Webinar overview

15.00 – 15.15 CET	Presentation by UKRI STFC
15.15 – 15.20 CET	Discussion #1
15.20 – 15.35 CET	Presentation by SINTEF
15.35 – 15.40 CET	Discussion #2
15.40 – 15.45 CET	Community synergy: Ideas from SINTEF
15.45 – 15.50 CET	Community synergy: Ideas from UKRI STFC
15.50 – 16.00 CET	Discussion #3

Webinar overview

15.00 – 15.15 CET	**Presentation by UKRI STFC**
15.15 – 15.20 CET	Discussion #1
15.20 – 15.35 CET	Presentation by SINTEF
15.35 – 15.40 CET	Discussion #2
15.40 – 15.45 CET	Community synergy: Ideas from SINTEF
15.45 – 15.50 CET	Community synergy: Ideas from UKRI STFC
15.50 – 16.00 CET	Discussion #3

Martin T. Horsch, Silvia Chiacchiera,
Michael A. Seaton, Ilian T. Todorov
STFC Daresbury Laboratory
UK Research and Innovation

Ontologies, data management and interoperability for digital marketplaces

VIMMP
VIRTUAL MATERIALS MARKETPLACE

EMMC Webinar

24th November 2020

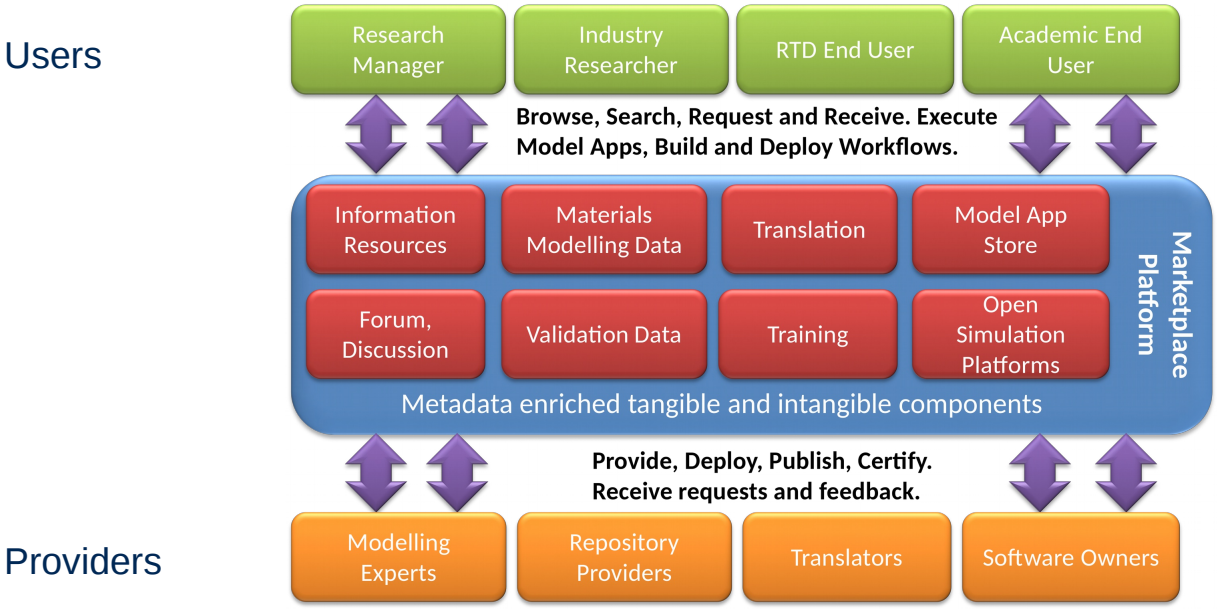
Part 1

Science and Technology Facilities Council

VIMMP
VIRTUAL MATERIALS MARKETPLACE

VIMMP – Virtual Materials Marketplace

VIMMP and **MarketPlace** are sibling H2020 projects[1] developing digital marketplaces, i.e., platforms to facilitate exchanges between providers and users in the area of materials modelling. Below, we show a graphical summary of the VIMMP concept.



An open platform. The interoperability of components relies on semantic standards (i.e., a **set of ontologies**).

[1] The projects sites are: https://www.vimmp.eu/ and https://www.the-marketplace-project.eu/

Standardization: a spectrum of possibilities

Standards can be intended for humans or for machines.
Can be at syntactic, semantic or pragmatic levels.
Moreover, the semantic spectrum goes from (unstructured) vocabularies, via taxonomies, to ontologies.

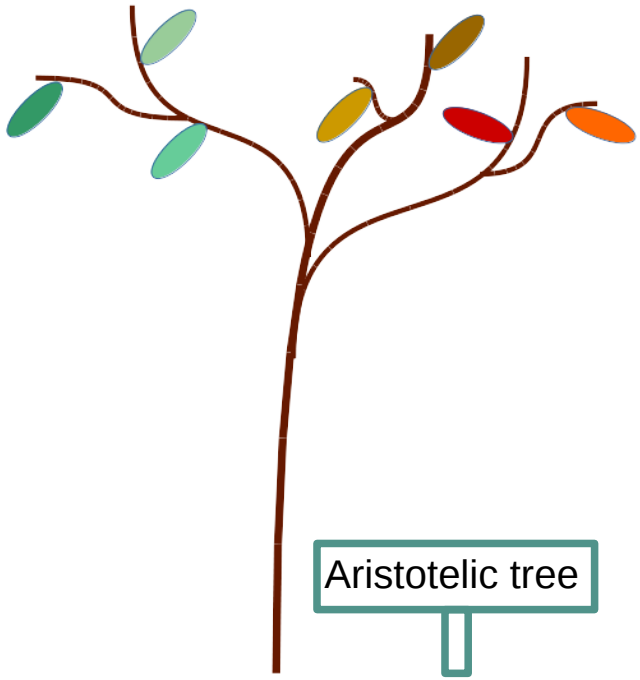
Standards undoubtedly support interoperability.
However, different design choices can be made:

- **Less expressive languages (e.g., XML-based)**:

  PRO: can be handled with multiple technologies/tools; typically lighter and faster.
  CONTRA: limited expressivity.

- **Richer languages (e.g., OWL)**:

  PRO: can describe more complex relations.
  CONTRA: rely on less widespread technologies; typically heavier to handle.

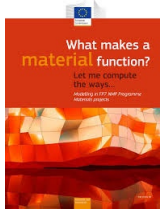


Standardization (2): recent efforts in materials modelling

Our work connects to community-guided assets:

MODA – MOdelling DAta

→ A template to describe simulation workflows (user case, model, solver, processor); CWA 17284.



RoMM – Review of Materials Modelling

What makes a material function? Let me compute the ways..., Anne F. de Baas (ed), 6th version, 2017.

→ Includes a classification of models according to their granularity (electronic, atomistic, mesoscopic, continuum)

EMMO – European Materials & Modelling Ontology

E. Ghedini *et al*, 2020; https://github.com/emmo-repo/EMMO .

→ A top-level ontology for applied sciences

Ontologies (in a nutshell)

What are they?

In philosophy, Ontology is the “science of what is”.

In information science, an ontology is a formal (machine-readable) representation of knowledge within a certain domain. It identifies the categories (“classes”) that exist in the domain and the relations between them.

Why are they useful?

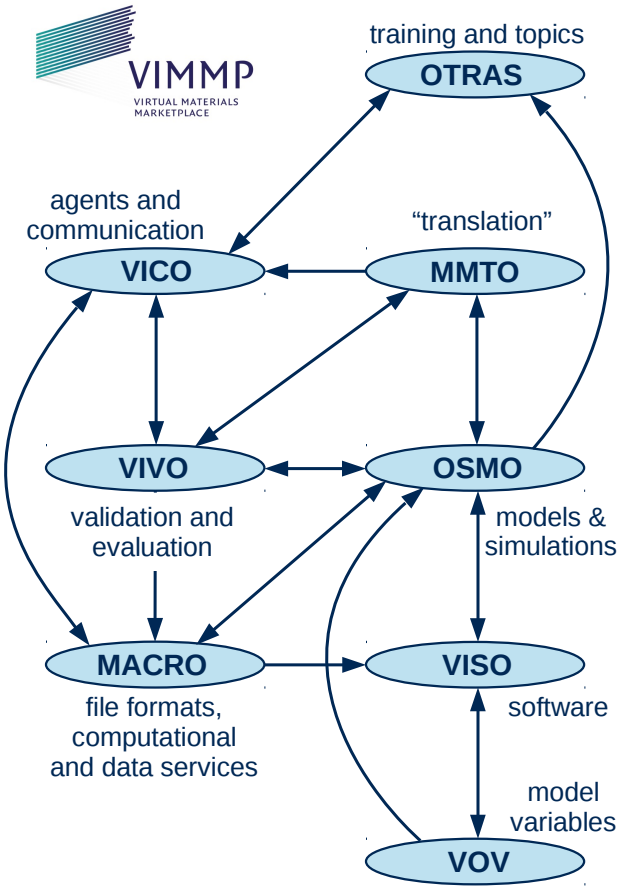
Ontologies allow 1) automatic reasoning, 2) easier exchange of information across heterogeneous sources.

What is the bigger picture?

The context is that of semantic technologies and semantic interoperability. Notably, the Semantic Web concept, an evolution of the World Wide Web that is based on semantics rather than ad-hoc links between resources (e.g., web-pages) was proposed in the 1990s.

Ontologies on the VIMMP marketplace

In VIMMP we have developed **a set of 8 ontologies**, covering all aspects of the marketplace.



How do we use them?

To guide data ingest, then search and browsing. Internally, they are the base of interoperability of the marketplace compontents.

How will this help/affect users and providers?

Users will indirectly see them via the available keywords and search criteria and results.

They will not be frozen, there will be a policy to allow users/providers to request for extensions.

Providers can choose down to which level of detail to adhere to the proposed common language: the deeper the adherence, the deeper the interoperability with other services.

Ontologies on the VIMMP marketplace

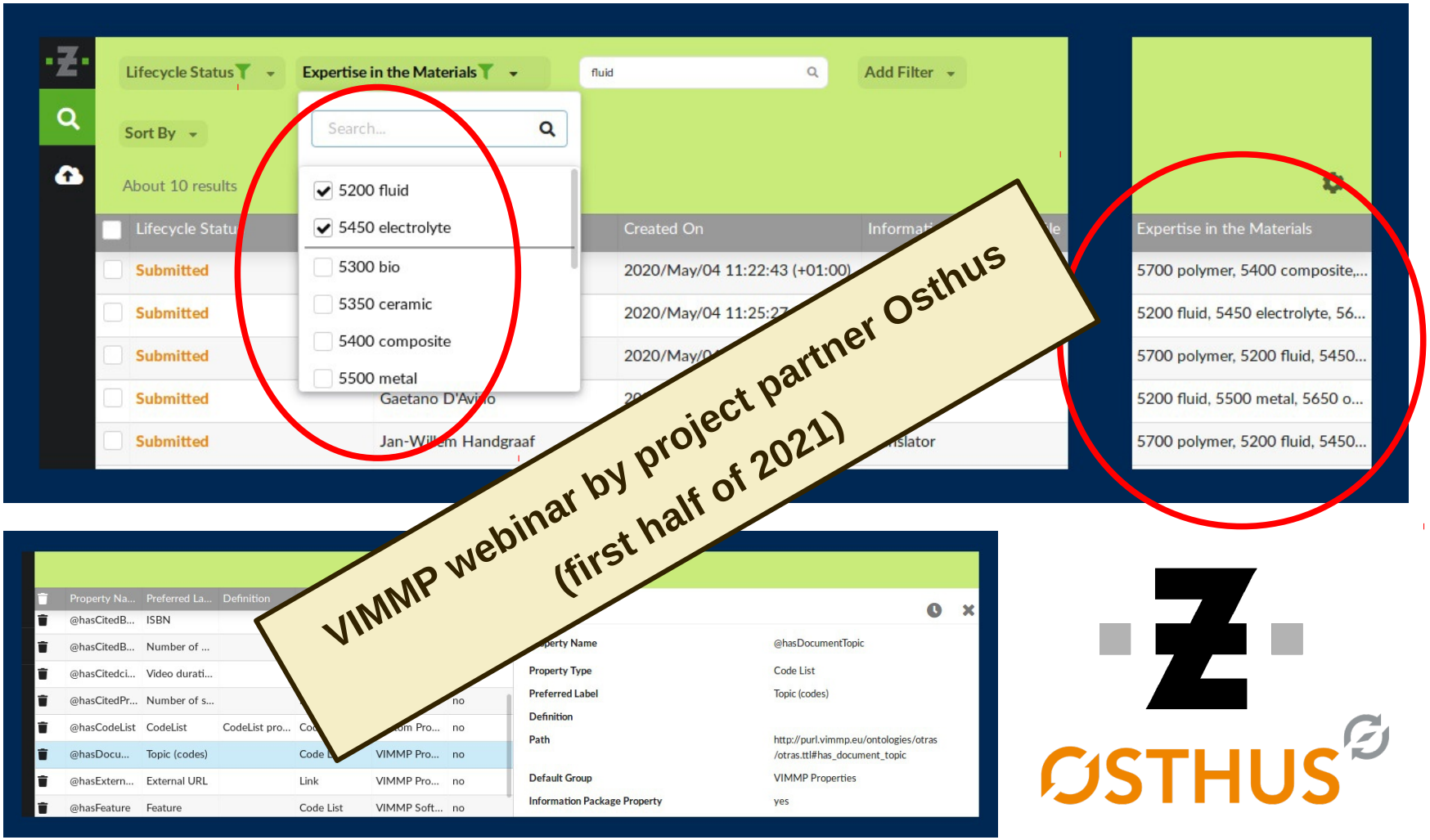
In VIMMP we have developed **a set of 8 ontologies**, covering all aspects of the marketplace.

Available under LGPL-v3 license

https://www.vimmp.eu/?p=349

Future releases also on: https://gitlab.cc-asp.fraunhofer.de/mic61299/vimmp_ontology

Data management on the VIMMP backend

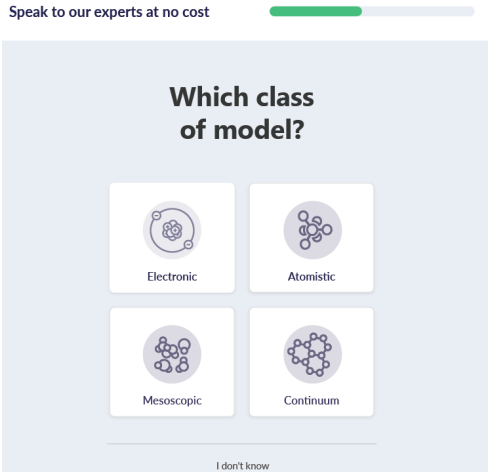


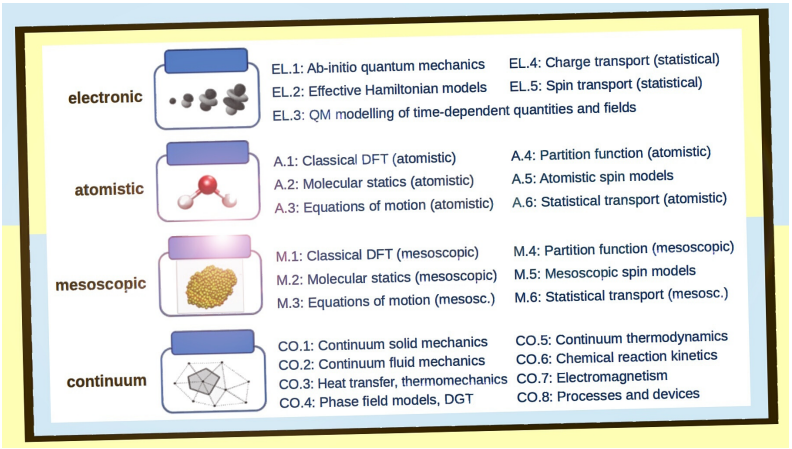
OSTHUS

Metadata-supported data ingest and retrieval

mm_topic_basic (codes **1XXX** and **2XXX**):
Basic prerequisites for materials modelling.
mm_topic_computational (codes **3XXX**):
Computational and numerical aspects of materials modelling.
mm_topic_data (codes **4XXX**):
Data science and technology aspects.
mm_topic_materials (codes **5XXX**):
Topics related to fluid and solid materials.
mm_topic_social (codes **6XXX**):
Social, economic, and community aspects.
mm_topic_theoretical (codes **7XXX**):
Theory (non-computational aspects).
mm_topic_interdisciplinary (codes **8XXX**)
mm_topic_side (codes **9XXX**):
Topics from other disciplines.

- 3100, 7100 electronic
  - 3120, 7120 physical equation EL.1
  - 3130, 7130 physical equation EL.2
  - *etc.*
- 3200, 7200 atomistic and mesoscopic
  - 3220, 7220 equations A.1 and M.1
    - 3222, 7222 physical equation A.1
    - 3225, 7225 physical equation M.1
  - 3230, 7230 equations A.2 and M.2
    - *etc.*
  - *etc.*
- 3300, 7300 continuum
  - 3320, 7320 physical equation CO.1
  - 3330, 7330 physical equation CO.2
  - *etc.*



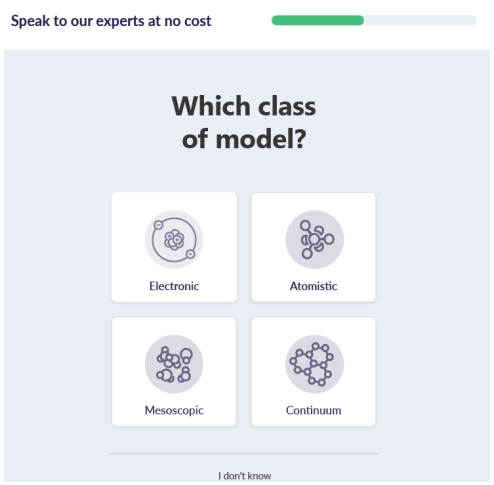


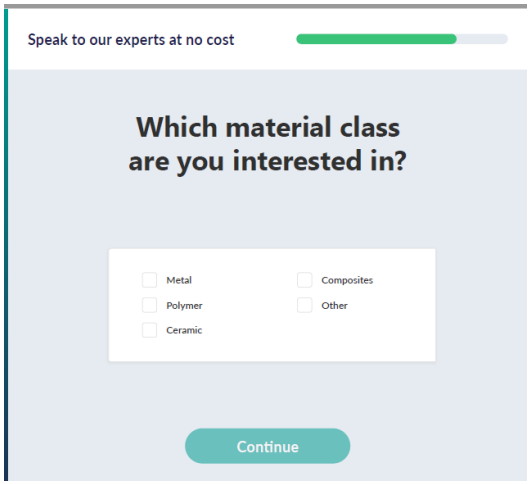
Metadata-supported data ingest and retrieval

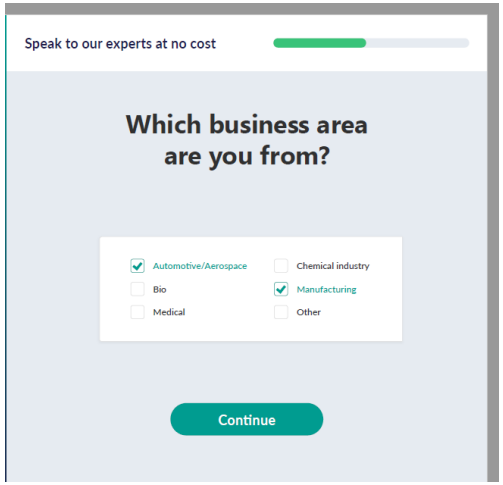
mm_topic_basic (codes **1XXX** and **2XXX**):
Basic prerequisites for materials modelling.
mm_topic_computational (codes **3XXX**):
Computational and numerical aspects of materials modelling.
mm_topic_data (codes **4XXX**):
Data science and technology aspects.
mm_topic_materials (codes **5XXX**):
Topics related to fluid and solid materials.
mm_topic_social (codes **6XXX**):
Social, economic, and community aspects.
mm_topic_theoretical (codes **7XXX**):
Theory (non-computational aspects).
mm_topic_interdisciplinary (codes **8XXX**)
mm_topic_side (codes **9XXX**):
Topics from other disciplines.

under 61XX: industrial

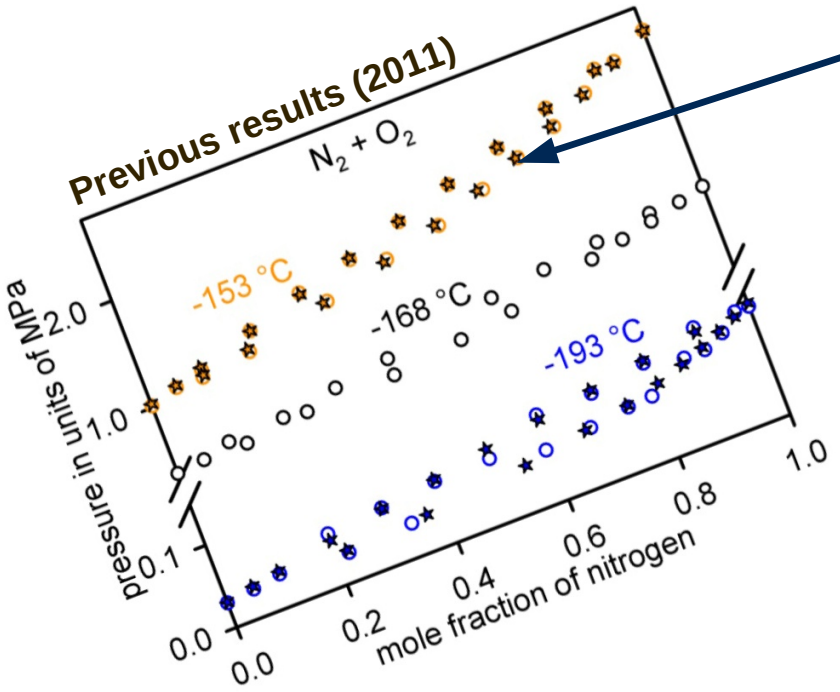
- 6120 chemical
- 6130 petrochemical
- 6140 transport
  - 6142 aerospace
  - 6144 automotive
  - 6148 railway
- 6150 biotechnology
- 6155 food
- 6160 medicine
- 6165 paper
- 6170 electrical
- 6175 machinery
- 6180 metal (basic and fabricated)
- 6190 special topics







Provenance description of simulation results

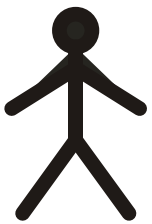


What values did *x* and *p* have?

How was the data point obtained?

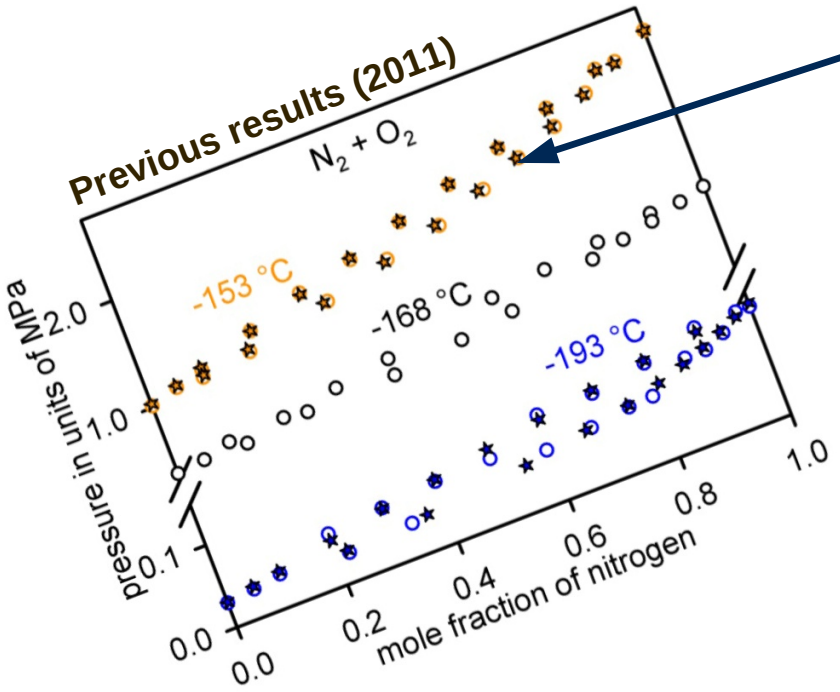
What is the margin of error, how was the error defined, and what software (or experimental setup) was used?

ask the person who carried out the work back in 2011



"I remember."

Provenance description of simulation results

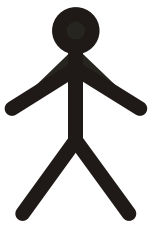


What values did *x* and *p* have?

How was the data point obtained?

What is the margin of error, how was the error defined, and what software (or experimental setup) was used?

ask the person who carried out the work back in 2011



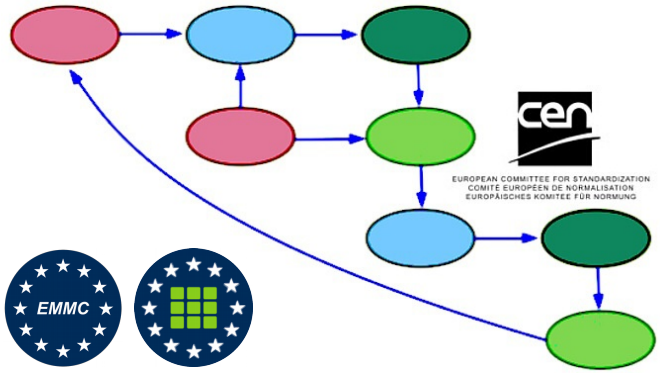
"I remember.
Haha, joke. Of course I don't."

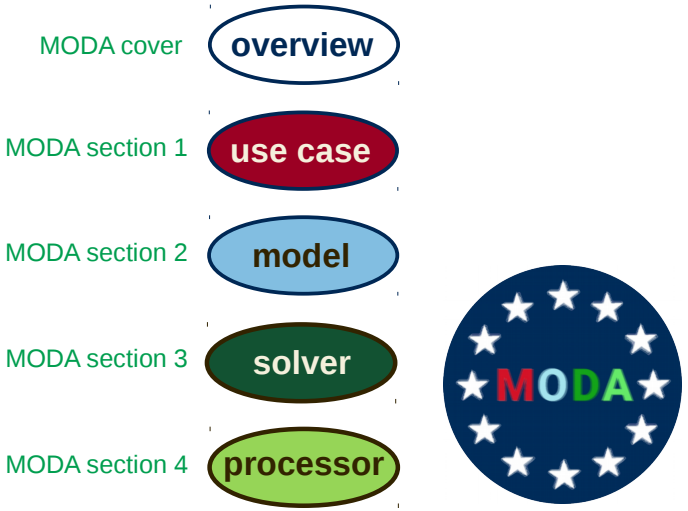
Good practice in managing research data:

Make all data findable, accessible, interoperable, and reusable (FAIR).

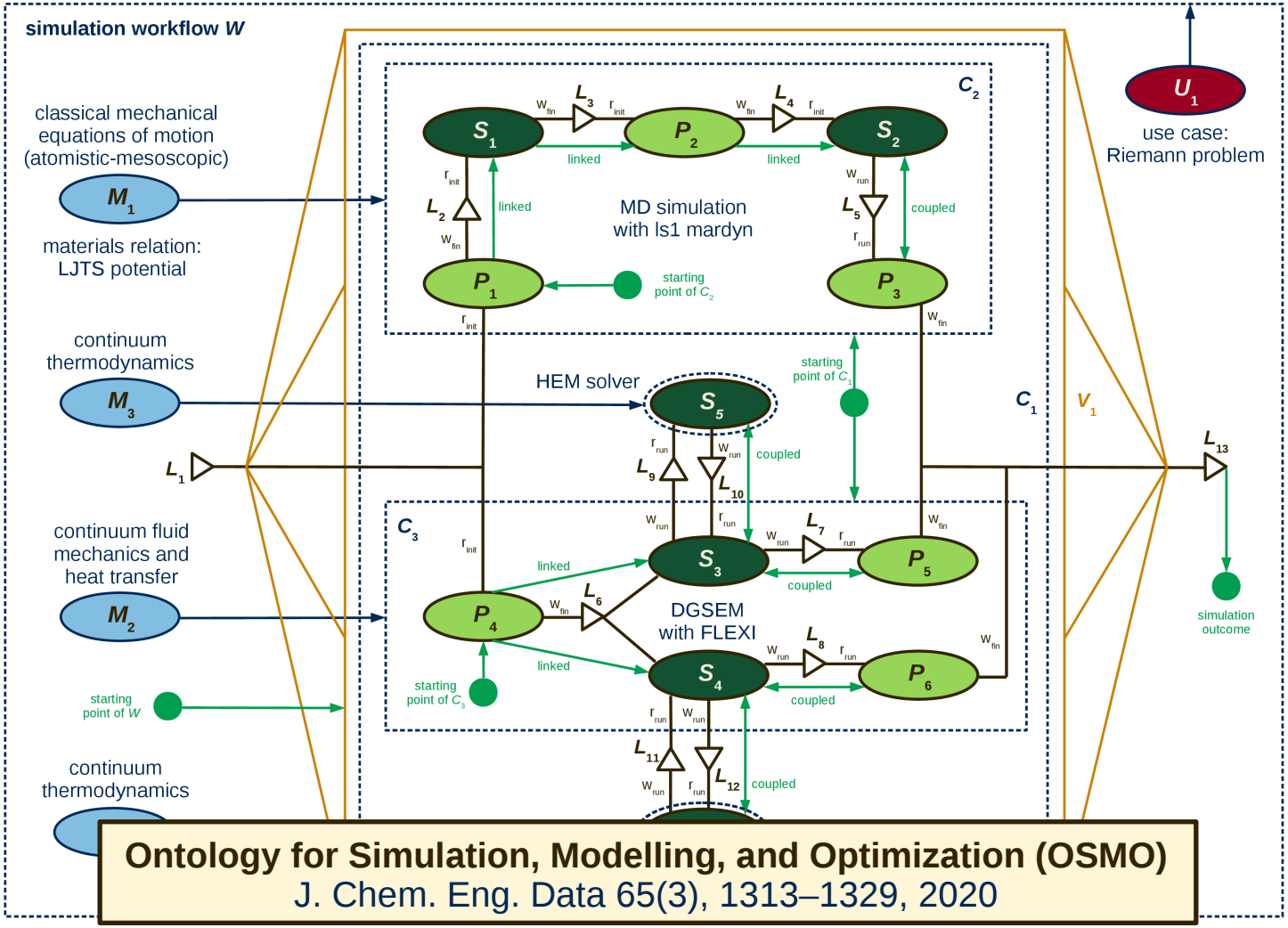
Provenance description of simulation results: MODA

3	Solver and Computational translation of the specifications		
3.1	Numerical Solver	*Please give name andtype of the solver* *e.g. Monte Carlo, SPH, FE, ...iterative, multi-grid, adaptive,...*	
3.2	Software tool	*Please give the name and if this is your own code, please specify if it can be shared with an evt link to website/publication.*	
3.3	Time step	*If applicable, please give the time step used in the solving operations.* *This is the numerical time step and this is not the same as the time lapse of the case to be simulated (see 1.4)*	
3.4	Computational Representation	Physics Equation, Material Relations, Material	*Computational representation of the physics equation, materials relation and material.* *There is no need to repeat user case info.* *"Computational" means that this only needs to be filled in when your computational solver represents the material, properties, equation variables, in a specific way.*
3.5	Computational boundary conditions	*If applicable.* *Please note hat these can be translations of the physical boundary conditions set in the user case or they can be pure computational. (e.g. a unit cell with mirror b.c. to simulate an infinite domain).*	
3.6	Additional Solver Parameters	*Please specify pure internal numerical solver details (if applicable), like* • *Specific tolerances* • *Cut-offs, convergence criteria* • *Integrator options*	





Provenance description of simulation results: OSMO

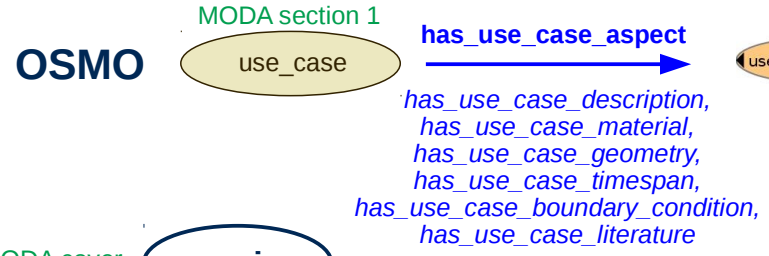


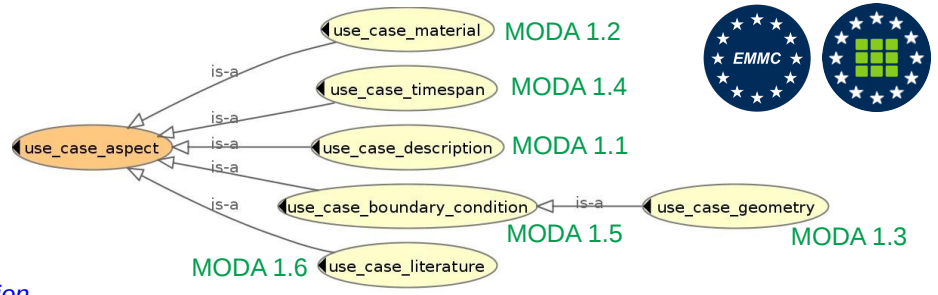
Ontology for Simulation, Modelling, and Optimization (OSMO)
J. Chem. Eng. Data 65(3), 1313–1329, 2020

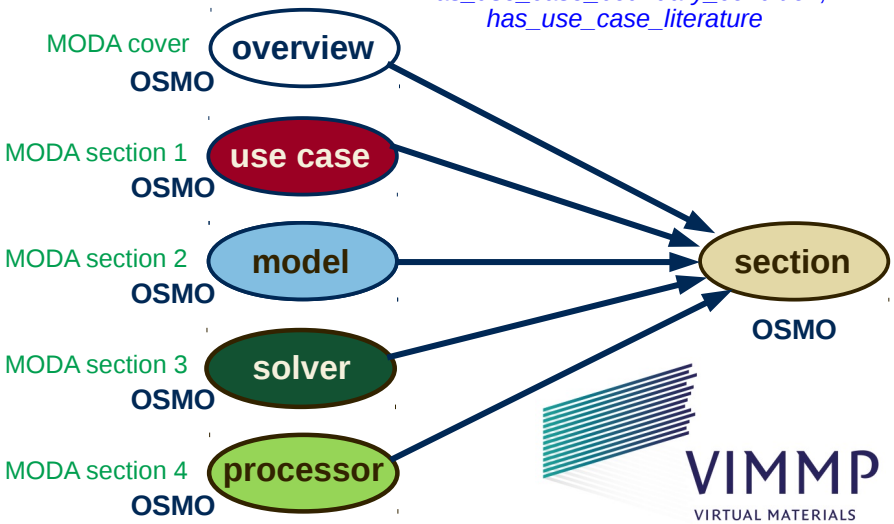
OSMO-based **provenance description** as an extension of the MODA workflow meta-data standard:

For all elements of the graph notation, there are corresponding concepts and relations from the ontology OSMO.

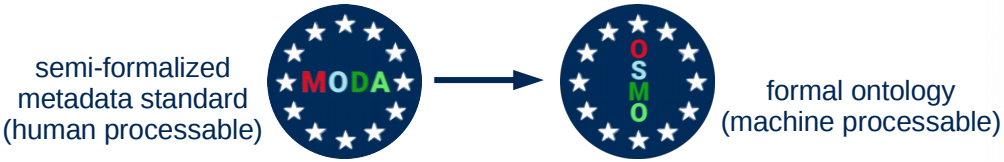
Provenance description of simulation results: OSMO



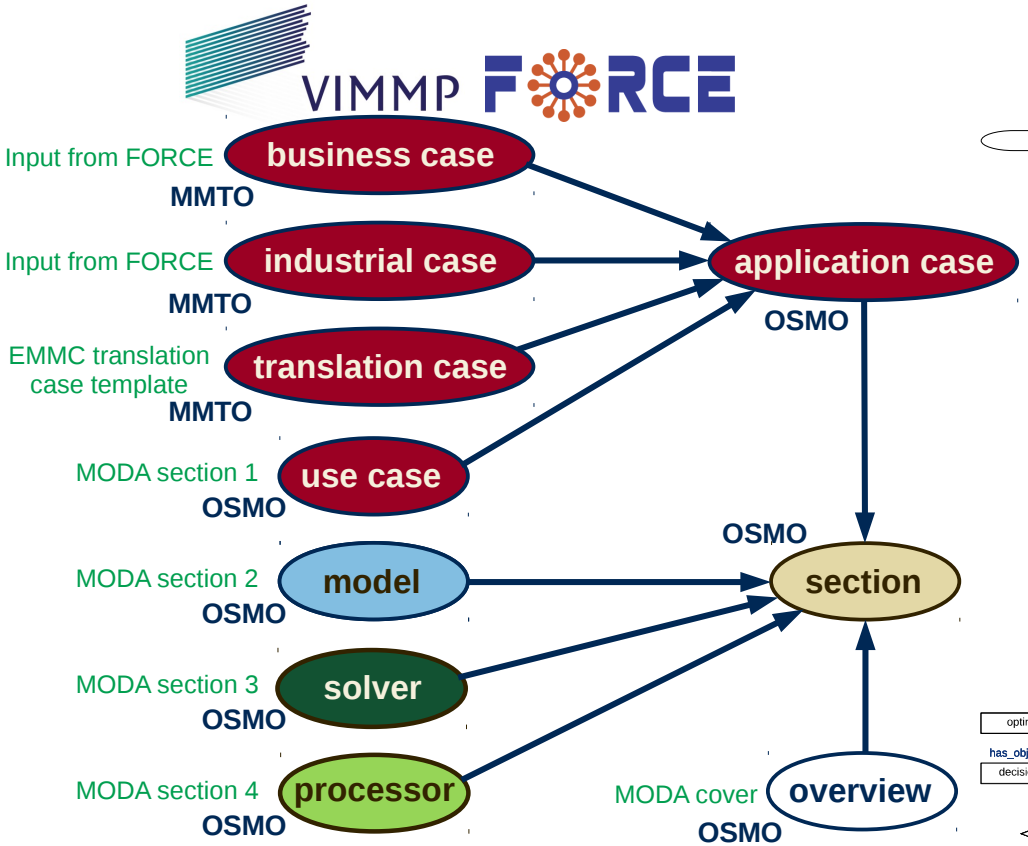


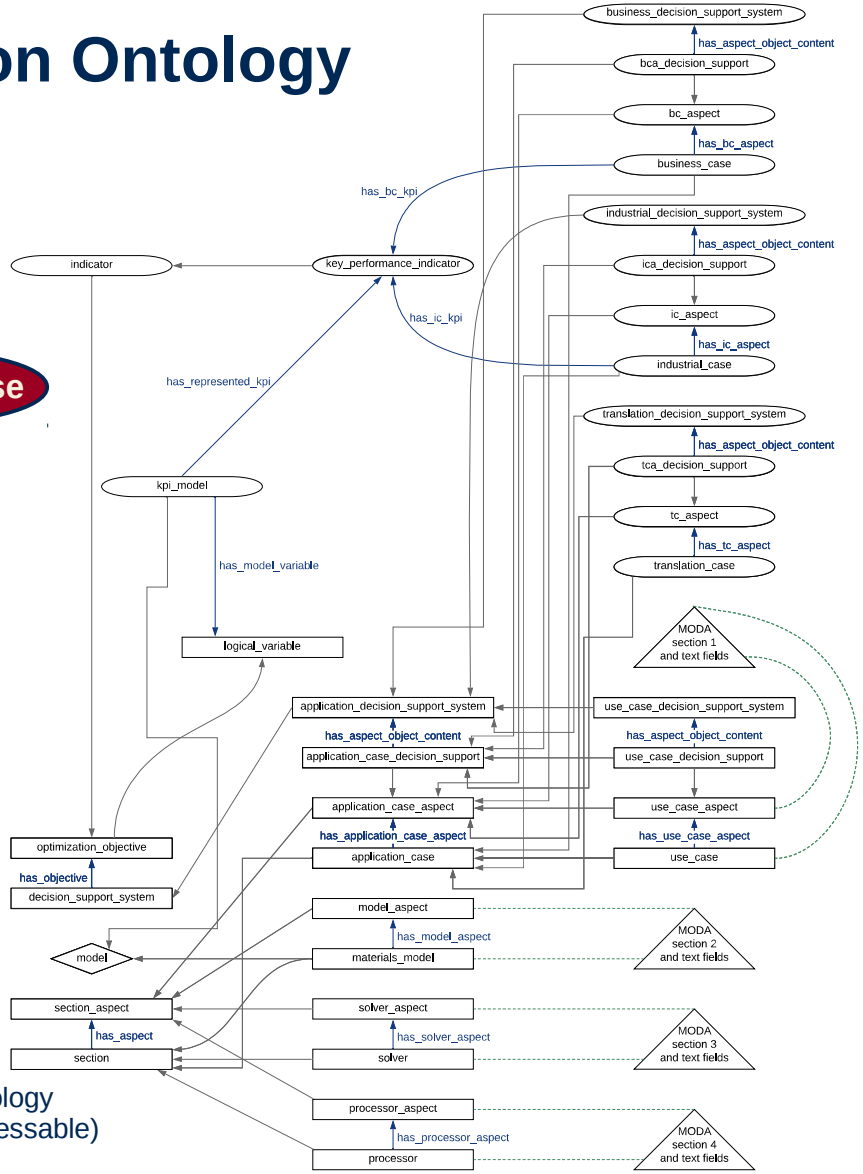


3	Solver and Computational translation of the specifications		
3.1	Numerical Solver		*Please give name andtype of the solver* *e.g. Monte Carlo, SPH, FE, ...iterative, multi-grid, adaptive,...*
3.2	Software tool		*Please give the name and if this is your own code, please specify if it can be shared with an evt link to website/publication.*
3.3	Time step		*If applicable, please give the time step used in the solving operations.* *This is the numerical time step and this is not the same as the time lapse of the case to be simulated (see 1.4)*
3.4	Computational Representation	Physics Equation, Material Relations, Material	*Computational representation of the physics equation, materials relation and material.* *There is no need to repeat user case info.* *"Computational" means that this only needs to be filled in when your computational solver represents the material, properties, equation variables, in a specific way.*
3.5	Computational boundary conditions		*If applicable.* *Please note hat these can be translations of the ph... user case or they can be pure computational. (e.g... simulate an infinite domain).*
3.6	Additional Solver Parameters		*Please specify pure internal numerical solver deta...* • *Specific tolerances* • *Cut-offs, convergence criteria* • *Integrator options*



Materials Modelling Translation Ontology



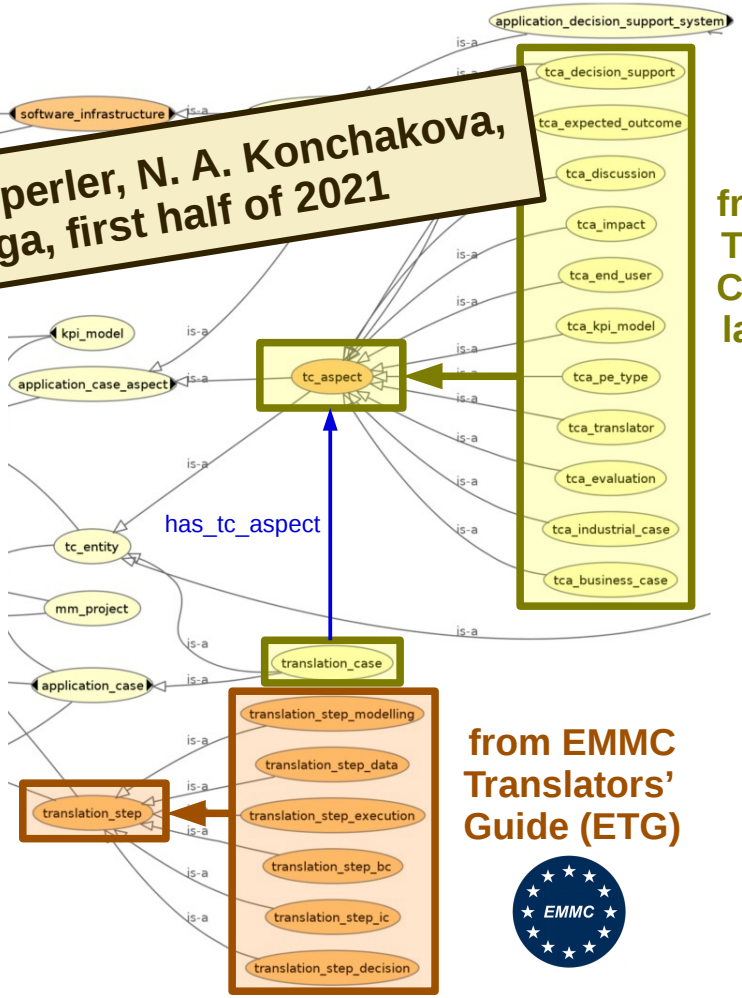


semi-formalized metadata standard (human processable)

formal ontology (machine processable)

Materials Modelling Translation Ontology

"Translation:" Webinar by A. Simperler, N. A. Konchakova, K.-L. Choy, I. Pagonabarraga, first half of 2021

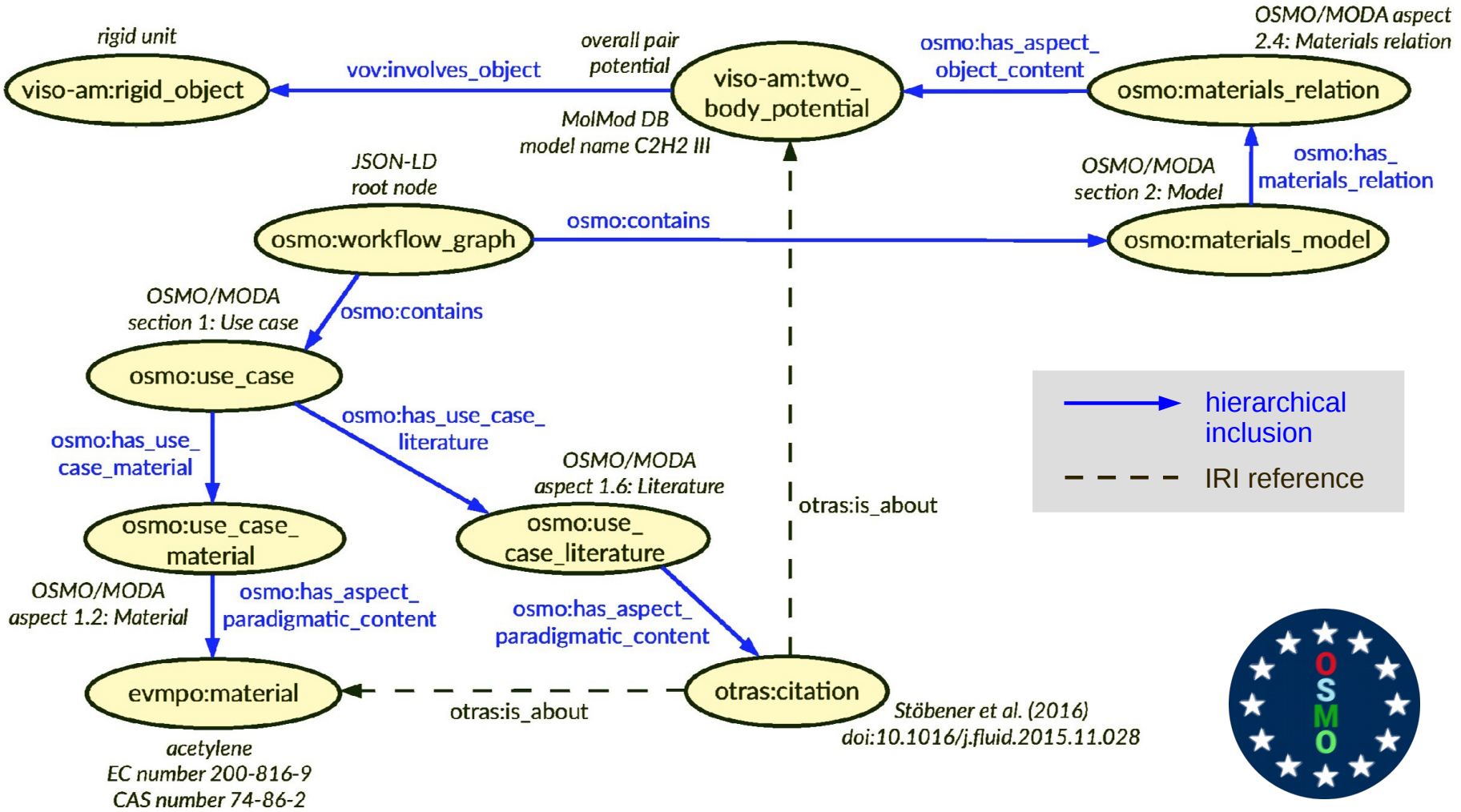


from EMMC Translation Case Template (ETCT)

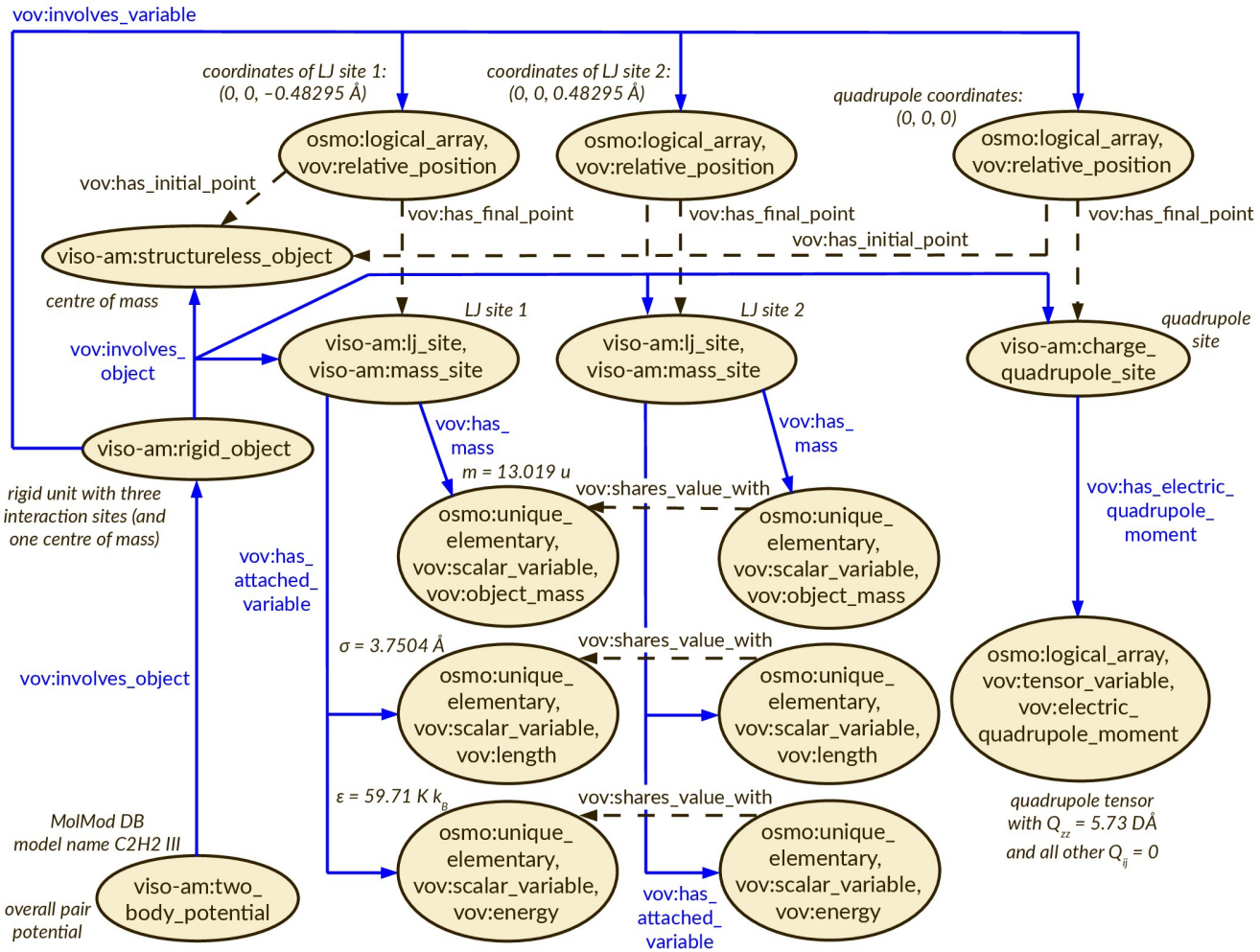
from EMMC Translators' Guide (ETG)

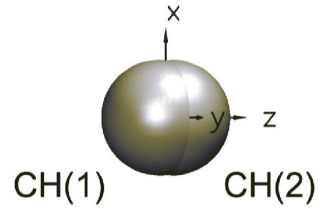
M. T. Horsch, S. Chiacchiera, M. A. Seaton, I. T. Todorov, B. Schembera, P. Klein, N. A. Konchakova, "Pragmatic interoperability and translation [...]," *Proceedings of DAMDID/RCDL*, **2020**.

Knowledge graph for a molecular model: MODA/OSMO



Knowledge graph for a molecular model: VOV/OSMO



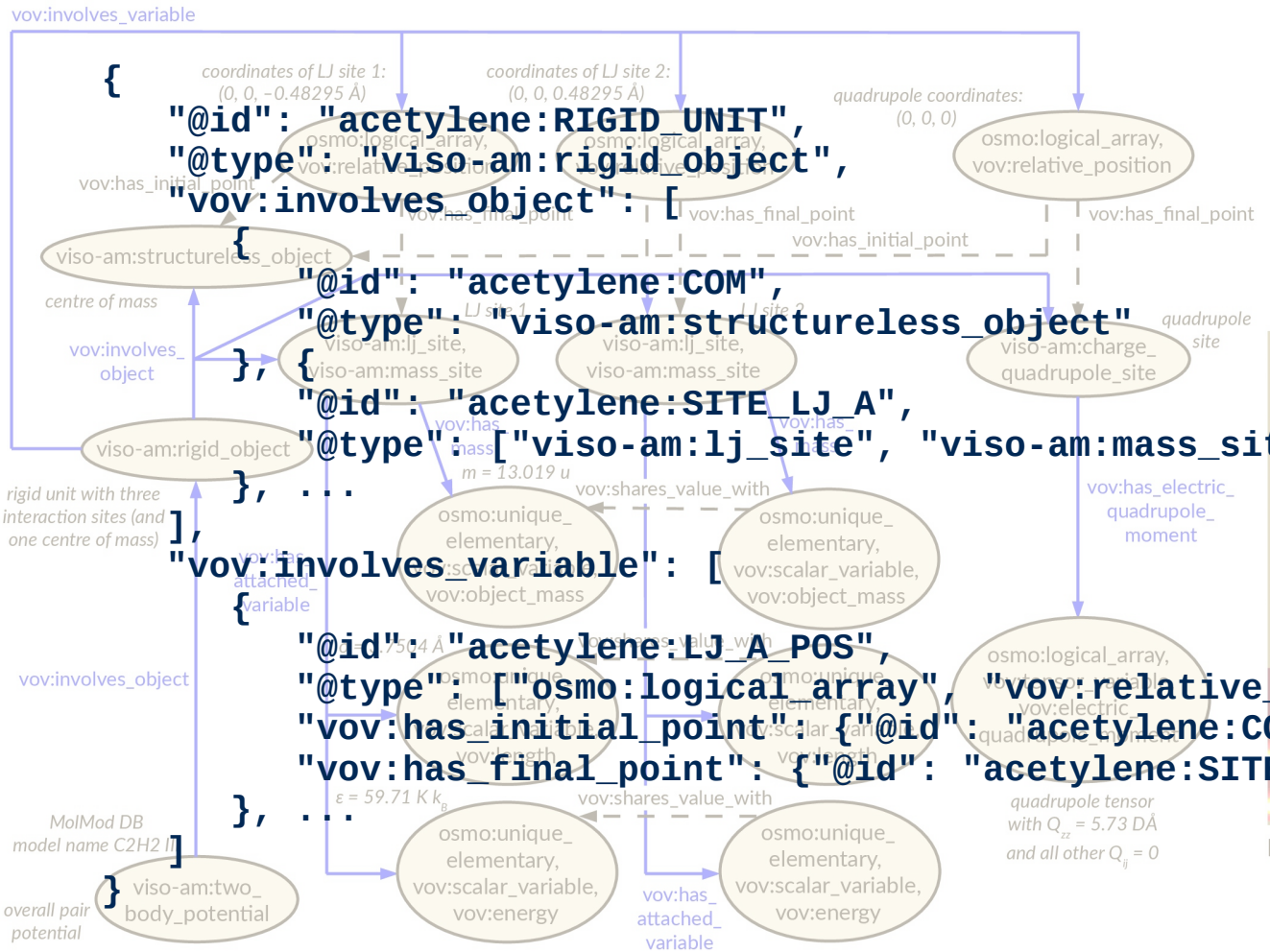


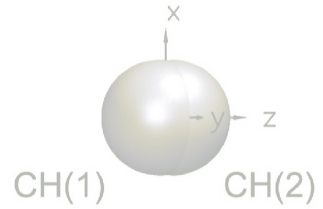
MolMod DB
(Molecular Model Database)

http://molmod.boltzmann-zuse.de/

pair potentials for
over 150 molecular fluids

Knowledge graph for a molecular model: JSON-LD



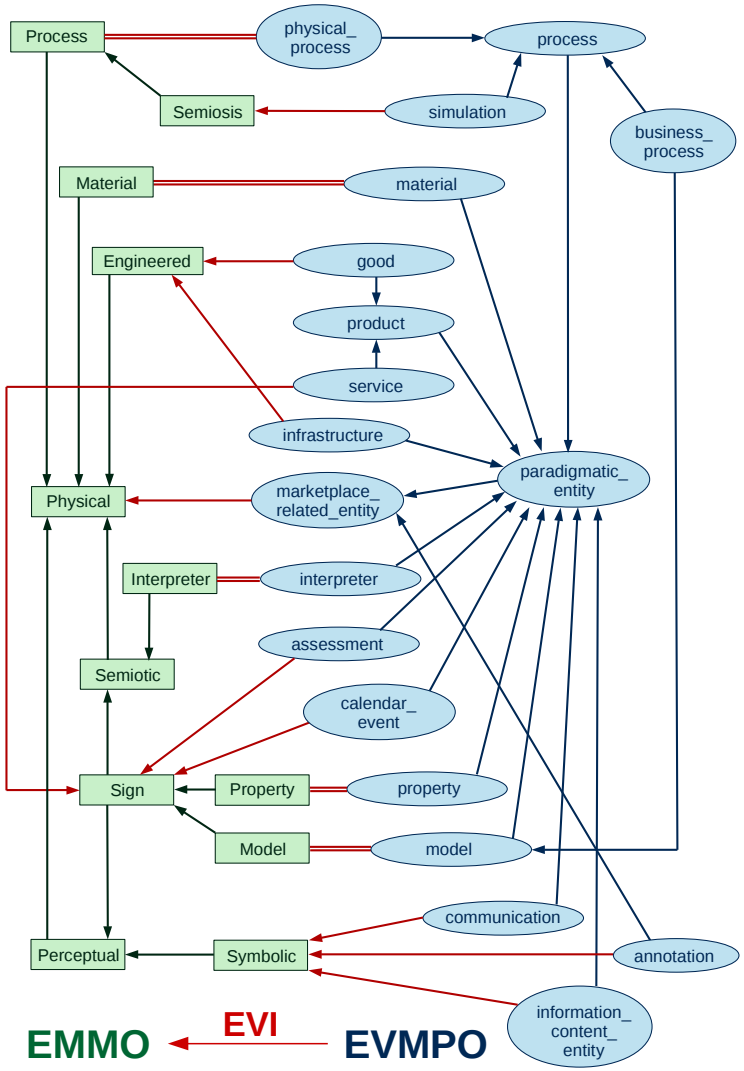


MolMod DB
(Molecular Model Database)

http://molmod.boltzmann-zuse.de/

pair potentials for
over 150 molecular fluids

Alignment of domain ontologies with the EMMO



(0) **annotation** (non-paradigmatic fundamental category), *i.e.*, anything in the knowledge graph that is not under (1) – (11)

(1) **assessment**, *i.e.*, a proposition on accuracy or performance or an expression of trust

(2) **calendar_event**, *i.e.*, a meeting or activity that is scheduled or can be scheduled; from W3C iCal ontology

(3) **communication**, *i.e.*, a message or part of a message (*e.g.*, an attachment) that is communicated

(4) **information_content_entity** from the Information Artifact Ontology; *e.g.*, a journal article, a data set, or a graph

(5) **infrastructure**, *i.e.*, a digital platform infrastructure, *e.g.*, data access, hardware, or software

(6) **interpreter**, *i.e.*, an item that can carry out a semiosis, as formalized by Peirce & the EMMO, creating an interpretant

(7) **material**, *i.e.*, an amount of substance & part of an object

(8) **model**, *i.e.*, a representamen that represents an object by direct similitude or within a mathematical framework

(9) **process**, *i.e.*, temporal evolution of one or multiple entities

(10) **product**, *i.e.*, a good or service that can be traded

(11) **property**, *i.e.*, a representamen that is determined as an interpretant by observation, involving a specific observer

Alignment of domain ontologies with the EMMO

Relations covered by the European Materials and Modelling Ontology[1] (EMMO)

1) **Taxonomy**: Conceptual hierarchy (subclass relation)

2) **Semiotics**: Representation of physical entities by signs

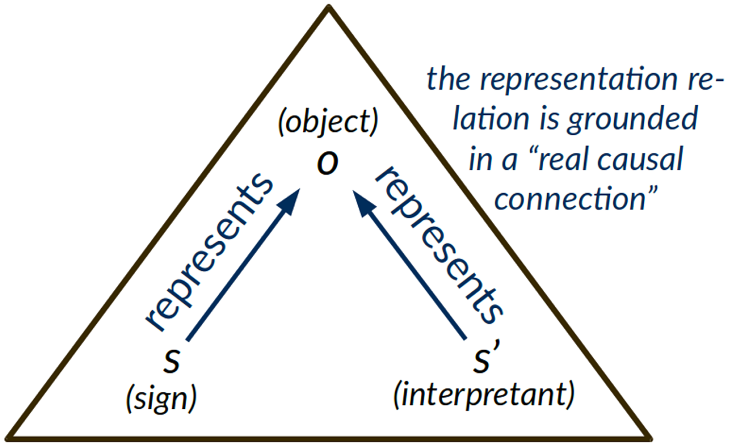
3) **Mereotopology**: Spatiotemporal parthood and connectivity

"represents" or **"is sign for"** will be abbreviated by **S**

semiosis

C. S. Peirce

Peircean semiotics



the semiosis, a process by which a new representamen, the interpretant, is created

[1]E. Ghedini, J. Friis, A. Hashibon, G. J. Schmitz, G. Goldbeck, *et al.*, **2020**; http://emmc.info/emmo-info/.

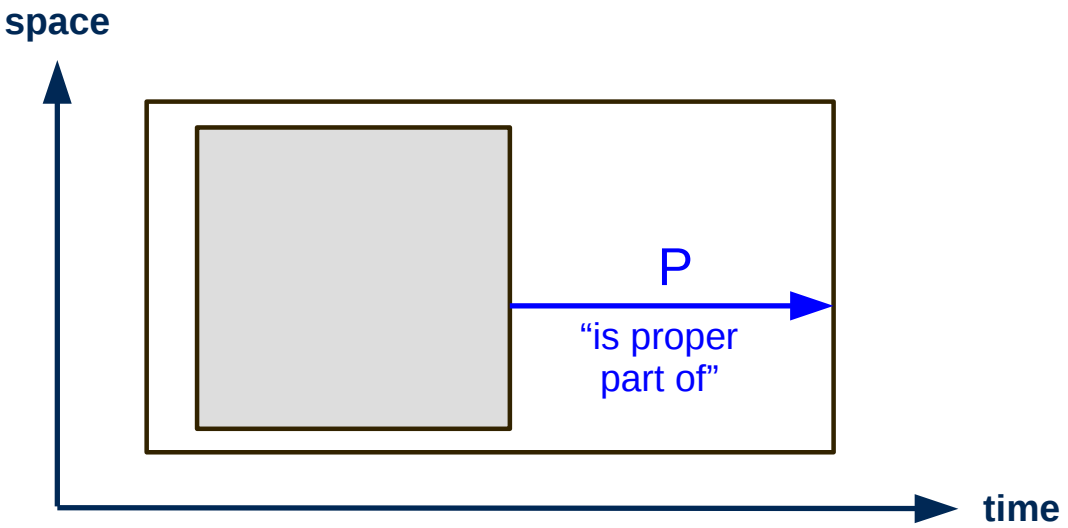
Alignment of domain ontologies with the EMMO

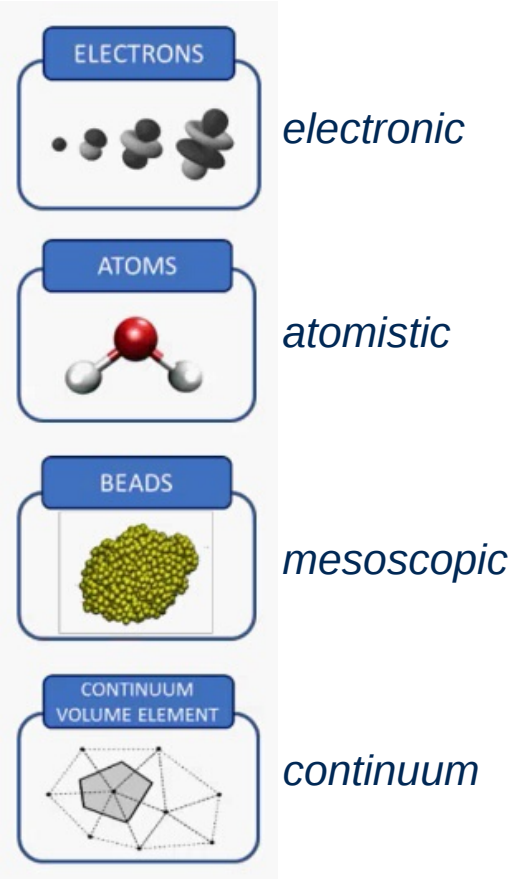
Relations covered by the European Materials and Modelling Ontology[1] (EMMO)

1) **Taxonomy**: Conceptual hierarchy (subclass relation)

2) **Semiotics**: Representation of physical entities by signs

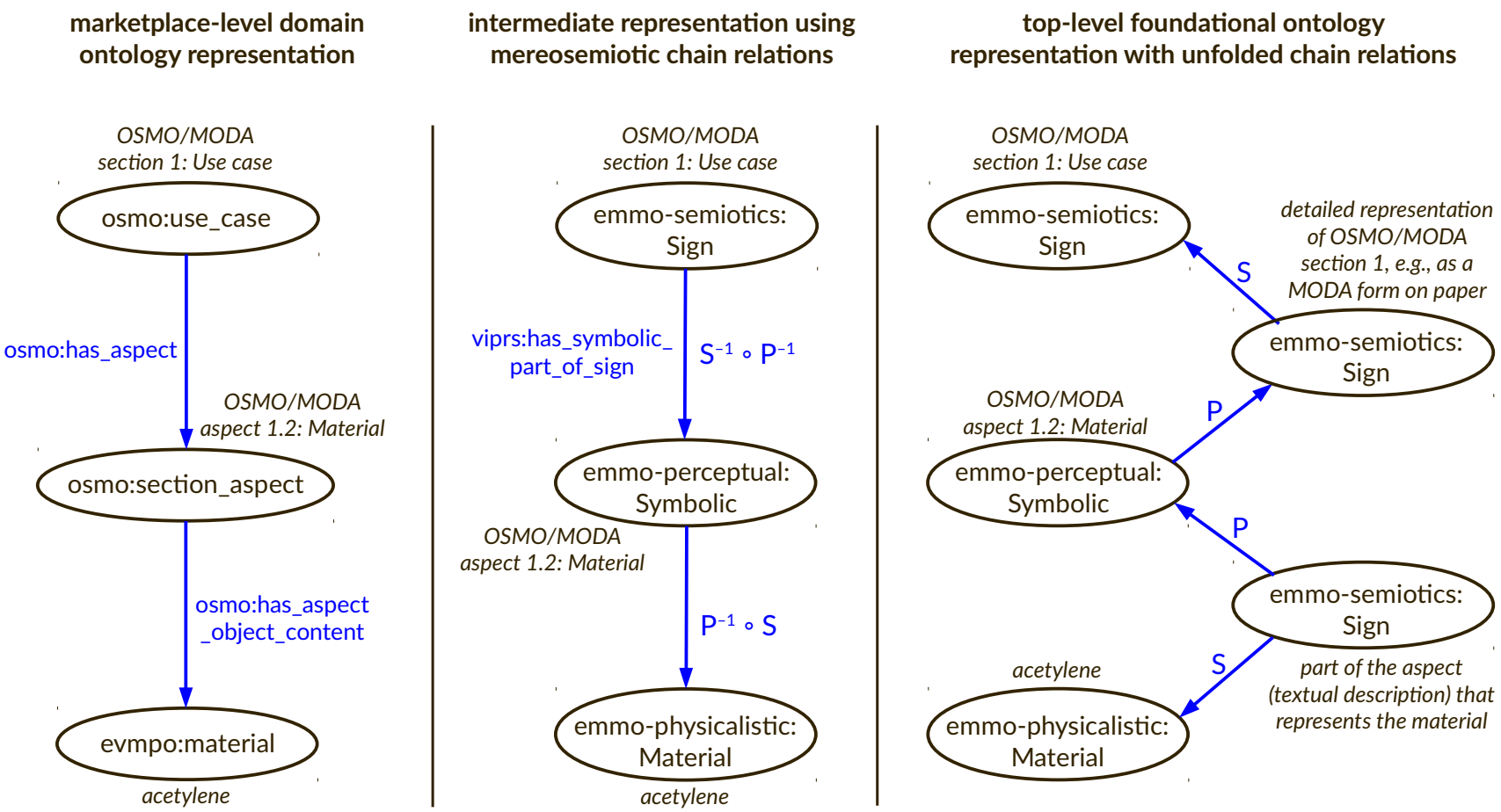
3) **Mereotopology**: Spatiotemporal parthood and connectivity





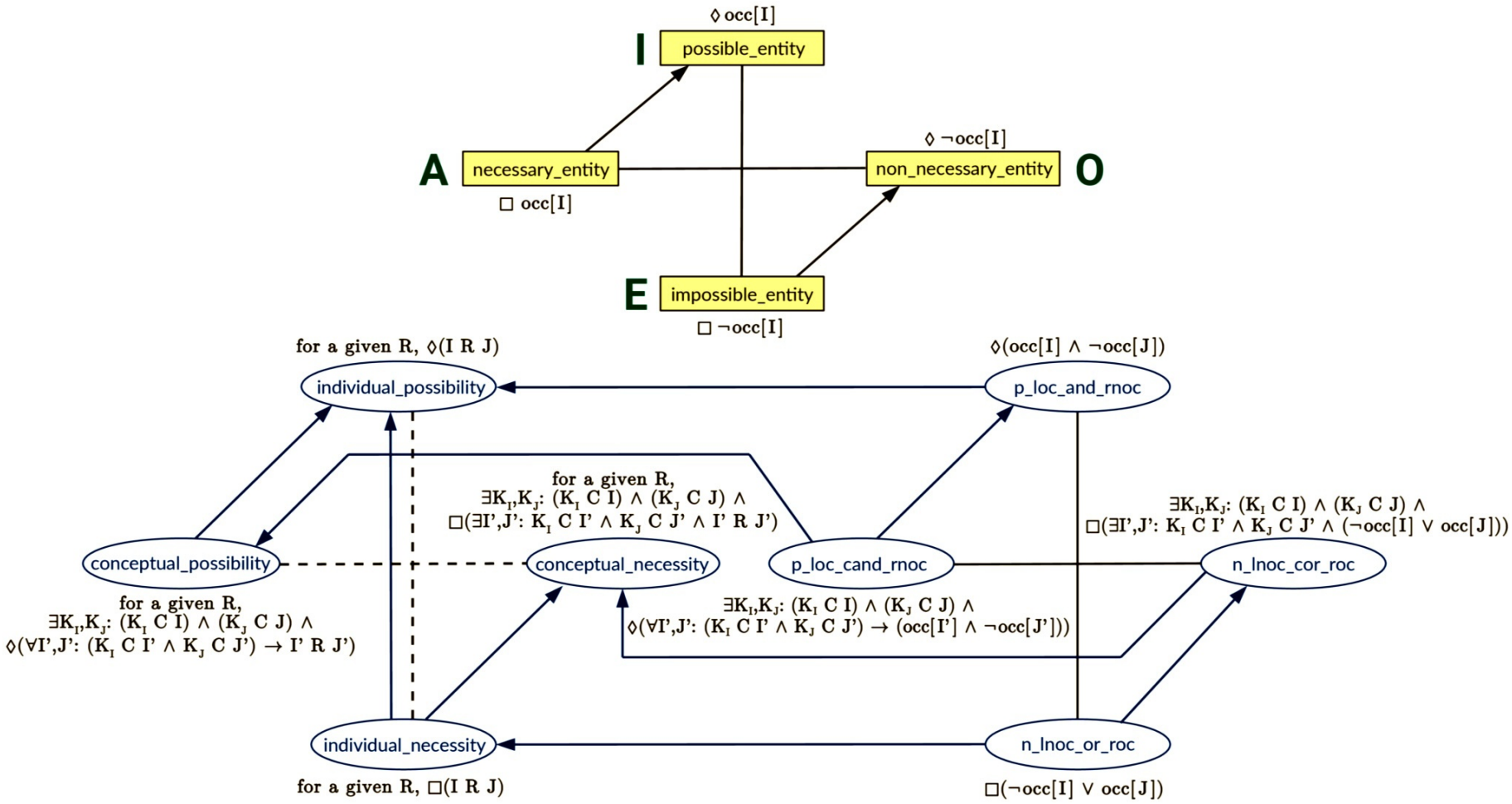
[1]E. Ghedini, J. Friis, A. Hashibon, G. J. Schmitz, G. Goldbeck, *et al.*, **2020**; http://emmc.info/emmo-info/.

Alignment of domain ontologies with the EMMO[1]



[1]M. T. Horsch, S. Chiacchiera, W. L. Cavalcanti, B. Schembera, *Data Technology in Materials Modelling*.

VIMMP Primitives (VIPRS) and modal relations[1]



[1]M. T. Horsch, S. Chiacchiera, W. L. Cavalcanti, B. Schembera, *Data Technology in Materials Modelling*.

Science and Technology Facilities Council

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 760907.

Thank you for your attention!

(Next: 5 minutes for discussion, then SINTEF)

Webinar overview

15.00 – 15.15 CET	Presentation by UKRI STFC
15.15 – 15.20 CET	**Discussion #1**
15.20 – 15.35 CET	Presentation by SINTEF
15.35 – 15.40 CET	Discussion #2
15.40 – 15.45 CET	Community synergy: Ideas from SINTEF
15.45 – 15.50 CET	Community synergy: Ideas from UKRI STFC
15.50 – 16.00 CET	Discussion #3

Webinar overview

15.00 – 15.15 CET	Presentation by UKRI STFC
15.15 – 15.20 CET	Discussion #1
15.20 – 15.35 CET	Presentation by SINTEF
15.35 – 15.40 CET	Discussion #2
15.40 – 15.45 CET	Community synergy: Ideas from SINTEF
15.45 – 15.50 CET	**Community synergy: Ideas from UKRI STFC**
15.50 – 16.00 CET	Discussion #3

Ideas for possible synergies between marketplaces

Problem: how to combine the marketplaces efforts and databases, for example so that they can be queried simultaneously?

Business level - “How do users register/get access?”

Not the focus of this talk.

Technical level - “How can users look for things?”

Beside a GUI (for manual upload and search), the VIMMP database will be accessible *via* a REST API, a common approach for web services. The metadata are typically exchanged in JSON format.

Content level - “What can users look for?”

The details depend of course on each database structure: which type of data is stored, what properties are given.

GUI: Graphical User Interface
REST API: Representational state transfer (REST) Application Programming Interface (API)

Ideas for possible synergies between marketplaces (2)

Problem: how to combine the marketplaces efforts and databases, for example so that they can be queried simultaneously?

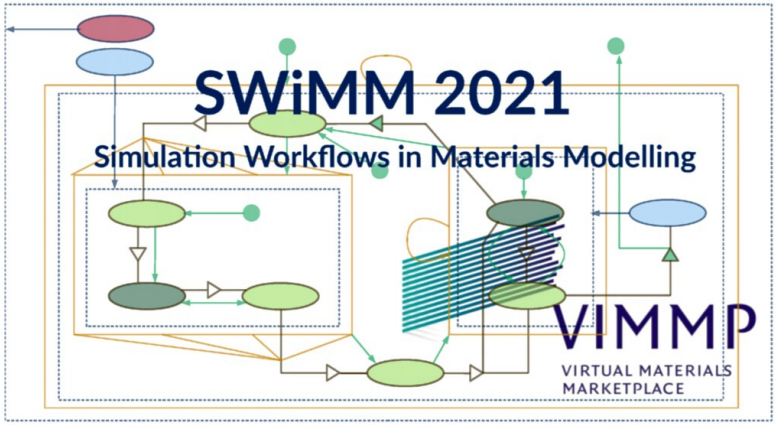
→ Focusing on the **content level**, some possibilities for integration:

- Building on the EVMPO idea: agree on a minimal set of concepts for which data will be stored and visible in the API. E.g., “translator” profiles, including “topic of expertise”.
- Otherwise (or for legacy work) provide a sort of mediation schema: a third model, with mappings to the two to be integrated.

Broadening the discussion:

- Coordinate with prospective users/providers of the marketplaces, e.g. the open simulation platforms, to identify prototypical queries from their side. This can also help us design the databases.

CECAM school in March 2021 supported by VIMMP



Simulation Workflows in Materials Modelling

15th – 19th March 2021

CECAM HQ
École Polytechnique Fédérale de Lausanne

https://www.cecam.org/workshop-details/27

cecam
Centre Européen de Calcul Atomique et Moléculaire

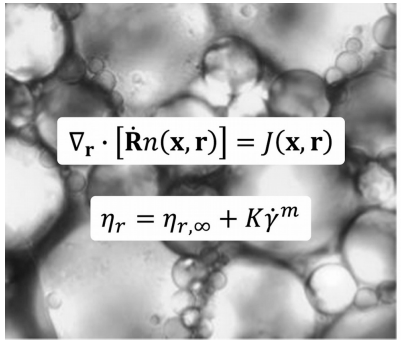
1. Salome and YACS: An integration platform for workflows
2. Industral-accuracy data-driven model parameterization
3. Autotuning, load balancing, and task based parallelization
4. Semantic interoperability and ontology-driven technology
5. European Materials and Modelling Ontology
6. The Pyiron IDE for simulation workflows
7. The atomic simulation environment Python library
8. Complex workflows with AiiDA and Materials Cloud

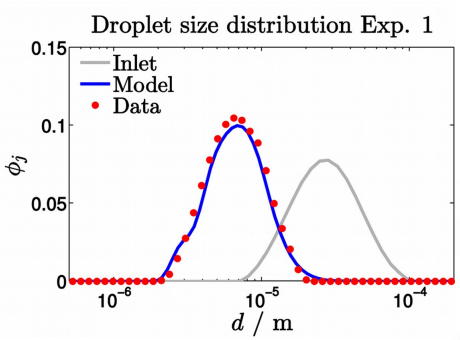
Next webinar

The VIMMP multiscale-modelling workflow for the simulation of food emulsions

Gianluca Boccardo, Adam Kowalski, Marco Trofa, and Marco Ferrari

Wednesday, 9th December 2020, 10.00 CET





VIMMP
VIRTUAL MATERIALS MARKETPLACE

EMMC

POLITECNICO DI TORINO 1859 1906

Unilever

Science and Technology Facilities Council

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 760907.

Thank you
for
your attention!

(Next: 10 minutes for discussion)

Webinar overview

15.00 – 15.15 CET	Presentation by UKRI STFC
15.15 – 15.20 CET	Discussion #1
15.20 – 15.35 CET	Presentation by SINTEF
15.35 – 15.40 CET	Discussion #2
15.40 – 15.45 CET	Community synergy: Ideas from SINTEF
15.45 – 15.50 CET	Community synergy: Ideas from UKRI STFC
15.50 – 16.00 CET	**Discussion #3**