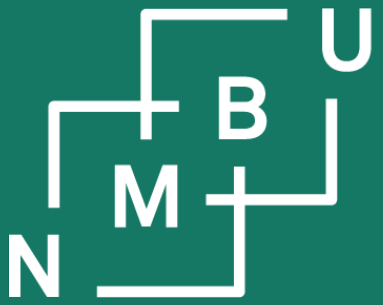


Norges miljø- og biovitenskapelige universitet

Standardized documentation of workflows and epistemic metadata for digitalization and interoperability in materials modelling

M. T. Horsch,[1,2] S. Chiacchiera,[2] F. Diewald,[3] P. Klein,[3] N. A. Konchakova,[4] P. Neumann,[5] H. A. Preisig,[6] B. Schembera,[7] S. Stephan,[8] I. T. Todorov,[2] and J. Vrabec[9]

[1]Norweg. Univ. Life Sciences, [2]UKRI STFC Daresbury Laboratory, [3]Fraunhofer ITWM, [4]Helmholtz-Zentrum Hereon, [5]Helmut Schmidt Univ. Hamburg, [6]Norweg. Univ. Science & Technology, [7]Univ. Stuttgart, [8]TU Kaiserslautern, [9]TU Berlin

eSSENCE-EMMC Meeting 2022 1st June 2022 Uppsala, Ångströmlaboratoriet

Noregs miljø- og biovitskaplege universitet

Platform interoperability principles and gaps

What are the interoperability requirements of digital platforms and research data infrastructures as regards relevant epistemological propositions?

How should we address this specifically within molecular and multiscale modelling?

VIMMP
VIRTUAL MATERIALS MARKETPLACE
H2020 GA no. 760907

VIPCOAT
H2020 GA no. 952903

DOME 4.0
H2020 GA no. 953163

ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS
H2020 GA no. 958371

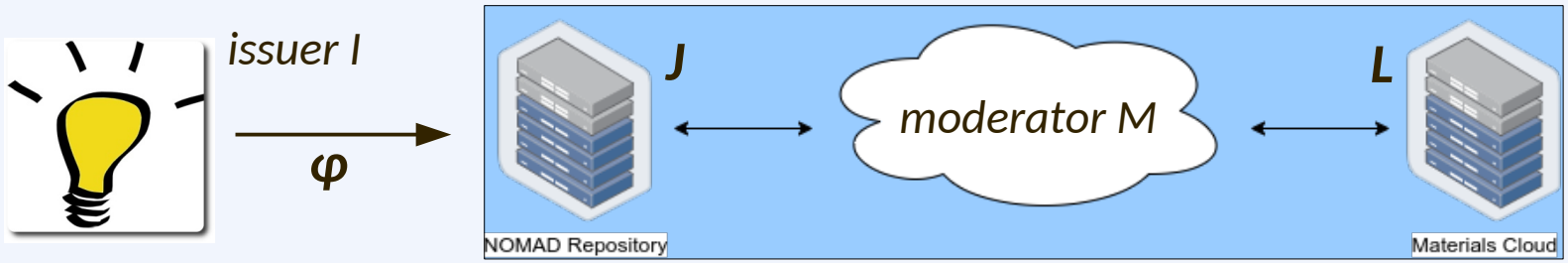
eSSENCE-EMMC Meeting

1st June 2022

Uppsala, Ångström Lab

Communication of knowledge

Scientific knowledge is a kind of knowledge (or else, little will qualify as knowledge). Research data infrastructures store and exchange scientific knowledge.



Scenario requiring epistemological formalization:

- "*M* asserts and approves $\varphi'(I, J, L, \varphi)$," where $\varphi'(I, J, L, \varphi)$ is given by:
- "The scientific knowledge φ, previously issued by a source *I*, has been communicated by the knowledge base *J* to the knowledge base *L*."
- *J*, *L*, and *M* have a justified true belief in φ'.
- φ is a justified tenable assertion, by the standards applied to *I* by *M*.

Norwegian University of Life Sciences

Epistemic opacity

Epistemic opacity (Humphreys, 2011): A cognitive "process is **epistemically opaque** relative to a cognitive agent *X* at time *t* just in case *X* does not know at *t* all of the **epistemically relevant elements** of the process."

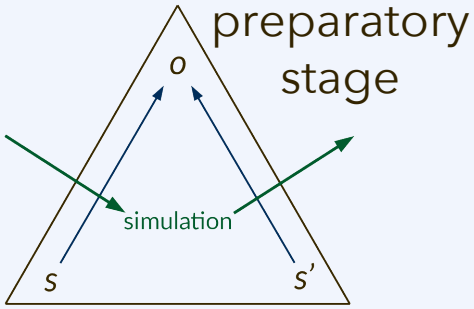
Epistemic metadata: Information that should be included in an adequate response to the queries "what **knowledge claims** have been formulated on the basis of the given data?" and "what exactly is the relation between the knowledge claims, their proponents, and the data?"

European AI Act proposal: "To address the **opacity** that may make certain AI systems **incomprehensible to or too complex for natural persons**, a certain degree of transparency should be required for high-risk AI systems.[1] Users should be able to interpret the system output and use it appropriately. High-risk AI systems should therefore be accompanied by **relevant documentation**".

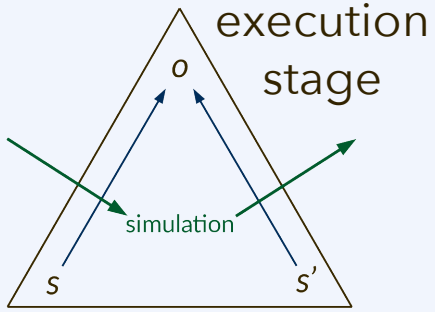
[1]Systems with "high risk" include all "safety components" related to "water, gas, heating, and electricity."

The aim: Epistemic FAIRness

Epistemic opacity and darkness of data can be countered by **epistemic FAIRness**, *i.e.*, FAIR provision of all the **relevant epistemic metadata** via digital infrastructures. Such infrastructures must permit reevaluating processes and results.

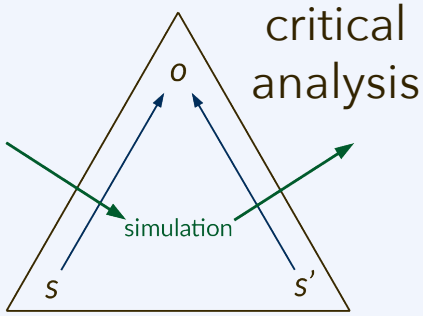


ex ante
Planning



in actu
Steering

„Reflexion im Vollzug“[1]



ex post
Evaluation

„Reflexion des Vollzugs“[1]

Q: 1. How were the data obtained – what is the **data provenance**?
2. What do the data say – what **knowledge claims** do we base on the data?
3. Why should we accept them – what is their **epistemic grounding**?

[1]K. Tulatz, *Epistemologie als Reflexion wissenschaftlicher Praxen*, **2018**.

Epistemically FAIR materials modelling

Norwegian University of Life Sciences

Priorities (**DORIC principles**) following doi:10.5281/zenodo.4571052

D	O	R	I	C
diversify technology	**observe** practices	have **realistic** objectives	**incentivize** open data	**co-design** data and workflows

The aim of the present work is to permit **communicating epistemic metadata** by developing a semantic artefact that fits into the pre-existing environment.

Cognitive processes are a very broad category[1] by which semantics about research practices and workflows can be formalized with a **mid-level ontology**.

[1]See for example a recent review by Elkobaisi *et al.* on ontologization of human emotional responses, *SN Computer Science* 3, 282, doi:10.1007/s42979-022-01116-x, **2022**.

Noregs miljø- og biovitskaplege universitet

European platforms in materials modelling

We are at the transition from H2020 to Horizon Europe.

What overall situation emerges as we approach the end point of H2020 NMBP work on materials digitalization?

VIMMP
VIRTUAL MATERIALS MARKETPLACE
H2020 GA no. 760907

VIPCOAT
H2020 GA no. 952903

DOME 4.0
H2020 GA no. 953163

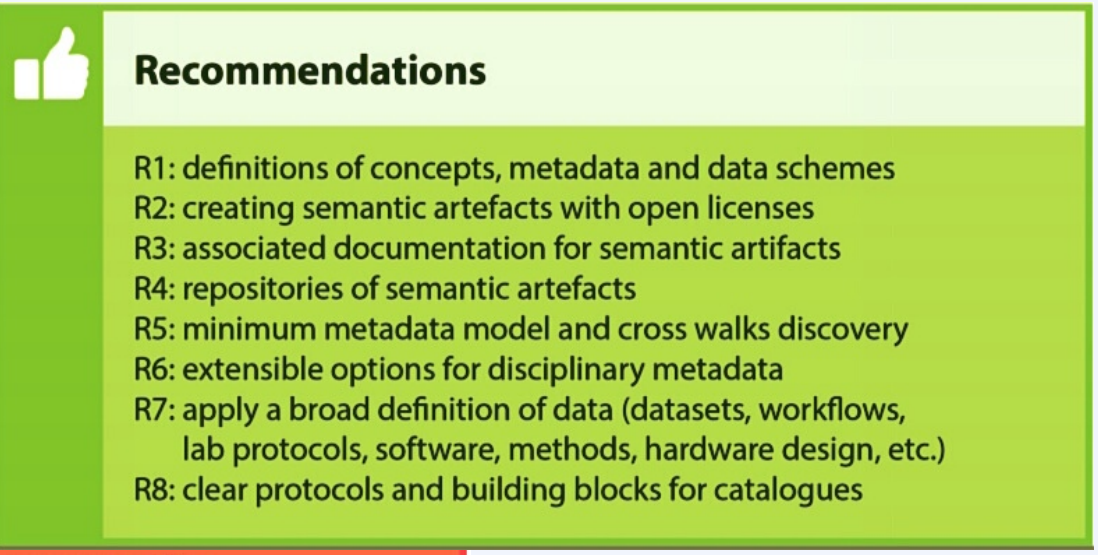
ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS
H2020 GA no. 958371

eSSENCE-EMMC Meeting

1st June 2022

Uppsala, Ångström Lab

EOSC Interoperability Framework[1]

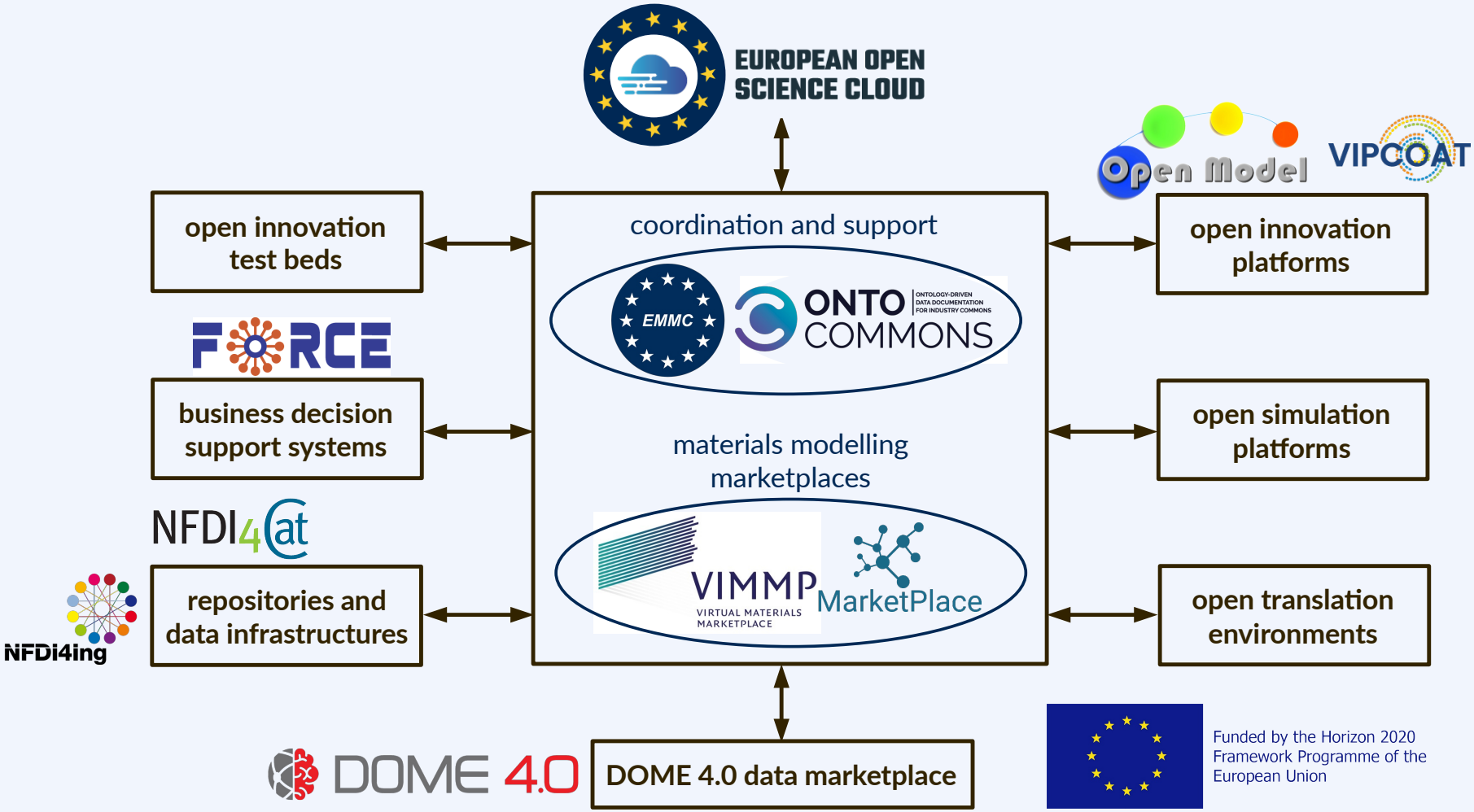


EUROPEAN OPEN SCIENCE CLOUD

[1]O. Corcho *et al.*, EOSC Interoperability Framework, doi:10.2777/620649, **2021**.

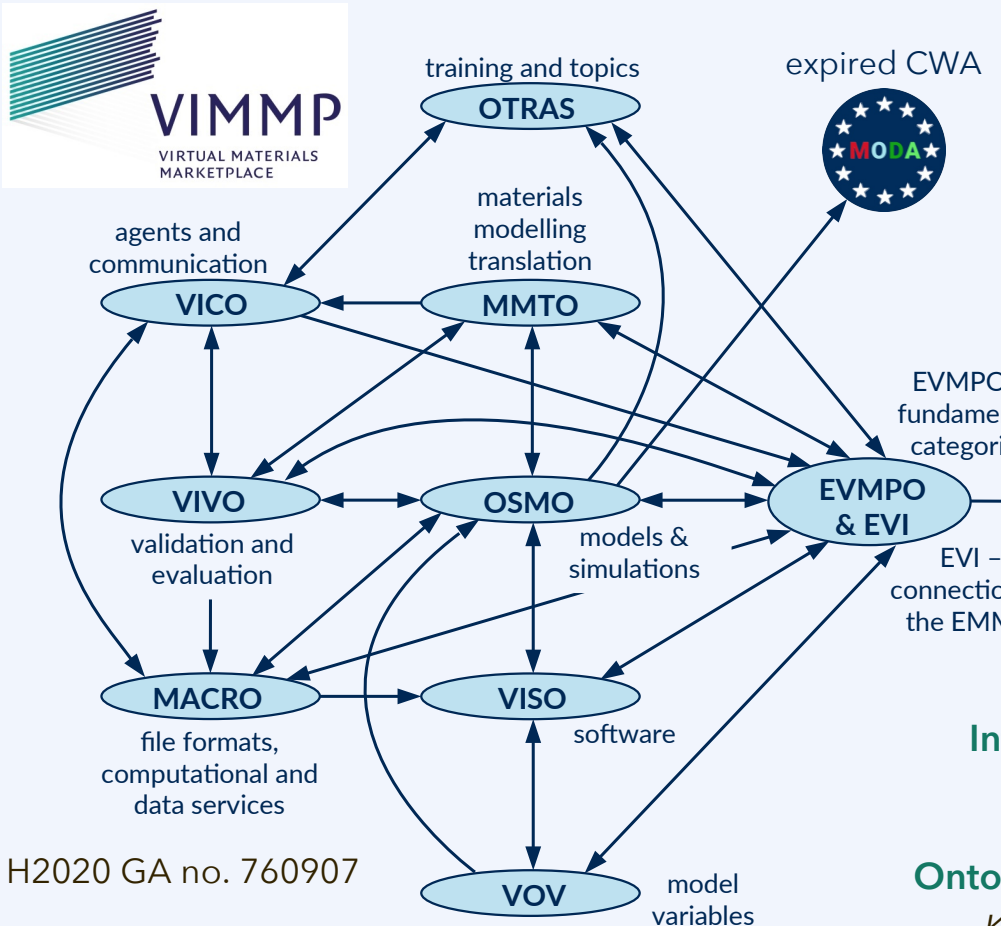
Norwegian University of Life Sciences

Ecosystem of digital platforms



Funded by the Horizon 2020 Framework Programme of the European Union

VIMMP system of ontologies



H2020 GA no. 760907

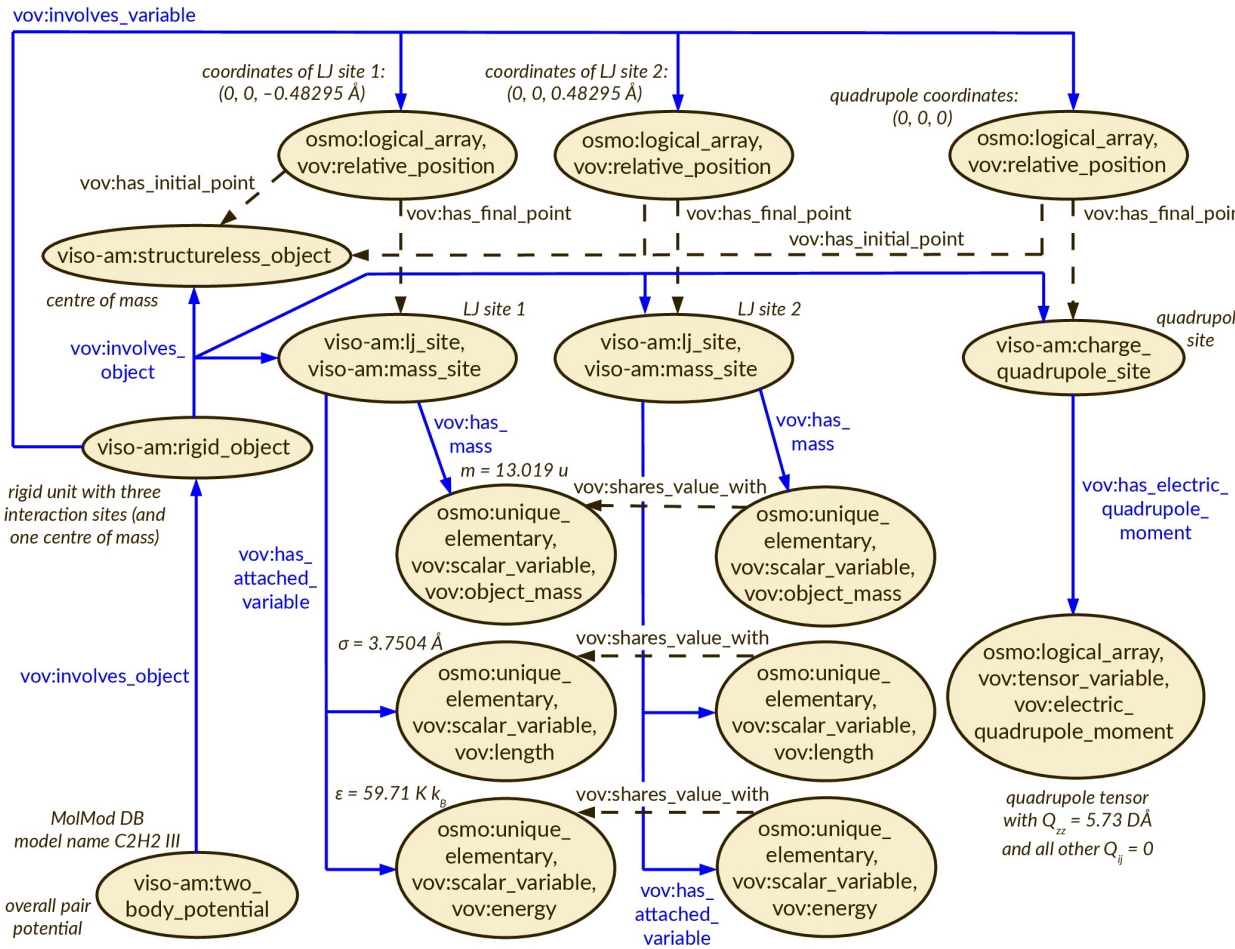
The **final EMMO release** had been announced for mid-2018. Four years (!) later, there are only beta versions. There are major changes between EMMO drafts.

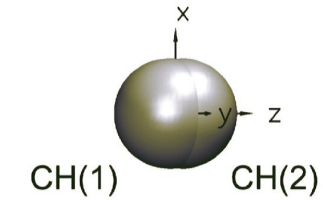
The Horizon Europe CL4 work programme continues to mention EMMO compliance.

Introduction and VIMMP ontology release
doi:10.5281/zenodo.3936795, **2021**

Ontologies for the Virtual Materials Marketplace
KI – Künstliche Intelligenz 34(3), 423–428,
doi:10.1007/s13218-020-00648-9, **2020**

Molecular modelling knowledge graph





MolMod DB (Molecular Model Database)

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

pair potentials for over 150 molecular fluids

[1]S. Stephan, M. Horsch, *et al.*, *Mol. Sim.* 45, 806–814, **2019**. [2]M. Horsch, S. Chiacchiera, *et al.*, *Proc. ISWC*, **2020**.

Noregs miljø- og biovitskaplege universitet

Documentation of research workflows

The novel platforms in materials digitalization all use ontology-based semantic technology.

How about epistemic metadata? What progress has been made in documenting data provenance?

VIMMP
VIRTUAL MATERIALS MARKETPLACE
H2020 GA no. 760907

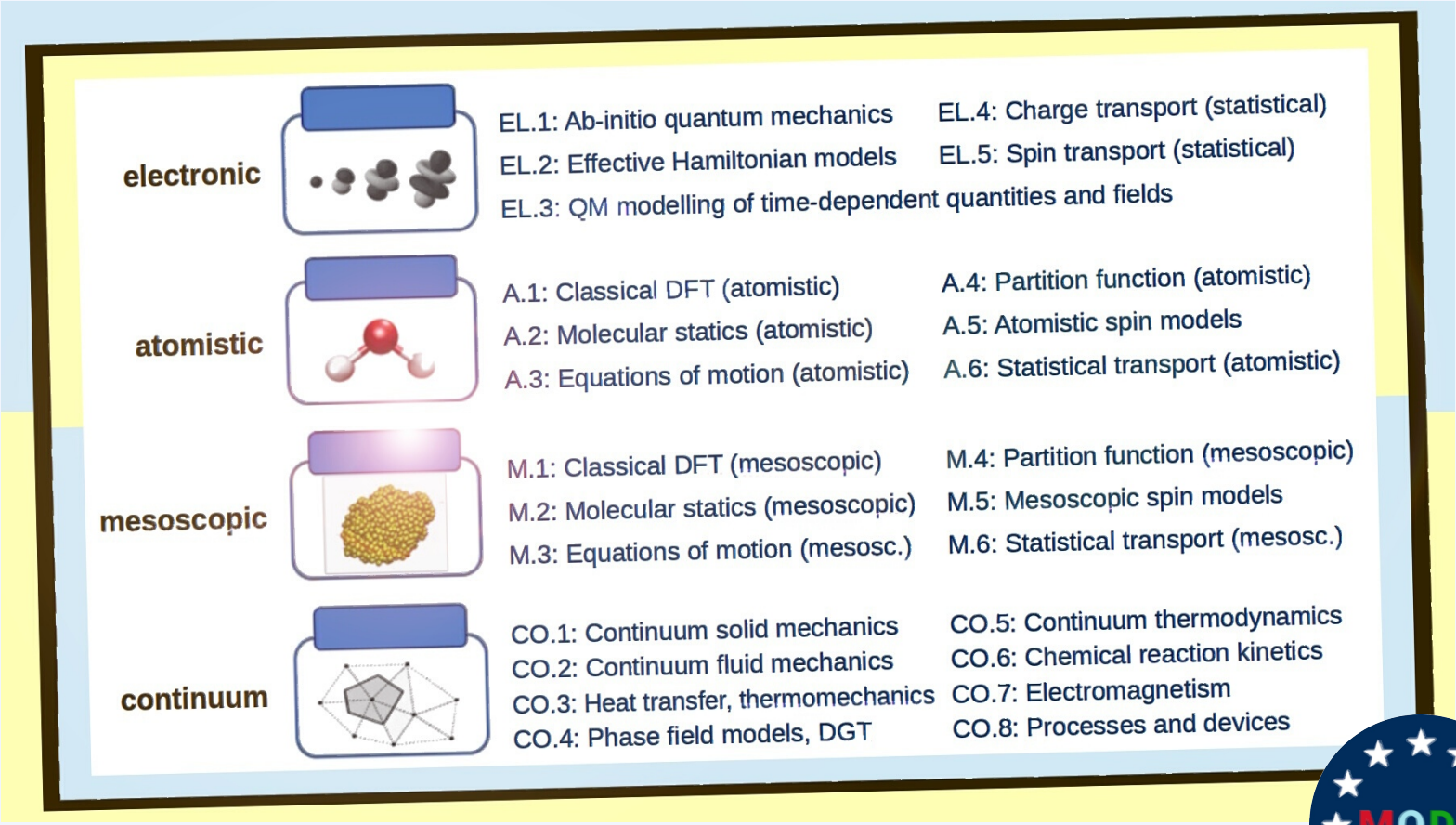
VIPCOAT
H2020 GA no. 952903

DOME 4.0
H2020 GA no. 953163

ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS
H2020 GA no. 958371

eSSENCE-EMMC Meeting | 1st June 2022 | Uppsala, Ångström Lab

RoMM/MODA: Closed epistemic space

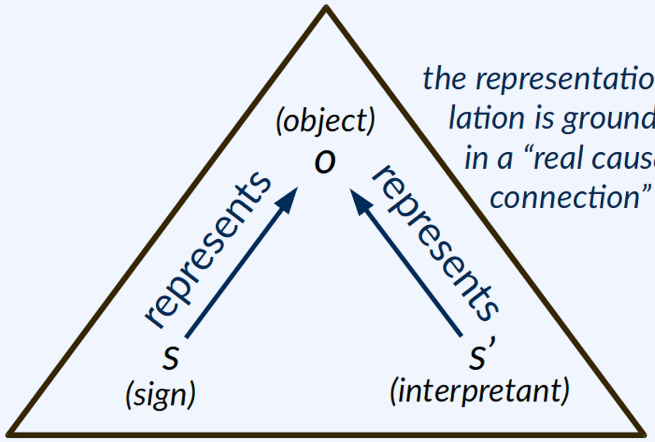


[1] *Journal of Chemical & Engineering Data* 65, 1313, doi:10.1021/acs.jced.9b00739, **2020**.
[2] A. F. de Baas (ed.), *What Makes a Material Function?*, ISBN 978-92-79-63185-6, **2017**.

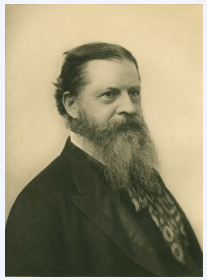
Mereotopology and Peircean semiotics

Norwegian University of Life Sciences

Peircean semiotics



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



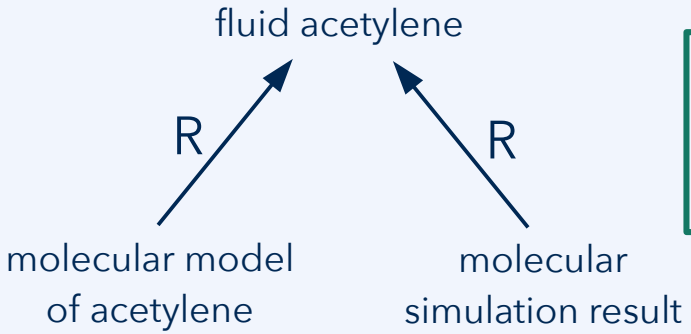
C. S. Peirce

Elementary Multiperspective Material Ontology[1, 2]

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

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

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



"represents" or **"is sign for"** is here abbreviated by R

[1]H. A. Preisig *et al.*, doi:10.23967/wccm-eccomas.2020.262, no. 262 in *Proc. ECCOMAS 2020*, **2021**.

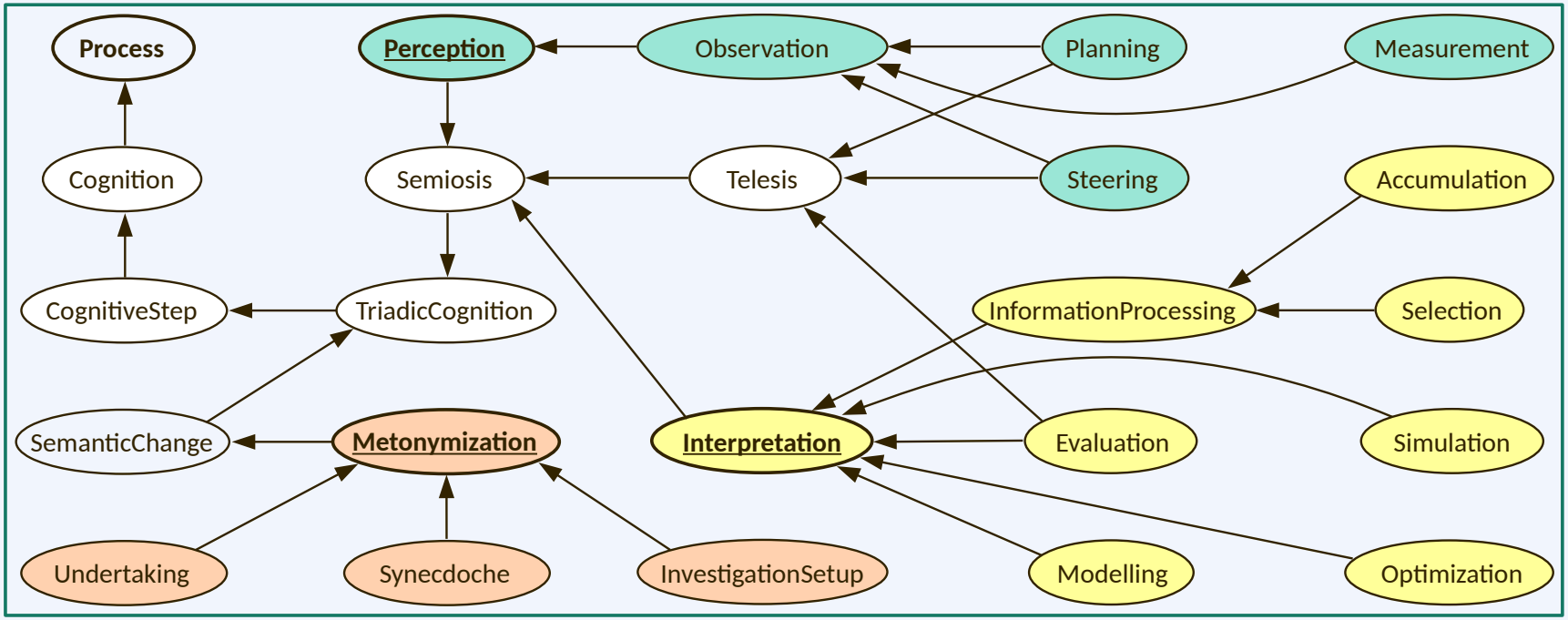
[2]S. Clark *et al.*, *Adv. Energ. Mat.* 12(17), 2102702, doi:10.1002/aenm.202102702, **2022**.

Mid-level ontology for mereosemiotics

Norwegian University of Life Sciences

PIMS-II is a mid-level ontology for scientific workflows as cognitive processes.[1, 2]

perception requires participation
(and overlap) of the perceived object



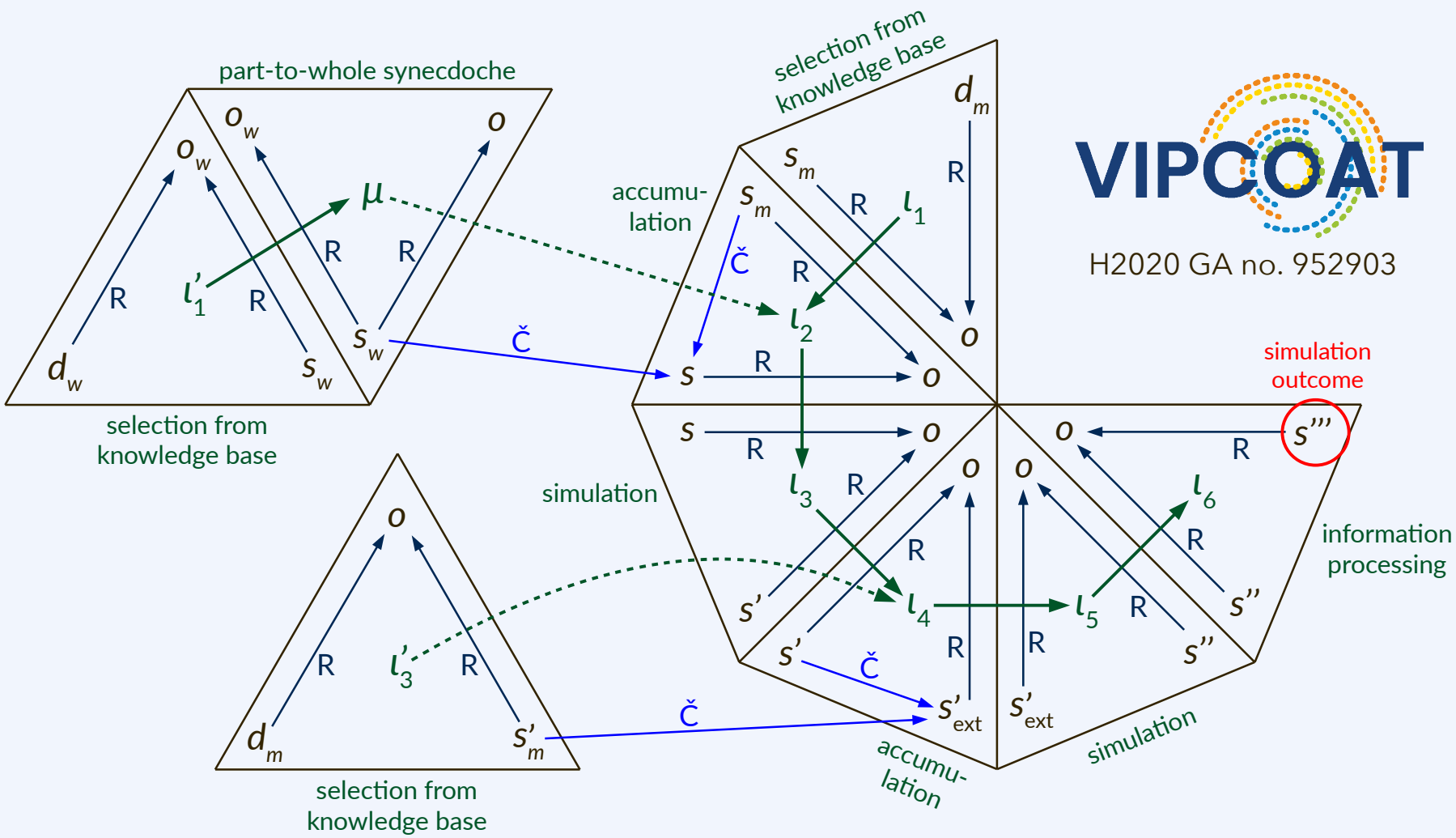
metonymization preserves the „real causal connection" (Peirce) between the sign and its old & new referents

interpretation and **metonymization** do not entail physical participation of the referents

[1]M. T. Horsch, no. 3 in *Proc. JOWO 2021*, **2021**.

[2]P. Klein *et al.*, no. 26 in *Proc. JOWO 2021*, **2021**.

Workflows as cognitive processes[1]



[1]P. Klein *et al.*, no. 26 in *Proc. JOWO 2021*, **2021**.

Noregs miljø- og biovitskaplege universitet

Now, what are the main challenges and gaps?

In research data provenance and workflow documentation, a new standardization effort is needed. The MODA CWA has expired.

We need a shift toward making the knowledge claims machine-actionable, not mainly workflows.

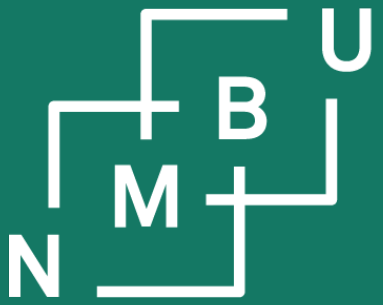
There, standardization must permit both semantic and epistemic heterogeneity.

VIMMP
VIRTUAL MATERIALS MARKETPLACE
H2020 GA no. 760907

VIPCOAT
H2020 GA no. 952903

DOME 4.0
H2020 GA no. 953163

ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS
H2020 GA no. 958371



Norges miljø- og biovitenskapelige universitet

Standardized documentation of workflows and epistemic metadata for digitalization and interoperability in materials modelling

M. T. Horsch,[1,2] S. Chiacchiera,[2] F. Diewald,[3] P. Klein,[3] N. A. Konchakova,[4] P. Neumann,[5] H. A. Preisig,[6] B. Schembera,[7] S. Stephan,[8] I. T. Todorov,[2] and J. Vrabec[9]

[1]Norweg. Univ. Life Sciences, [2]UKRI STFC Daresbury Laboratory, [3]Fraunhofer ITWM, [4]Helmholtz-Zentrum Hereon, [5]Helmut Schmidt Univ. Hamburg, [6]Norweg. Univ. Science & Technology, [7]Univ. Stuttgart, [8]TU Kaiserslautern, [9]TU Berlin

eSSENCE-EMMC Meeting 2022 1st June 2022 Uppsala, Ångströmlaboratoriet