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

Digitalization in materials modelling: The Virtual Materials Marketplace

VIMMP
VIRTUAL MATERIALS MARKETPLACE

ReaxPro Conference

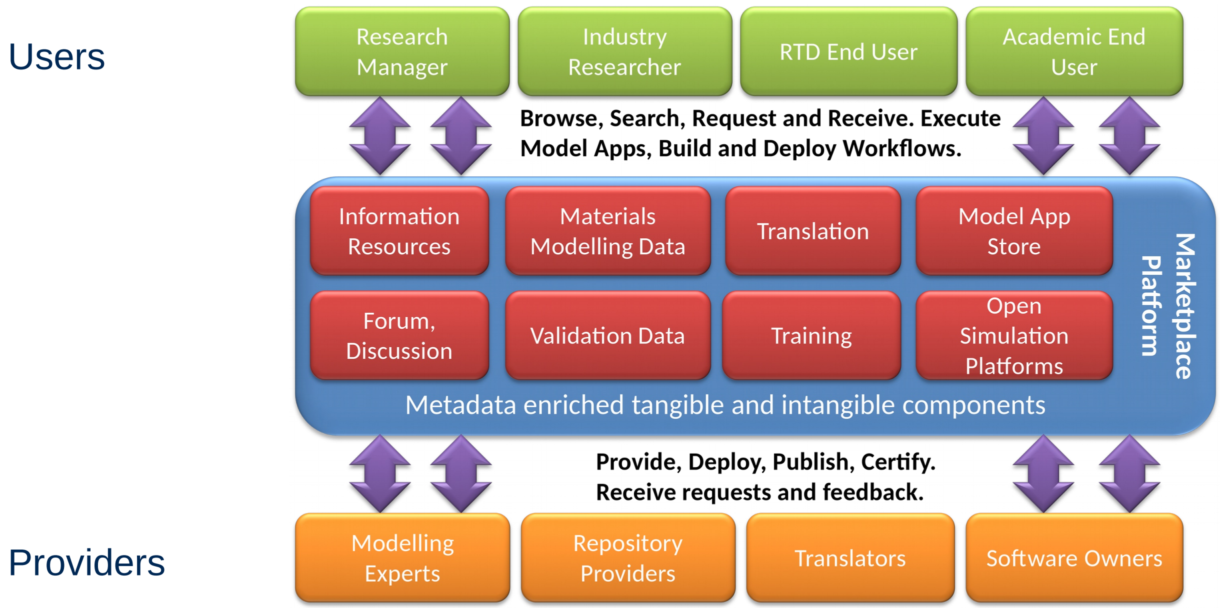
10th July 2020

UKRI 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 undoubtly 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.

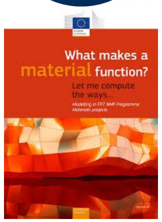
- **Richer languages (e.g., OWL)**:

  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?

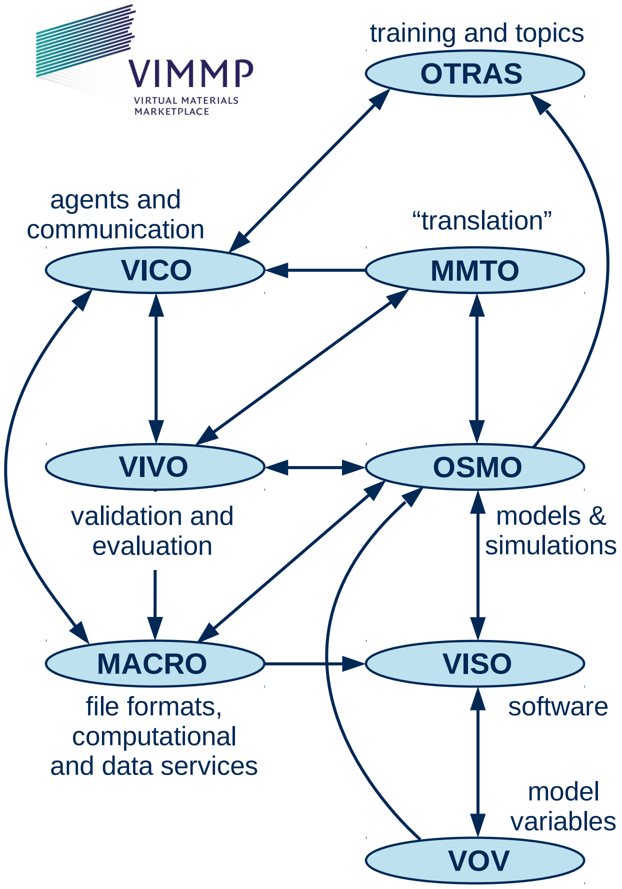
Ontologies allow 1) automatic reasoning, 2) easier exchange of information across heterogeneous sources.

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.

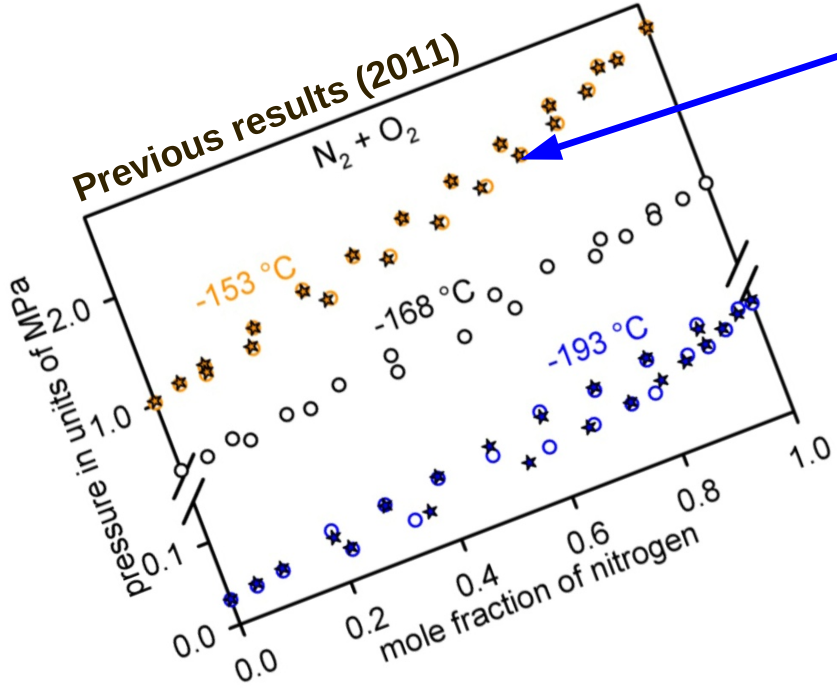
How will this help/affect users and providers?

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.

Provenance description of thermophysical data

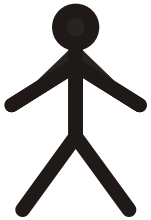


What values did *x* and *p* have?

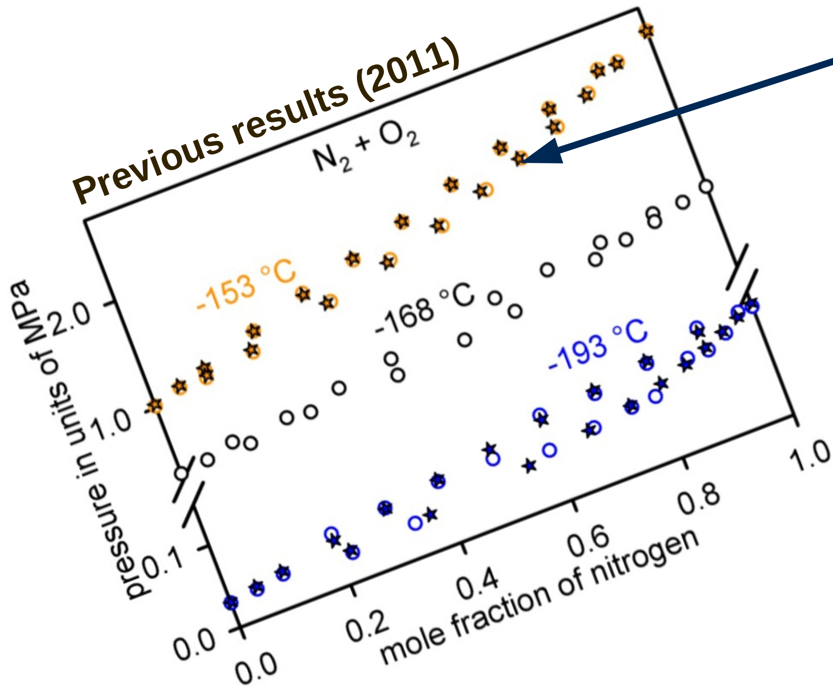
How was the data point obtained?

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



Provenance description of thermophysical data

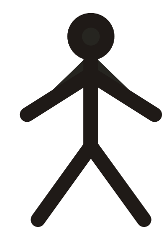


What values did *x* and *p* have?

How was the data point obtained?

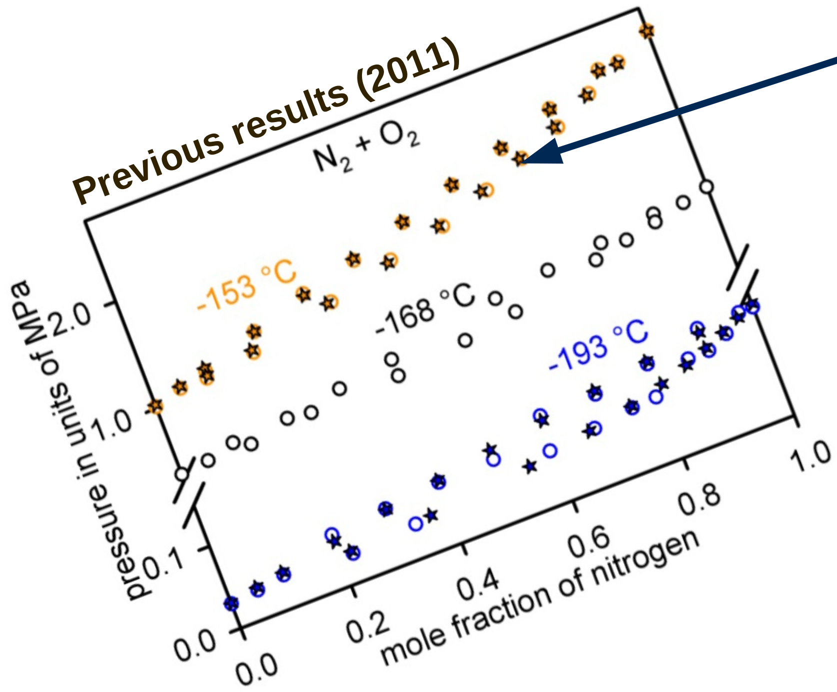
What is the margin of error, how was the error defined, and what software (or experimental setup) was used?

ask the person who carried out the work back in 2011



"I remember."

Provenance description of thermophysical data

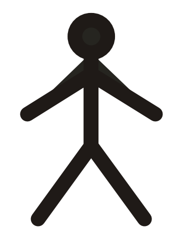


What values did *x* and *p* have?

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."

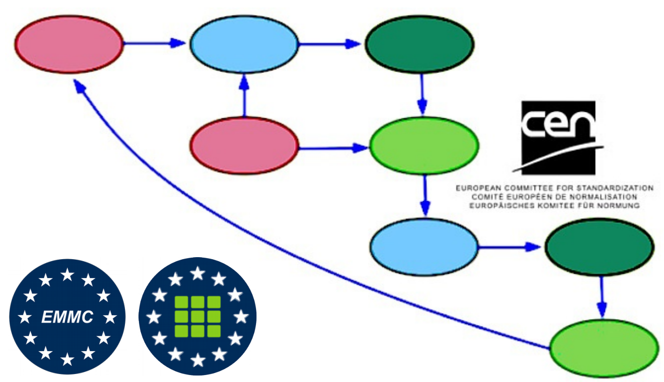
Good practice in handling research data:

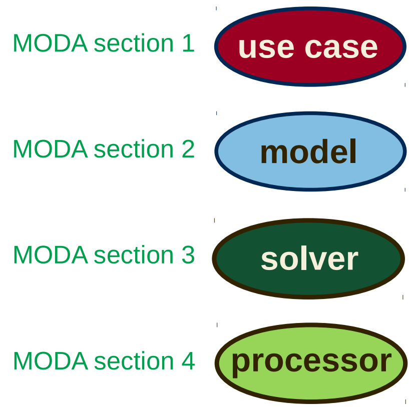
Keep all data findable, accessible, interoperable, and reusable (FAIR).

Provenance description of simulation results

1	Aspect of the User Case/System to be Simulated	
1.1	Aspect of the User Case to be Simulated	Describe the aspects of the user case textually. No modelling information should appear in this box. This case could also be simulated by other models in a benchmarking operation! The information in this chapter can be end-user information, measured data, library data etc. It will appear in the pink circle of your workflow picture. Simulated input which would have been calculated by another model should not be included (but in chapter 2.4) Also the result of pre-processing necessary to translate the user case specifications to values for the physics variables of the entities van be documented here.
1.2	Material	Describe the chemical composition, ...and the values used for properties and from which database these are taken. If pre-processing was needed please specify the methodology.
1.3	Geometry	Size, form, picture of the system (if applicable) Note that computational choices like simulation boxes are to be documented in chapter 3.
1.4	Time Lapse	Duration of the case to be simulated. This is the duration of the situation to be simulated. This is not the same as the computational times to be given in chapter 3.
1.5	Manufacturing Process or In-Service Conditions	If relevant, please list the conditions to be simulated (if applicable). These can be boundary, initial and global conditions. E.g. heated walls, external pressures and bending forces. Please note that these might appear as terms in the PE or as boundary conditions, and this will be documented in the relevant chapters. Note: These conditions will be expressed in physics relations in Ch 2.4 Please specify the values used for parameters and from which database these are taken. If pre-processing was needed please specify the methodology.
1.6	Publication on this Data	Publication documenting the simulation with this single model (if available and if not already included in the overall publication).

MODA workflow description



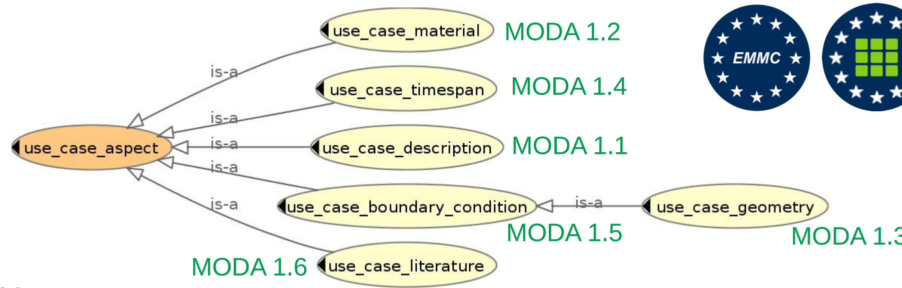


Provenance description of simulation results

MODA section 1

OSMO use_case **has_use_case_aspect** →

has_use_case_description,
has_use_case_material,
has_use_case_geometry,
has_use_case_timespan,
has_use_case_boundary_condition,
has_use_case_literature



MODA section 1 — **use case**

MODA section 2 — **model**

MODA section 3 — **solver**

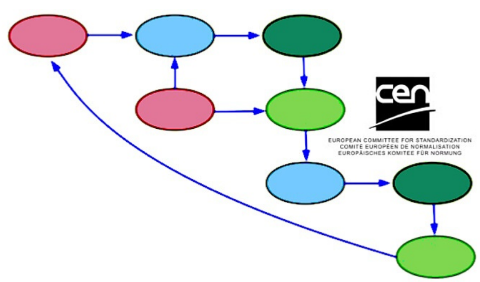
MODA section 4 — **processor**

"sections"

OSMO

"aspects"

"graphs"



1	**Aspect of the User Case/System to be Simulated**	
1.1	**Aspect of the User Case to be Simulated**	*Describe the aspects of the user case textually.* *No modelling information should appear in this box. This case could also be simulated by other models in a benchmarking operation! The information in this chapter can be end-user information, measured data, library data etc. It will appear in the pink circle of your workflow picture. Simulated input which would have been calculated by another model should not be included (but in chapter 2.4)* *Also the result of pre-processing necessary to translate the user case specifications to values for the physics variables of the entities van be documented here.*
1.2	**Material**	*Describe the chemical composition, …and the values used for properties and from which database these are taken. If pre-processing was needed please specify the methodology.*
1.3	**Geometry**	*Size, form, picture of the system (if applicable)* *Note that computational choices like simulation boxes are to be documented in chapter 3.*
1.4	**Time Lapse**	*Duration of the case to be simulated.* *This is the duration of the situation to be simulated. This is not the same as the computational times to be given in chapter 3.*
1.5	**Manufacturing Process or In-Service Conditions**	*If relevant, please list the conditions to be simula… These can be boundary, initial and global condi…* *E.g. heated walls, external pressures and be… Please note that these might appear as term… conditions, and this will be documented in t…* *Note: These conditions will be expressed in… Ch 2.4 Please specify the values used for paramete… are taken. If pre-processing was needed pleas…*
1.6	**Publication on this Data**	*Publication documenting the simulation with this … and if not already included in the overall publication).*

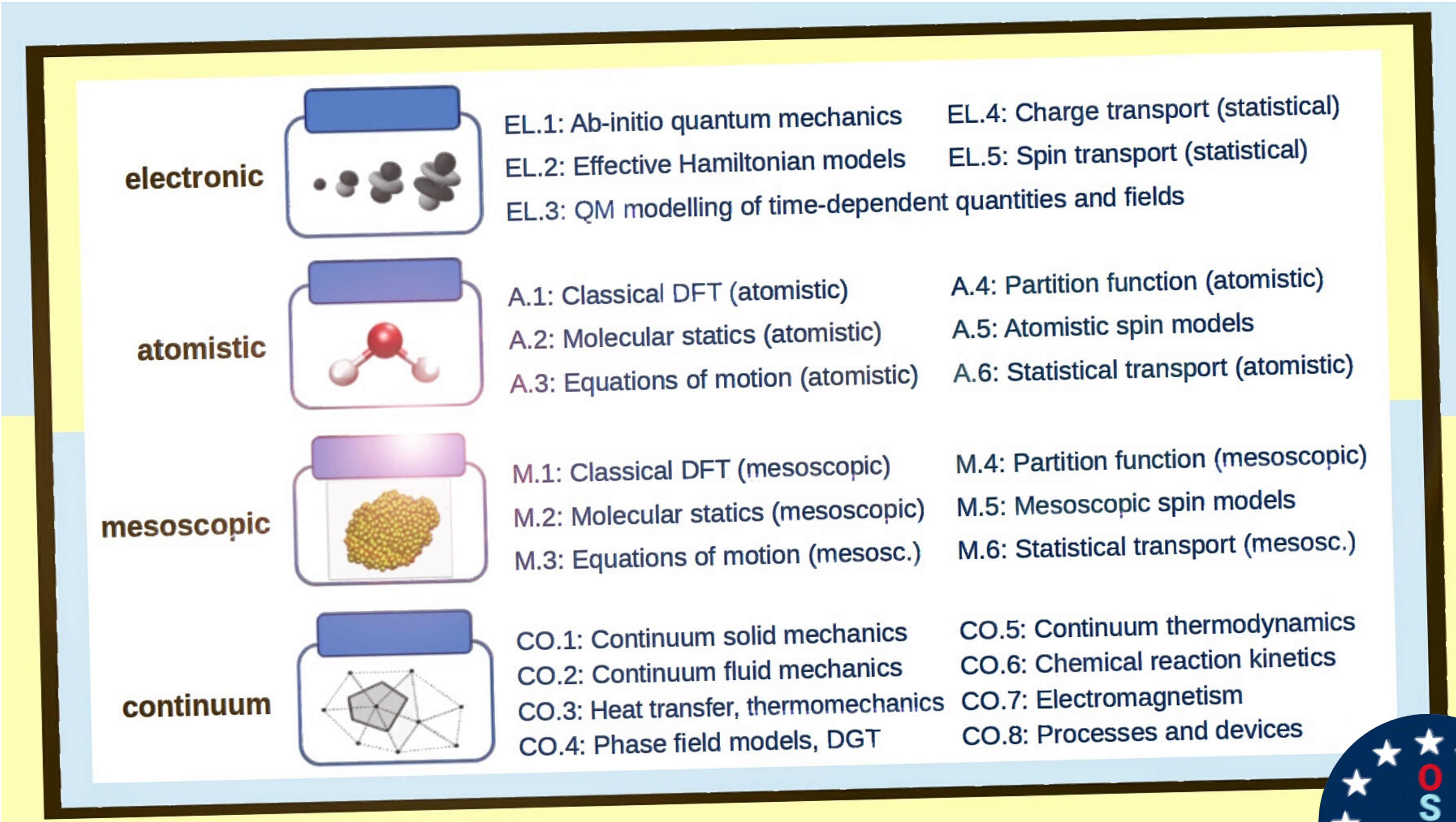
Ontology for Simulation, Modelling, and Optimization

J. Chem. Eng. Data 65(3), 1313–1329, 2020

Science and Technology Facilities Council

VIMMP VIRTUAL MATERIALS MARKETPLACE

Provenance description of simulation results

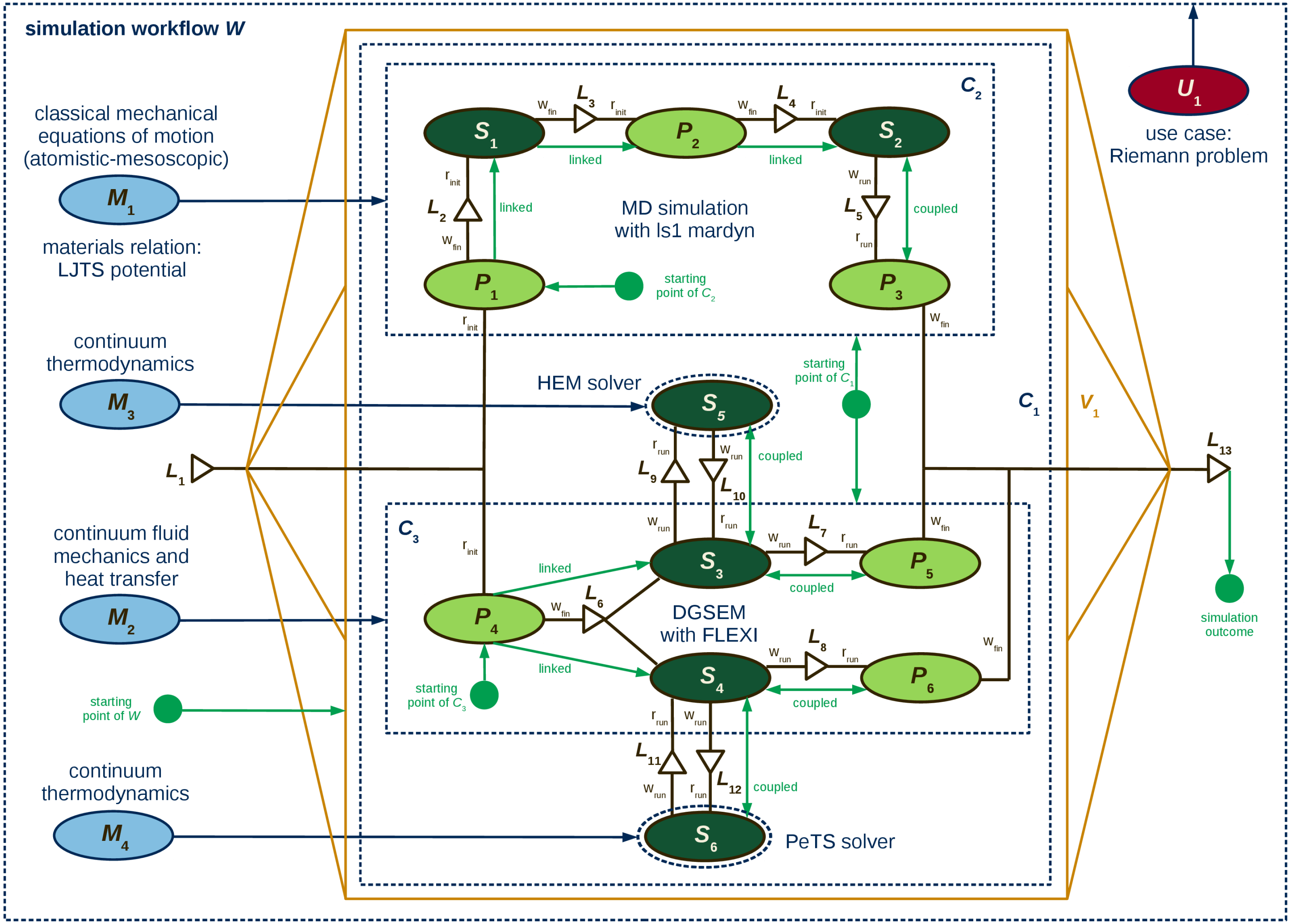


Ontology for Simulation, Modelling, and Optimization
J. Chem. Eng. Data 65(3), 1313–1329, 2020

Science and Technology Facilities Council

VIMMP VIRTUAL MATERIALS MARKETPLACE

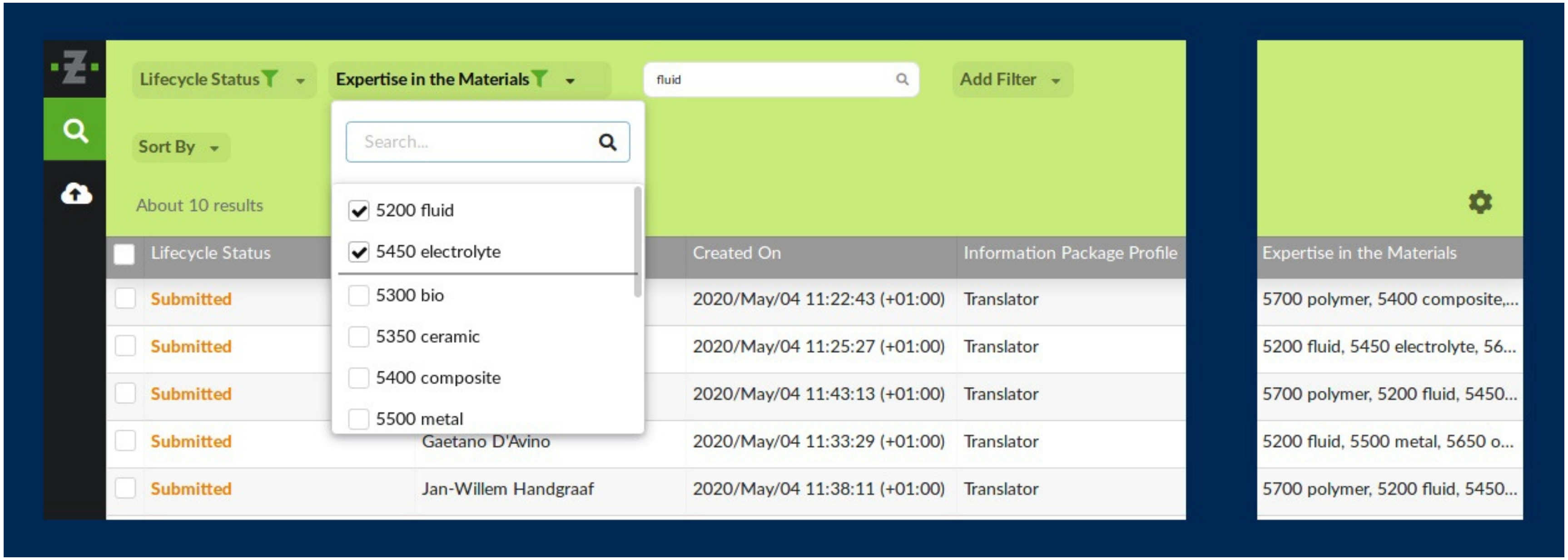
Provenance description of simulation results

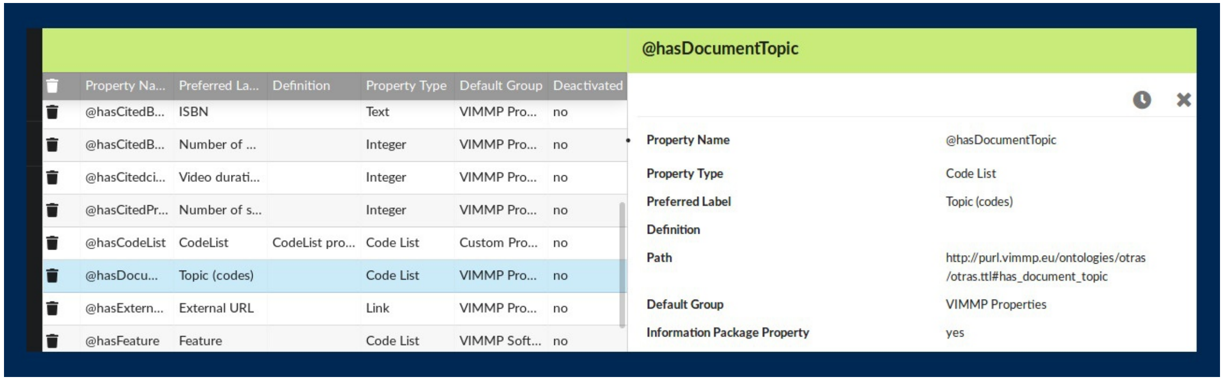


OSMO-based **logical data transfer (LDT)** extension of MODA:

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

Data management on the VIMMP backend





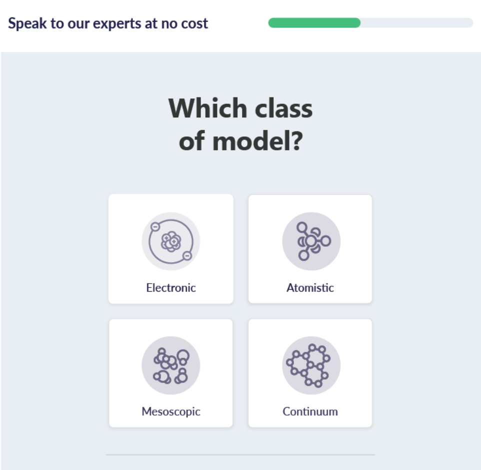
Metadata-supported data ingest and retrieval

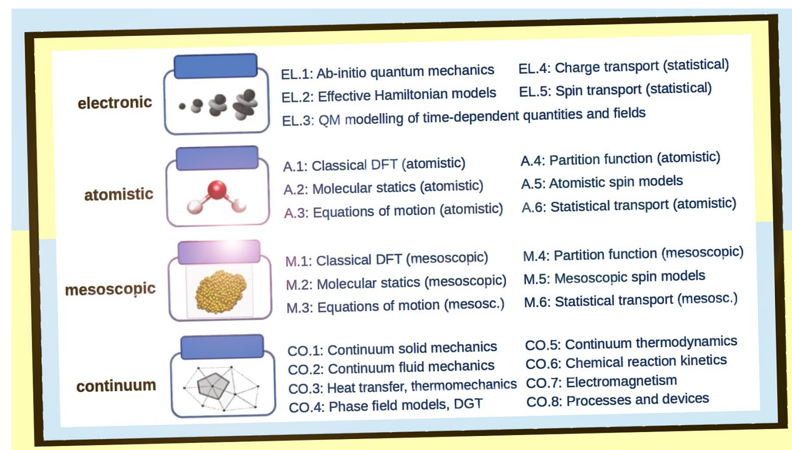
mm_topic_basic (codes **1XXX** and **2XXX**):
Basic prerequisites for materials modelling.
mm_topic_computational (codes **3XXX**):
Computational and numerical aspects of materials modelling.
mm_topic_data (codes **4XXX**):
Data science and technology aspects.
mm_topic_materials (codes **5XXX**):
Topics 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.*



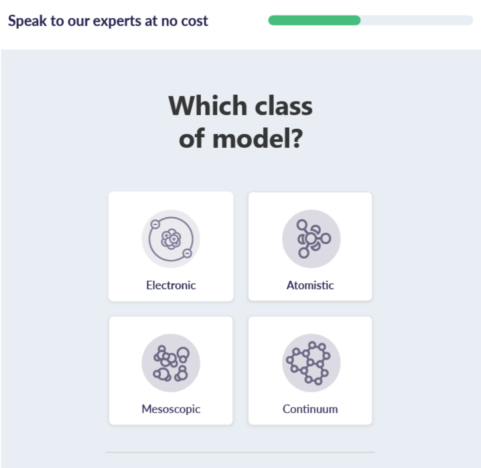


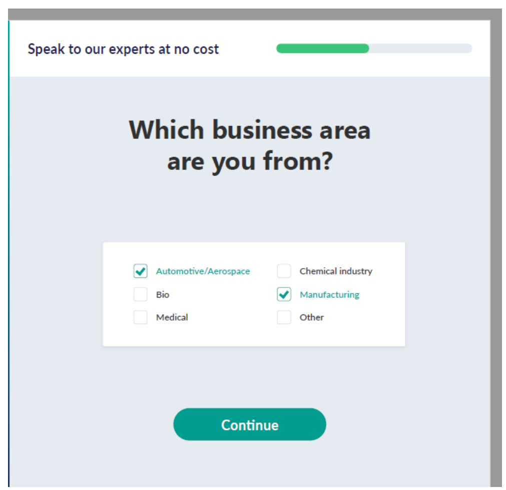
Metadata-supported data ingest and retrieval

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

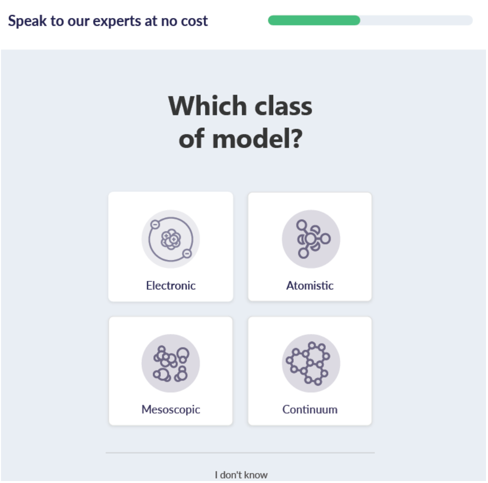


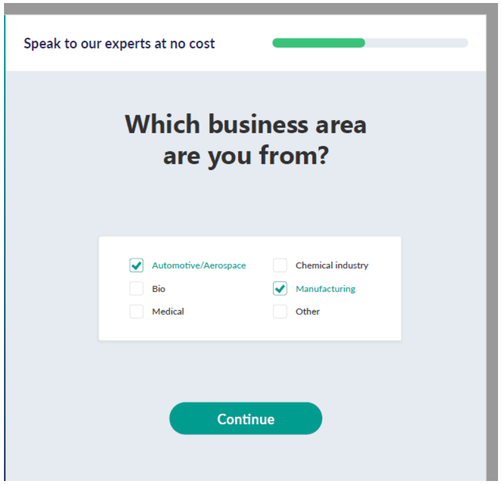


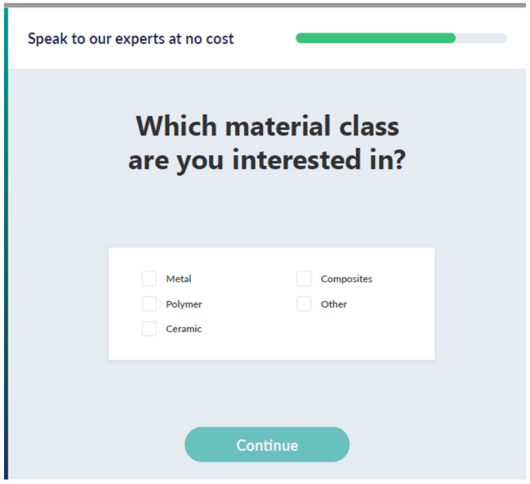
Metadata-supported data ingest and retrieval

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.

- 5100 general
- 5200 fluid
- 5300 bio
- 5350 ceramic
- 5400 composite
- 5450 electrolyte
- 5500 metal
- 5550 mineral
- 5600 nano
- 5650 organic
- 5700 polymer
- 5750 semiconductor
- 5800 ultracold
- 5850 unstable
- 5900 special topics







Community involvement in collaboration with the EMMC

Ontology governance, continuous development and maintenance

https://emmc.eu/

European Materials Modelling Council (EMMC ASBL)

The non-profit association EMMC ASBL was created in 2019 to ensure the continuity, growth, and sustainability of community activities for modellers, materials data scientists, software owners, materials modelling translators, and manufacturers in Europe. The EMMC regards the **integration of materials modelling and digitalization** as critical for an advancement of industrial process and product design.

EMMC Focus Area on Digitalization

In computational engineering, digitalization encompasses aspects of representing, managing, accessing, and utilizing digital information about products, components, materials, their behaviour, and their processing.

Community involvement in collaboration with the EMMC

Ontology governance, continuous development and maintenance

https://emmc.eu/

Community feedback and involvement:

- Use of the EMMC Forum as a collaborative workspace for communication on implementation and extension.
- Moderation of ontology feedback and review processes by the EMMC Task Group on Digital Marketplaces, which is to be launched soon by the EMMC Focus Areas on Digitalization and Interoperability.

EMMC Focus Area on Digitalization

In computational engineering, digitalization encompasses aspects of representing, managing, accessing, and utilizing digital information about products, components, materials, their behaviour, and their processing.

Community-governed top-level interoperability layer

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

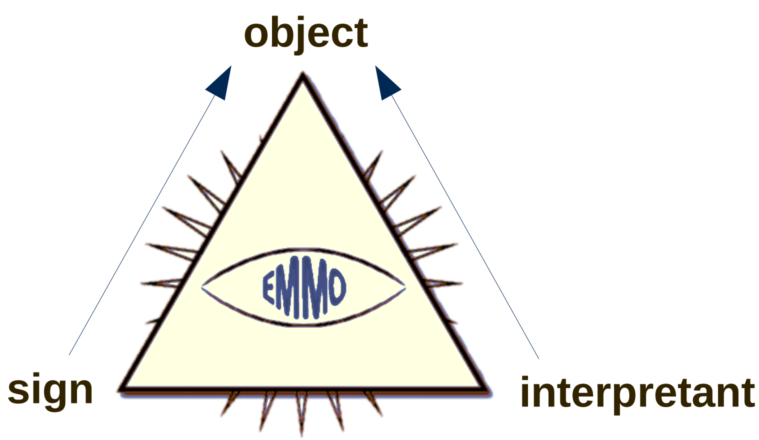
1) Taxonomy: Subclass relation (between classes)

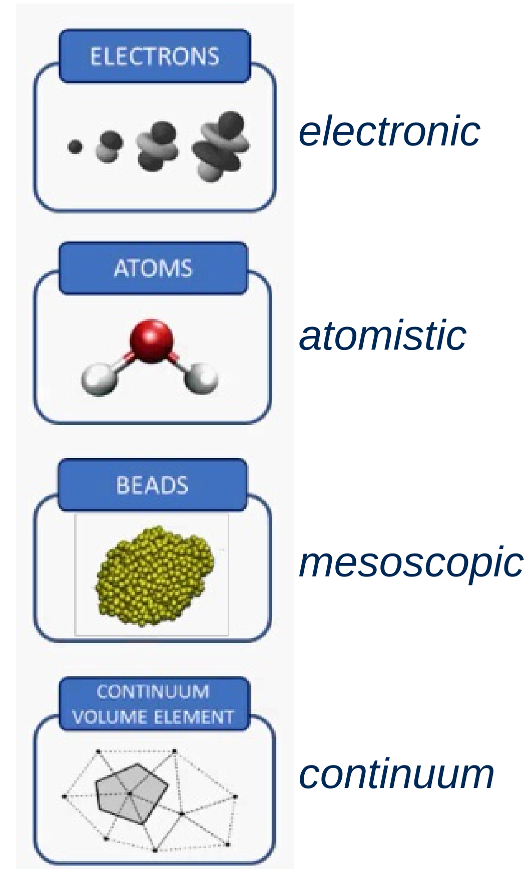
2) Semiotics: Representation of physical entities by signs

3) Mereotopology: Slicing, parthood, collection membership

semiosis

C. S. Peirce





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

Announcement: CECAM school on simulation workflows

Simulation Workflows in Materials Modelling

15th – 19th March 2021

CECAM HQ
École Polytechnique Fédérale de Lausanne

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

Centre Européen de Calcul Atomique et Moléculaire

1. Salome and YACS: An integration platform for workflows
2. Atomic Simulation Environment (ASE)
3. Semantic interoperability and ontologies for workflows
4. Pyiron: An IDE for simulation workflows
5. Data-driven models from high throughput simulation
6. AiiDA + Materials Cloud informatics platform
7. Autotuning, load balancing, and task based parallelization

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