-processing of the modelling results "wipes out" the microstructure and represents the coating in its dynamical state as a homogenised pseudo phase whith the capability to model the leaching process from the coat into the defect above the metallic substrate.
- **Model 4**: If a defect occurs and is exposed to humidity, water condenses in the defect area. The pseudo phase leaching model is then coupled to a second CFD model responsible for simulating the transport of inhibitor molecules throuhg the water in the defect to the substrate surface, where they adsorb to form a protective layer. We model the complete system with the pseudo phase representation and use this simulation to optimise the overall coating system.

This list, in the sense of MODA, should be part of row 5 in the MODA overview, Fig. 2, as the rationale behind the workflow in a full-blown description.

A graphical representation of the present simulation workflow using the MODA standard is depicted in Fig. 3.

In the language of MODA, we use a so-called loose coupling of models since the models are executed consecutively. We note that more model coupling modes are described in the MODA standard which are not discussed here. However, the findings in this work apply *mutatis mutantis* to these coupling modes as well.

In Fig 4, we briefly sketch for Model 3 only the more detailed MODA part.

A full MODA description of each model in any MODA overview section calls for many more details. In particular, the mathematical equations, their boundary and initial conditions need to be fully specified, together with a numerical solver used in the MODA workflow. A full discussion of these details is beyond the scope of this work.

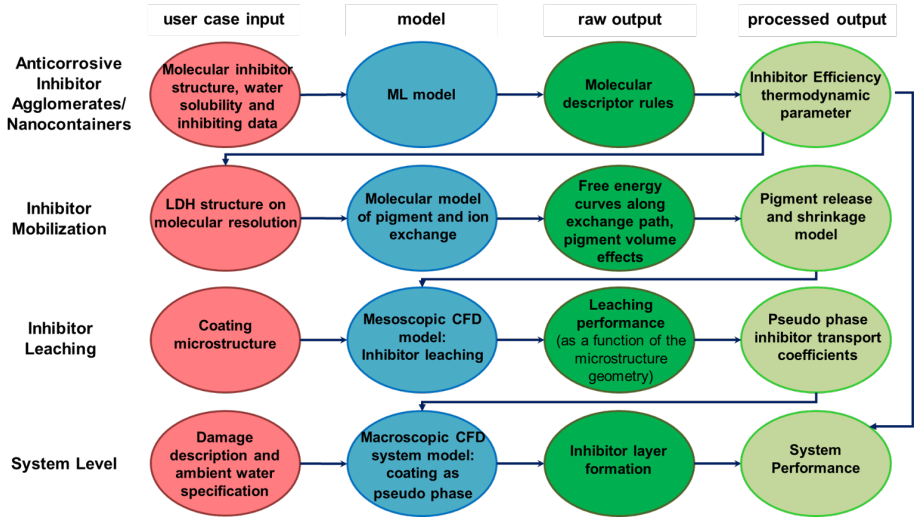

Fig. 3. MODA fiche overview for a simulation workflow: interaction between models of inhibitor selection, inhibitor leaching and inhibitor performance at a system level.

1	Aspect of the User Case/System to be Simulated	
1.1	Aspect of the User Case to be simulate	Transport through and leaching from the coating matrix into the water filled crack.
1.2	Material	Coating consists of binder and pigments, formulation data from standard data bases.
1.3	Geometry	Taken from SEM / synchrotron images or generated by stochastic geometry
1.4	Time Lapse	The leaching process may take minutes, up to approximately 1h.
1.5	Manufacturing process or in-service conditions	Ambient conditions and temperature variations +/- 40 °C.
1.6	Publication on this data	No publication on leaching of novel active protective coating, data acquisition part of the VIPCOAT project

2	Generic Physics Of The Model Equation		
2.1	Model type and name	Transport in porous media.	
2.2	Model entity	Scalar field.	
2.3	Model Physics/ Chemistry equation	**Equation**	Diffusion equation.
		Physical quantities	Inhibitor concentration.
2.4	Materials relations	**Material Relation (MR)**	Functional relation of chemical potential of inhibitor in water.
		Physical quantities/ descriptors for each MR	Inhibitor diffusion constant and chemical potential in water (experimental). From model 2: inhibitor release/mobilization model from pigment (LDH).
2.5	Physics formulation of the conditions	Ambient pressure and temperature, ambient air constitution and influence of pH.	
2.6	Simulated input	Pseudo-phase model.	

3	Solver and Computational translation of the specifications		
3.1	Numerical Solver	Finite volumes (explicit or implicit in time)	
3.2	Software tool	ParPac	
3.3	Time step	Will be adapted to problem stiffness.	
3.4	Computational Representation	Physics Equation, Material Relations, Material	Diffusion equation with source term from Model 2 on top of the pigment surface, discretized using cubical finite volumes.
3.5	Computational boundary conditions	Robin boundary conditions on water box face parallel to coat surface.	
3.6	additional Solver Parameters	• Timestep to be used.	

4	POST PROCESSING	
4.1	The processed output	Descriptor rules for pseudo-phase model of coating as a leaching entity.
4.2	Methodologies	Regression analysis and homogenization.
4.3	Margin Of Error	Given the overall uncertainties related to the initialization, a sensitivity analysis with trustworthy eigenvalues/eigenvectors of the Fisher matrix is appreciated.

Fig. 4. Sketch of the MODA sections describing Model 3, used for simulating the inhibitor leaching, of the MODA overview, Fig. 2, as one subsequence in the workflow 3.

It shoud be noted that the modelling workflow in Fig. 3 accoring to the MODA standard suffers in particular from the absense of a precice semantic meaning of the arrows, which is one motivation to introduce in this work a different graphical notation described in the next section for the use case in this work.

2.3. Physical topology of models and their interfaces

The physical topology is a graphical method to capture the essentials of model supported reasoning. It shows the materials modelling process as a network of extensive-quantity-exchanging capacities. The capacities may be

distributed, meaning the intensive properties are a function of the position or lumped if that is not the case. They are also associated with a relative time scale, namely event, dynamic or constant/static. The surroundings of a process are modelled by a set of infinite capacities, which – in analogy to thermodynamics – we call "reservoirs." Only the intensive properties are known for reservoirs, and the total extensive quantities, *i.e.*, mass and energy, are not balanced. The intensive properties of the reservoir make the embedded process change its state. This distinction between extensive and intensive variables and their materials modelling roles is absent from MODA.

In the demonstration case, *cf.* Section 2.1, we use reservoirs (infinitely large with constant intensive properties) to model the electrolyte resources and distributed systems for the water layer on top, the defect, the combination of coating matrix and electrolyte-containing pores and the inhibitor layer. The adsorption of the inhibitor to the substrate is modelled as a point capacity where the adsorption takes place and distributed system for the inhibitor layer (Fig. 5). For more details on the graphical modelling language and its relation to MODA we refer to [31]. The main criticism of MODA here is that there are no control structures in the MODA workflow description, leading to an ambiguous mapping to executable workflows.

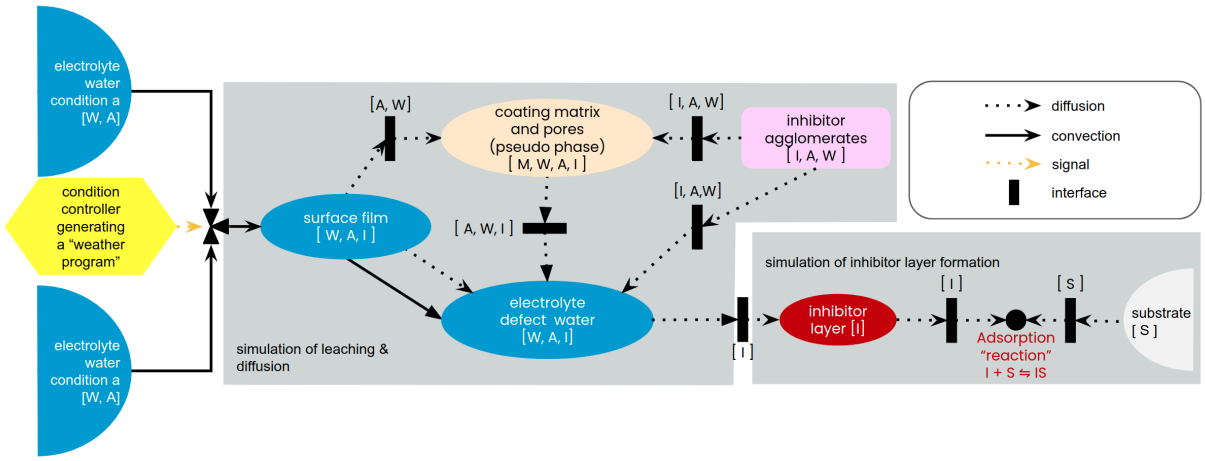

Fig. 5. A possible physical topology for the coating process. The "weather" is here simulated by mixing two streams of electrolyte, allowing for changing concentration as well as switching on/off the supply. The coating matrix and the pores are combined to form a "pseudo phase," which implies that the overall volume normalises properties. The inhibitor leaches into the pores filled with the watery electrolyte in the matrix, thus the pseudo phase, forming an inhibiting layer. The large grey boxes suggest a possible simulation arrangement. Thus, one would simulate the leaching and the diffusion in the crack and the pores together. The building of the inhibitor layers would be solved in a second task.

The inhibitor is stored in a sandwich-type of structure, which captures the inhibitor between layers of support. The support is not exchanging material with its environment, as shown in Fig. 6. Its sole purpose is to contain the inhibitor through a static electrical field, which we do not show in this exhibition. One may view these particles like cookies with a filling of inhibitor. The release of the inhibitor is an ion-exchange process in which the inhibitor ion is replaced by either water or electrolyte components that cause corrosion of the substrate metal, with chlorine being the main one. The production of the inhibitor yields agglomerates of inhibitor-loaded cookies, which yield a cumulative interface of the inhibitor particles with the water phase in the pseudo phase (coating matrix and water in the pores combined) (Fig. 5).

2.4. *OSMO: The ontologized variant of the MODA tables*

For simulation workflows described following the MODA metadata standard, an extensive corpus of examples is available and growing due to the endorsement by the EMMC and incorporation into a coherent line of development from Horizon 2020. The Ontology for Simulation, Modelling, and Optimisation (OSMO [20–23]), one of the domain ontologies from the Virtual Materials Marketplace (VIMMP) project [22, 42–44], is the ontology version of MODA: All metadata provided through MODA can be carried over 1:1 into OSMO and can be integrated into

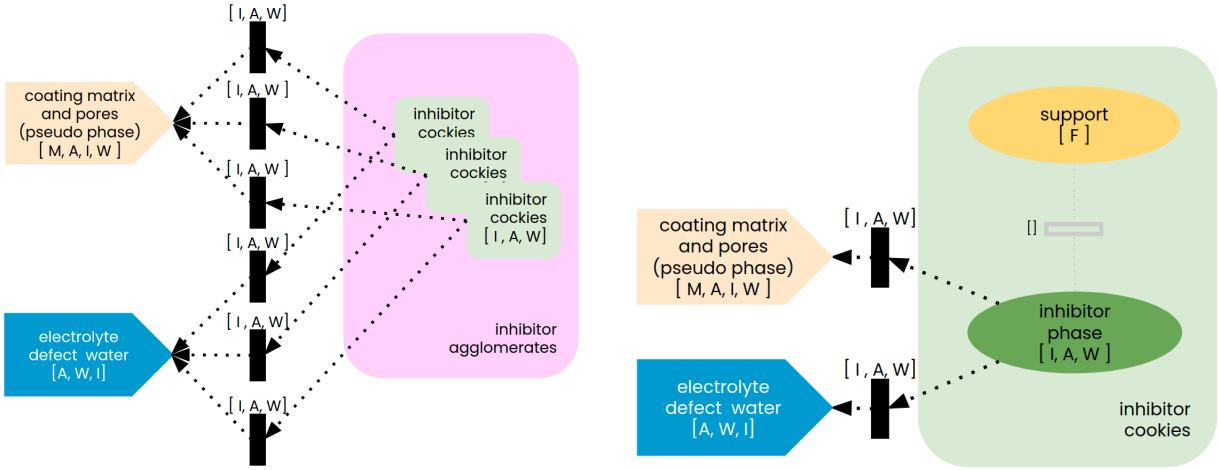

Fig. 6. **Left**: Expansion of the "inhibitor agglomerates" from Fig. 5, showing three representative "cookies" as an example. **Right:** Single cookie, with the support material and the in-between sandwiched ion layer of inhibitor (LDH), active components like chlorine anions, and water.

any Resource Description Framework (RDF) based semantic technology. Thereby, OSMO becomes an ontology that is consistent with MODA and can be employed to annotate simulation results and exchange information on simulation workflows without being tied to any specific file format. Using MODA alone, this would be impossible since MODA itself is not an ontology and does not have any specified correspondence with RDF triples. Following the structure from MODA, OSMO section types include simulation_overview, use_case (generalized to application_case to also cover metadata on *translation* processes as specified by the EMMC [21]), materials_model, solver, and processor.

Additionally, OSMO provides workflow documentation features that go substantially beyond MODA. In particular, the graphical modelling workflow description formalism from MODA [19] is affected by serious shortcomings and ambiguities [20], particularly in view of an executable workflow deployment. One of these problems is that the arrows in a MODA workflow graph (such as Fig. 2) are not equipped with well-defined semantics. For experts within a certain subdiscipline, the arrows will usually have an intuitive meaning – but that is insufficient for making the associated research data FAIR to all stakeholders. Furthermore, attaching semantics to the MODA arrows by an automated post-processing step is impossible since a single MODA graph can represent a variety of very different workflows [20, therein, Fig. 4]. This property completely excludes the potential to automatise the generation of executable workflows on top of materials modelling tools directly from a MODA workflow graph. With OSMO, which also extends the graphical representation of MODA (cf. Fig. 7), one provides additional information on coupling and linking between workflow elements and the meaning and internal structure of the exchanged variables.

ProMo's topology approach captures, compared MODA, more information about the structure of the model. Specifically, the physical topology reflects time scales, distribution effects (intensities being a function of the spatial coordinates), and what is present in the individual parts and transported between them. Also, model control structures are included, reflecting the switching of models as is commonly the case when the number of phases changes. ProMo is based on reductionism applied to specific application domains. The smallest granules are defined as the fundamental entities. In contrast to the MODA, we thus do not attempt a global definition but have it application-domain dependent. The equations representing the individual base entity's behaviour are constructed from a minimal set of fundamental quantities, usually the ones defining the configuration space for thermodynamics and mechanics. This approach leads to internally consistent models following the physical base concepts, mainly the conservation principles. ProMo's topology approach removes the main shortcomings of the MODA, and with given complete information, automatic processing of the information is enabled.

The part of the ProMo workflow shown in Fig. 7 corresponds to the third row in Fig. 3, including the way it is coupled to the elements in the second and the fourth row. The elements of the LDT graph directly correspond to individuals and relations from the OSMO-based ontological representation [20, therein, Tab. 4]. In particular, it is

made clear whether the executable sections of a workflow (*i.e.*, solvers and processors) are *coupled* synchronously or *linked* asynchronously, or whether they exhibit concurrency. Logical resources, the triangles in Fig. 7, can be annotated to provide detailed information on the data exchanged between sections, and it is made explicit whether the logical access to the data occurs upon initialisation, upon termination, or at runtime.

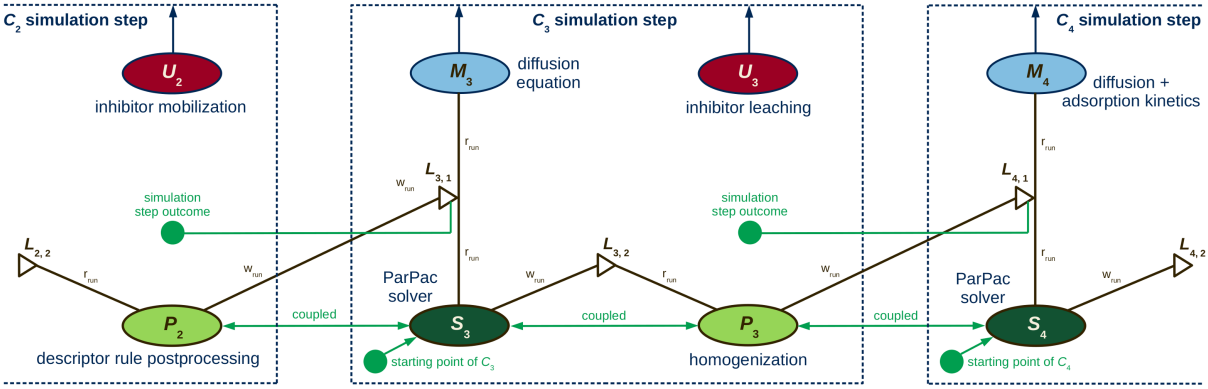

Fig. 7. OSMO logical data transfer (LDT) graph [20] for the VIPCOAT scenario, centered on the model for inhibitor leaching (simulation step C_3), which corresponds to the gray box in the middle of the topology in Fig. 5. In the present approach, the topology is the main source from which a (human) translator would design the simulation workflow.

2.5. *The workflow as a cognitive process: Workflow diagrams enriched with semantics*

The *Physicalistic Interpretation of Modelling and Simulation – Interoperability Infrastructure* (PIMS-II) is a mid-level ontology for cognitive processes. In line with major design choices underlying the EMMO top-level ontology, PIMS-II is based on mereotopology [45–47] and Peircean semiotics [48–50], which are combined to *mereosemiotics* as a coherent ontological paradigm [26, 51]. PIMS-II is formalised both in OWL2 description logic [27] as well as by a series of axioms in modal first-order logic [26]; it also includes a graphical notation, *cf.* Fig. 8. Both the EMMO and PIMS-II implement mereosemiotics, which facilitates an alignment between the two ontological levels. Furthermore, by design, PIMS-II is also well aligned with the *metadata4ing* ontology from NFDI4Ing, the German National Research Data Infrastructure for Engineering [4].

In PIMS-II, following Peirce's approach, an elementary cognitive step is a process that is conceptualised as a *triad*. Building on a previously established representation relation between a *sign* (or representamen) and an *object* (the referent of the representamen), each cognitive step introduces a new, third element, called the *interpretant* in the case of a *semiosis* as conceived by Peirce [48], by which an additional representation relation comes into being. In Fig. 8 and similar diagrams, these triads are visualised as triangles, where the three triadic elements are situated in the corners; blue arrows denote representation relations from the representamen to the referent, and green arrows denote dependency relations between cognitive steps. As in the case of OSMO LDT graphs, the elements of the PIMS-II based graphs all visualise individuals and relations from the associated RDF-based knowledge graph (compare Fig. 12). For the present example, we divided the distributed systems in Fig. 5 into two CFD simulations modelling the behaviour of the pseudo-phase and the leaching process. Additionally, the leaching process is considered by a molecular simulation of a single LDH cookie.

A final, *minimal* representation of the workflow is here given in the form of a simplified Petri net, which uses *gates* instead of transitions to impose causality: An input/output mapping can only be executed if all inputs are available as indicated in Fig. 9. Therefore, each computational step is equipped with an input gate. The gate functions as a transition in a place/transition net following Petri [52]: The computation is only started if all inputs are present. In addition, input data must be persistent. Thus, the gate-compute mechanism guarantees causality. In this way, the present minimal representation is close to well-established varieties of Petri nets such as condition-event nets [53]. Workflows would commonly step in time, which implies that the computational sequences are repeated.

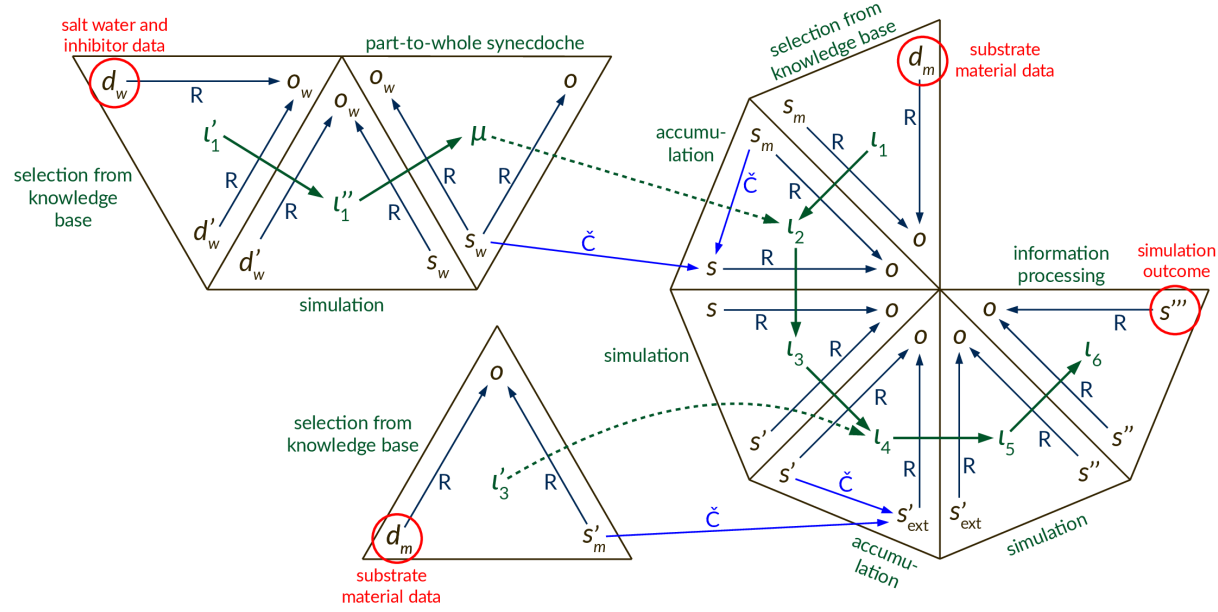

Fig. 8. Modelling and simulation workflow represented as a cognitive process, based on Peircean semiotics, using the graphical notation associated with PIMS-II [26]; arrows labelled Č denote the PIMS-II relation isSemioticallyConstitutiveOf, arrows labelled R denote isRepresentamenFor, unlabelled solid arrows denote directlyGrounds, and unlabelled dashed arrows denote directlyPrecedesStep [27]. See also the correspondence between this diagram and the higher-level representation of the same workflow by a simplified Petri net (Fig. 9).

Thus computational loops are formed, which must be initialised in the beginning. For this purpose, an additional element, namely an input-selection switch, needs to be introduced, which will change from initial conditions to taking the initial conditions for the next step as the final condition of the previous time step. One may also implement a split, where a signal is passed to more than one gate. An example is the starting point, which sends the starting signal to all three activities. Mapping PIMS-II workflows to condition-event nets serve multiple purposes, namely 1) concurrency and dependency analysis for optimising the massively-parallel or distributed deployment and for estimating the computational resources required by workflows, 2) alignment and communication between different platforms (*e.g.*, simulation hubs or research data infrastructures) and, eventually, 3) the integration of MODA/OSMO workflows and other process models via PIMS-II, and domain ontologies aligned with PIMS-II or the EMMO, into ProMo [17, 31, 54].

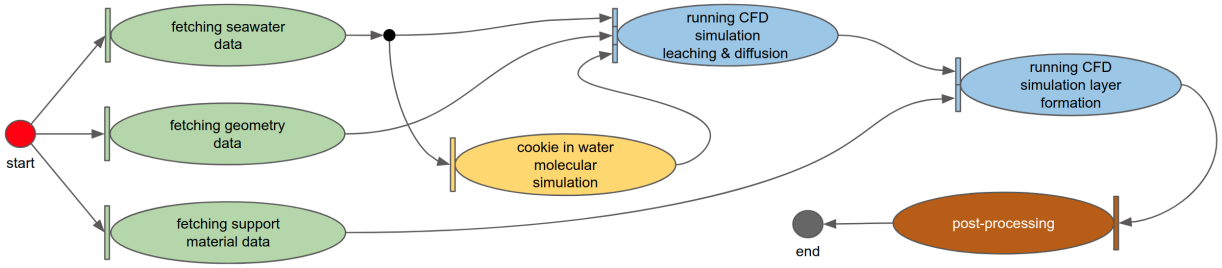

Fig. 9. A slightly simplified Petri-like net representation of the workflow leaving out the "weather" simulation section. Each computational step has an import gate. A computation can only be started once all the inputs are available. Data at the input are persistent. A computational step may have several outputs allowing to trigger other computational steps. The workflow is for one time-step only and assumes that the leaching model is computed only once during each step. Adding time-stepping requires a loop, which needs to be initiated by introducing an input switch as an additional control element. The post-processing step includes activities like graphical representation of the result or the computing of performance criteria.

3. MODA to EMMO crosswalk

3.1. Knowledge graph transformation systems (KGTS)

By a *crosswalk* (applying EOSC terminology [5]), we refer to any algorithm, tool, or specification by which instances of one semantic artefact are systematically mapped to instances of another [55]. This mapping includes conventional ontology alignments obtained as solutions of the ontology matching problem [56–59], but can go beyond that, as it is indeed necessary for workflow and provenance metadata in computational engineering; *n.b.*, this is not due to any specific complexity of the underlying disciplinary matters. Instead, it is a consequence of requirements by the European Commission, which simultaneously endorses both MODA [19] and the EMMO [16–18] as metadata standards to be applied to the same domain of knowledge, requiring a crosswalk between them. In the crosswalk from MODA to the EMMO, it is an intermediate step that turns out to be crucial to the quality of the outcome. On the one hand, MODA has a direct correspondence with OSMO, *cf.* Section 2.4. On the other hand, the mid-level ontology PIMS-II is constructed to be close to the top-level ontology EMMO [26, 27]. However, there is a substantial mismatch between the way in which information is arranged in MODA and the system of relations provided by the EMMO. This fact is reflected in deviations between the structure of knowledge graphs that would correspond to each other in OSMO (close to MODA) and PIMS-II (close to the EMMO). Consequently, employing conventional item-by-item ontology alignments would lead to a major loss of knowledge. Instead, the present work explores a route based on knowledge graph transformation systems (KGTS), *i.e.*, graph transformation systems [60, 61] applied to knowledge graphs based on RDF triples; the present knowledge graphs can all be assumed to comply with $\mathcal{SROIQ}(\mathcal{D})$ description logic [62, 63].

Graph transformation is among the formalisms that are occasionally (though not very frequently) used in semantic-web architectures [64–68], including work on ontology merging [65] and dynamics of knowledge bases [67]. To assess the viability of KGTS for crosswalks between workflow representations in computational molecular engineering, a candidate fragment of a KGTS was developed for OSMO as the source ontology and PIMS-II as the target ontology. Two selected rewriting rules from this fragment are shown in Fig. 10. In all rules, newly created vertices and edges exclusively instantiate concepts and relations from the target ontology, and *each rule deletes at least one instance of a concept or relation from the source ontology*. It substantially restricts the expressive capacity in comparison with graph grammars in general, which are Turing-complete; *n.b.*, however, that the expressive capacity of conventional ontology alignments is still strictly included, while termination after a linear number of transformation steps is guaranteed. It reduces the problem to $\mathcal{O}(mn)$ instances of the graph isomorphism problem, where n is the size of the source graph and m is the number of rules (usually, a constant). While that problem is not known to be solvable in polynomial time, it has been shown to be quasipolynomial [69]. It is also the use case for which engines of semantic technology software (*e.g.*, SPARQL endpoints) are best optimized. Since race conditions can occur between critical pairs of rewriting rules [70], the outcome is not in general uniquely defined; if required, uniqueness of the mapping can be enforced by imposing an order of precedence between applicable rules.

The rules from Fig. 10 illustrate how, particularly by specifying multi-node shape constraints (implementable straightforwardly in SPARQL or SHACL) as a source pattern, the KGTS can retain information that would be lost in a conventional alignment based on immediate conceptual or relational subsumptions: The connection between a simulation (*i.e.*, a semiosis, following Peircean semiotics on which the EMMO is based) and the simulated object is one of the foundational elements for the EMMO, and hence for PIMS-II, which it is crucial to preserve. However, in MODA, and hence in OSMO, it is *not the simulation but the solver* that is immediately represented as a section (corresponding to a page in MODA), and information on the simulated object is *not directly associated with the solver* at all, but with a different section, namely, the use case. Graph-based patterns can take such indirect connections into account.

3.2. Application to the use case scenario

This is illustrated here for the VIPCOAT use case from Section 2.1, applying the candidate KGTS to the part of the workflow consisting of the CFD simulation of leaching and diffusion and the preceding data-combination processing step. The source graph, shown in Fig. 11 (top), corresponds to an annotation by metadata following

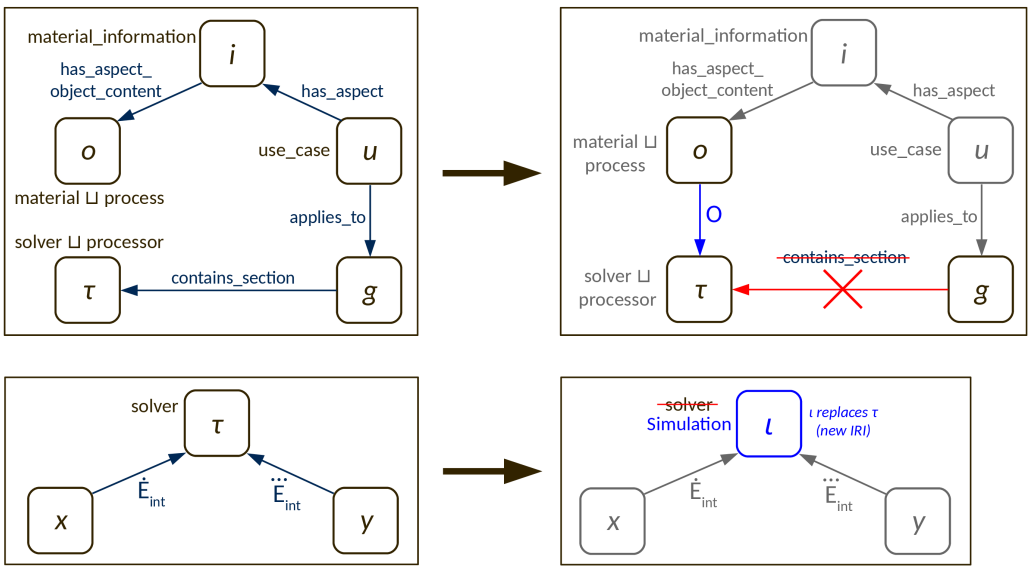

Fig. 10. Two rewriting rules selected from a knowledge graph transformation system (KGTS) that aligns OSMO [20–23, 43, 44] with PIMS-II [26, 27]. Arrows labelled $\dot{E}_{\text{int}}$ denote the PIMS-II relation isSignInInterpretation, arrows labelled $\ddot{E}_{\text{int}}$ denote isInterpretantInInterpretation, and the arrow labelled O denotes isReferentIn. The left-hand side specifies the pattern upon which the applicability of the rule is conditional. The right-hand side specifies the graph which with it is to be replaced; grey colour: Vertices and labels unaffected by the transformation step. **Top:** Where the pattern on the left-hand side is detected, the edge labelled contains_section is deleted and an edge labelled O is created. **Bottom:** Where a vertex (individual) instantiating the concept solver has any edges labelled $\dot{E}_{\text{int}}$ and $\ddot{E}_{\text{int}}$, representing input and output from a simulation step, it is replaced with a *new* vertex, *i.e.*, an individual with a new IRI is introduced, satisfying all properties and occurring in the same assertions as the old individual; then for the new vertex, the instantiation of solver (from OSMO) is deleted, and an assertion is created by which it instantiates Simulation (from PIMS-II). This replacement is necessary because the solver and the simulation cannot be the same individual, but it is desirable for the new element to retain the role of the old element in several respects; other rewriting rules can be applied subsequently to correct any undesired properties that have been carried over from the source individual.

OSMO, *e.g.*, as it would be obtained by digitalizing the MODA input provided by a user of a simulation hub or a research data infrastructure (see Fig. 7). The target graph, obtained by applying a sequence of rewriting rules, is shown at the bottom. The KGTS crosswalk succeeds at retaining the most relevant features from the source graph. This includes information on the dependency (linking) between the two steps of the workflow, the data items communicated from step to step, and the relation between the representing elements and their referent, the simulated leaching and diffusion process. The individuals shared by both knowledge graphs, *i.e.*, nodes that do not undergo deletion or replacement (M_3, o, s, s', s_m, s_w, and U_3) during this crosswalk, are highlighted with double-line borders. The capability to replace source individuals with new individuals (as in the second rule from Fig. 10) substantially increases the viability of the crosswalk; *e.g.*, where MODA/OSMO indicates the presence of a solver in a workflow, this almost always means that in the corresponding PIMS-II workflow there should be a Simulation (at EMMO level, a Semiosis). However, an ontology alignment that would subsume one under the other, osmo:solver $\sqsubseteq$ emmo:Semiosis, would be incorrect; the solver and the semiosis are not the same individual, it is only the presence of one element in the source graph that indicates the presence of the other elements in the target graph. Such correspondences go beyond what can usually be realized by conventional ontology matching.

4. Human intervention: Metadata curation and workflow stewardship

While metadata tools can assist in digitalizing and annotating data, it is advisable to include human support in the process. This includes data stewardship [8, 71, 72] and data curation [73, 74], functions that overlap but differ in nuance. Analyzing requirements from high-performance computing *cf.* Schembera and Durán [32], reveals

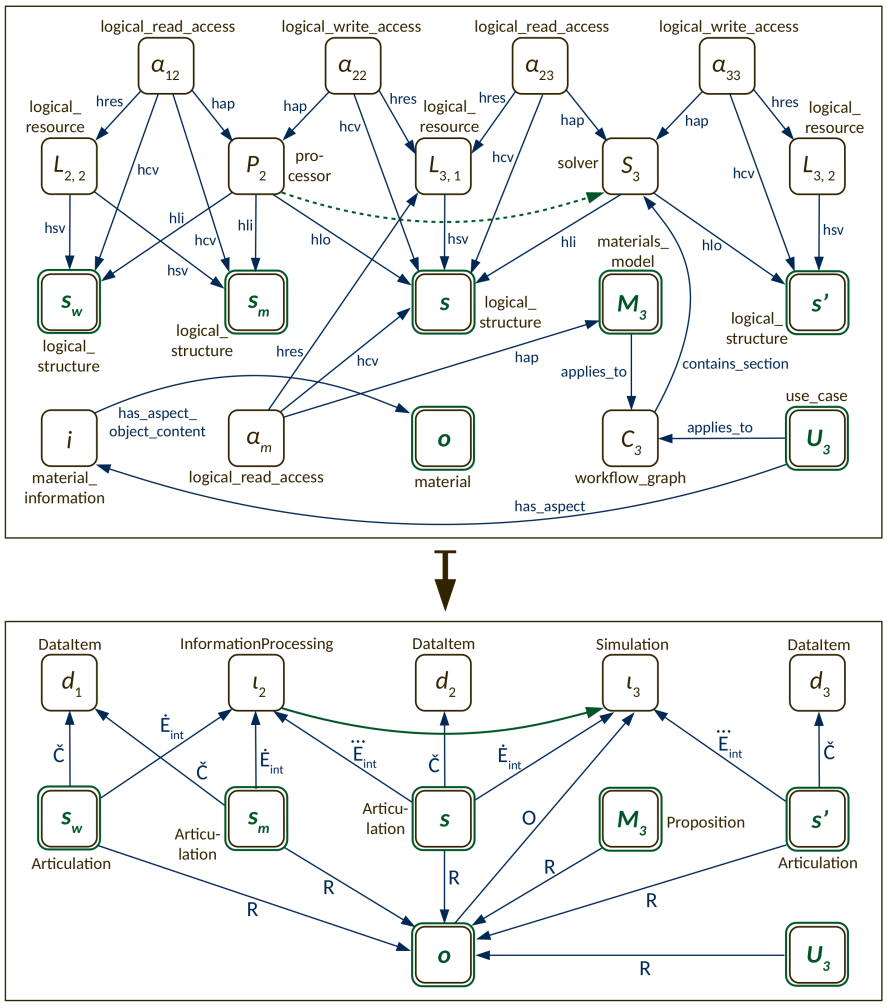

Fig. 11. Source and target knowledge graphs from a potential application of a KGTS implementation of the OSMO to PIMS-II crosswalk. **Top:** Source knowledge graph, containing four of the sections (P_2, U_3, M_3, and S_3) and three of the logical resources ($L_{2,2}$, $L_{3,1}$, and $L_{3,2}$) shown in Fig. 7. Abbreviations for relations from OSMO: "hap" for has_access_point, "hcv" for has_carried_variable, "hli" for has_logical_input, "hlo" for has_logical_output, "hres" for has_resource, and "hsv" for has_stored_variable; the unlabelled dashed arrow denotes is_linked_to. **Bottom:** Target annotation in PIMS-II as it would be constructed automatically by an automated crosswalk implementation. Arrows labelled Č denote the PIMS-II relation isSemioticallyConstitutiveOf, arrows labelled $\dot{E}_{int}$ denote isSignInInterpretation, arrows labelled $\ddot{E}_{int}$ denote isInterpretantInInterpretation, arrows labelled O denote isReferentIn, arrows labelled R denote isRepresentamenFor, and the unlabelled solid arrow denotes directlyGrounds.

that technical and organizational (or "ethical") tasks interrelate, which is best addressed by advancing the role of a Scientific Data Officer (SDO) as a career. The job description for an SDO encompasses data stewardship and curation, legal and procedural control of good practices, and user management [32]. Here, an SDO might reannotate the outcome from the KGTS (Fig. 11, bottom) as illustrated in Fig. 12. Thereby, three kinds of improvements are made: First, instantiations of concepts and relations are made more specific (*e.g.*, from InformationProcessing to Accumulation). Second, the graph structure is simplified by eliminating unnecessary nodes. Third, helpful additional edges are created; here, the model employed in simulation ι_3 occurs in two ways: The data item s and the proposition m. This connection is made explicit by stating that s articulates m (relation $\lhd_a$).

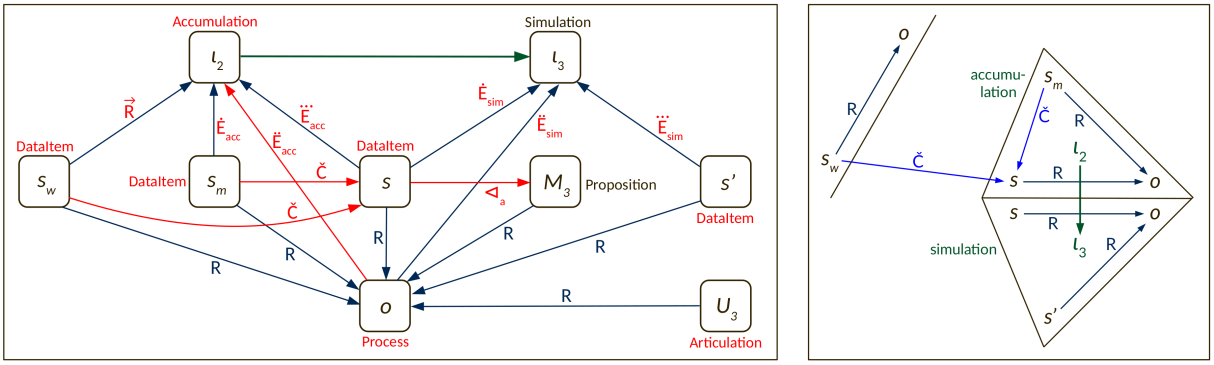

Fig. 12. **Left:** Improved target graph as it might be obtained from an autogenerated target graph as in Fig. 11 (bottom) by the stewardship and curation work of a Scientific Data Officer [32]; arrows labelled $\check{\mathsf{C}}$ denote the PIMS-II relation isSemioticallyConstitutiveOf, the arrow labelled $\dot{\mathsf{E}}_{\mathsf{acc}}$ denotes isAccumulationInputIn, the arrow labelled $\ddot{\mathsf{E}}_{\mathsf{acc}}$ denotes isObjectInAccumulation, the arrow labelled $\dddot{\mathsf{E}}_{\mathsf{acc}}$ denotes isAccumulationOutputIn, the arrow labelled $\dot{\mathsf{E}}_{\mathsf{sim}}$ denotes isSimulationInputIn, the arrow labelled $\ddot{\mathsf{E}}_{\mathsf{sim}}$ denotes isSimulatedIn, the arrow labelled $\dddot{\mathsf{E}}_{\mathsf{sim}}$ denotes isSimulationOutputIn, arrows labelled R denote isRepresentamenFor, the arrow labelled $\vec{R}$ denotes isIncludedInAccumulation, the arrow labelled $\lhd_{\mathsf{a}}$ denotes articulates, and the unlabelled solid arrow denotes directlyGrounds. Therein, the interpretation ι_3 is the leaching simulation step discussed in detail above. **Right:** The corresponding part of Fig. 8.

5. Conclusion

In this work, we apply an ontology-based toolset to support the *translation* process as specified and recommended by the European Materials Modelling Council (EMMC). A minimalistic graphical modelling language with only a few construction elements can represent highly complex model topologies of industrially relevant processes. The present work demonstrates this for active protective coating development, applicable to many different surfaces in their respective operational environment.

The connection between a MODA-based and an EMMO-based annotation of workflows constitutes a challenge. The present work resolves by constructing a context-sensitive crosswalk between the two semantic artefacts. The present knowledge graph transformation system (KGTS) candidate fragment produces acceptable results even without human intervention. Despite illustrating the potential for automatization of a translation process and the viability of the suggested approach, we do not advise deploying complex crosswalks unsupervised. Requirements for human oversight apply particularly strongly to the challenge of mapping information content from one metadata schema to another. The difficulty is precisely due to semantic heterogeneity: Conceptual schemes *appear to be incommensurable* unless a crosswalk has already been accepted as valid by the community of its users (which is an organizational task, requiring an agreement) or mappings are approved on a case-by-case basis, requiring an explicit control and affirmation *each time*. Therefore, it is strictly impossible to substantiate the validity of crosswalks purely by formal verification. If high standards of correctness are to be met (which is certainly not always the case since approximate annotation is often good enough), human supervision by a translator or an SDO will be advisable.

The present work demonstrates that it is possible to devise ontology-based model construction tools for the specific example of active protective coatings. In the future, we plan to explore more industrially relevant application areas in a similar spirit, most probably using a generalization of the approach used in this work to attain much broader translation capabilities. Thus, this work opens a route for performing translation tasks, supported by a (semi-)automated ontologisation of the translation model artifacts, that result in simulation platform agnostic workflows as long as the simulation platform supports the EMMO. Humans working as translators in the multiscale materials modelling eco-system of the EMMC are therefor not necessarily expected to be ontology experts.

Acknowledgments The co-authors P.K., H.A.P., and N.K. acknowledge funding from the Horizon 2020 research and innovation programme of the EU by grant agreement no. 952903, VIPCOAT, and H.A.P. also acknowledges funding from Horizon 2020 by GA no. 760173, MarketPlace. The authors acknowledge permission by Frederico Maia, Smallmatek Lda, Portugal, to use Figure 1. This work was facilitated by activities of Inprodat e.V., Kaiserslautern.

References

[1] P. Klein, H.A. Preisig, M.T. Horsch and N. Konchakova, Application of an ontology based process model construction tool for active protective coatings: Corrosion inhibitor release, in: *Proceedings of JOWO 2021*, CEUR-WS, 2021, p. 26, http://ceur-ws.org/Vol-2969/ (workshop proceedings version of the present article).

[2] N. Konchakova, P. Klein, E. Lidorikis, A. Laskarakis, W.L. Cavalcanti and J. Friis, Open Innovation in Horizon Europe, position paper, VIPCOAT, MUSICODE and OpenModel, 2022. doi:10.5281/zenodo.5848551.

[3] Q. Shao, A. Makradi, D. Fiorelli, A. Mikdam, W. Huang, H. Hu and S. Belouettar, Material twin for composite material microstructure generation and reconstruction based on statistical continuum theory, *Composites C* **7** (2022), 100216. doi:10.1016/j.jcomc.2021.100216.

[4] T. Bronger, É. Demandt, I. Heine, A. Kraft, N. Preuß and A. Schwarz, Die Nationale Forschungsdateninfrastruktur für die Ingenieurwissenschaften (NFDI4Ing), *Bausteine Forschungsdatenmanagement* **2021**(2) (2021), 110–123. doi:10.17192/bfdm.2021.2.8329.

[5] O. Corcho, M. Eriksson, K. Kurowski, M. Ojsteršek, C. Choirat, M. Sanden and F. Coppens, EOSC Interoperability Framework, technical report, EOSC, 2021. doi:10.2777/620649.

[6] J. Bicarregui, Building and sustaining data infrastructures: Putting policy into practice, technical report, Wellcome Trust, 2016, doi:10.6084/m9.figshare.4055538.v2.

[7] S. Collins, F. Geneva, N. Harrower, S. Hodson, S. Jones, L. Laaksonen, D. Mietchen and R. Petrauskaité, *Turning FAIR into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data*, EC, Brussels, 2018. ISBN 978-92-79-96547-0.

[8] B. Mons, *Data Stewardship for Open Science*, CRC, Boca Raton, 2018. ISBN 978-1-4987-5317-3.

[9] M.D. Wilkinson, M. Dumontier, S.-A. Sansone, L.O. Bonino da Silva Santis, M. Prieto, D. Batista, P. McQuilton, T. Kuhn, P. Rocca-Serra, M. Crosas and E. Schultes, Evaluating FAIR maturity through a scalable, automated, community-governed framework, *Scientific Data* **9** (2019), 174, doi:10.1038/s41597-019-0184-5.

[10] G. Guizzardi, Ontology, ontologies and the "I" of FAIR, *Data Intelligence* **2**(1–2) (2020), 181–191. doi:10.1162/dint_a_00040.

[11] G. Gygli and J. Pleiss, Simulation foundry: Automated and F.A.I.R. molecular modeling, *Journal of Chemical Information and Modeling* **60**(4) (2020), 1922–1927. doi:10.1021/acs.jcim.0c00018.

[12] M.T. Horsch, J. Francisco Morgado, G. Goldbeck, D. Iglezakis, N.A. Konchakova and B. Schembera, Domain-specific metadata standardization in materials modelling, in: *Proceedings of DORIC-MM 2021*, UKRI STFC, 2021, pp. 12–27. http://purl.org/net/epubs/work/50300311.

[13] D.J. Hand, *Dark Data: Why What You Don't Know Matters*, Princeton University Press, 2020. ISBN 978-0-6911-8237-7.

[14] B. Schembera, Like a rainbow in the dark: Metadata annotation for HPC applications in the age of dark data, *Journal of Supercomputing* **77** (2021), 8946–8966. doi:10.1007/s11227-020-03602-6.

[15] A. Corallo, A.M. Crespino, V. Del Vecchio, M. Lazoi and M. Marra, Understanding and defining dark data for the manufacturing industry, *IEEE Transactions on Engineering Management* (2022). doi:10.1109/TEM.2021.3051981.

[16] J. Francisco Morgado, E. Ghedini, G. Goldbeck, A. Hashibon, G.J. Schmitz, J. Friis and A. de Baas, Mechanical testing ontology for digital-twins: A roadmap based on EMMO, in: *Proceedings of SeDiT 2020*, CEUR-WS, 2020, p. 3. http://ceur-ws.org/Vol-2615/.

[17] H.A. Preisig, T.F. Hagelien, J. Friis, P. Klein and N. Konchakova, Ontologies in computational engineering, in: *Proceedings of WCCM-ECCOMAS 2020*, Scipedia, 2021, p. 262. doi:10.23967/wccm-eccomas.2020.262.

[18] E. Ghedini, G. Goldbeck, J. Friis et al., Elementary Multiperspective Material Ontology (EMMO), GitHub, 2018–2022. https://github.com/emmo-repo/EMMO.

[19] European Committee for Standardization, Materials Modelling: Terminology, Classification and Metadata, CEN Workshop Agreement (expired), 17284:2018 (E), CEN, 2018, Note: In accordance with CEN-CENELEC procedures, the validity of this CWA expired on 17th April 2021, three years after its publication by CEN. https://www.cencenelec.eu/media/CEN-CENELEC/CWAs/RI/cwa17284_2018.pdf.

[20] 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 & Engineering Data* **65**(3) (2020), 1313–1329. doi:10.1021/acs.jced.9b00739.

[21] M.T. Horsch, S. Chiacchiera, M.A. Seaton, I.T. Todorov, B. Schembera, P. Klein and N.A. Konchakova, Pragmatic interoperability and translation of industrial engineering problems into modelling and simulation solutions, in: *Proceedings of DAMDID 2020*, Springer, 2021, pp. 45–59. ISBN 978-3-030-81199-0.

[22] M.T. Horsch, S. Chiacchiera, W.L. Cavalcanti and B. Schembera, *Data Technology in Materials Modelling*, Springer, Cham, 2021. ISBN 978-3-030-68596-6.

[23] M.T. Horsch, D. Toti, S. Chiacchiera, M.A. Seaton, G. Goldbeck and I.T. Todorov, OSMO: Ontology for simulation, modelling, and optimization, in: *Proceedings of JOWO 2021*, CEUR-WS, 2021, p. 47. http://ceur-ws.org/Vol-2969/.

[24] N. Romanos, M. Kalogerini, E.P. Koumoulos, K. Morozinis, M. Sebastiani and C. Charitidis, Innovative data management in advanced characterization: Implications for materials design, *Materials Today Communications* **20** (2019), 100541. doi:10.1016/j.mtcomm.2019.100541.

[25] V.K. Mittal, S.C. Bailin, M.A. Gonzalez, D.E. Meyer, W.M. Barrett and R.L. Smith, Toward automated inventory modeling in life cycle assessment: The utility of semantic data modeling to predict real-world chemical production, *ACS Sustainable Chemistry & Engineering* **6**(2) (2018), 1961–1976. doi:10.1021/acssuschemeng.7b03379.

[26] M.T. Horsch, Mereosemiotics: Parts and signs, in: *Proceedings of JOWO 2021*, CEUR-WS, 2021, p. 3. http://ceur-ws.org/Vol-2969/.

[27] M.T. Horsch, Physicalistic Interpretation of Modelling and Simulation: Interoperability Infrastructure (PIMS-II), 2021–2022. http://www.molmod.info/semantics/pims-ii.ttl.

[28] D. Hristova-Bogaerds, P. Asinari, N.A. Konchakova, L. Bergamasco, A. Marcos Ramos, G. Goldbeck, D. Höche, O. Swang and G.J. Schmitz, EMMC Translators' Guide, technical report, EMMC, 2019. doi:10.5281/zenodo.3552260.
[29] P. Klein, N. Konchakova, D.G. Hristova-Bogaerds, M. Noeske, A. Simperler, G. Goldbeck and D. Höche, Translation in Materials Modelling: Process and Progress, technical report, OntoTRANS – FORCE, 2021. doi:10.5281/zenodo.4729917.
[30] M. Pezzotta, J. Friis, G.J. Schmitz, N. Konchakova, D. Höche, D. Hristova-Bogaerds, P. Asinari, L. Bergamasco and G. Goldbeck, Report on translation case studies describing the gained experience, technical report, EMMC, 2021. doi:10.5281/zenodo.4457849.
[31] H.A. Preisig, Ontology-based process modelling – with examples of physical topologies, *Processes* **9**(4) (2021), 592. doi:10.3390/pr9040592.
[32] B. Schembera and J.M. Durán, Dark data as the new challenge for big data science and the introduction of the scientific data officer, *Philosophy & Technology* **33** (2019), 93–115. doi:10.1007/s13347-019-00346-x.
[33] G.P. Rodrigo, M. Henderson, G.H. Weber, C. Ophus, K. Antypas and L. Ramakrishnan, ScienceSearch: Enabling search through automatic metadata generation, in: *Proceedings of the 14th IEEE International Conference on e-Science*, IEEE, 2018, pp. 93–104. ISBN 978-1-5386-9156-4.
[34] P. Visser, Novel totally chrome free corrosion inhibiting coating technology for protection of aluminium alloys, *Transactions of the IMF* **89**(6) (2011), 291–294. doi:10.1179/174591911X13172025236839.
[35] M.L. Zheludkevich, J. Tedim and M.G.S. Ferreira, "Smart" coatings for active corrosion protection based on multi-functional micro and nanocontainers, *Electrochimica Acta* **82** (2012), 314–323. doi:10.1016/j.electacta.2012.04.095.
[36] P. Visser, H. Terryn and J.M.C. Mol, On the importance of irreversibility of corrosion inhibitors for active coating protection of AA2024-T3, *Corrosion Science* **140** (2018), 272–285. doi:10.1016/j.corsci.2018.05.037.
[37] J. Aurich, F. Schneider, P. Mayer, B. Kirsch and H. Hasse, Oberflächenerzeugungs-Morphologie-Eigenschafts-Beziehungen, *Zeitschrift für wirtschaftlichen Fabrikbetrieb* **111**(4) (2016), 213–216. doi:10.3139/104.111502.
[38] VIPCOAT, Virtual Open Innovation Platform for Active Protective Coatings Guided by Modelling and Optimization (H2020 GA no. 952903), 2021–2025.
[39] S.K. Poznyak, J. Tedim, L.M. Rodrigues, A.N. Salak, M.L. Zheludkevich, L.F.P. Dick and M.G.S. Ferreira, Novel inorganic host layered double hydroxides intercalated with guest organic inhibitors for anticorrosion applications, *Applied Materials & Interfaces* **1**(10) (2009), 2353–2362. doi:10.1021/am900495r.
[40] F. Maia, J. Tedim, A.D. Lisenkov, A.N. Salak, M.L. Zheludkevich and M.G.S. Ferreira, Silica nanocontainers for active corrosion protection, *Nanoscale* **4** (2012), 1287–1298. doi:10.1039/C2NR11536K.
[41] T. Würger, D. Mei, B. Vaghefinazari, D.A. Winkler, S.V. Lamaka, M.L. Zheludkevich, R.H. Robert H. Meißner and C. Feile, Exploring structure-property relationships in magnesium dissolution modulators, *npj Materials Degradation* **5** (2021), 2, doi:10.1038/s41529-020-00148-z.
[42] VIMMP, Virtual Materials Marketplace (H2020 GA no. 760907), 2018–2022. https://vimmp.eu/.
[43] M.T. Horsch, S. Chiacchiera, M.A. Seaton, I.T. Todorov, K. Šindelka, M. Lísal, B. Andreon, E. Bayro 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 – Künstliche Intelligenz* **34**(3) (2020), 423–428. doi:10.1007/s13218-020-00648-9.
[44] VIMMP, VIMMP Ontologies, 2018–2022. https://gitlab.com/vimmp-semantics/vimmp-ontologies.
[45] N. Asher and L. Vieu, Toward a geometry of common sense: A semantics and a complete axiomatization of mereotopology, in: *Proceedings of IJCAI-95*, Morgan Kaufmann, 1995, pp. 846–852. ISBN 978-1-55860-363-9.
[46] A.G. Cohn and A.C. Varzi, Mereotopological connection, *Journal of Philosophical Logic* **32** (2003), 357–390. doi:10.1023/A:1024895012224.
[47] G.J. Schmitz, A phase-field perspective on mereotopology, *AppliedMath* **2**(1) (2022), 54–103. doi:10.3390/appliedmath2010004.
[48] C.S. Peirce, Logic as semiotic: The theory of signs, in: *Philosophical Writings of Peirce*, J. Buchler, ed., Dover, New York, 1955, pp. 98–119. ISBN 978-0-48620217-4.
[49] R. Marty, *L'Algèbre des Signes: Essai de Sémiotique Scientifique d'après C.S. Peirce*, John Benjamins, Amsterdam, 1990. ISBN 978-90-272-3296-0.
[50] J.F. Sowa, Signs and reality, *Applied Ontology* **10**(3–4) (2015), 273–284. doi:10.3233/AO-150159.
[51] M.T. Horsch, S. Chiacchiera, B. Schembera, M.A. Seaton and I.T. Todorov, Semantic interoperability based on the European Materials and Modelling Ontology and its ontological paradigm: Mereosemiotics, in: *Proceedings of WCCM-ECCOMAS 2020*, Scipedia, 2021, p. 297. doi:10.23967/wccm-eccomas.2020.297.
[52] C.A. Petri, Kommunikation mit Automaten, PhD thesis, TH Darmstadt, 1962.
[53] C.A. Petri, Concurrency as a basis of system thinking, in: *Proceedings of the 5th Scandinavian Logic Symposium*, Aalborg Universitetsforlag, 1979, pp. 143–162, ISBN 978-8-7730-7037-6.
[54] A.T. Elve and H.A. Preisig, From process graph to process simulation with proper model documentation, *Computer Aided Chemical Engineering* **43** (2018), 235–240. doi:10.1016/B978-0-444-64235-6.50042-5.
[55] T. Habermann, Mapping ISO 19115-1 geographic metadata standards to CodeMeta, *PeerJ Computer Science* **5** (2019), e174. doi:10.7717/peerj-cs.174.
[56] P. Ochieng and S. Kyanda, A statistically-based ontology matching tool, *Distributed and Parallel Databases* **36** (2018), 195–217. doi:10.1007/s10619-017-7206-0.
[57] J.F. de Souza, S.W. Matsui Siqueira and B. Nunes, A framework to aggregate multiple ontology matchers, *International Journal of Web Information Systems* **16**(2) (2020), 151–169. doi:10.1108/IJWIS-05-2019-0023.

[58] L. Zhou, M. Cheatham and P. Hitzler, Towards association rule-based complex ontology alignment, in: *Proceedings of JUST 2019*, Springer, 2020, pp. 287–303. ISBN 978-3-030-41406-1.
[59] C. Trojahn dos Santos, R. Vieira, D. Schmidt, A. Pease and G. Guizzardi, Foundational ontologies meet ontology matching: A survey, *Semantic Web* (2021). doi:10.3233/SW-210447.
[60] B. König and J. Stückrath, A general framework for well-structured graph transformation systems, in: *Proceedings of CONCUR 2014*, Springer, 2014, pp. 467–481. ISBN 978-3-662-44583-9.
[61] B. König, D. Nolte, J. Padberg and A. Rensink, A tutorial on graph transformation, in: *Graph Transformation, Specifications, and Nets*, R. Heckel and G. Taentzer, eds, Springer, Cham, 2018, pp. 83–104. ISBN 978-3-319-75395-9.
[62] I. Horrocks, O. Kutz and U. Sattler, The even more irresistible SROIQ, in: *Proceedings of KR'06*, AAAI, 2006, pp. 57–67. ISBN 978-1-57735-271-6.
[63] V. Telnov and Y. Korovin, Semantic web and knowledge graphs as an educational technology of personnel training for nuclear power engineering, *Nuclear Energy and Technology* **5**(3) (2019), 273–280. doi:10.3897/nucet.5.39226.
[64] H.Q. Yu and Y. Hong, Graph transformation for the semantic web: Queries and inference rules, in: *Proceedings of ICGT 2008*, Springer, 2008, pp. 511–513. ISBN 978-3-540-87404-1.
[65] M. Mahfoudh, L. Thiry, G. Forestier and M. Hassenforder, Algebraic graph transformations for merging ontologies, in: *Proceedings of MEDI 2014*, Springer, 2014, pp. 154–168. ISBN 978-3-319-11586-3.
[66] J.O. Batista, J.P.A. Almeida, E. Zambon and G. Guizzardi, Building correct taxonomies with a well-founded graph grammar, in: *Proceedings of RCIS 2021*, Springer, 2021, pp. 506–522. ISBN 978-3-030-75018-3.
[67] J.H. Brenas and A. Shaban-Nejad, Proving the correctness of knowledge graph update: A scenario from surveillance of adverse childhood experiences, *Frontiers in Big Data* **4** (2021), 660101. doi:10.3389/fdata.2021.660101.
[68] M.N. Roulofs and R. Vos, Automatically inferring technology compatibility with an ontology and graph rewriting rules, *Journal of Engineering Design* **32**(2) (2021), 90–114. doi:10.1080/09544828.2020.1860202.
[69] L. Babai, Graph isomorphism in quasipolynomial time, in: *Proceedings of STOC '16*, ACM, 2016, pp. 684–697. doi:10.1145/2897518.2897542.
[70] G. Grochau Azzi, A. Corradini and L. Ribeiro, On the essence and initiality of conflicts in $\mathcal{M}$-adhesive transformation systems, *Journal of Logical and Algebraic Methods in Programming* **109** (2019), 100482. doi:10.1016/j.jlamp.2019.100482.
[71] M.J. Cruz, S. Kurapati and Y. Turkyilmaz-van der Velden, The role of data stewardship in software sustainability and reproducibility, in: *Proceedings of IEEE e-Science 2018*, IEEE, 2018, pp. 1–8. doi:10.1109/eScience.2018.00009.
[72] R. Pergl, R. Hooft, M. Suchánek, V. Knaisl and J. Slifka, "Data Stewardship Wizard": A tool bringing together researchers, data stewards, and data experts around data management planning, *Data Science Journal* **18**(1) (2019), 59. doi:10.5334/dsj-2019-059.
[73] A.M. Tammaro and V. Casarosa, Who is the data curator? Defining a vocabulary, in: *Proceedings of IRCDL 2018*, Springer, 2018, pp. 249–255. ISBN 978-3-319-73164-3.
[74] A.M. Tammaro, K.K. Matusiak, F.A. Sposito and V. Casarosa, Data curator's roles and responsibilities: An international perspective, *Libri* **69**(2) (2019), 89–104. doi:10.1515/libri-2018-0090.