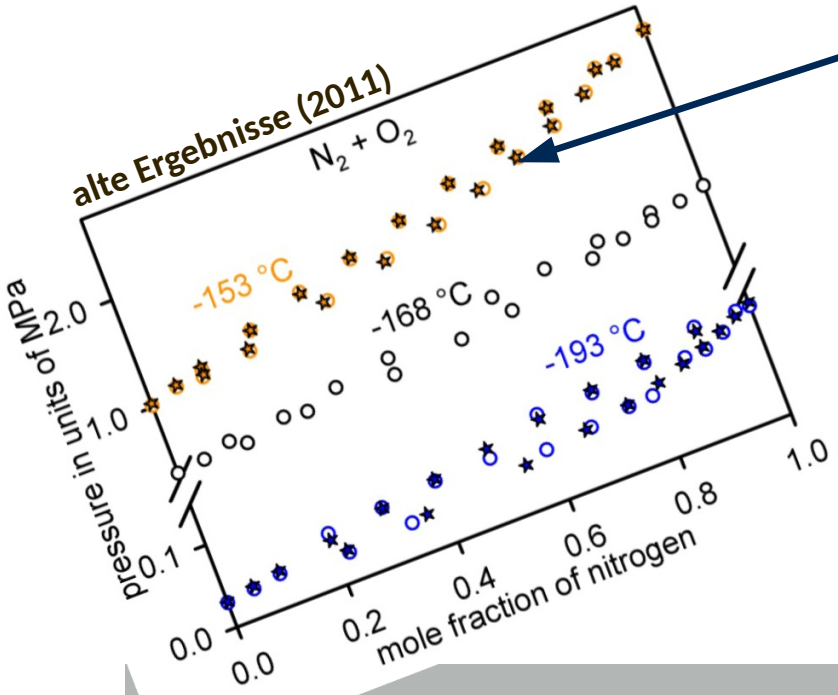
Digitalisierung in der Molekularen Thermodynamik

Martin Thomas Horsch

Motivation und Überblick

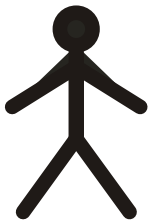


Welche Werte hatten *x* und *p* hier genau?

Wie wurde der Datenpunkt ermittelt?

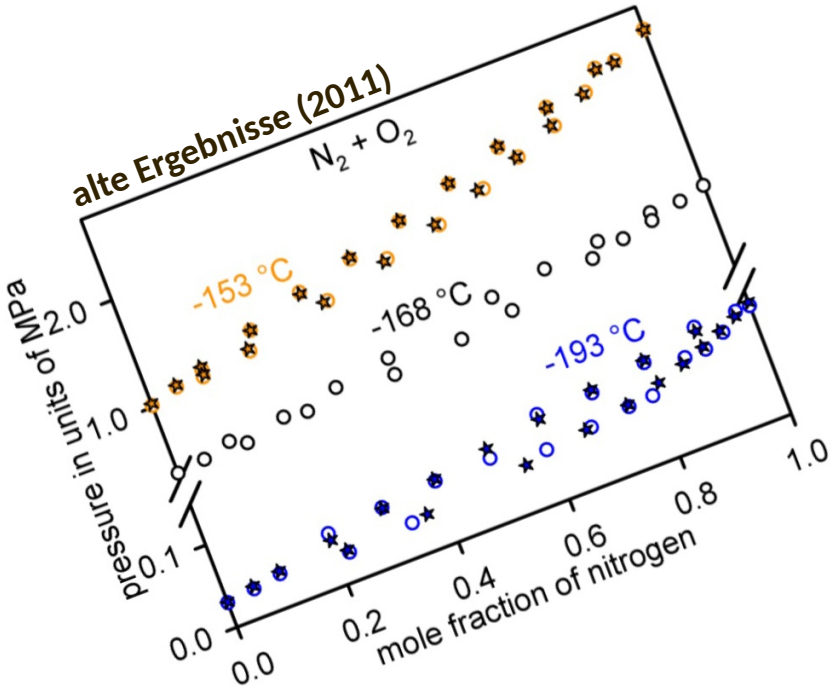
Wie groß ist der Fehler, wie war der Fehler genau definiert, und welche Software(version) wurde eingesetzt?

Nachfrage an damaligen Mitarbeiter



"Ich erinnere mich."

Motivation und Überblick

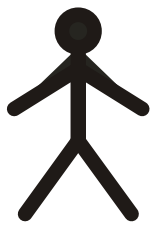


Welche Werte hatten *x* und *p* hier genau?

Wie wurde der Datenpunkt ermittelt?

Wie groß ist der Fehler, wie war der Fehler genau definiert, und welche Software(version) wurde eingesetzt?

Kompetenzfragen



"Ich erinnere mich.
Der Ordner ist wahrscheinlich in einer Kiste bei mir im Keller."

Gute Praxis im Umgang mit Forschungsdaten:

Diese sollen **auffindbar**, **zugänglich**, gemeinsam nutzbar **(interoperabel)** und dauerhaft **nachnutzbar** bleiben.

Motivation und Überblick

FAIR-Prinzipien des Datenmanagements

Gute Praxis im Umgang mit Forschungsdaten

Diese sollen **auffindbar**, **zugänglich**, gemeinsam nutzbar **(interoperabel)** und dauerhaft **nachnutzbar** bleiben.

F: Findability
A: Accessibility
I: Interoperability
R: Reusability

Motivation und Überblick

FAIR-Prinzipien des Datenmanagements

Gute Praxis im Umgang mit Forschungsdaten

Warum jetzt auf einmal?

Warum erst jetzt und nicht schon 1990?

F: Findability
A: Accessibility
I: Interoperability
R: Reusability

Organisation und Projekte

(1) Motivation und Überblick
(2) **Organisation und Projekte**
(3) Semantische Technologie
(4) Datenprovenienz
(5) Wissensrepräsentation
(6) Molekulare Modelle
(7) Workflowentwicklung
(8) Perspektive

Nationale Forschungsdateninfrastruktur (NFDI)

nfdi4chem
NFDI4Cat
NFDi4ing

Organisation und Projekte

Science and Technology Facilities Council

VIMMP
VIRTUAL MATERIALS MARKETPLACE

http://vimmp.eu/

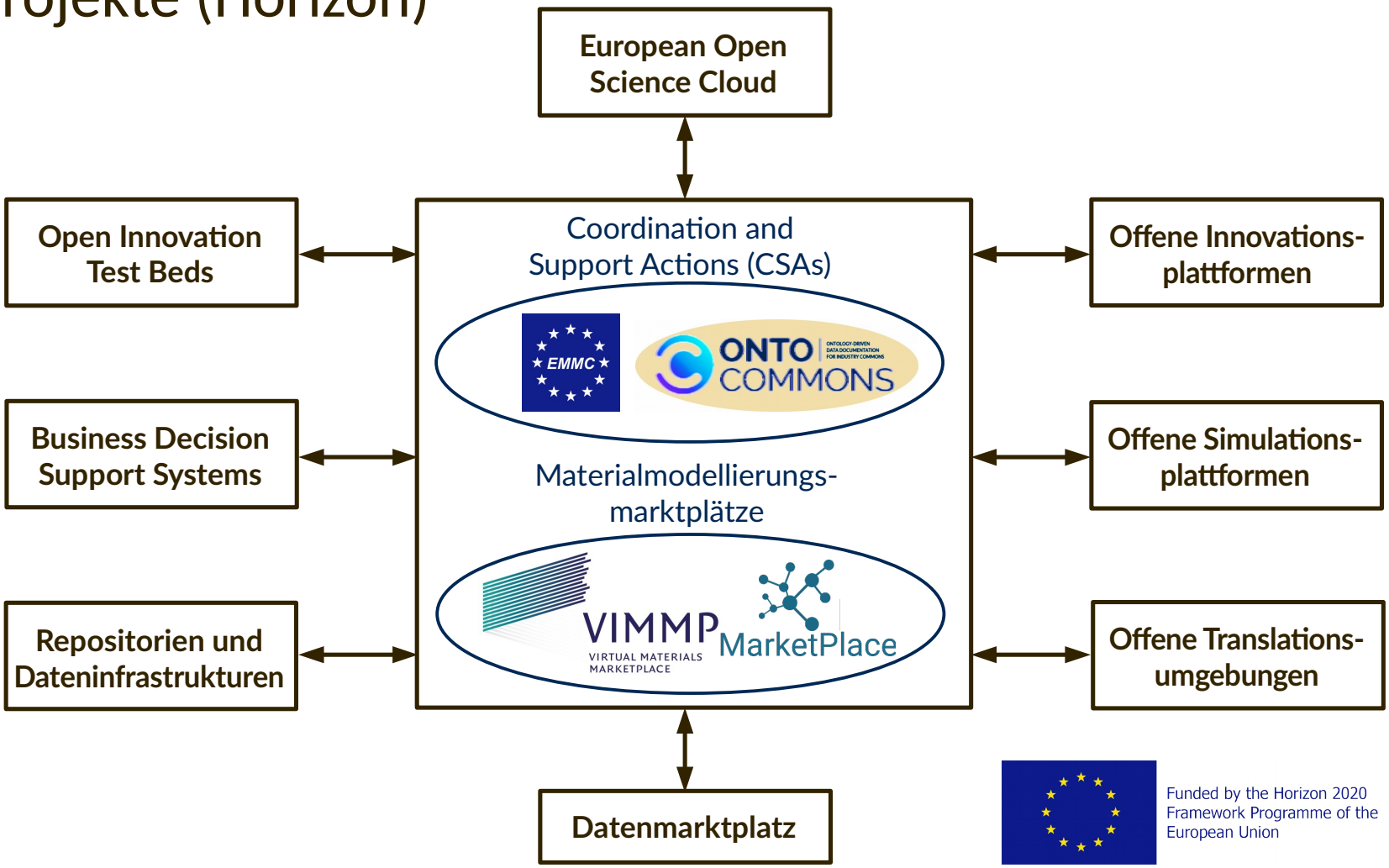
- Horizon 2020 project
  - Innovation action, grant agreement *no.* 760907
  - H2020 (NMBP-25-2017)
  - 4 years project – started on 1st January 2018

Objective: To support accelerating innovation in manufacturing industries by using electronic, atomistic, mesoscopic, and continuum materials modelling.

engineering challenges → **modelling & simulation** → **actionable decisions**

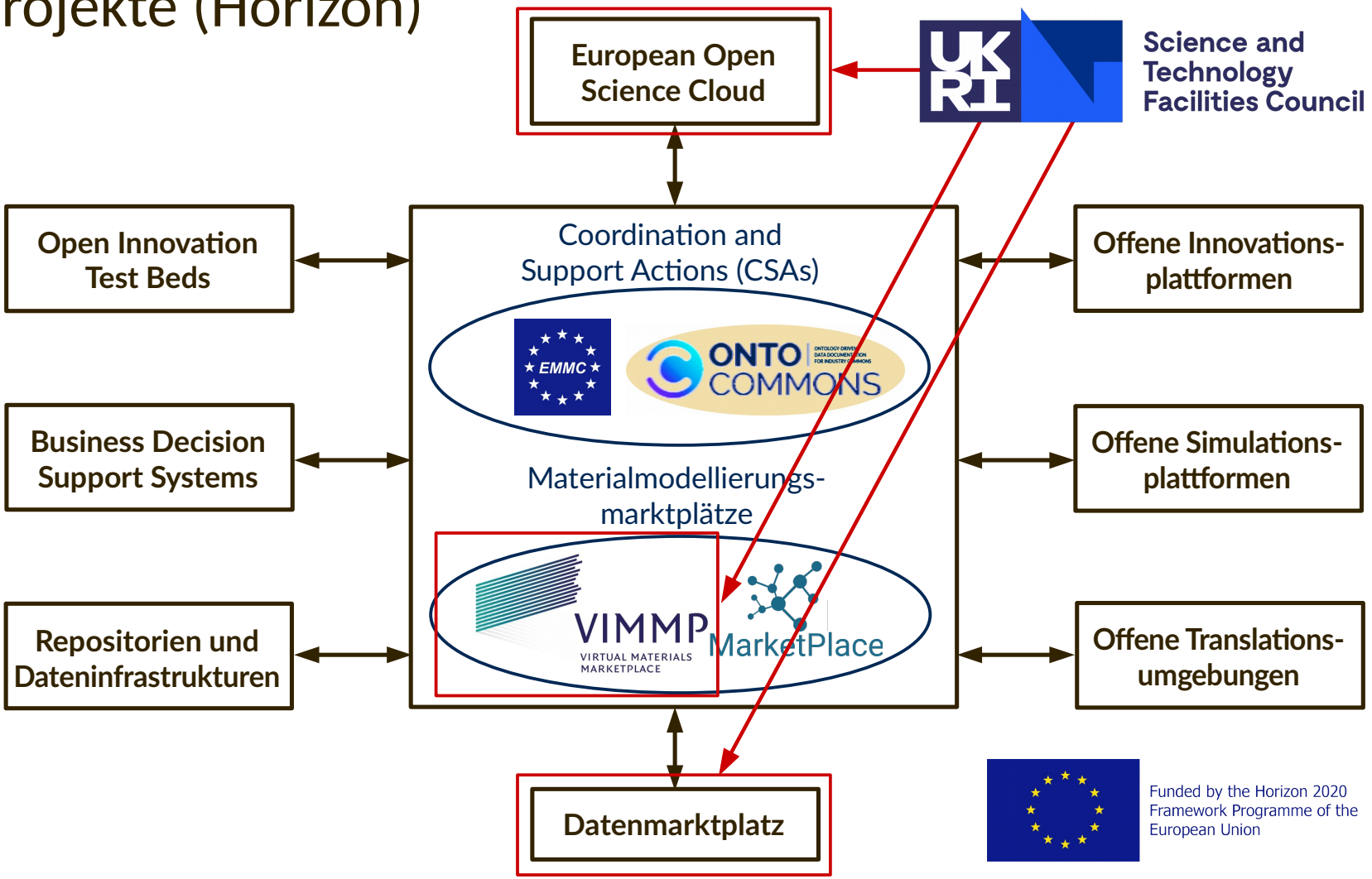
materials modelling "translation"

Projekte (Horizon)



Funded by the Horizon 2020 Framework Programme of the European Union

Projekte (Horizon)



Organisation und Projekte

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.

Organisation und Projekte

RESEARCH DATA ALLIANCE

https://www.rd-alliance.org/

Innovationszentrum für Prozessdatentechnik (Inprodat e.V.)

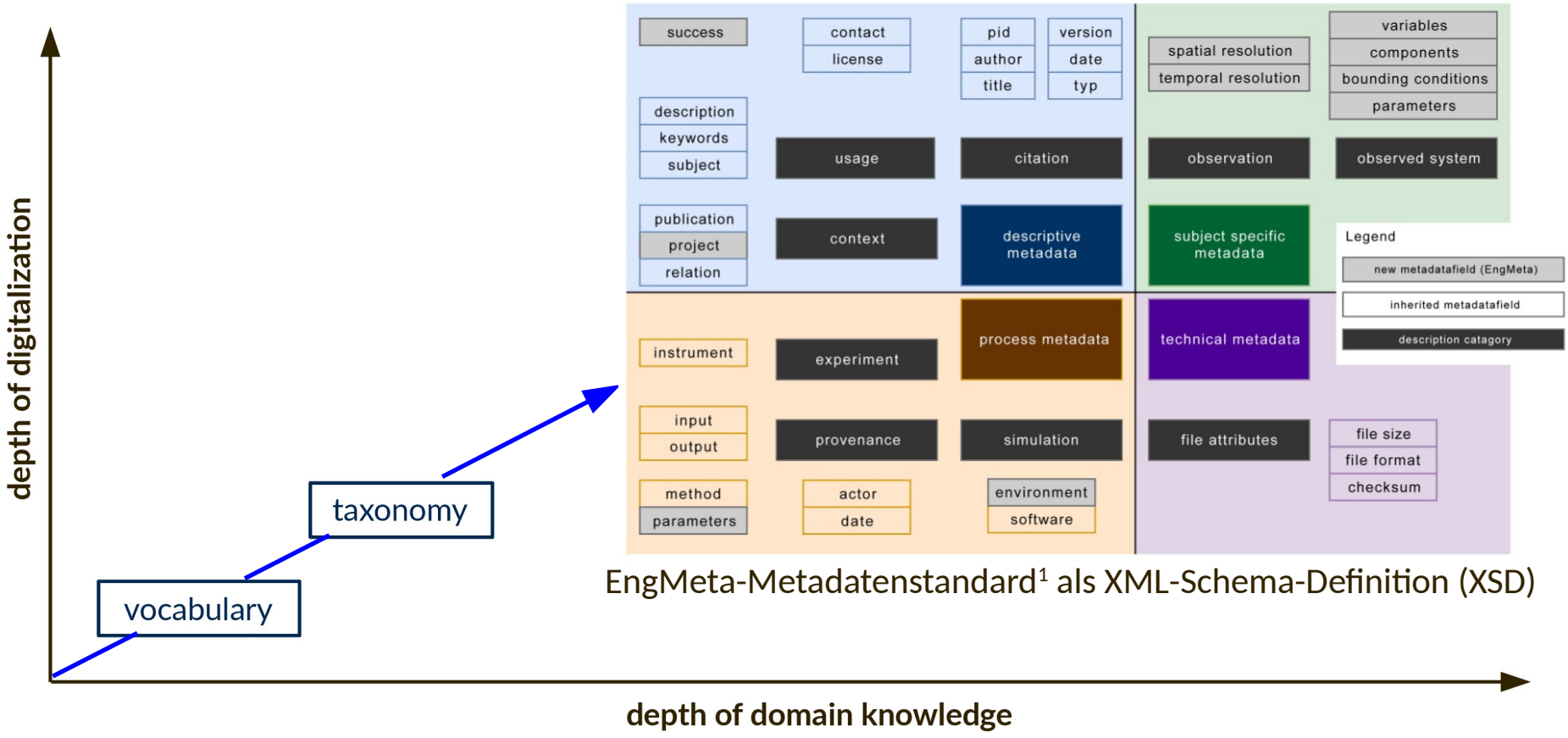
INPRODAT

Prozessdatentechnik ≡ Datentechnik ⊓ Verfahrenstechnik

http://www.inprodat.de/

https://zenodo.org/communities/inprodat/

Semantische Technologie

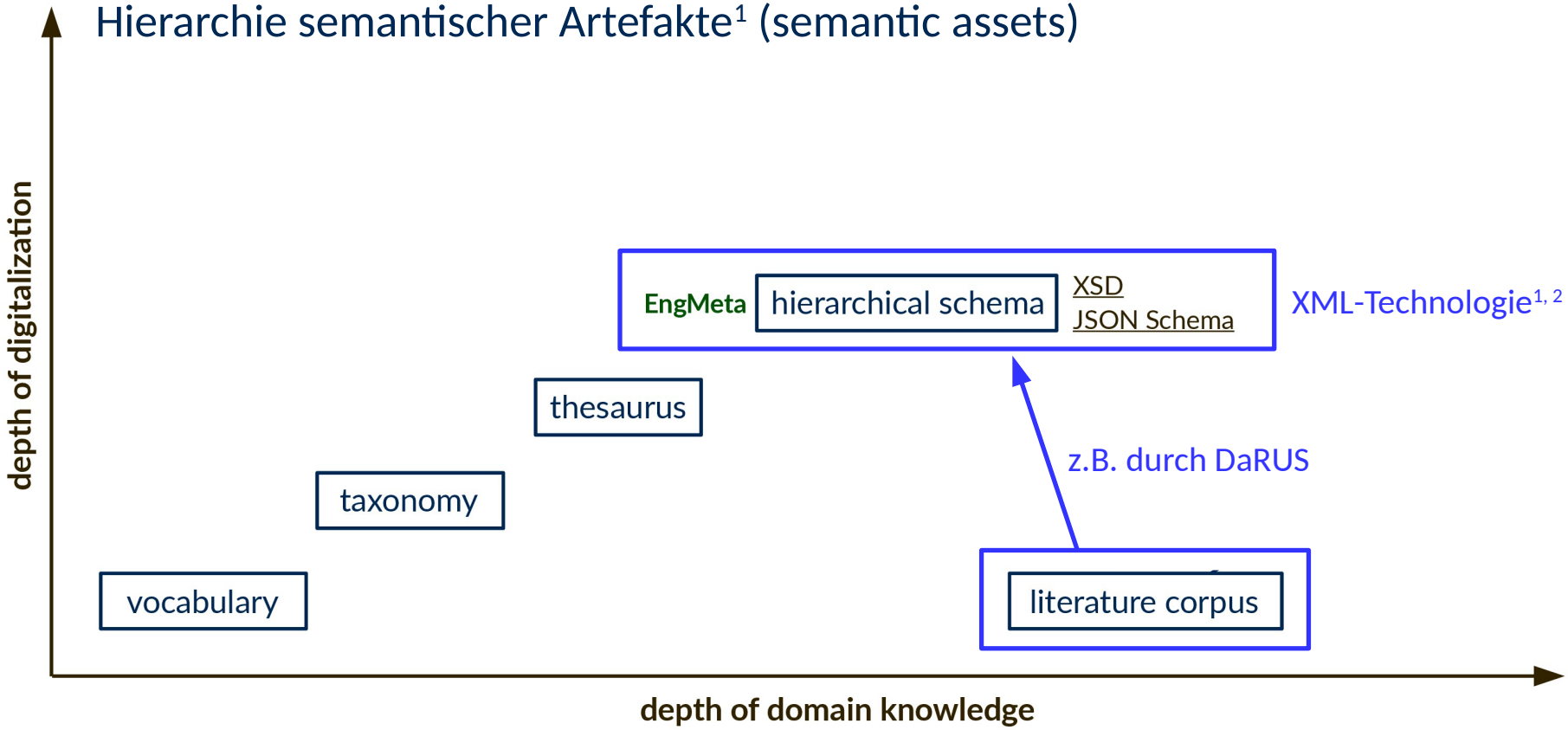


EngMeta-Metadatenstandard[1] als XML-Schema-Definition (XSD)

[1]B. Selent, H. Kraus, N. Hansen, B. Schembera, A. Seeland, D. Iglezakis, doi:10.11588/heibooks.598.c8422, **2020**.

Semantische Technologie

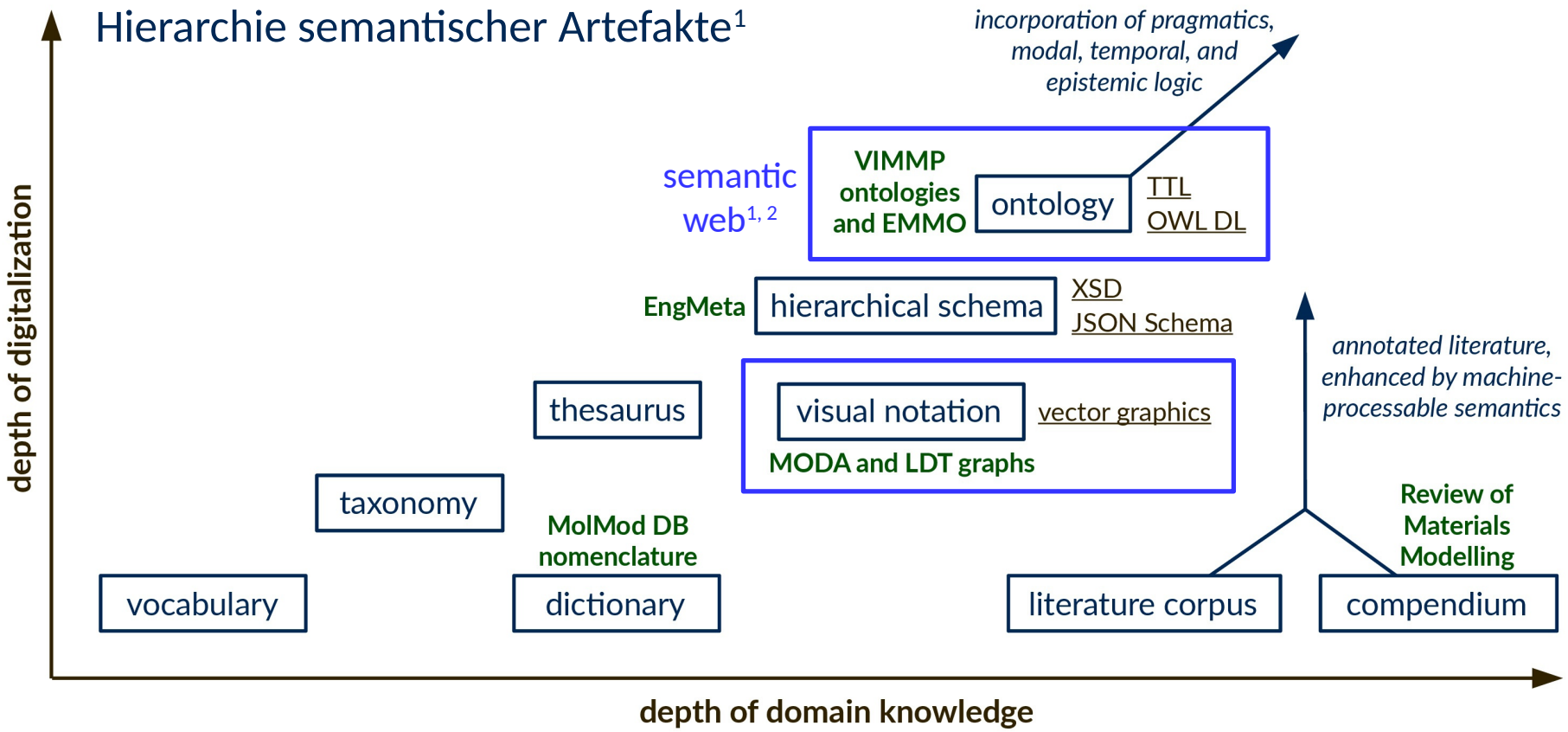
Hierarchie semantischer Artefakte[1] (semantic assets)



[1]M. T. Horsch, S. Chiacchiera, W. L. Cavalcanti, B. Schembera, *Data Technology in Materials Modelling*, Springer, **2021**.

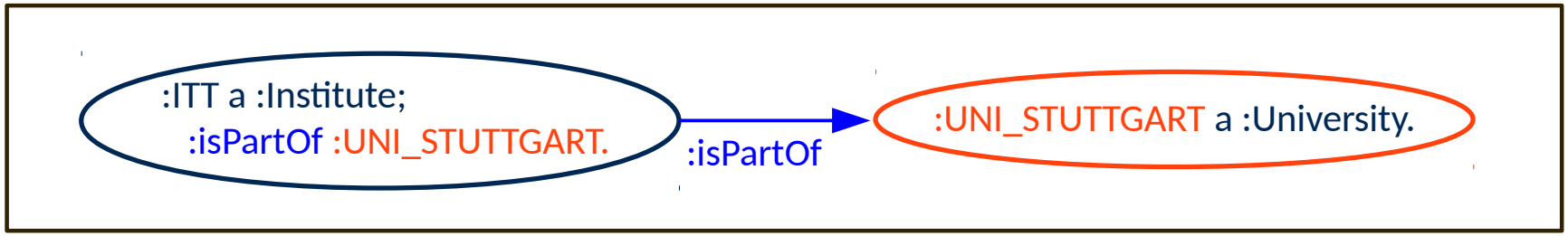
[2]B. Selent, H. Kraus, N. Hansen, B. Schembera, A. Seeland, D. Iglezakis, doi:10.11588/heibooks.598.c8422, **2020**.

Semantische Technologie



[1]M. T. Horsch, S. Chiacchiera, W. L. Cavalcanti, B. Schembera, *Data Technology in Materials Modelling*, Springer, **2021**.
[2]E. Ghedini, J. Friis, A. Hashibon, G. J. Schmitz, G. Goldbeck, *et al.*, **2021**; http://emmc.info/emmo-info/.

Semantische Technologie

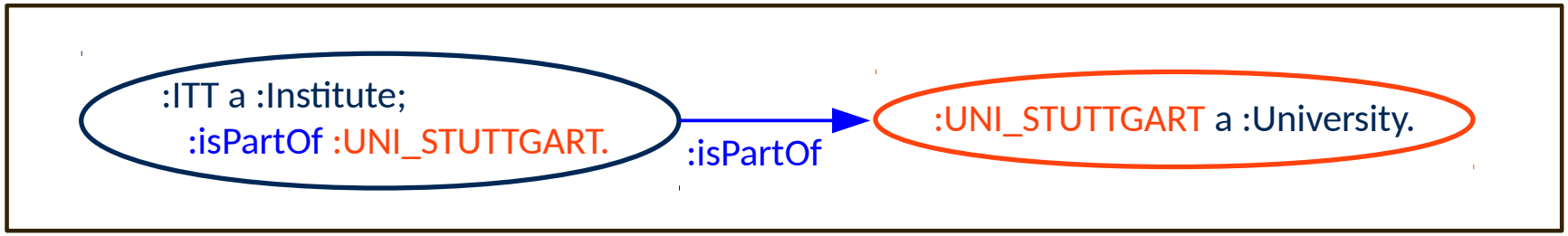


Wissensgraph; auch: Szenario oder assertional box (ABox)

:ITT :isPartOf :UNI_STUTTGART.

RDF-Tripel, bestehend aus Subjekt, Prädikat und Objekt

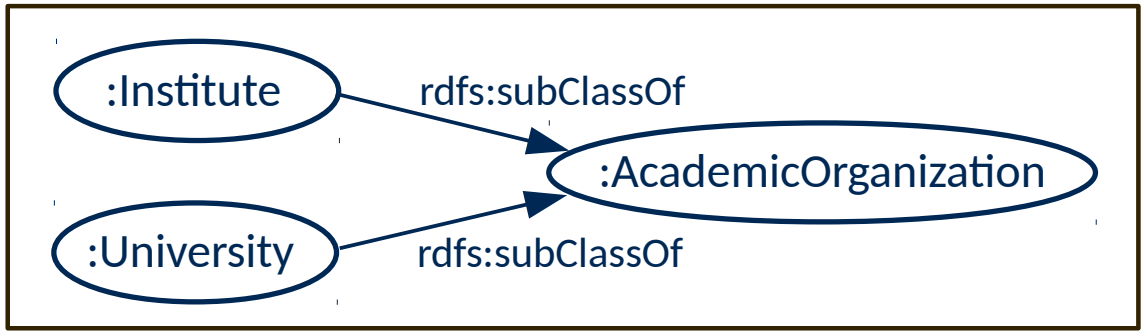
Semantische Technologie



Wissensgraph; auch: Szenario oder assertional box (ABox)

:ITT :isPartOf :UNI_STUTTGART.

RDF-Tripel, bestehend aus Subjekt, Prädikat und Objekt



Taxonomie als Teil einer Ontologie

auch: TBox (terminological box)

Provenienz von Simulationsdaten

Entwicklungslinie semantischer Standards des EMMC

RoMM VI → **MODA** → **CWA 17284** → **Ontologien**

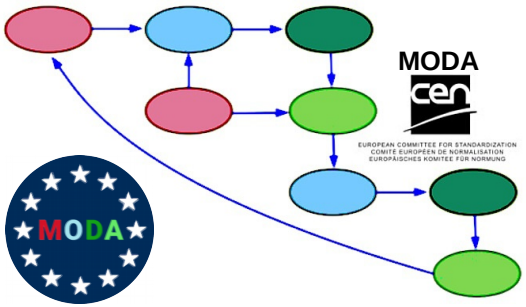
Terminologie bzw. Vokabular in Buchform

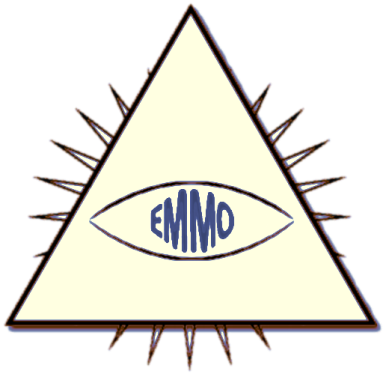
MODA-Simulationsworkflows (semiformal)

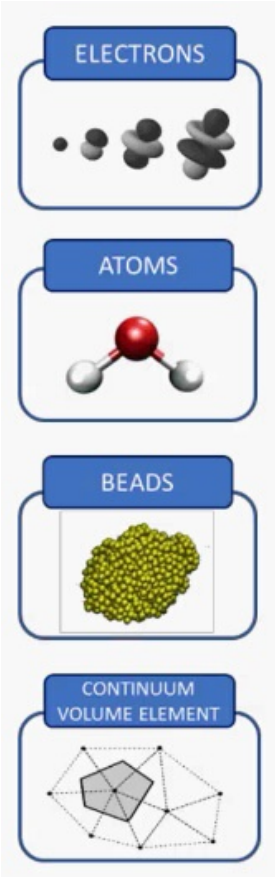
CEN workshop agreement

EMMO: Top-Level-Ontologie

Domänenontologien







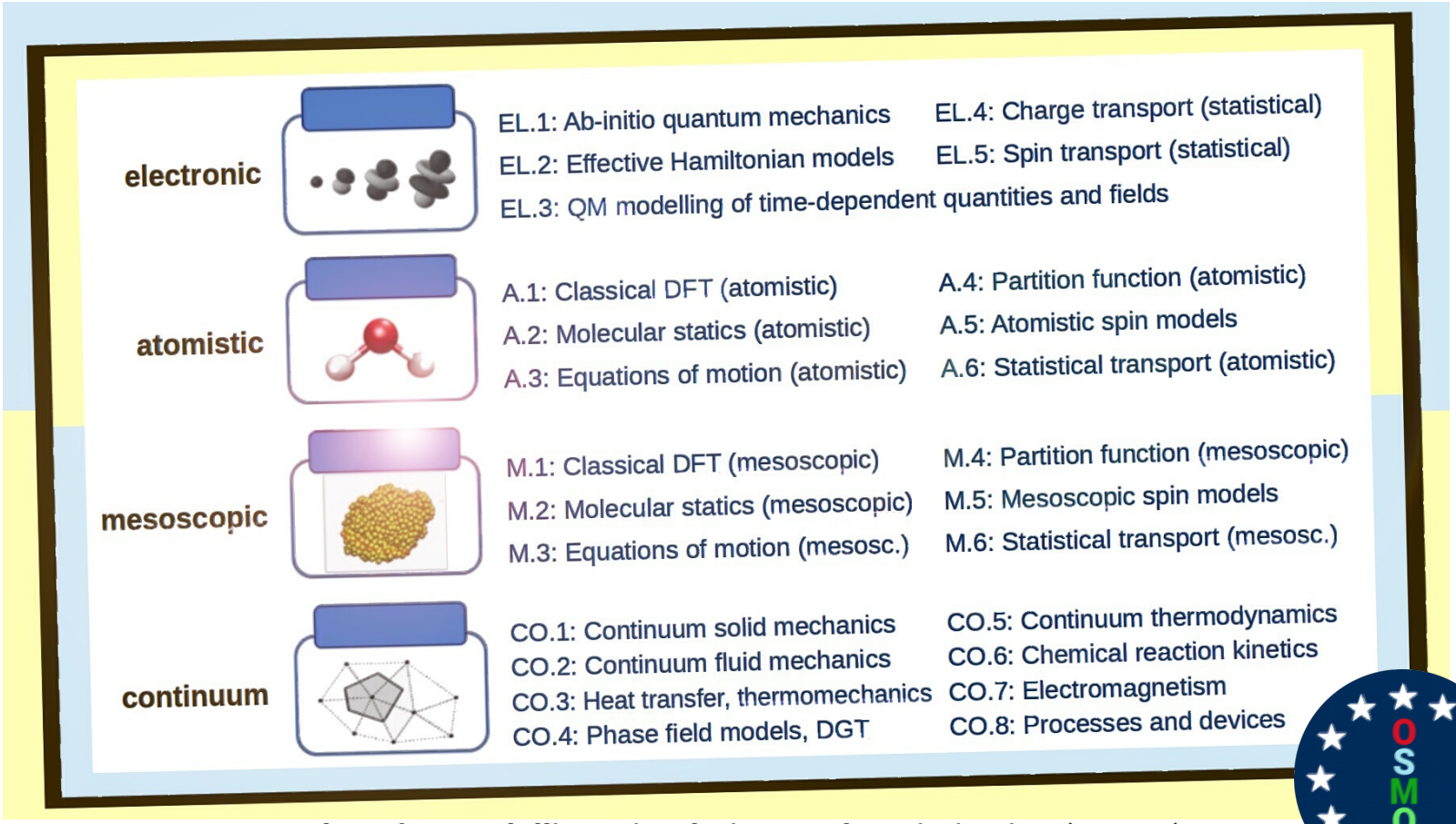
electronic

atomistic

mesoscopic

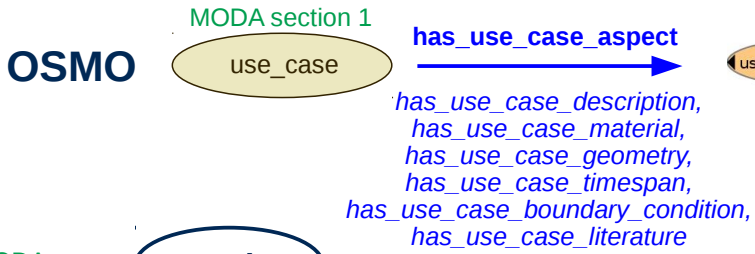
continuum

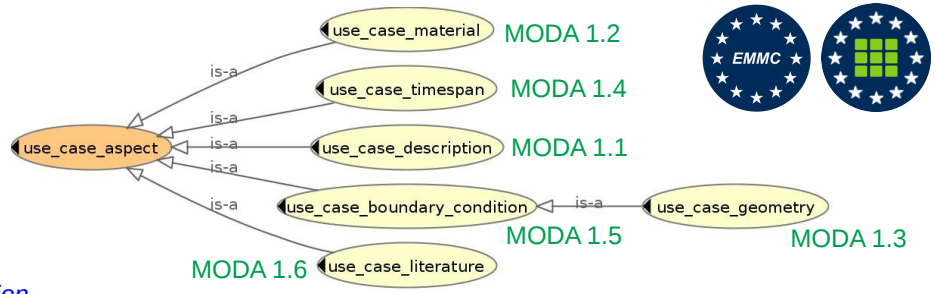
Provenienz von Simulationsdaten

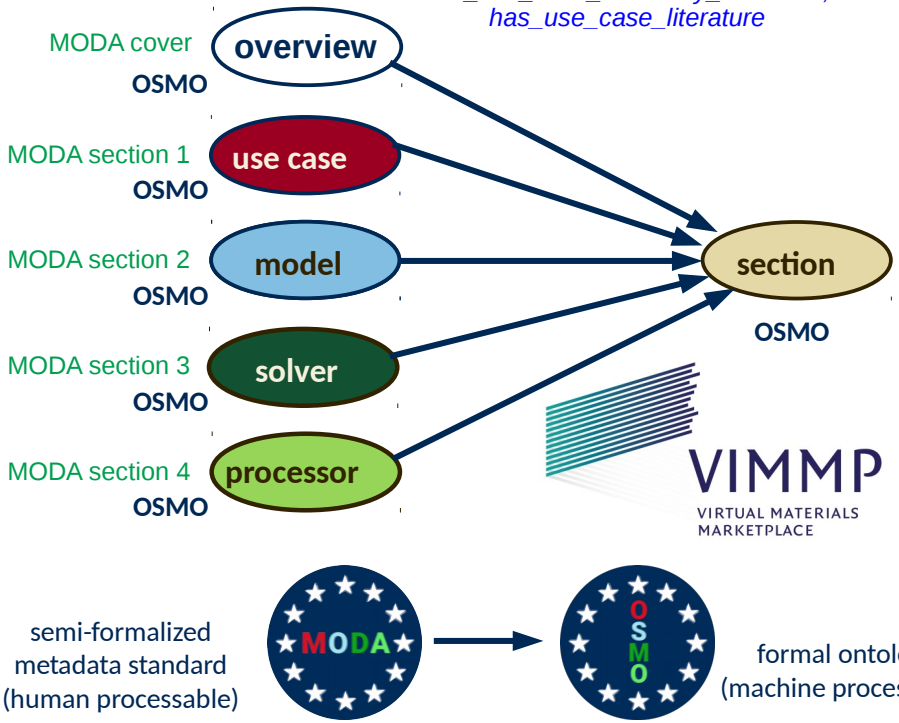


Ontology for Modelling, Simulation, and Optimization (OSMO)
J. Chem. Eng. Data 65(3), 1313–1329, **2020**

Provenienz von Simulationsdaten

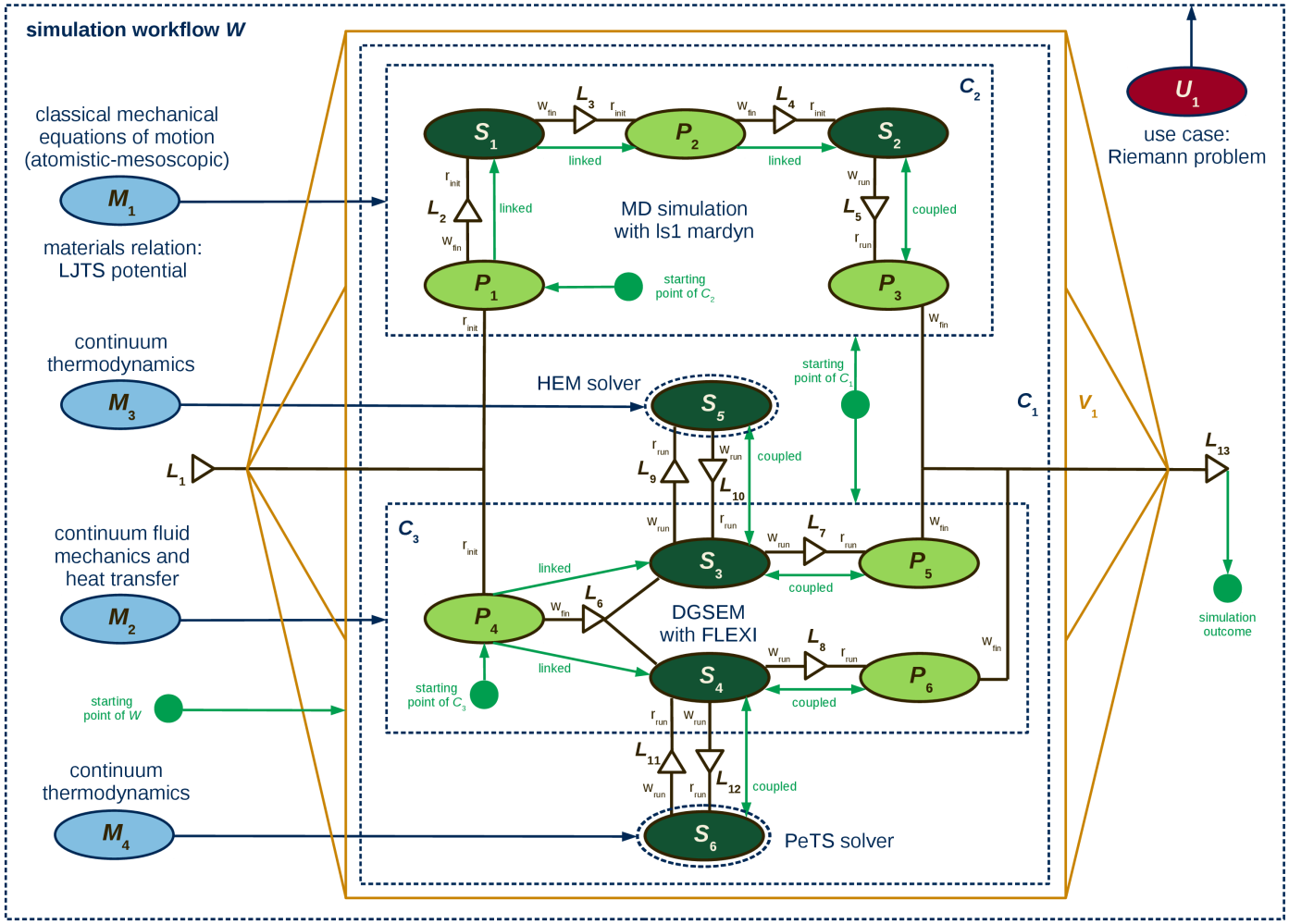






3	Solver and Computational translation of the specifications		
3.1	Numerical Solver		*Please give name andtype of the solver e.g. Monte Carlo, SPH, FE, ...iterative, multi-grid, adaptive,...*
3.2	Software tool		*Please give the name and if this is your own code, please specify if it can be shared with 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 pl... 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*

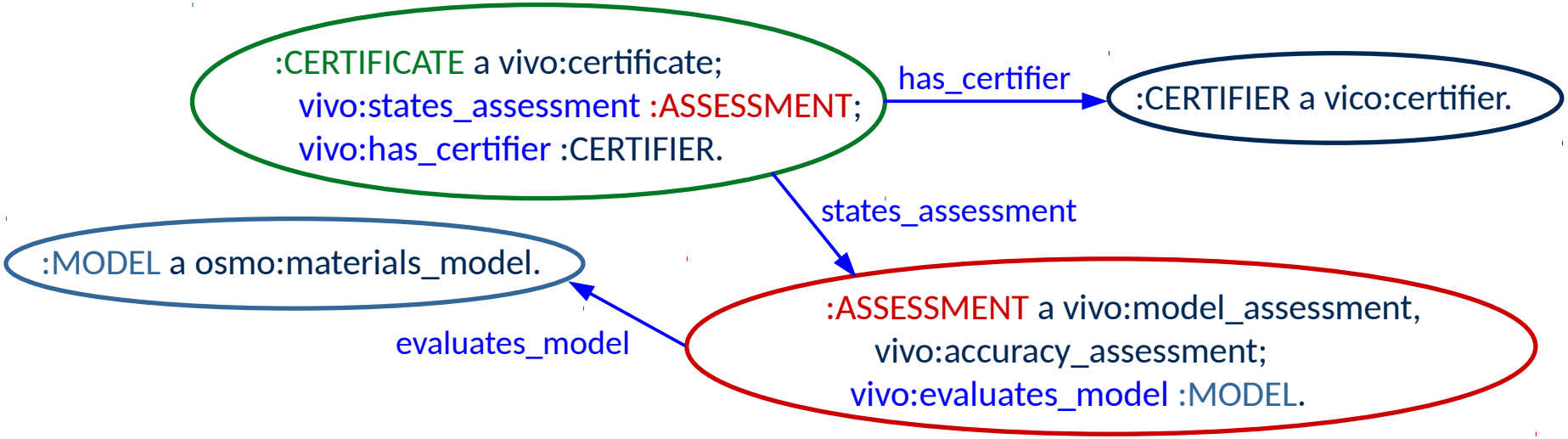
Provenienz von Simulationsdaten



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.

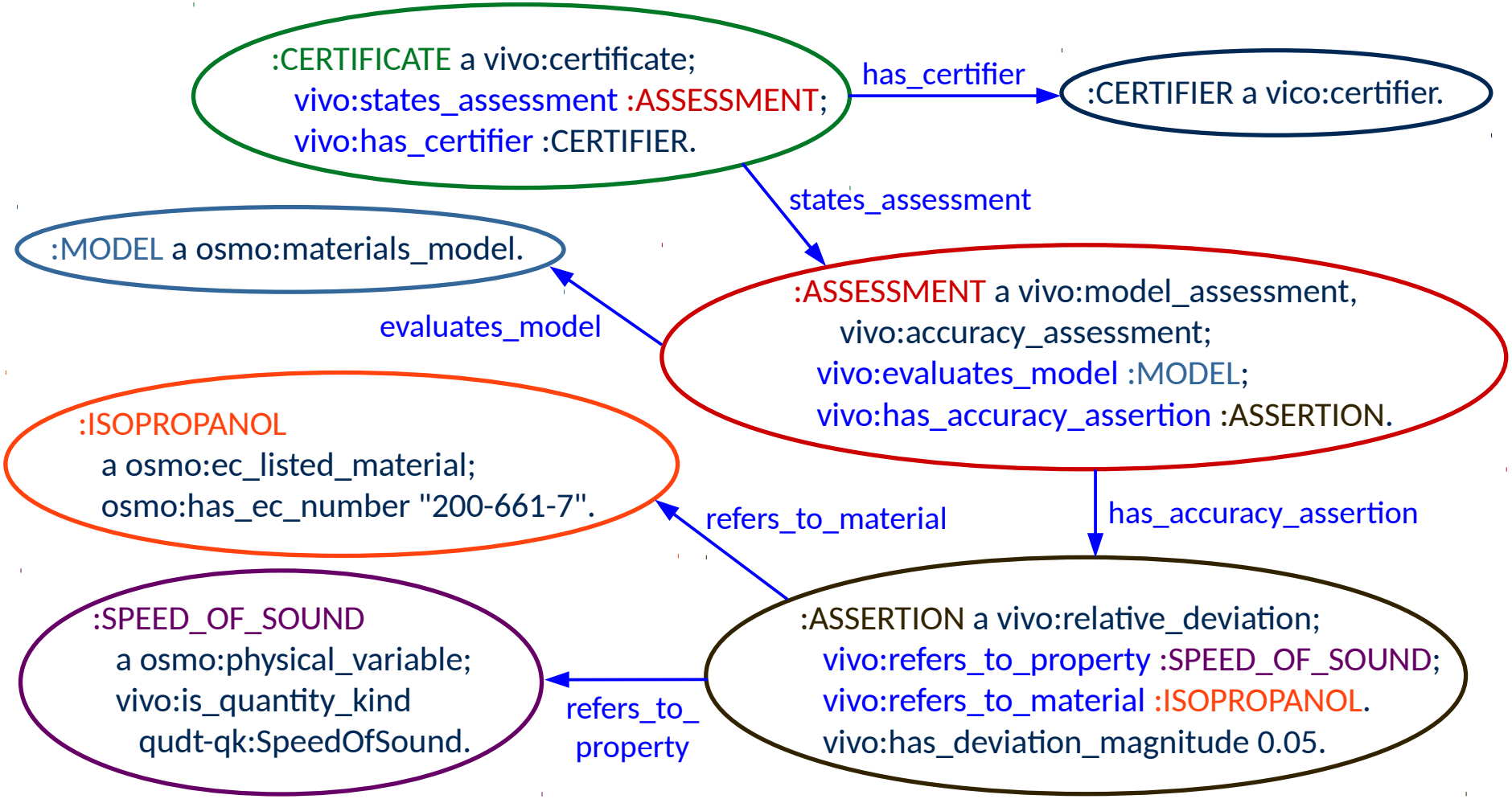
Wissensrepräsentation durch Ontologien



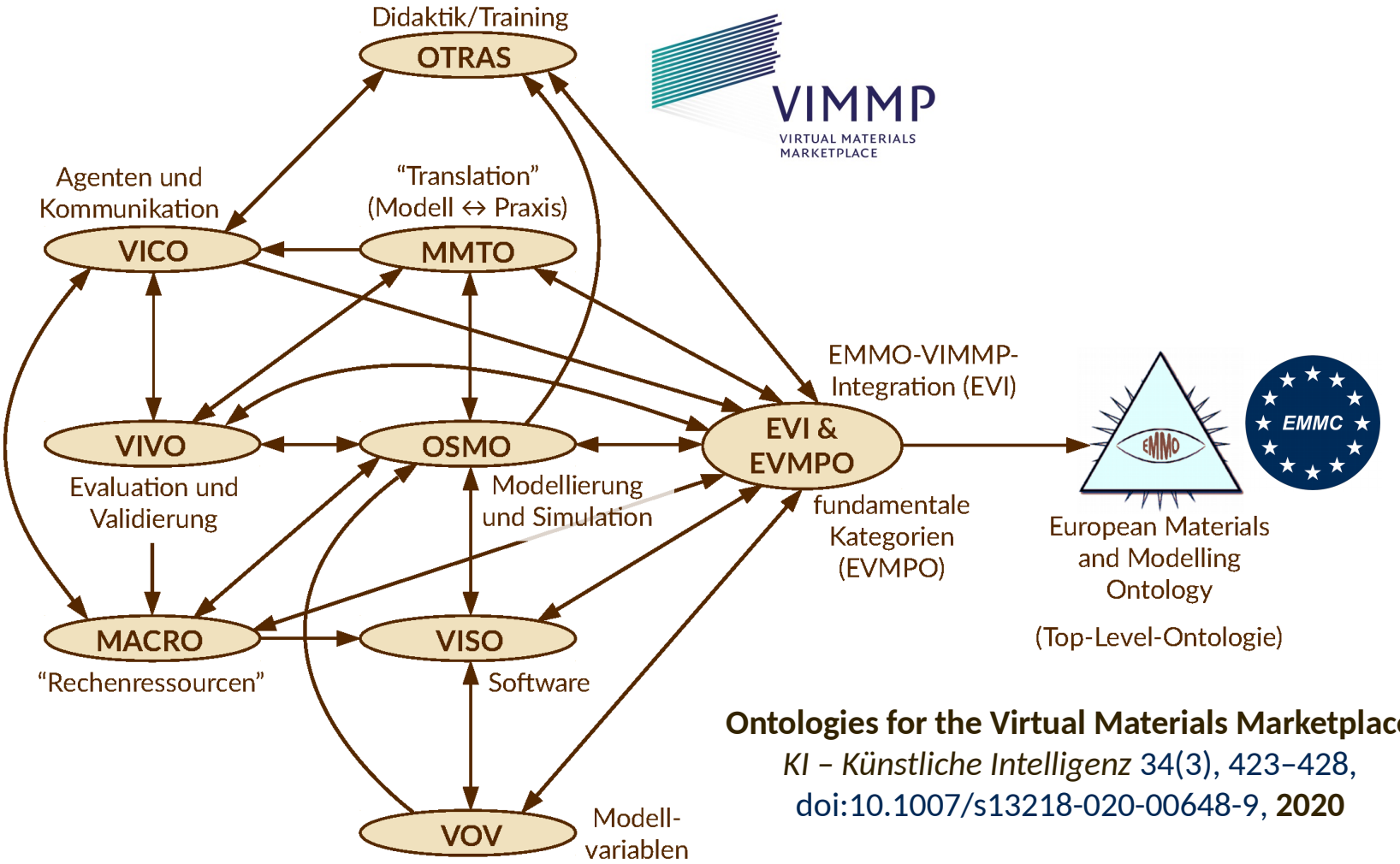
Der Gutachter :CERTIFIER gibt ein Zertifikat (die IRI des Zertifikats ist :CERTIFICATE) mit einem Gutachten (die IRI des Gutachtens ist :ASSESSMENT) heraus, das die Evaluation des Modells :MODEL zum Gegenstand hat.

Das Gutachten gibt den relativen Fehler des Modells für die Schallgeschwindigkeit von Isopropanol mit 5% an.

Wissensrepräsentation durch Ontologien

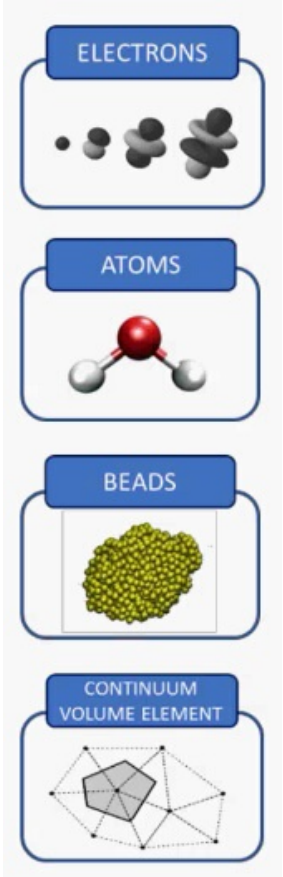


Wissensrepräsentation durch Ontologien



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

Wissensrepräsentation durch Ontologien



electronic

atomistic

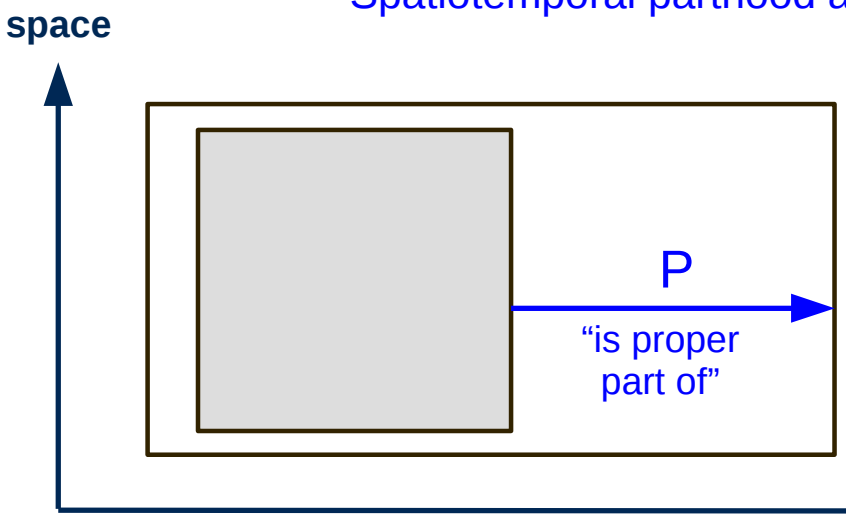
mesoscopic

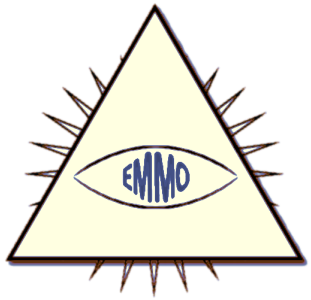
continuum

European Materials and Modelling Ontology[1]

1) Taxonomy:
Conceptual hierarchy (subclass relation)

2) Mereotopology:
Spatiotemporal parthood and connectivity

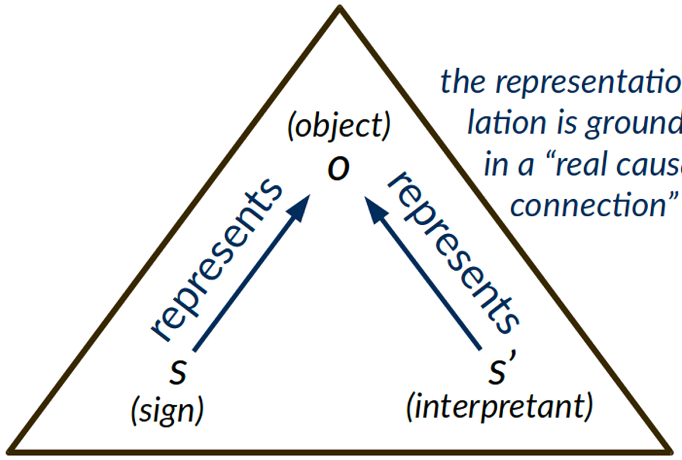




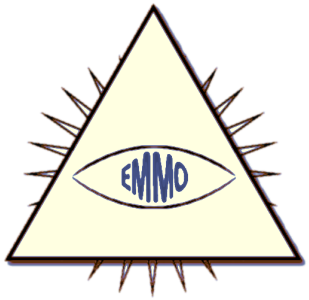
[1]E. Ghedini, J. Friis, A. Hashibon, G. J. Schmitz, G. Goldbeck, *et al.*, **2021**; http://emmc.info/emmo-info/.

Wissensrepräsentation durch Ontologien

Peircean semiotics



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



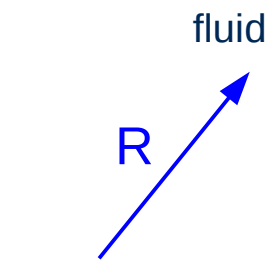
C. S. Peirce

European Materials and Modelling Ontology

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

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

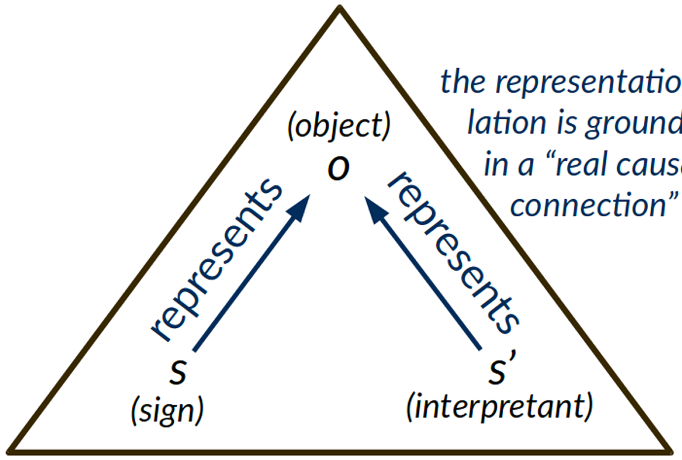
3) **Semiotics**:
   Representation of physical entities by signs



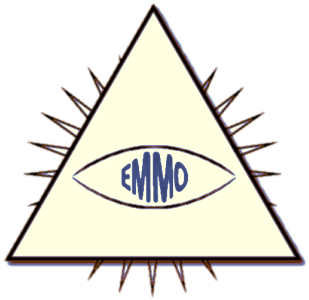
"represents" or **"is sign for"** is here abbreviated by **R**

Wissensrepräsentation durch Ontologien

Peircean semiotics



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



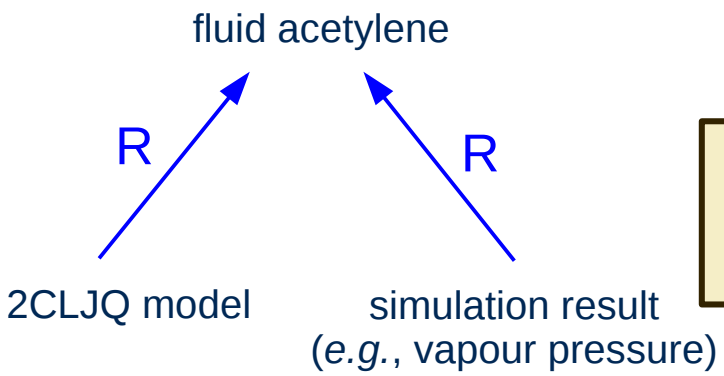
C. S. Peirce

European Materials and Modelling Ontology

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

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

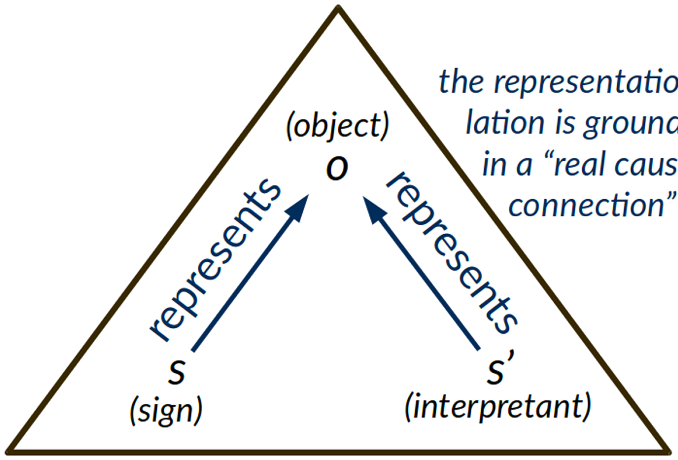
3) **Semiotics**:
   Representation of physical entities by signs



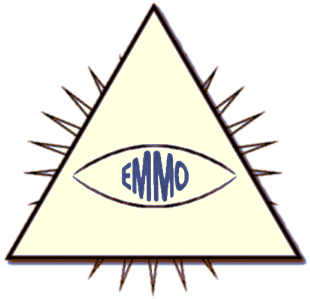
"represents" or **"is sign for"** is here abbreviated by **R**

Wissensrepräsentation durch Ontologien

Peircean semiotics



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

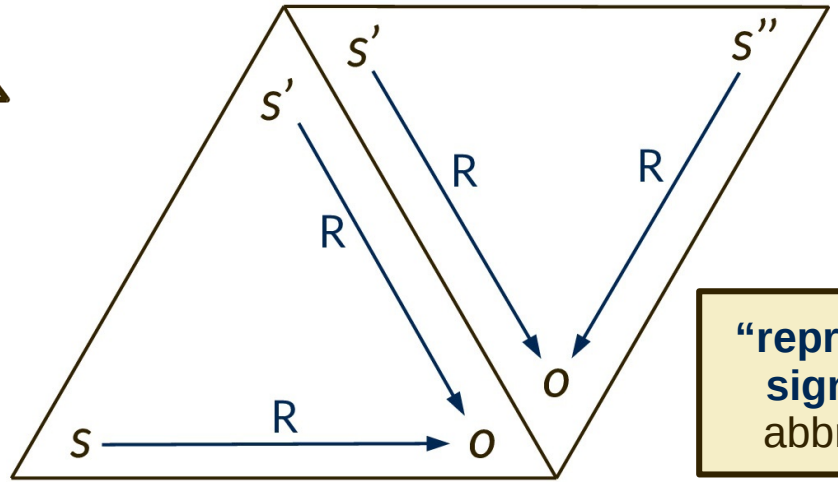


C. S. Peirce

Cognitive process (example):

- First, experimental data *s* for the material *o* are used to parameterize a model, obtaining model *s'*.
- Then, a simulation is done using model *s'*, yielding the simulation result *s''* (which also represents *o*).

In Peircean semiotics, a representation relation is carried over from one cognitive step (*i.e.*, triad) to the next.



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

Wissensrepräsentation durch Ontologien

Counterfactuals

Example:

"We simulate the energy consumption in 2020 assuming that COVID had not occurred" (while we know and assert that it did.)

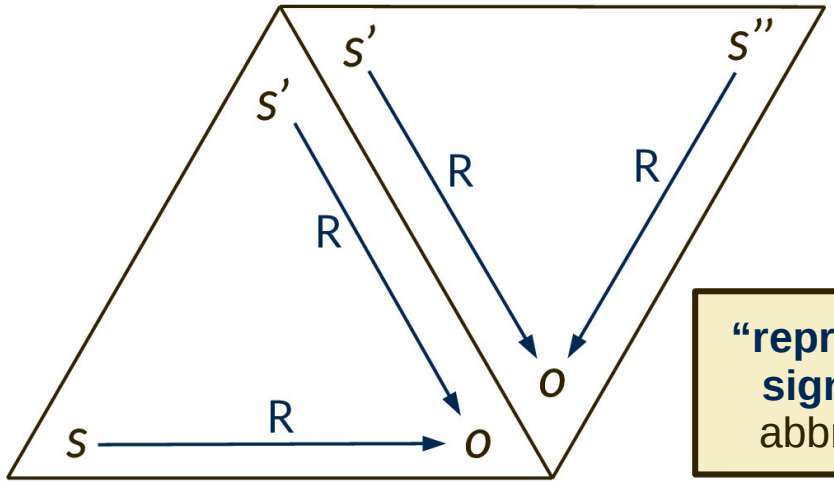
How can something *represent*, *or be a sign for, something that we know not to exist*, in an ontology that follows the basic approach from the EMMO?

Similar: Simulations of mutually exclusive possible events, *e.g.*, in *optimization*.

Cognitive process (example):

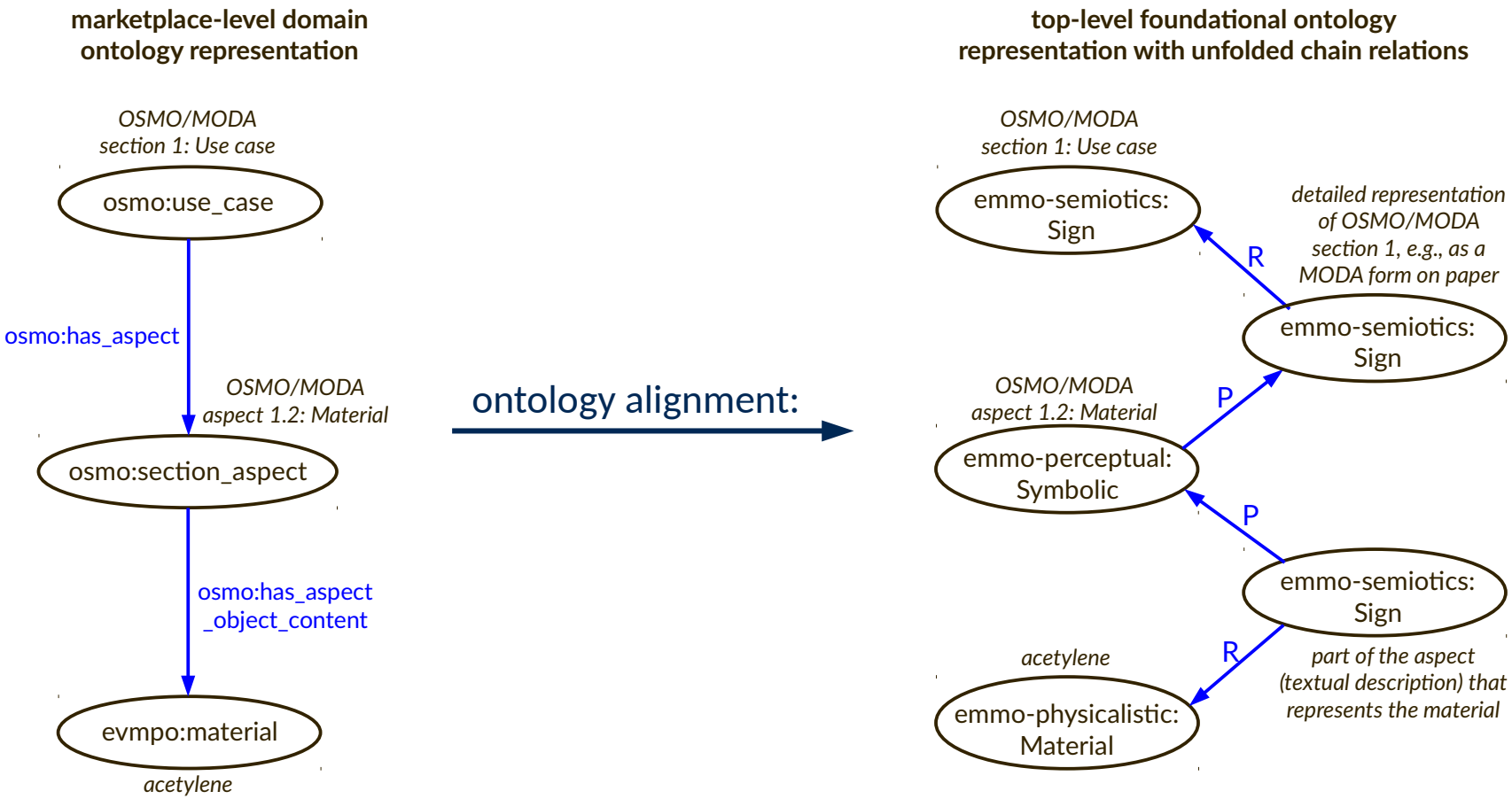
- First, experimental data *s* for the material *o* are used to parameterize a model, obtaining model *s'*.
- Then, a simulation is done using model *s'*, yielding the simulation result *s''* (which also represents *o*).

In Peircean semiotics, a representation relation is carried over from one cognitive step (*i.e.*, triad) to the next.



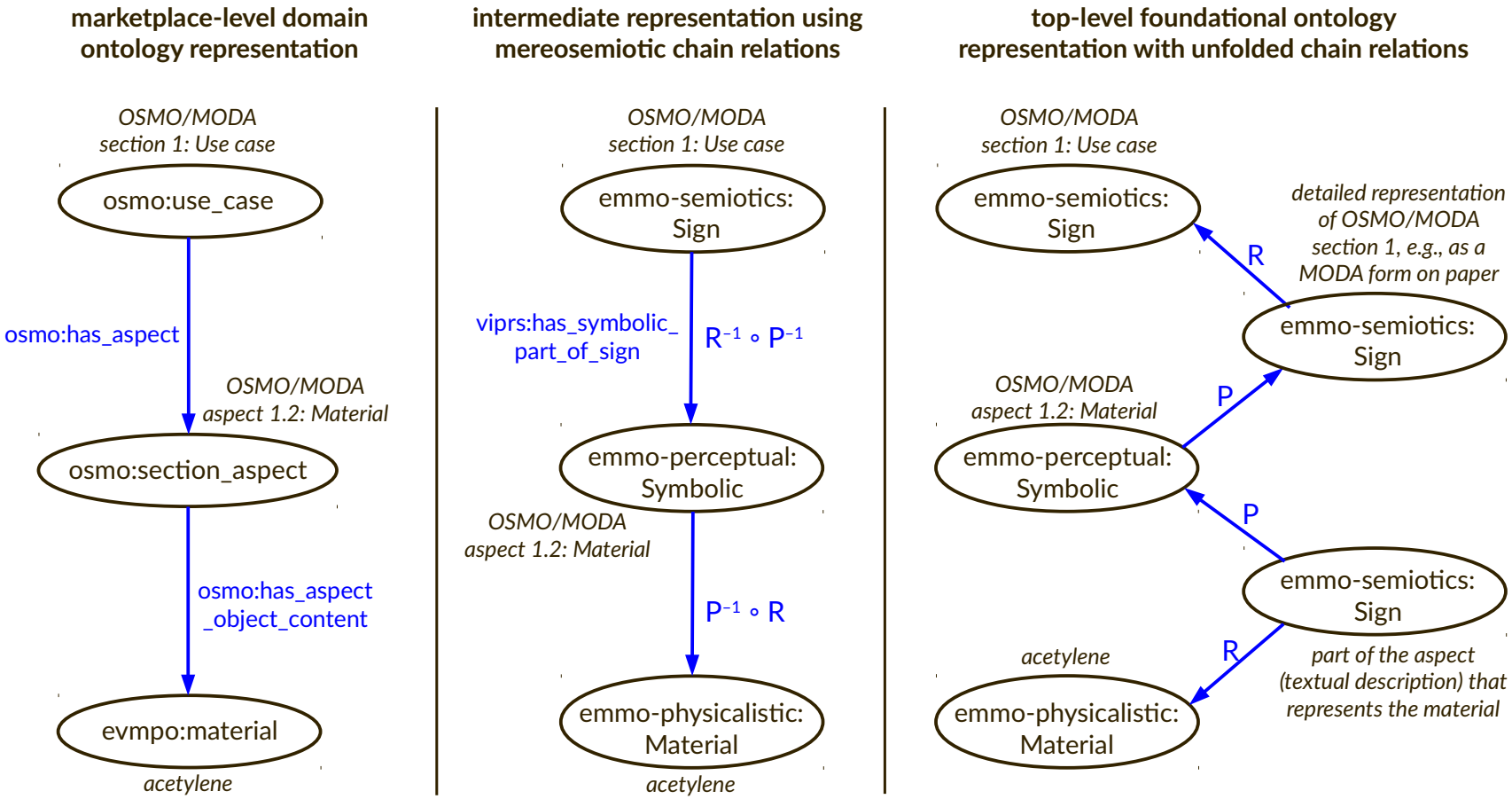
"represents" or **"is sign for"** is here abbreviated by **R**

Wissensrepräsentation durch Ontologien[1]



[1]M. T. Horsch, S. Chiacchiera, W. L. Cavalcanti, B. Schembera, *Data Technology in Materials Modelling*, Springer, **2021**.

Wissensrepräsentation durch Ontologien[1]



[1]M. T. Horsch, S. Chiacchiera, W. L. Cavalcanti, B. Schembera, *Data Technology in Materials Modelling*, Springer, **2021**.

Molekulare Modelldatenbank (MolMod DB)

Geometrie

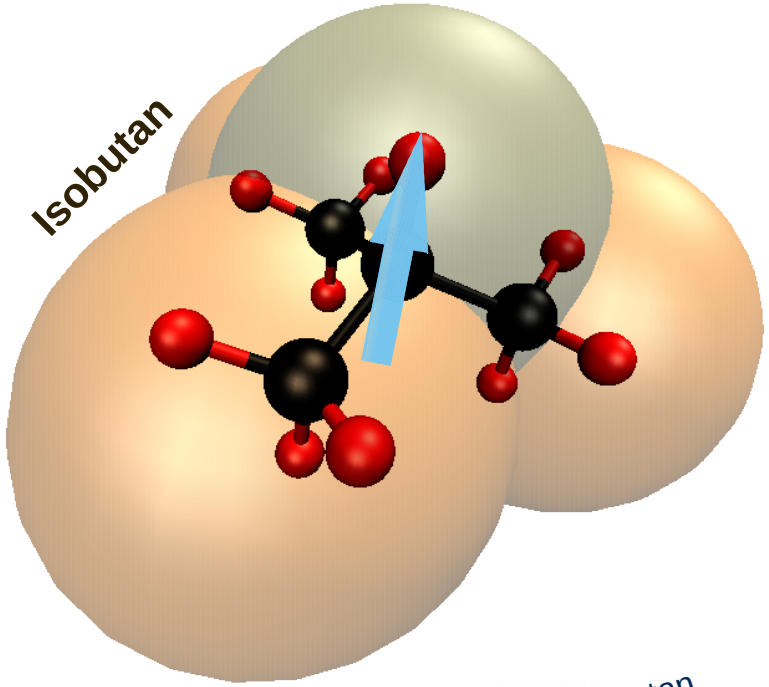
Position der Wechsel-
wirkungszentren

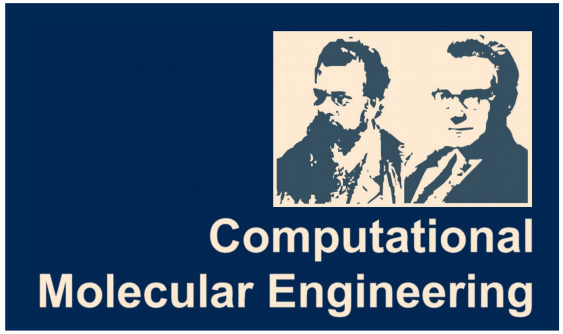
Dispersion und Repulsion

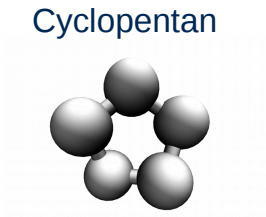
Lennard-Jones-Potential:
Längen- und Energieparameter

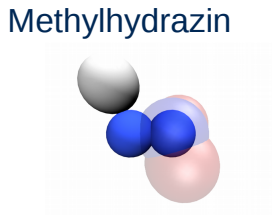
Elektrostatik

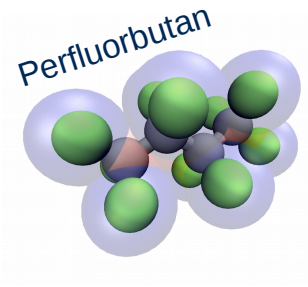
Punktpolaritäten
(Ladung, Dipol, Quadrupol):
Stärke und Orientierung











Molekulare Modelldatenbank (MolMod DB)

Geometrie

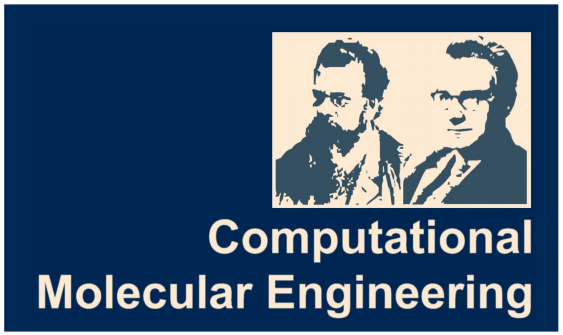
Position der Wechsel-
wirkungszentren

Dispersion und Repulsion

Lennard-Jones-Potential:
Längen- und Energieparameter

Elektrostatik

Punktpolaritäten
(Ladung, Dipol, Quadrupol):
Stärke und Orientierung

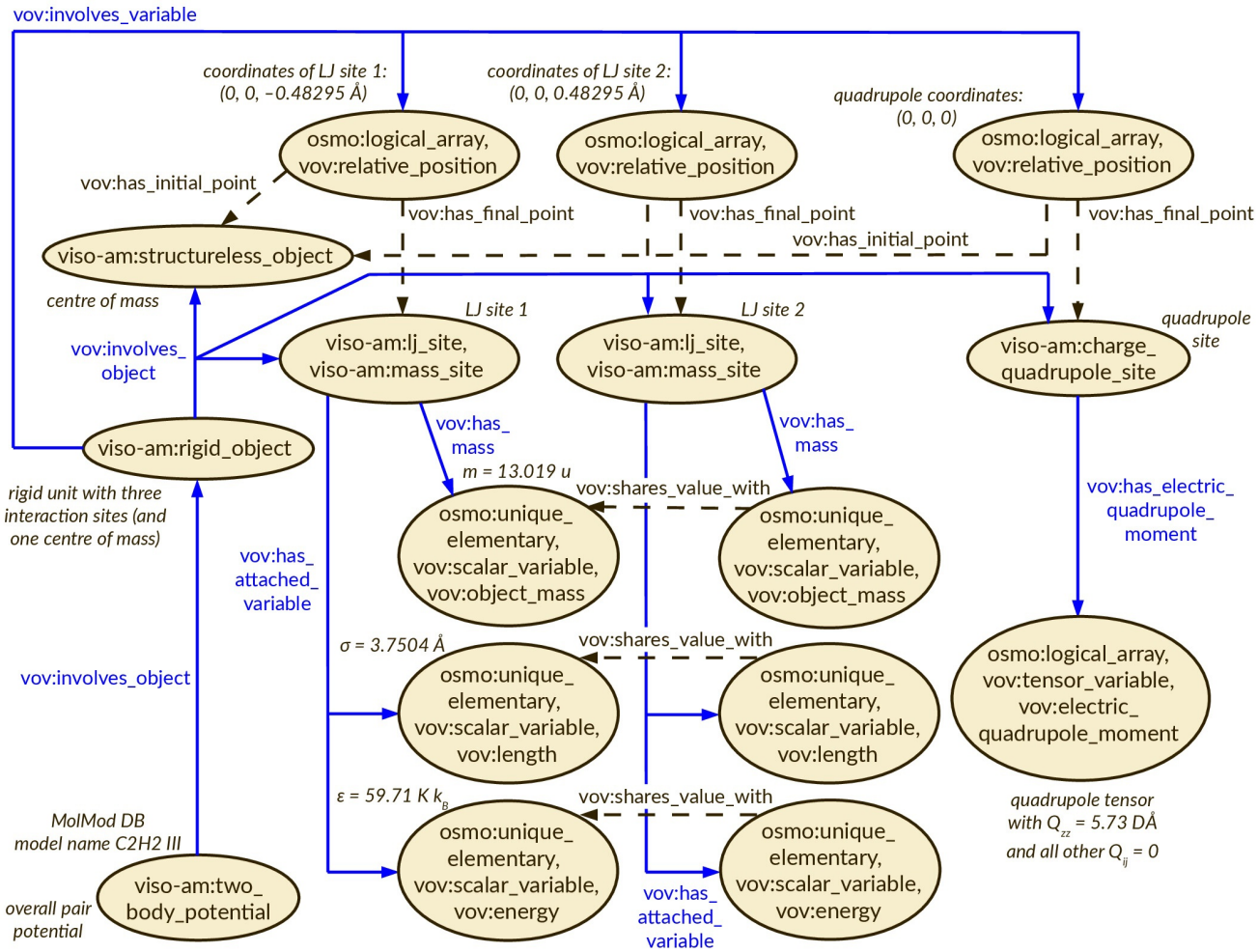


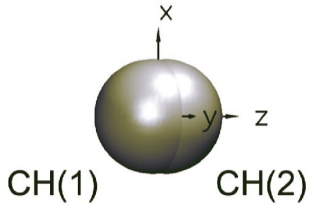
Molekulare Modelldatenbank

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

Paarpotentiale für über 150 molekulare Fluide

Molekulare Modelldatenbank (MolMod DB)



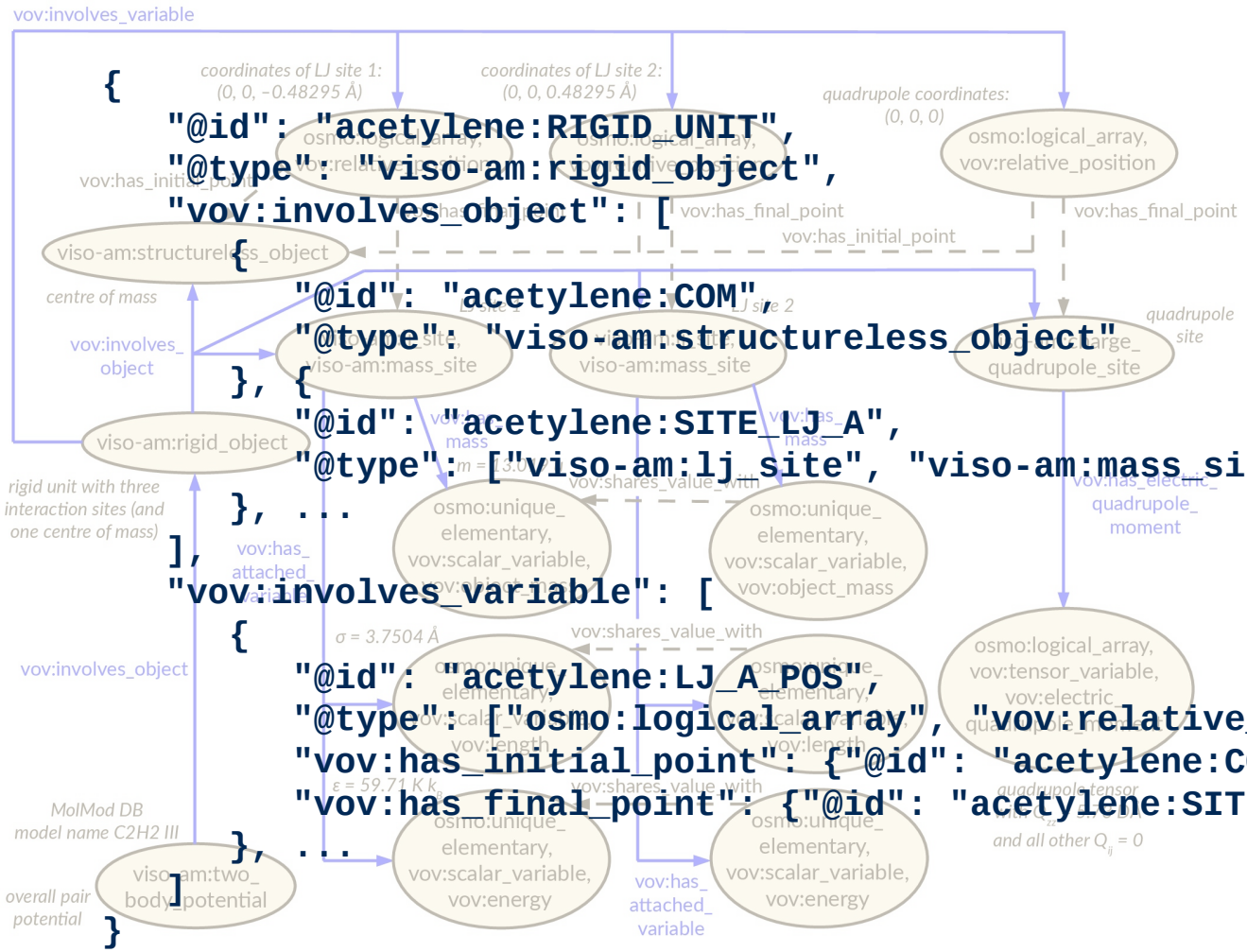


MolMod DB (Molecular Model Database)

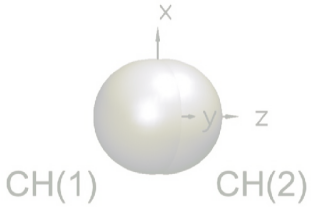
http://molmod.boltzmann-zuse.de/

pair potentials for over 150 molecular fluids

Molekulare Modelldatenbank (MolMod DB)




```
{
    "@id": "acetylene:RIGID_UNIT",
    "@type": "viso-am:rigid_object",
    "vov:involves_object": [
        {
            "@id": "acetylene:COM",
            "@type": "viso-am:structureless_object"
        }, {
            "@id": "acetylene:SITE_LJ_A",
            "@type": ["viso-am:lj_site", "viso-am:mass_site"], ...
        }, ...
    ],
    "vov:involves_variable": [
        {
            "@id": "acetylene:LJ_A_POS",
            "@type": ["osmo:logical_array", "vov:relative_position"],
            "vov:has_initial_point": {"@id": "acetylene:COM"},
            "vov:has_final_point": {"@id": "acetylene:SITE_LJ_A"}, ...
        }, ...
    ]
}
```



MolMod DB
(Molecular Model Database)

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

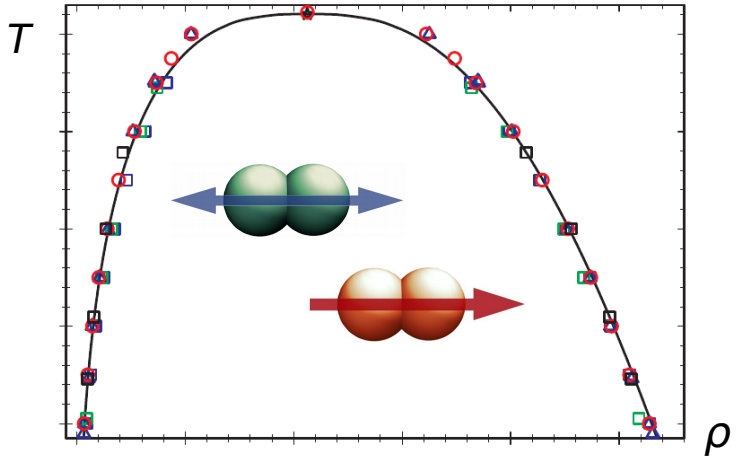
pair potentials for
over 150 molecular fluids

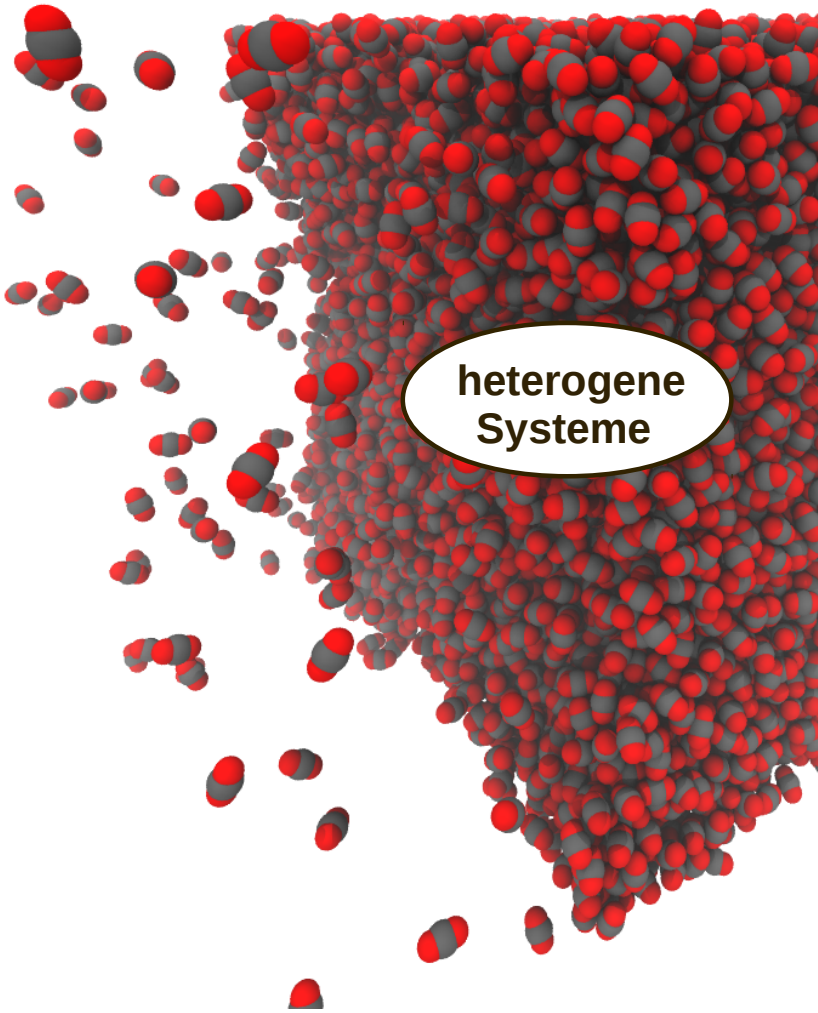
Modellparametrierung durch multikritierielle Optimierung

Anwendungsszenario: Phasengleichgewichtsdaten

- Molekulare Modelle aus der Literatur wurden in der Regel an Eigenschaften der homogenen Phasen (Bulk) angepasst, nicht aber an Grenzflächeneigenschaften.
- Verbreitete Modellklassen, die für zahlreiche Fluide eingesetzt werden, sind 2CLJD und 2CLJQ. Diese haben vier Modellparameter.

homogene Systeme



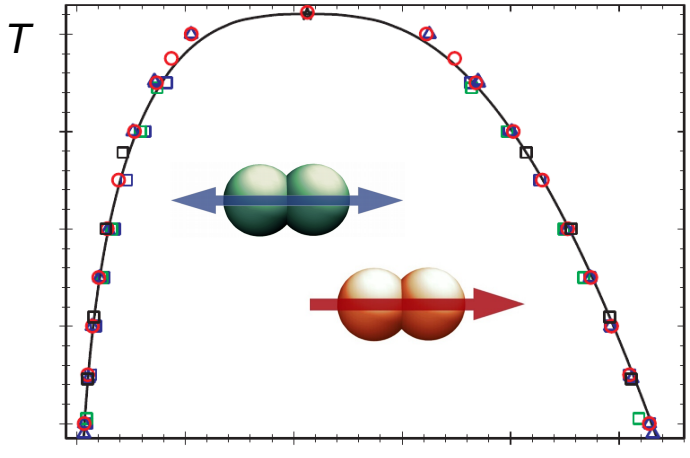


Modellparametrierung durch multikritierielle Optimierung

Anwendungsszenario: Phasengleichgewichtsdaten

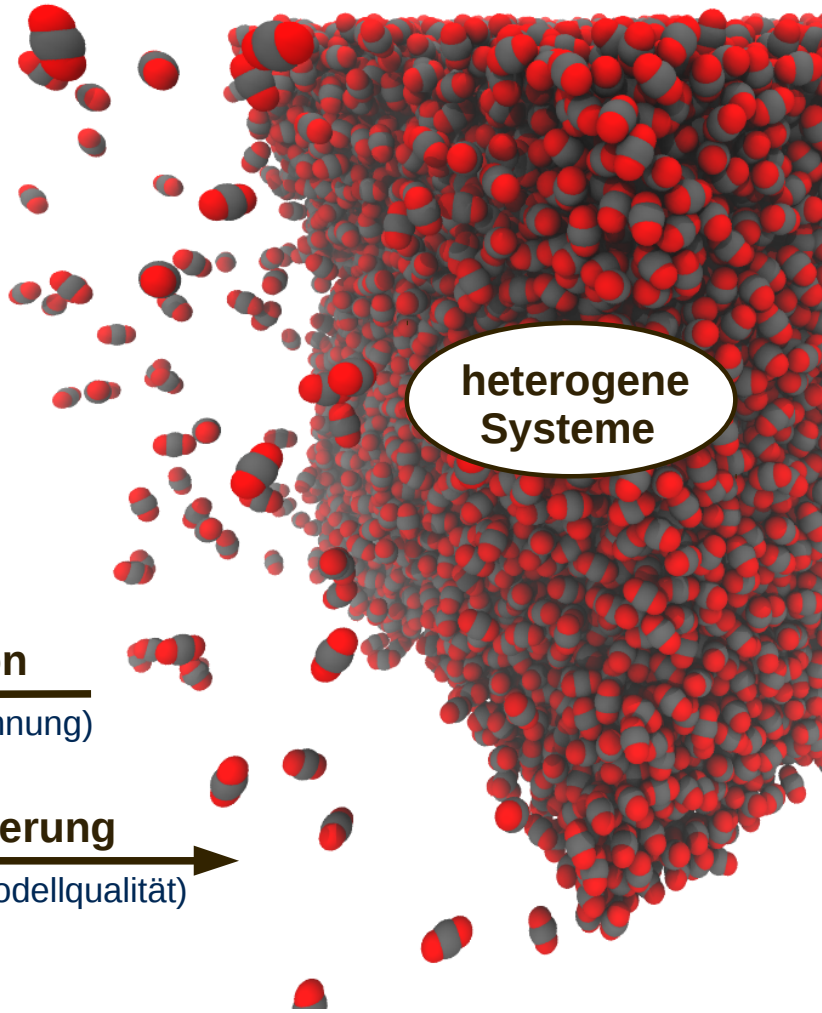
- Molekulare Modelle aus der Literatur wurden in der Regel an Eigenschaften der homogenen Phasen (Bulk) angepasst, nicht aber an Grenzflächeneigenschaften.
- Verbreitete Modellklassen, die für zahlreiche Fluide eingesetzt werden, sind 2CLJD und 2CLJQ. Diese haben vier Modellparameter.

homogene Systeme



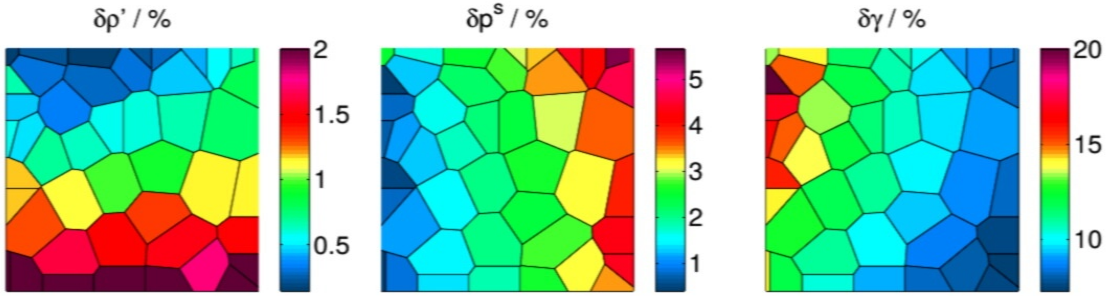
Korrelation
(Oberflächenspannung)

Reparametrierung
(Verbesserung der Modellqualität)



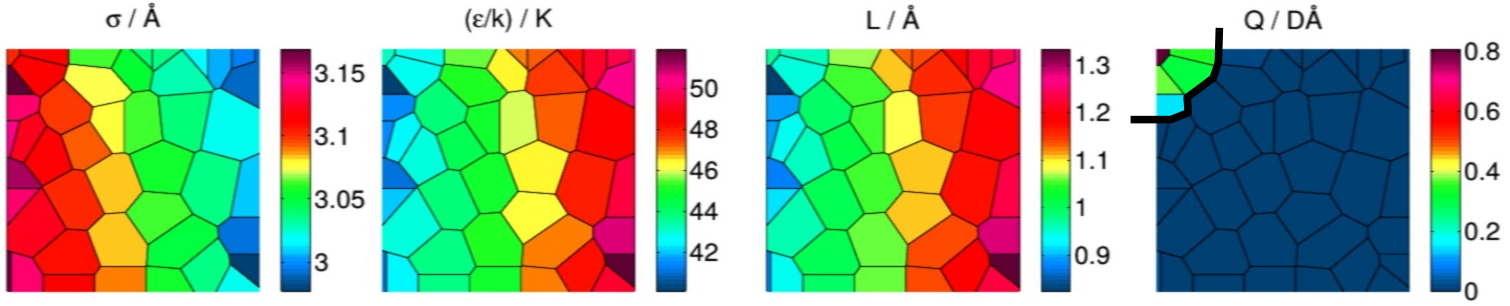
Modellparametrierung durch multikritierielle Optimierung

Visualisierung der Paretofront durch **selbstorganisierte Patchplots**[1]



Zielfunktions-raum

Pareto-optimale 2CLJQ-Modelle für Sauerstoff



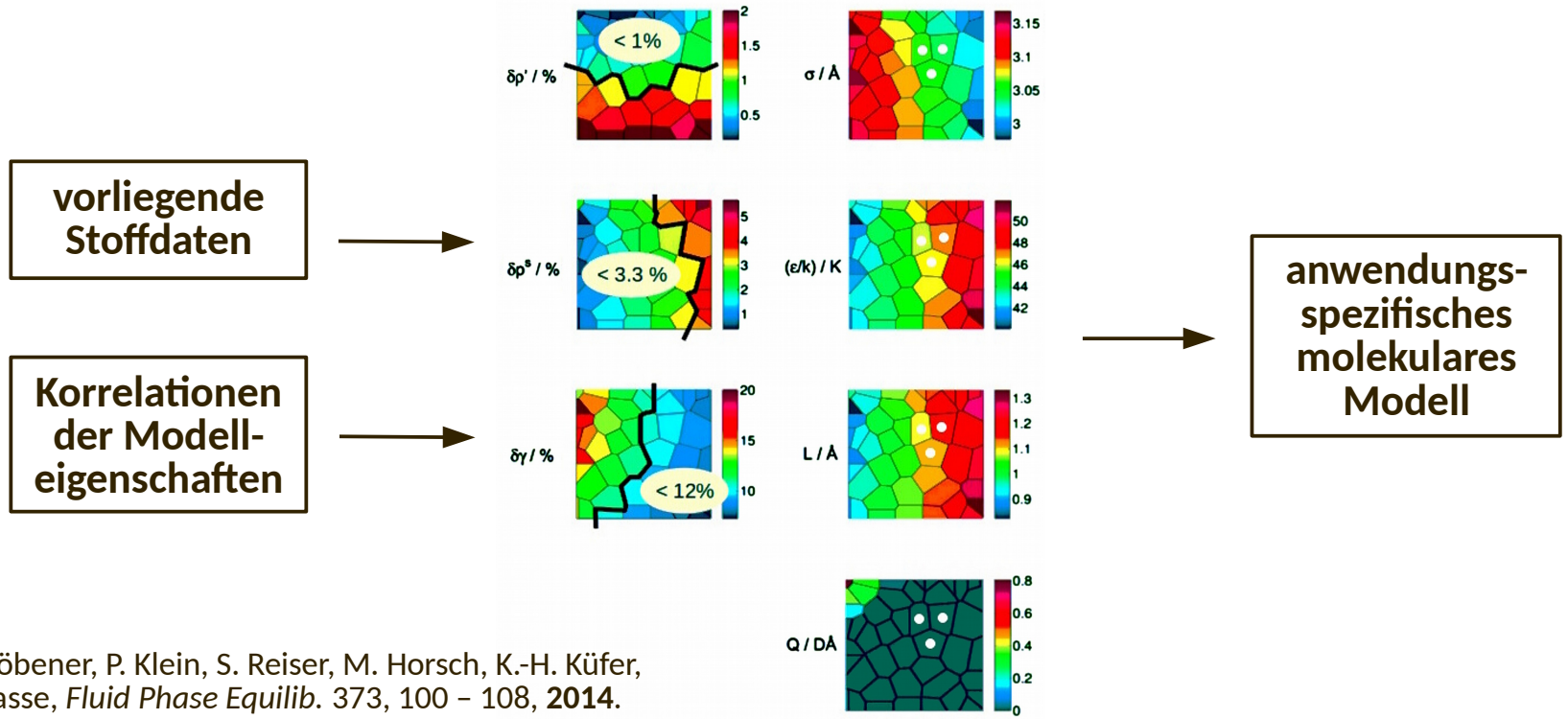
Parameter-raum

[1]K. Stöbener, P. Klein, M. Horsch, K. Küfer, H. Hasse, *Fluid Phase Equilib.* 411, 33 – 42, **2016**.

Modellparametrierung durch multikritierielle Optimierung

“Design eines molekularen Modells in fünf Minuten”

- Nutzergesteuerte **multikriterielle Optimierung** durch Auswahl aus der Paretofront.[1]
- Voraussetzung: Vorhandene Stoffdaten, charakterisierte Modellklasse (z.B. 2CLJQ).



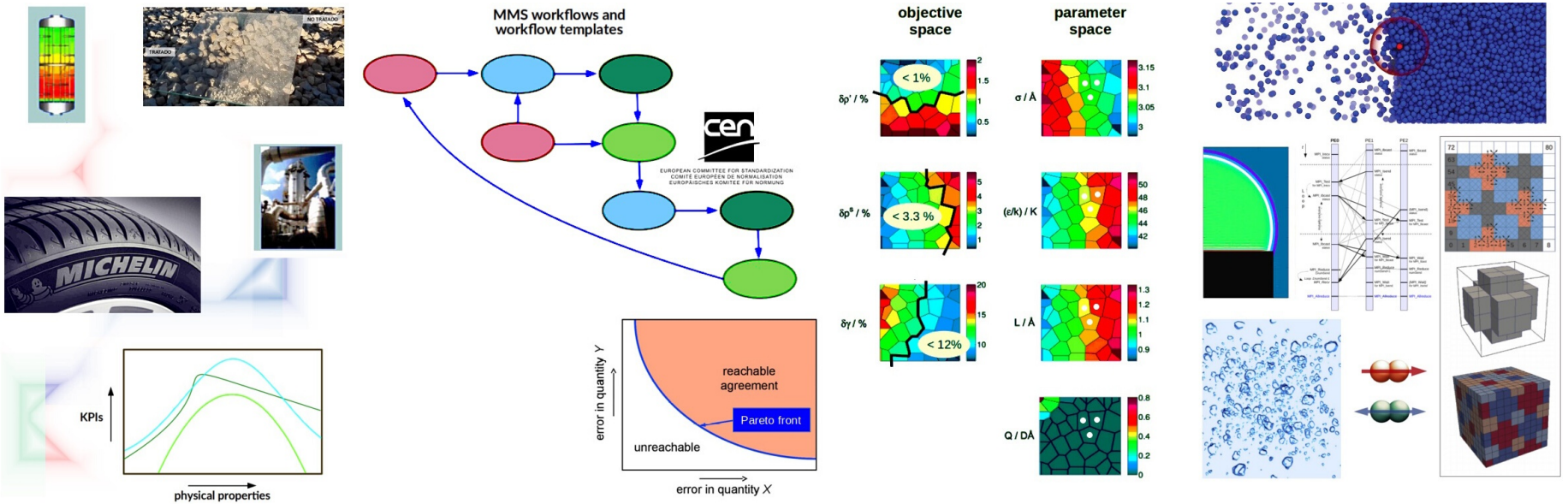
[1]K. Stöbener, P. Klein, S. Reiser, M. Horsch, K.-H. Küfer, H. Hasse, *Fluid Phase Equilib.* 373, 100 – 108, **2014**.

Workflowentwicklung: Coupling & Linking

Entwicklung in der Community, inkl. eigene Vorarbeiten:

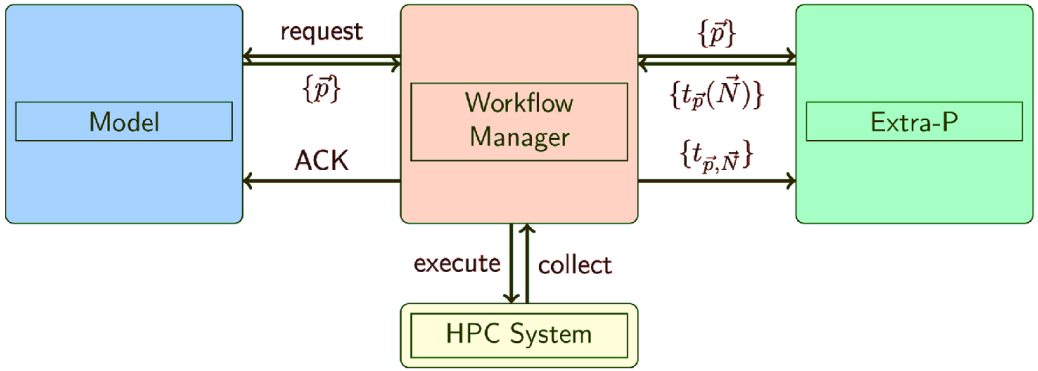
Integration molekulare Simulation + molekulare EOS + Dichtegradiententheorie

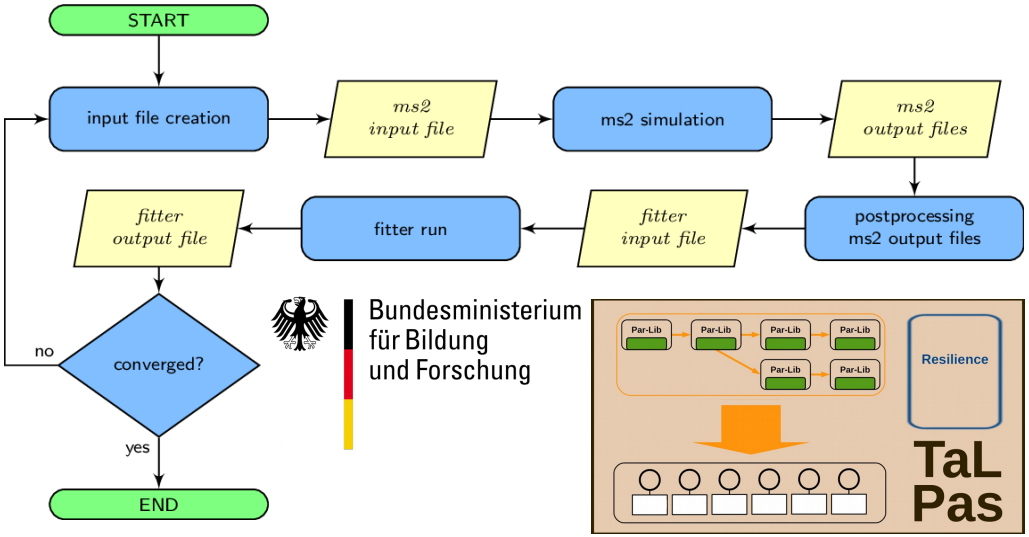
Methoden zur Modellparametrierung und Fehleranalyse



BDSS — workflow design — model design — solvers and API

Workflowentwicklung: Coupling & Linking





Bundesministerium für Bildung und Forschung

Workflow-Management-System aus dem BMBF-Projekt TaLPas (2017 – 2020)

Integrierte MD-Codes:

ls1 mardyn

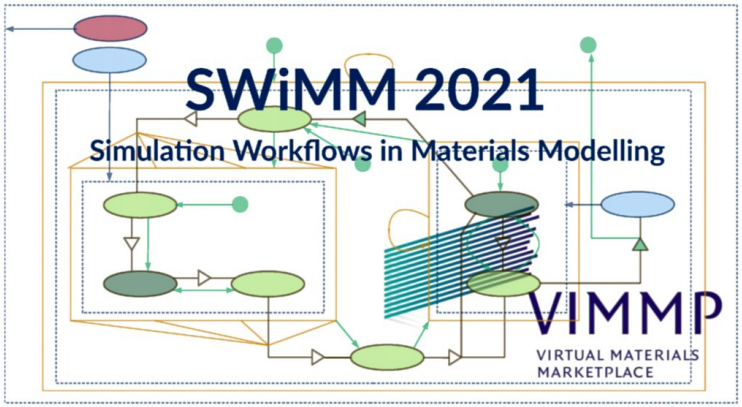
https://www.ls1-mardyn.de/

ms2

https://www.ms-2.de/

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

Simulation Workflows in Materials Modelling (CECAM)



Simulation Workflows in Materials Modelling

(SWiMM 2021)

15th – 26th March 2021

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

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

cecam
Centre Européen de Calcul Atomique et Moléculaire

Perspektive für die Digitalisierung in der Thermodynamik

Prioritäten („DORIC-Prinzipien“) nach doi:10.5281/zenodo.4571052

D **diversify** technologies

O **observe** practices

R **realistic** objectives

I **incentivize** open data

C **co-design** data with solvers

Perspektive für die Digitalisierung in der Thermodynamik

(1) Motivation und Überblick
(2) Organisation und Projekte
(3) Semantische Technologie
(4) Datenprovenienz
(5) Wissensrepräsentation
(6) Molekulare Modelle
(7) Workflowentwicklung
(8) Perspektive

Vgl. Selent, Kraus, Hansen, Schembera, Seeland, Iglezakis, **2020.**

D **diversify** technologies

O **observe** practices

R **realistic** objectives

I **incentivize** open data

C **co-design** data with solvers

Input von Anwendern/Entwicklern gesucht (kurzfristig)

Community event am 15. März 2021 (Registrierung bis morgen, 10. März 2021!)

Community input in Vorbereitung des DORIC-MM 2021 Workshop am 7. Juni 2021

Input von Anwendern/Entwicklern gesucht (kurzfristig)

Community event am 15. März 2021 (Registrierung bis morgen, 10. März 2021!)

Domains of knowledge from Industry Commons (20 selected domains)

	materials and properties	manufacturing and processes
disciplines 401-01, 401-02, 401-06, and 405-05	surface morphology, mechanical components	machining, additive manufacturing, *etc.*
discipline 401-04	polymers	polymer technology
discipline 402-02	solid materials	solid material synthesis and characterization
discipline 402-03	woven and nonwoven materials *etc.*	textile manufacturing
disciplines 403-01 and 404-02	thermodynamic properties of fluids	thermal separation unit operations & processes

	materials and properties	manufacturing and processes
discipline 403-02	reactants, catalysts, products	chemical reactors and kinetics
discipline 403-03	particles, granular media	processes with granular/ particulate media
discipline 403-04	products, reactants, *etc.* in biotech/food industry	manufacturing in biotech/food industry
disciplines 404-01 and 404-02	refrigerants	refrigeration technology
disciplines 404-03 and 404-04	rheological and transport properties of fluids	compressors, turbines, aerodynamics, *etc.*

https://ontocommons.eu/doric-mm-2021/

DORIC - MM 2021

Digitalisierung in der Molekularen Thermodynamik

Martin Thomas Horsch