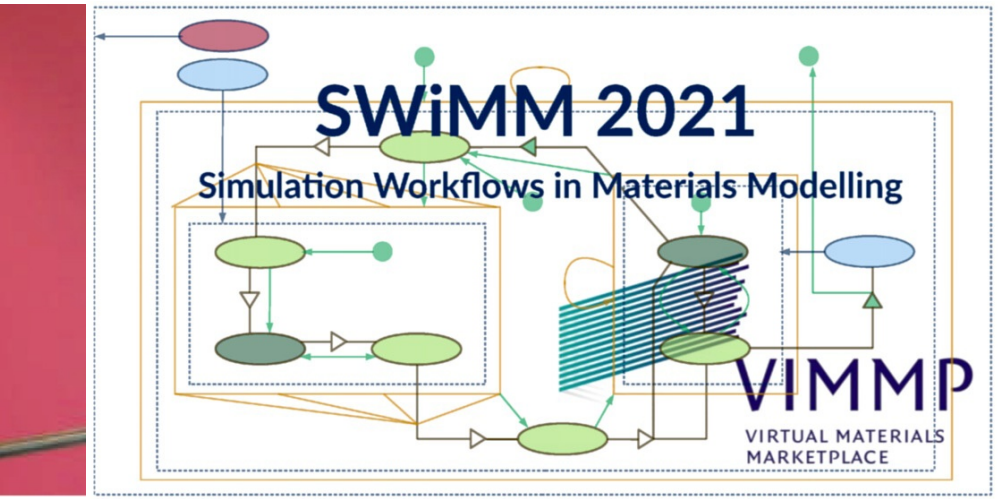
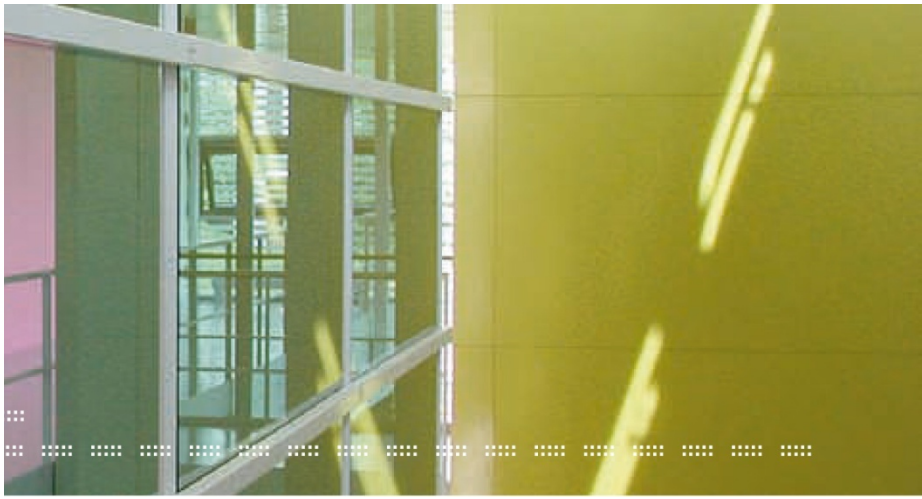
Semantic interoperability and ontology-driven technology for simulation workflows

Silvia Chiacchiera and Martin Thomas Horsch





The meaning of … coffee

An Italian bar in Germany:

- Silvia: “Un caffè macchiato, per favore.” [A *caffè macchiato*, please.]
- Barman: “Un caffè macchiato… che significa?” [A *caffè macchiato*… what does it mean?]

The meaning of … coffee

An Italian bar in Germany:

- Silvia: “Un caffè macchiato, per favore.” [A *caffè macchiato*, please.]
- Barman: “Un caffè macchiato… che significa?” [A *caffè macchiato*… what does it mean?]

What is happening?

The barman needs a clarification, namely: which meaning do we refer to, the Italian one (as language suggests) or the German one (as the location suggests)? They differ, notably in the average size of the cup and drink.

The meaning of … coffee

An Italian bar in Germany:

- Silvia: “Un caffè macchiato, per favore.” [A *caffè macchiato*, please.]
- Barman: “Un caffè macchiato… che significa?” [A *caffè macchiato*… what does it mean?]

Let’s analyse this further:

1) The two subjects speak the same language (Italian)

2) A disambiguation is needed (to agree which convention is being followed)

3) They are in a bar: the first sentence implies a trading interaction between the speakers (e.g., this wouldn’t be the case in a friend’s place)

These correspond to: 1) syntactics, 2) semantics and 3) pragmatics (see next).

The meaning of … coffee

An Italian bar in Germany:

- Silvia: “Un caffè macchiato, per favore.” [A *caffè macchiato*, please.]
- Barman: “Un caffè macchiato… che significa?” [A *caffè macchiato*… what does it mean?]

From Merriam-Webster (following Charles Morris):

1) Syntactics: “a branch of semiotics that deals with the **formal relations between signs or expressions** in abstraction from their signification and their interpreters.”

2) Semantics: “the study of **meanings**; a branch of semiotics dealing with the **relations between signs and what they refer to** (…).”

3) Pragmatics: “a branch of semiotics that deals with the **relation between signs** or linguistic expressions **and their users**; a branch of linguistics [the study of human speech] that is concerned with the relationship of sentences to the **environment** in which they occur.”

According to Morris, these are the three branches of semiotics: “a general philosophical **theory of signs and symbols** that deals especially with their function **in both artificially constructed and natural languages.**”

The meaning of … data

Now: how does this picture look in the context of **data**?

Let's imagine we receive a file "sigma.dat" (on the right).

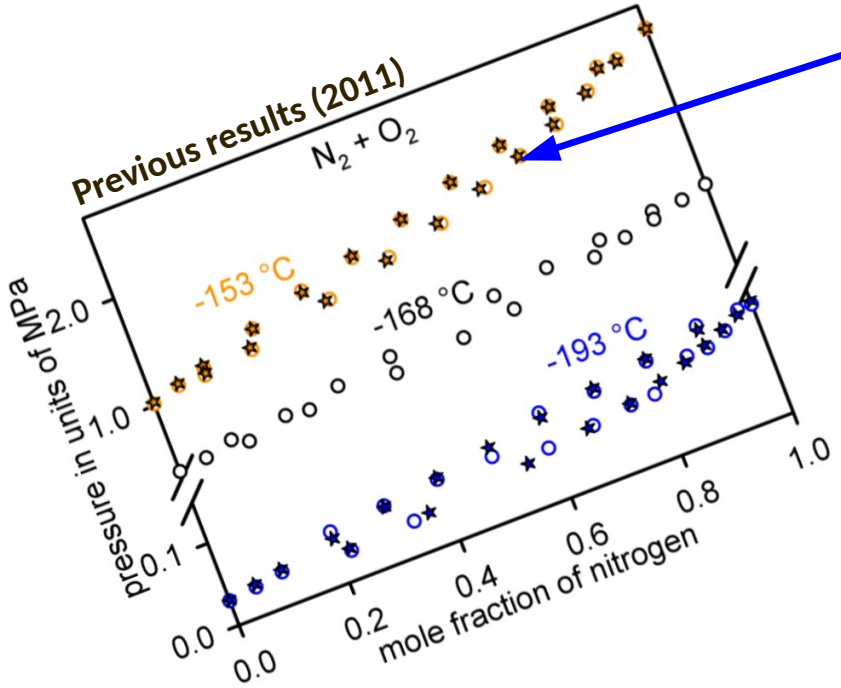
```
# Model 1
# A       sigma       sigma_err
40.0      1.17745     0.167
60.0      3.03579     0.3592
80.0      3.62384     0.3797
100.0     4.30474     0.3719


# Model 2
# A       sigma       sigma_err
40.0      1.25022     0.1238
60.0      2.75247     0.2723
80.0      4.05209     0.2691
100.0     4.05401     0.2726
```

1) Syntax/syntactics:
formal relations between signs → The file format (ASCII text file).

2) Semantics:
meaning, relations between signs and what they refer to → Info about the content: e.g., what each column and block means, the data provenance, etc.

3) Pragmatics:
relation between signs and their users/enviroment → E.g.: we type "rm sigma.dat" in a terminal. Depending on our rights on the file, it will be removed or not.

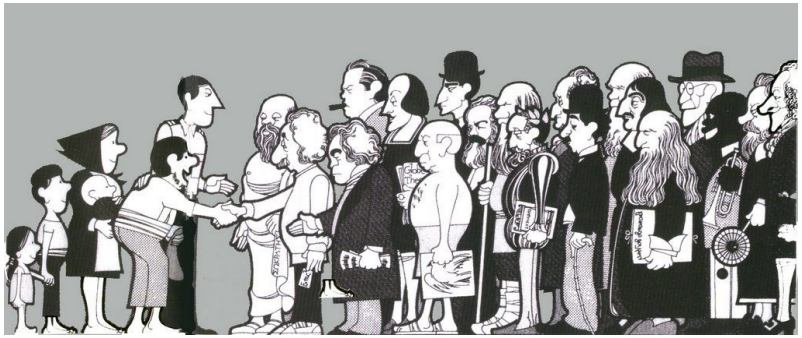
Why do we need semantic technology?



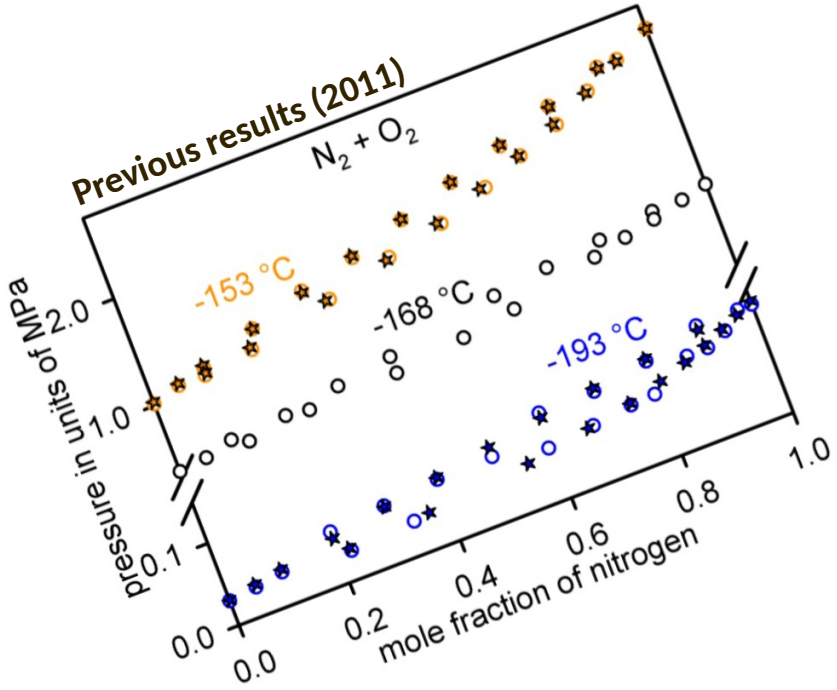
What values did *x* and *p* have?

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?



Why do we need semantic technology?

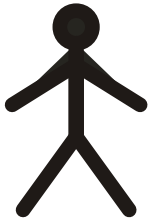


What values did *x* and *p* have?

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?

competency questions



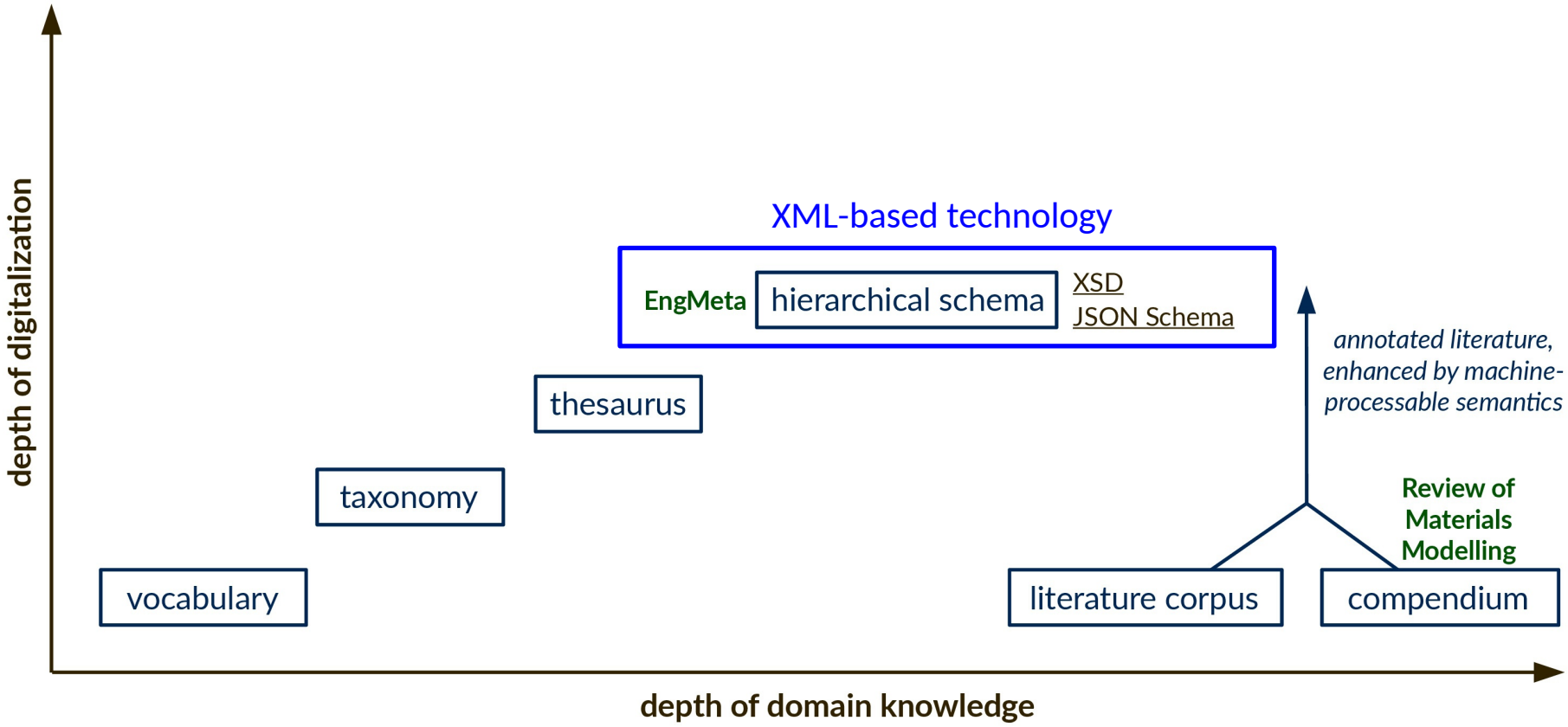
"I remember.
Haha, joke. Of course I don't."

Good practice in managing research data:

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

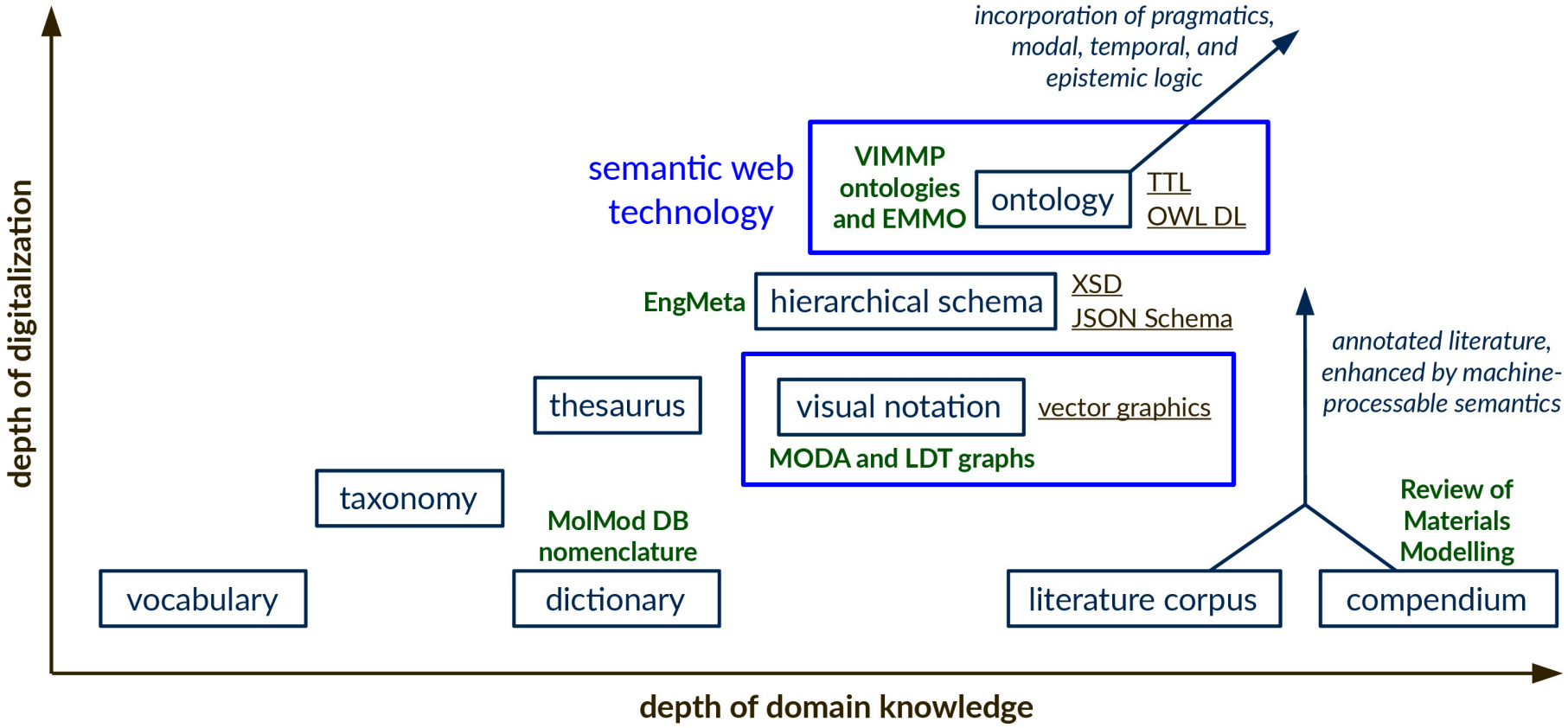
Metadata standardization

Hierarchy of semantic assets



Metadata standardization

Hierarchy of semantic assets



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 (in a nutshell)

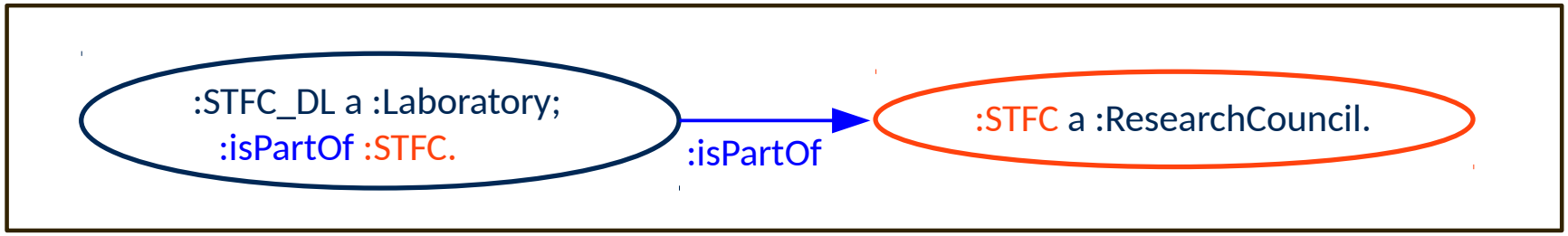
```
subject    a                 class_of_subject;
           has_property      first_object, second_object;
           other_property    another_object.
```

TTL format applied to OWL

Example

```
osmo:logical_access a owl:Class;
  rdfs:subClassOf evmpo:annotation;
  owl:disjointWith osmo:section_aspect;
  owl:disjointUnionOf
    (osmo:logical_read_access osmo:logical_write_access);
  rdfs:comment "connection from a section to a logical resource".
osmo:has_access_point a owl:ObjectProperty;
  rdfs:domain osmo:logical_access;
  rdfs:range osmo:section;
  rdfs:subPropertyOf viprs:satisfies_requirement_of.
osmo:has_resource a owl:ObjectProperty;
  rdfs:domain osmo:logical_access;
  rdfs:range osmo:logical_resource.
```

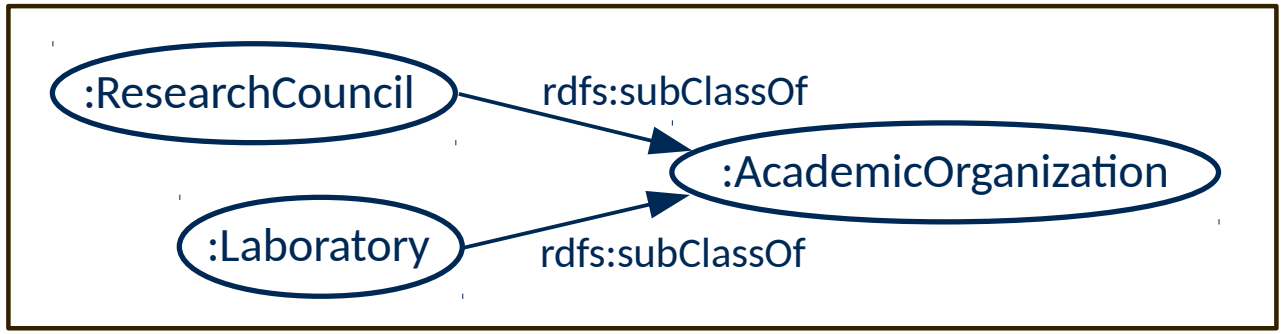
Ontologies (in a nutshell)



Knowledge graph; also: Scenario or assertional box (ABox)

:STFC_DL :isPartOf :STFC.

RDF triple, consisting of subject, predicate, and object



Taxonomy, part of the ontology

also: TBox (terminological box)

Ontology development in materials modelling

Community-governed development of metadata standards

RoMM VI → **MODA** → **CWA 17284** → **Ontologies**

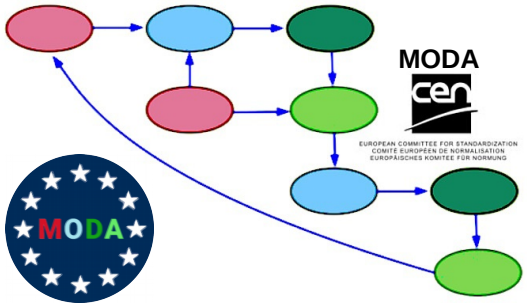
Review of Materials Modelling (compendium)

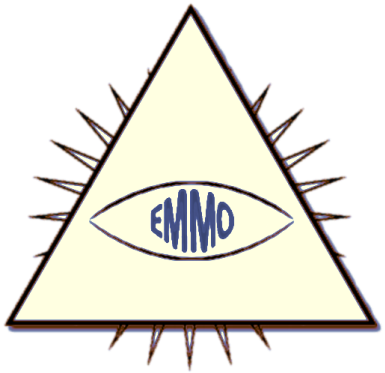
MODA ("Model Data")

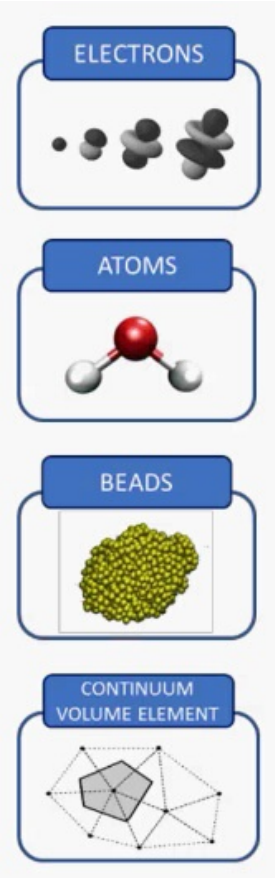
CEN workshop agreement

Domain ontologies

EMMO top-level ontology







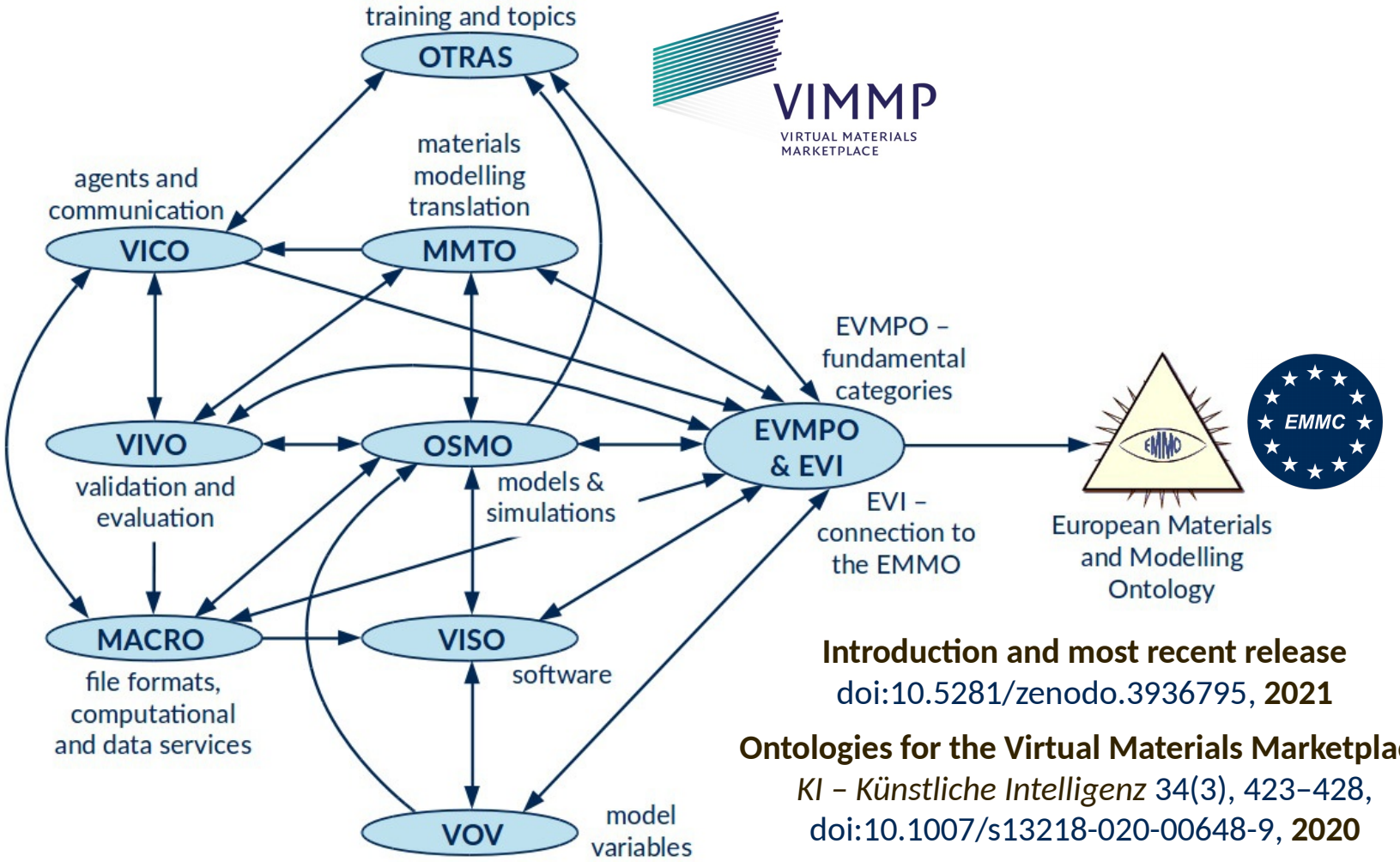
electronic

atomistic

mesoscopic

continuum

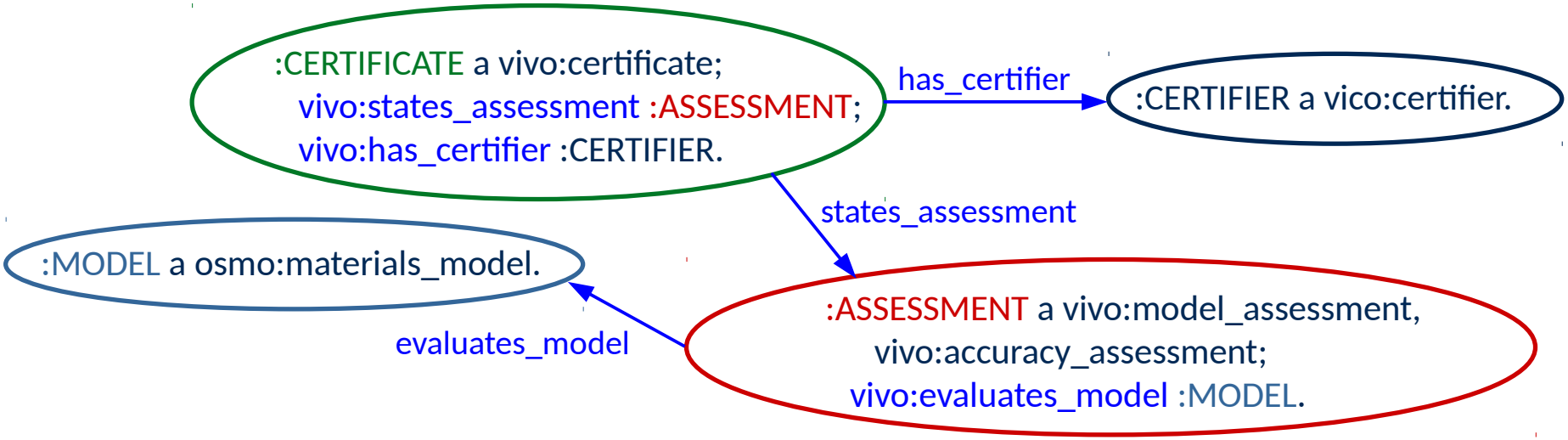
Ontology development in materials modelling



Introduction and most recent release
doi:10.5281/zenodo.3936795, **2021**

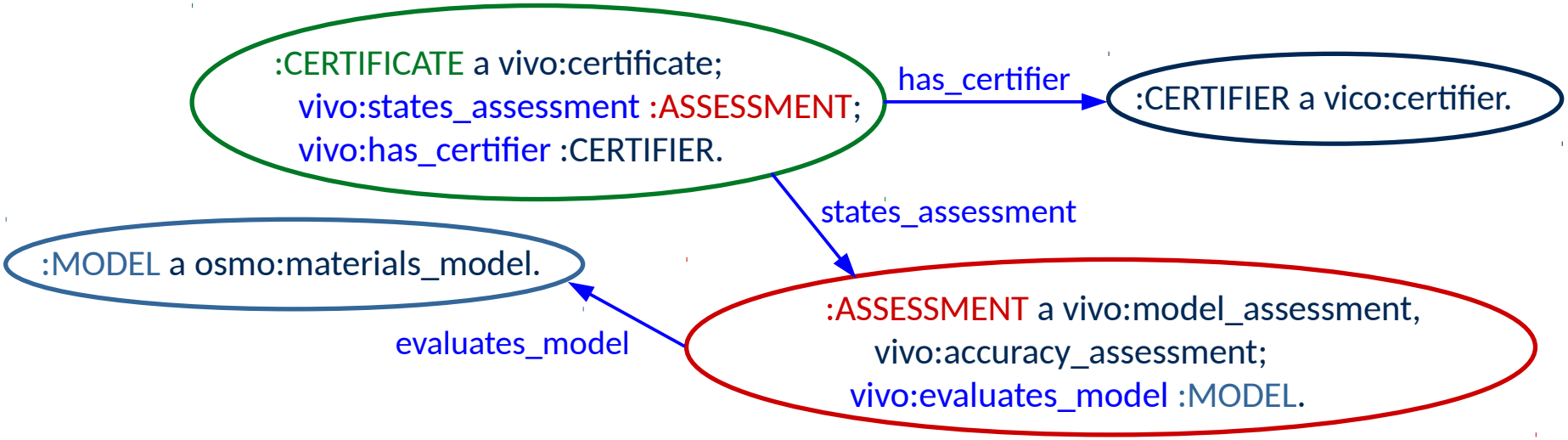
Ontologies for the Virtual Materials Marketplace
KI – Künstliche Intelligenz 34(3), 423–428,
doi:10.1007/s13218-020-00648-9, **2020**

Knowledge graphs



The certifier :CERTIFIER has issued a certificate (the IRI of which is :CERTIFICATE) stating a model accuracy assessment (the IRI of which is :ASSESSMENT) that evaluates the materials model :MODEL.

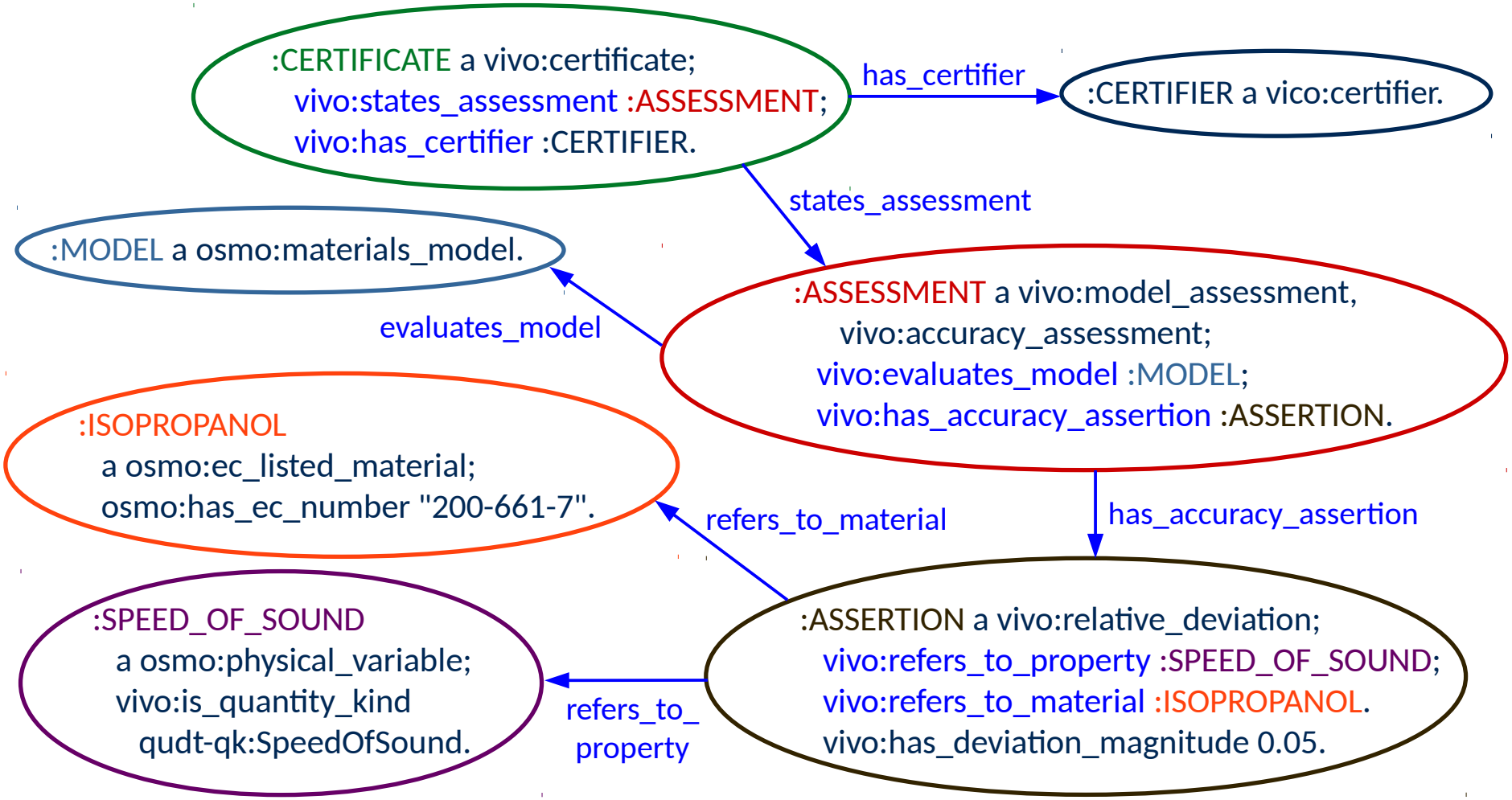
Knowledge graphs



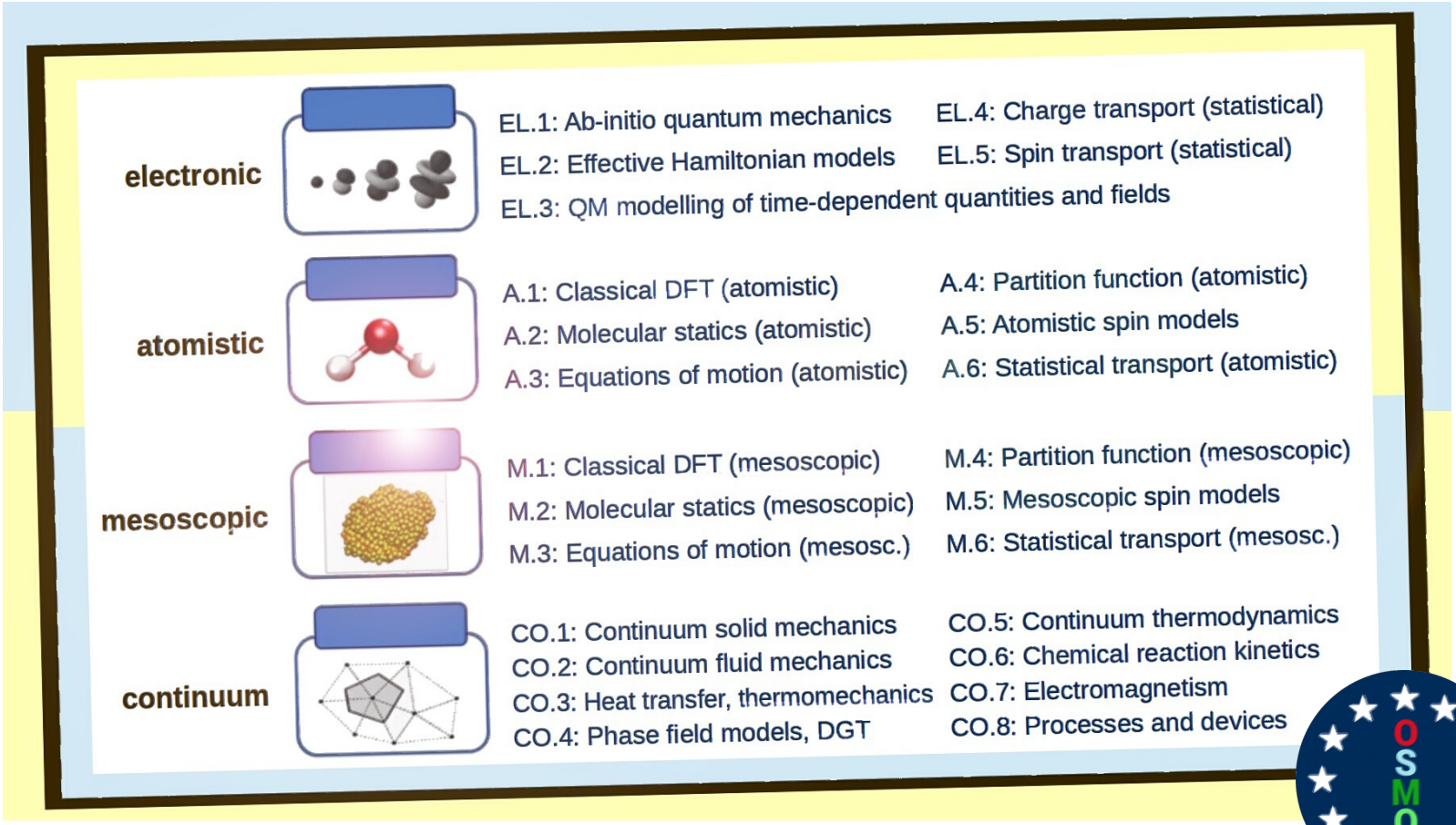
The certifier :CERTIFIER has issued a certificate (the IRI of which is :CERTIFICATE) stating a model accuracy assessment (the IRI of which is :ASSESSMENT) that evaluates the materials model :MODEL.

In the model accuracy assessment :ASSESSMENT, it is asserted that the materials model :MODEL has a relative error of 5% for the speed of sound of isopropanol.

Knowledge graphs



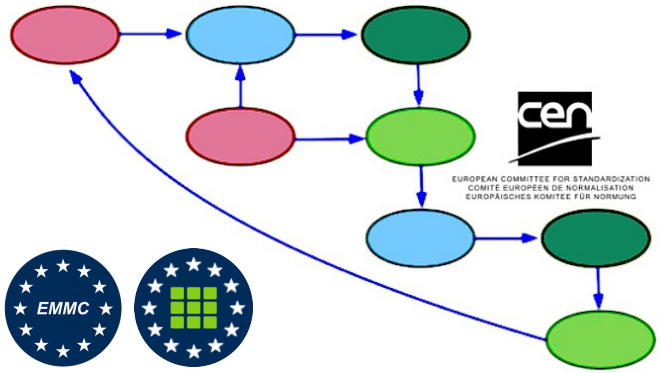
Simulation data provenance

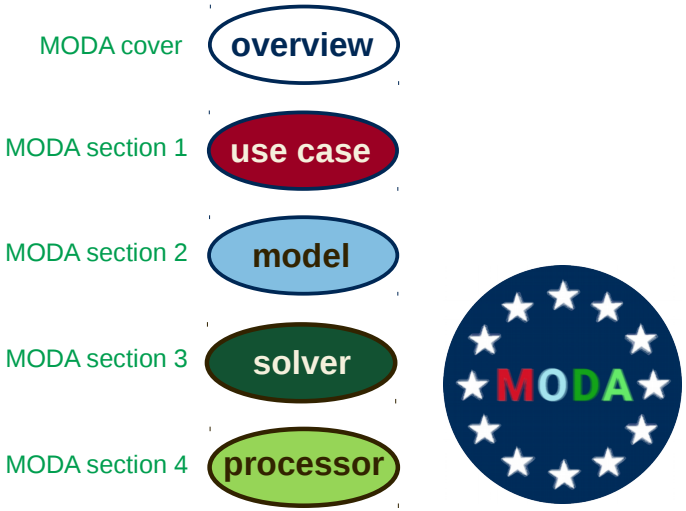


Ontology for Simulation, Modelling, and Optimization (OSMO)
J. Chem. Eng. Data 65(3), 1313–1329, **2020**

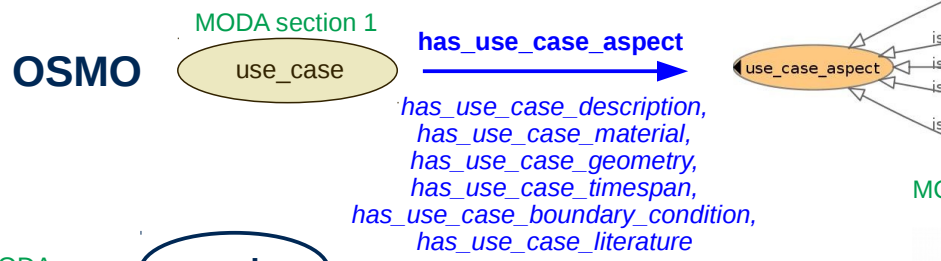
Simulation data provenance

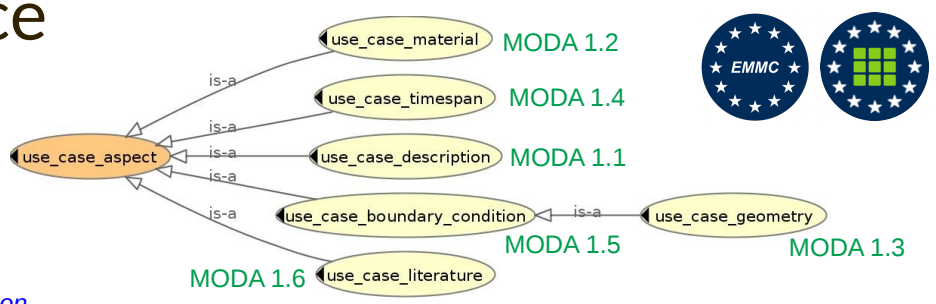
3	Solver and Computational translation of the specifications		
3.1	Numerical Solver	*Please give name andtype of the solver* *e.g. Monte Carlo, SPH, FE, ...iterative, multi-grid, adaptive,...*	
3.2	Software tool	*Please give the name and if this is your own code, please specify if it can be shared with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

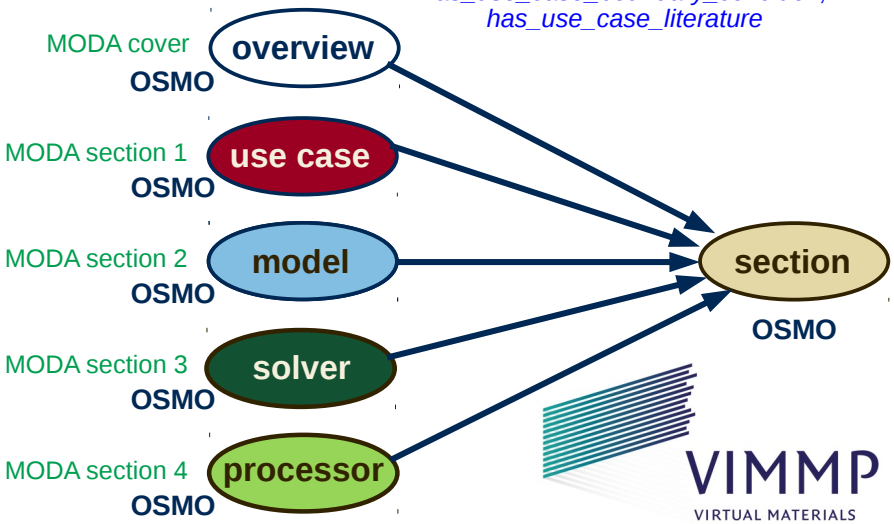




Simulation data provenance



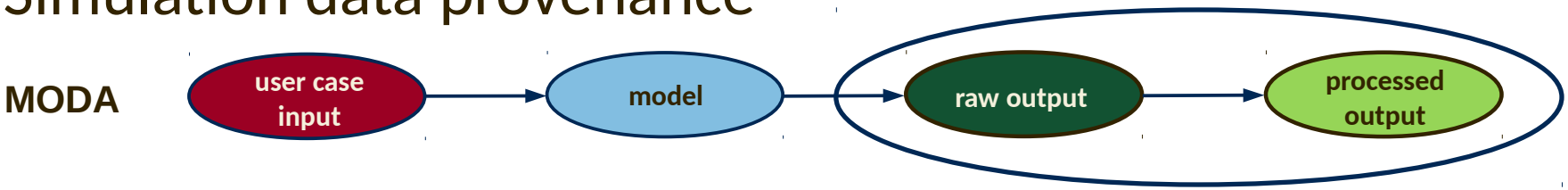




3	Solver and Computational translation of the specifications		
3.1	Numerical Solver		*Please give name andtype of the solver e.g. Monte Carlo, SPH, FE, ...iterative, multi-grid, adaptive,...*
3.2	Software tool		*Please give the name and if this is your own code, please specify if it can be shared with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

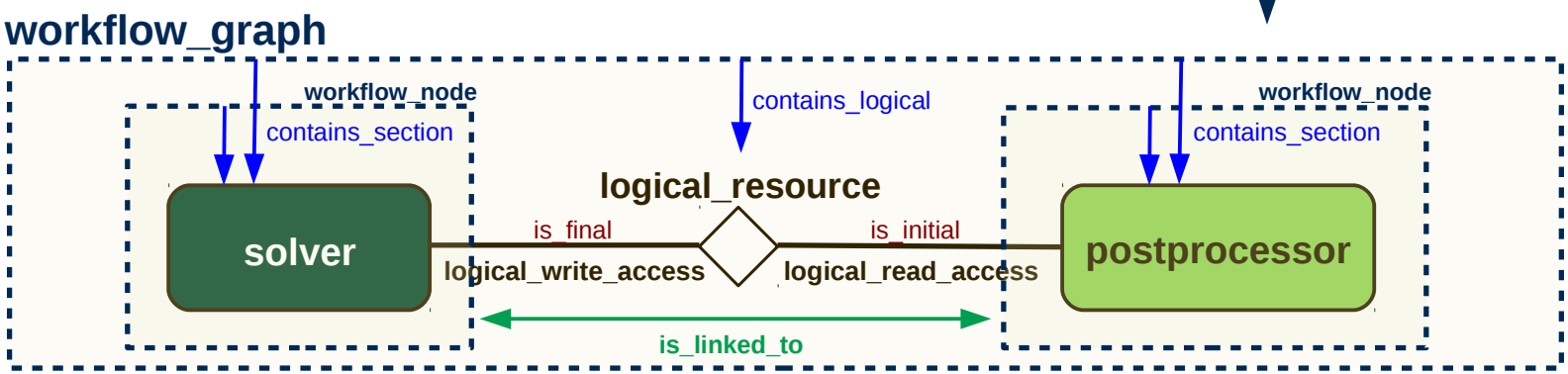
semi-formalized metadata standard (human processable) → formal ontology (machine processable)

Simulation data provenance



"What is the exact meaning of the blue arrows?"

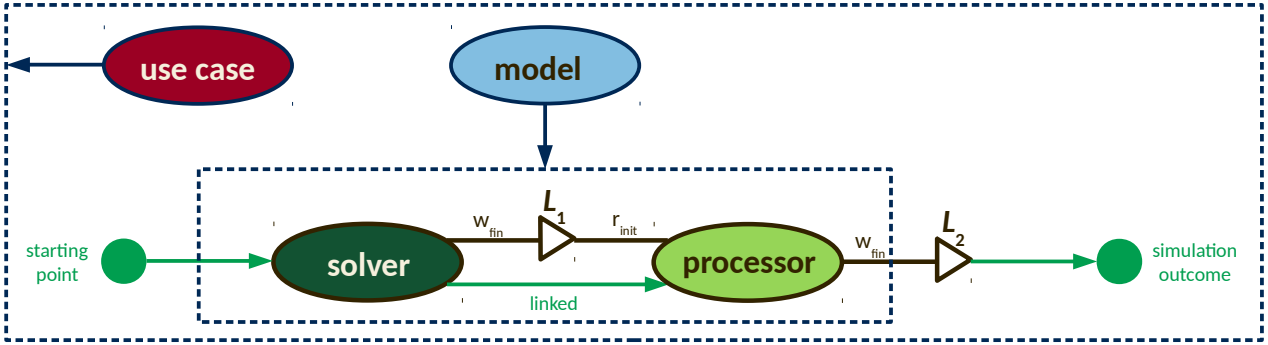
OSMO: Characterization of workflow semantics by logical data transfer (LDT) graphs



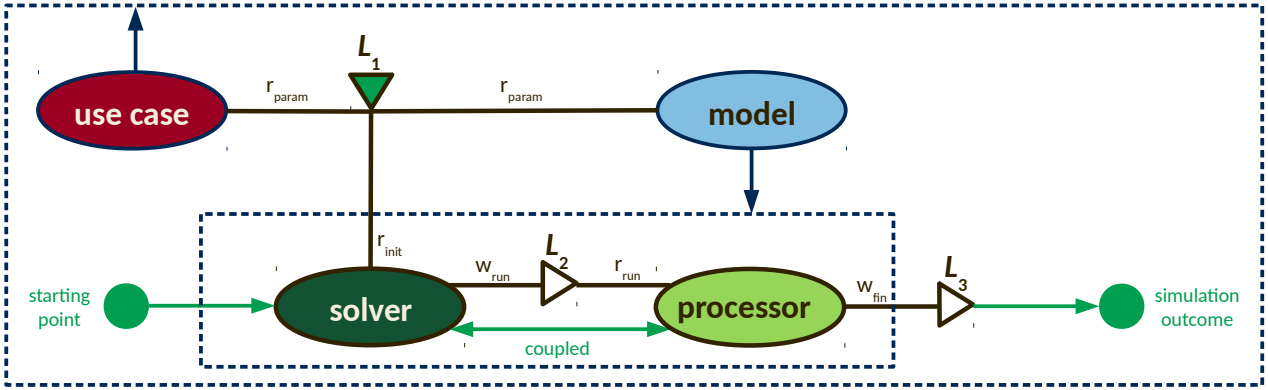
Simulation data provenance

MODA

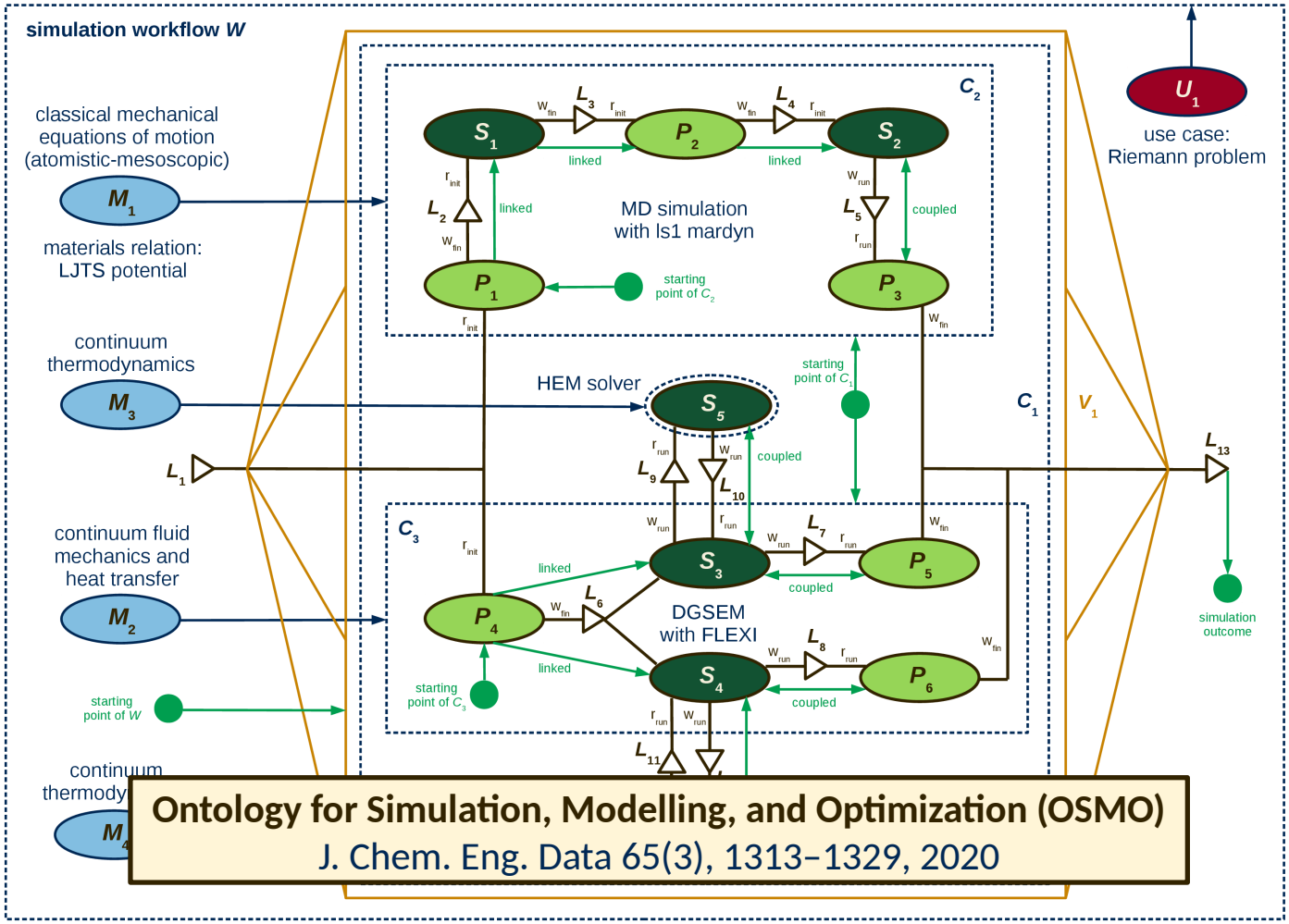
LDT case 1



LDT case 2



Simulation data provenance



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.

Annotating molecular models

Geometry

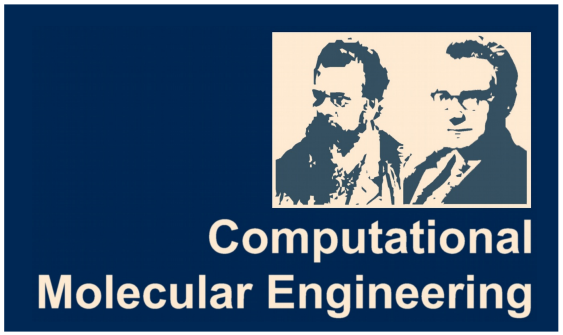
Types and positions
of interaction sites

Dispersion and repulsion

Lennard-Jones or Mie potential:
Size and energy parameters

Electrostatics

Point charge or multipole
(point dipole or quadrupole):
Magnitude and orientation

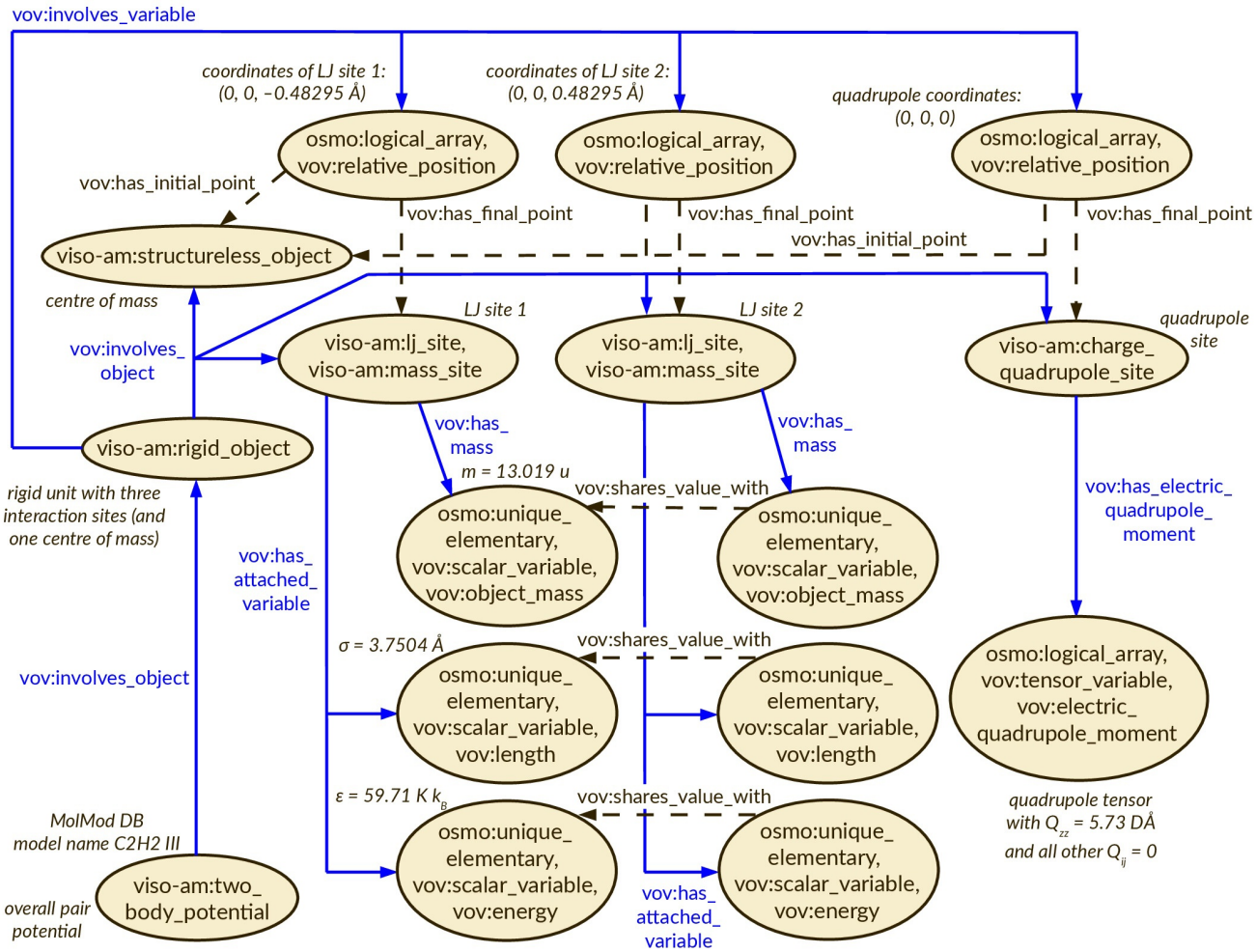


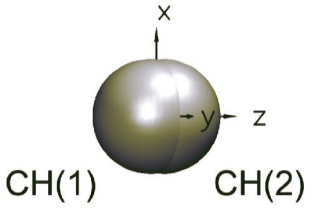
Molecular model database (MolMod DB)

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

Pair potentials for over 150 molecular fluids

Annotating molecular models



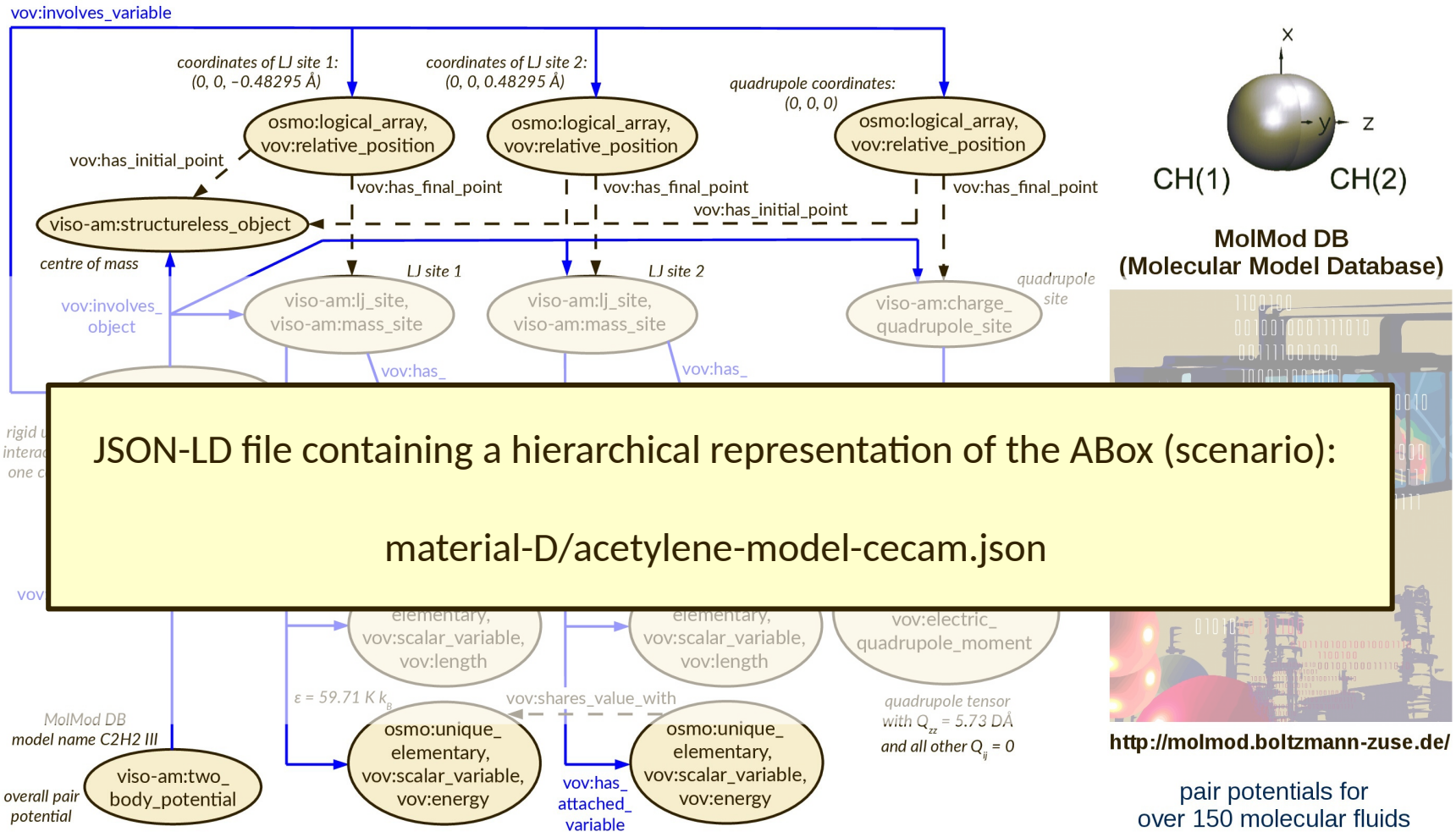


MolMod DB (Molecular Model Database)

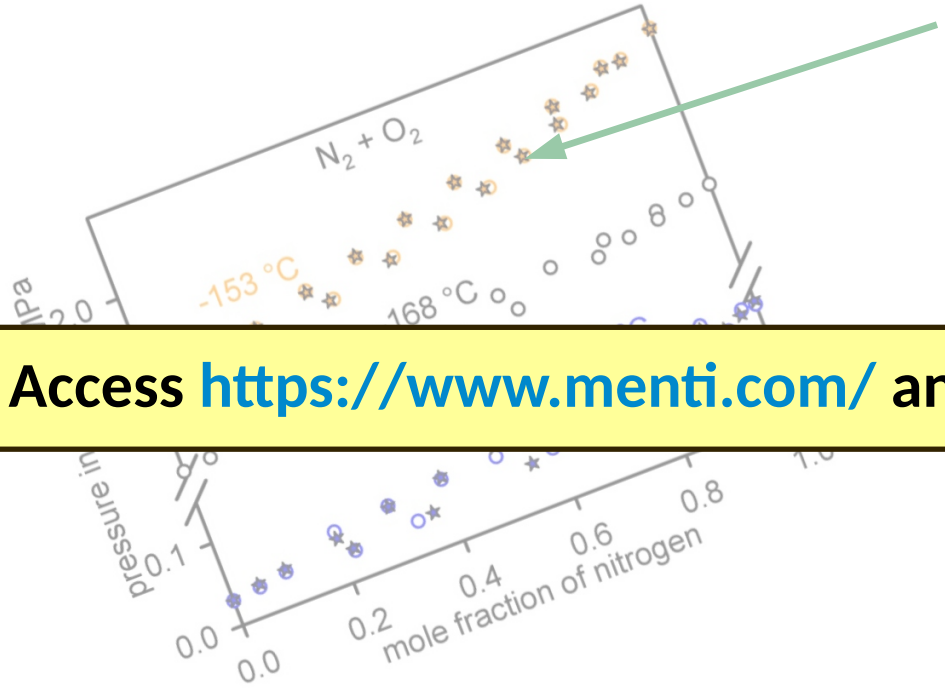
http://molmod.boltzmann-zuse.de/

pair potentials for over 150 molecular fluids

Annotating molecular models



Practical use of research data infrastructures



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

How was the data point obtained?

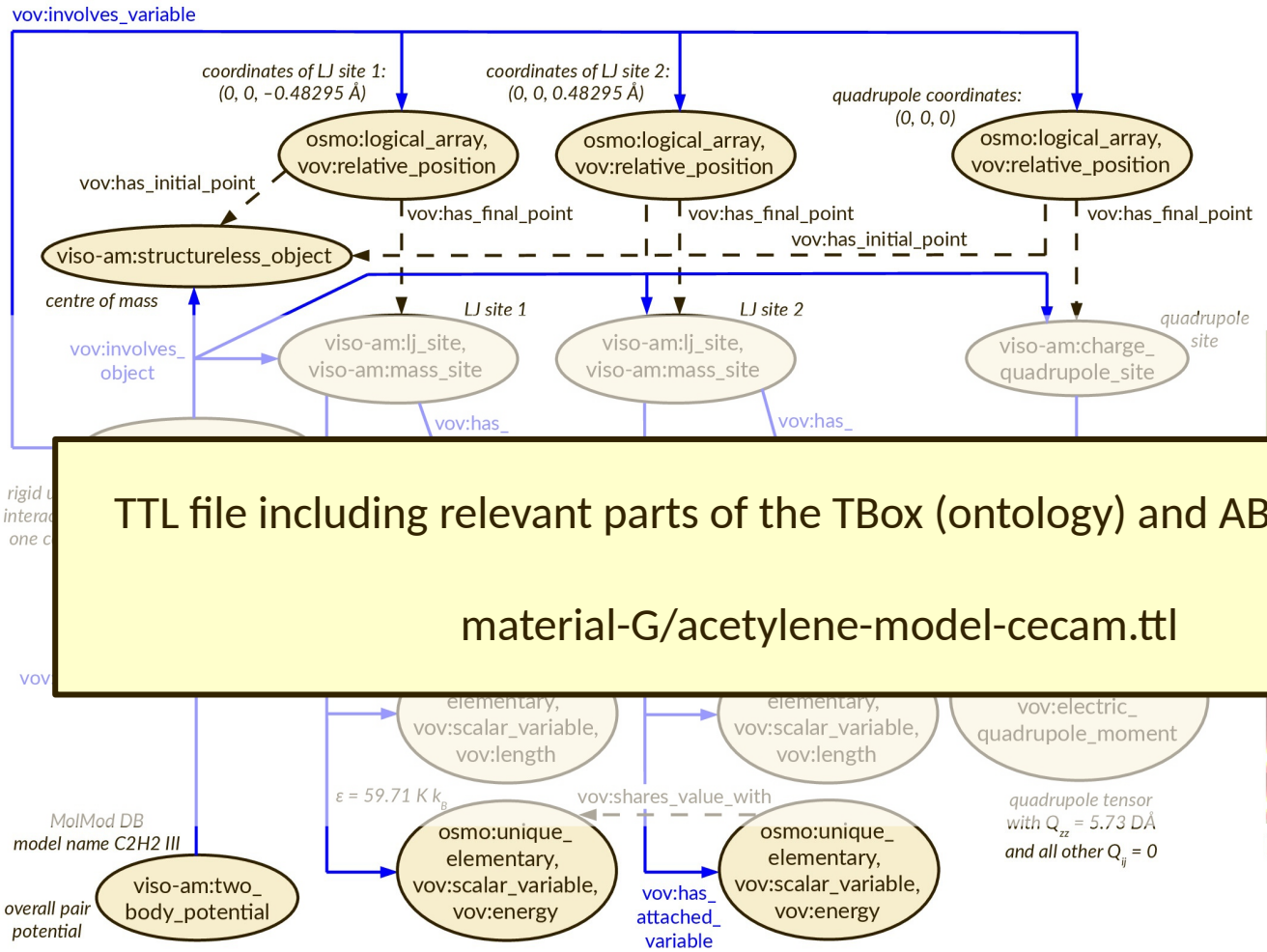
What is the margin of error, how was the error defined, and what software

Access https://www.menti.com/ and type the code 8878 5997.

competency questions

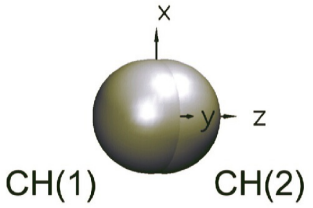
„What questions would you ask about data provenance when looking for simulation results in a research data infrastructure?“

Tools for working with ontologies

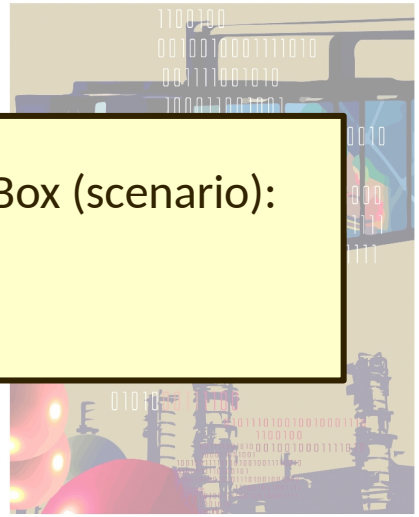


TTL file including relevant parts of the TBox (ontology) and ABox (scenario):

material-G/acetylene-model-cecam.ttl



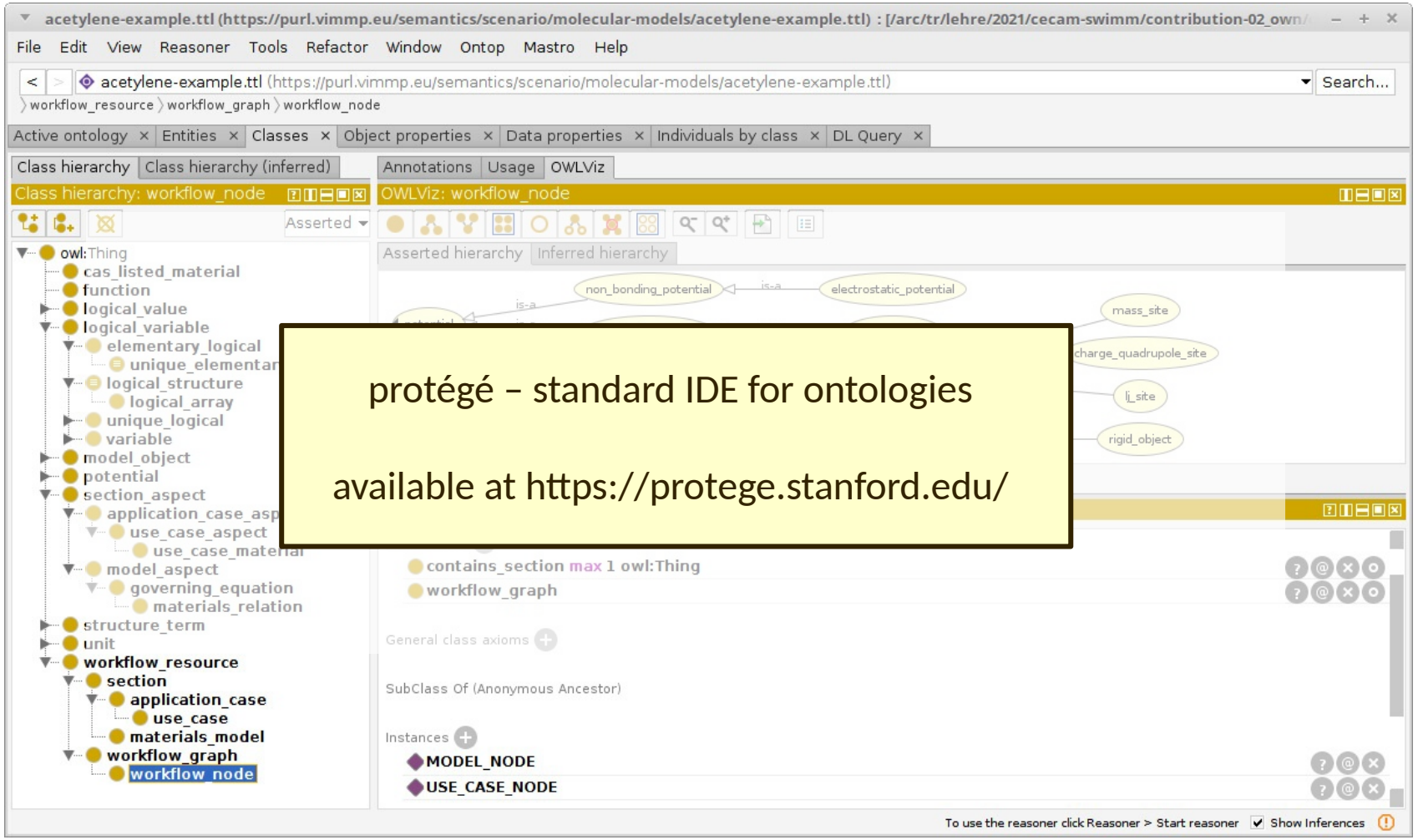
MolMod DB (Molecular Model Database)



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

pair potentials for over 150 molecular fluids

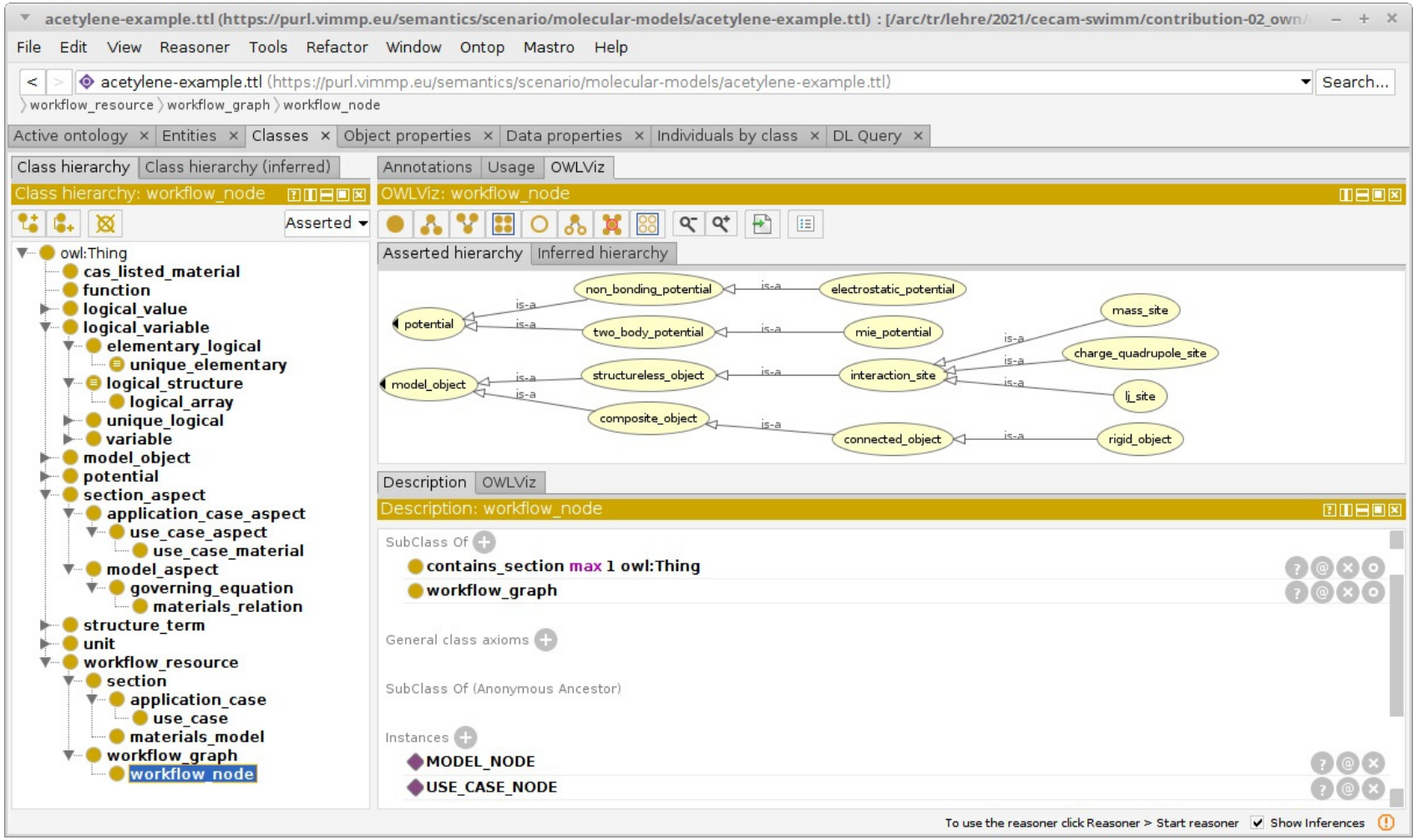
Tools for working with ontologies



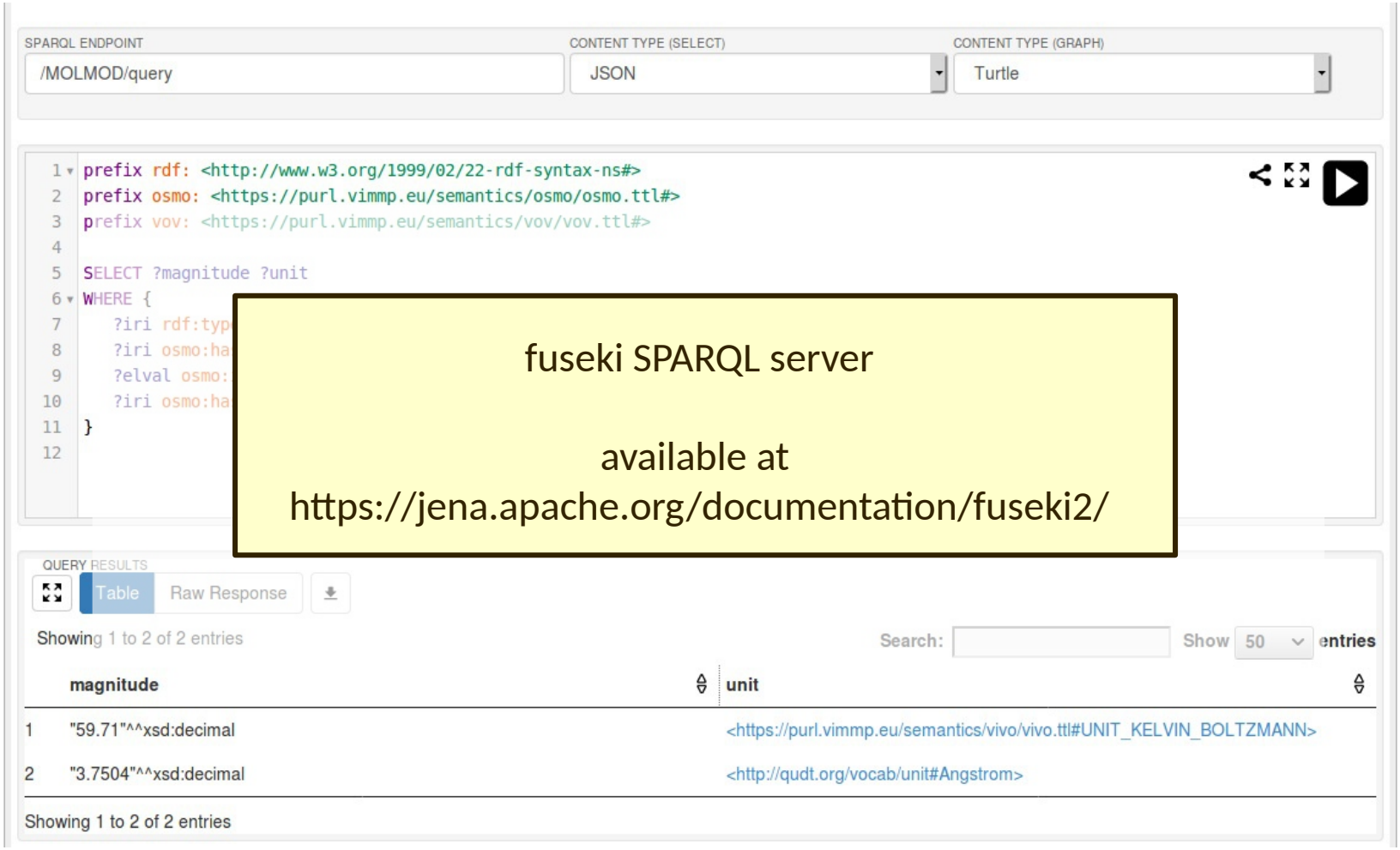
protégé – standard IDE for ontologies

available at https://protege.stanford.edu/

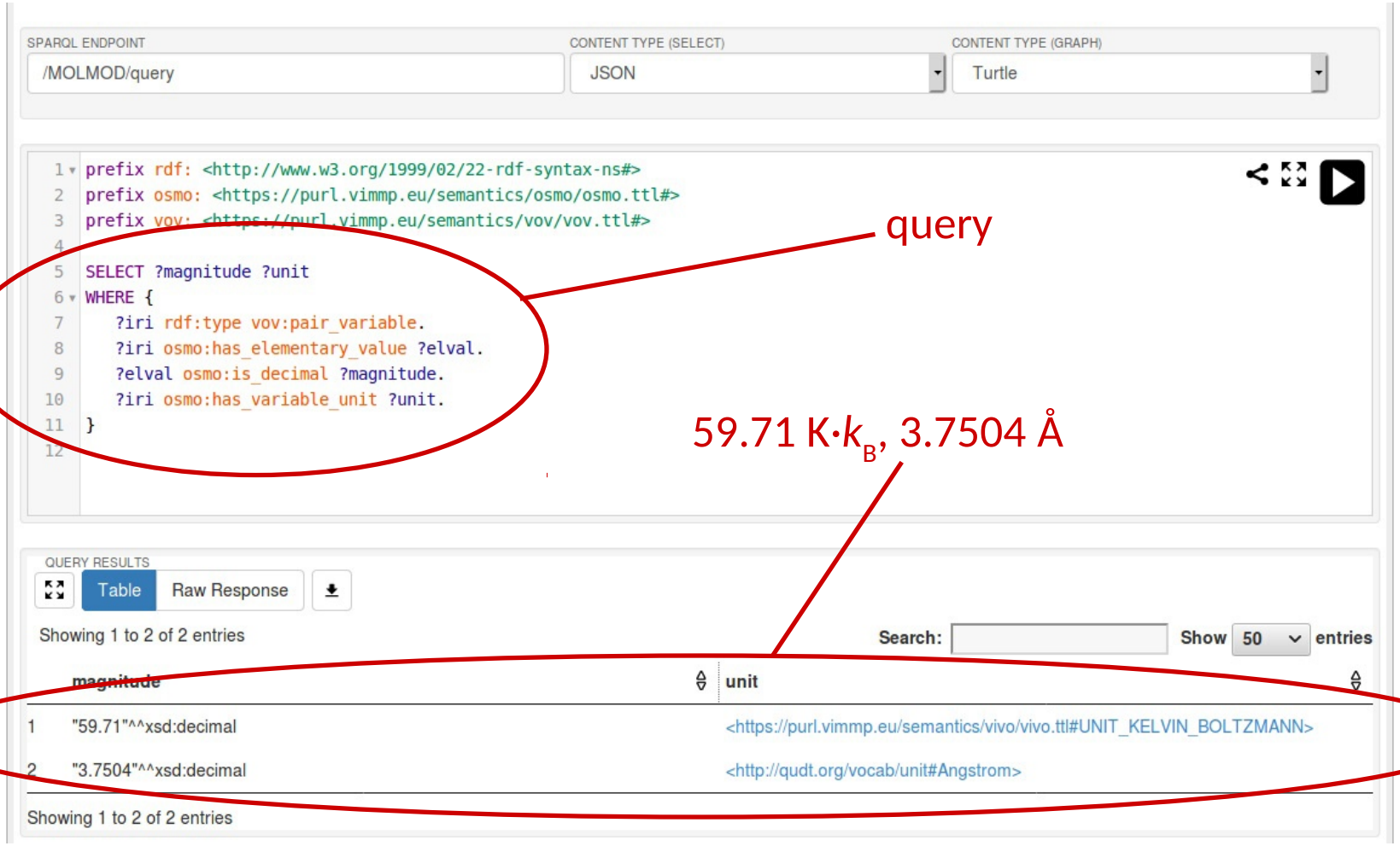
Tools for working with ontologies



Tools for working with ontologies



Tools for working with ontologies



Query languages

History: SQL (Structured Query Language) for relational databases

model_parameters

id	name	value	unit	category	reference
1	“sigma”	3.7504	Å	pair_variable	
2	“epsilon”	59.71	$K \cdot k_B$	pair_variable	
3	“site mass”	13.019	*u*	object_variable	
4	“quadrupole”		DÅ	tensor	276
5	“site 1 position”		Å	vector	178
6	“site 2 position”		Å	vector	179

category to be selected

pointers to IDs from other tables

Query languages

History: SQL (Structured Query Language) for relational databases

model_parameters

id	name	value	unit	category	reference
1	“sigma”	3.7504	Å	pair_variable	
2	“epsilon”	59.71	$K \cdot k_B$	pair_variable	
3	“site mass”	13.019	*u*	object_variable	
4	“quadrupole”		DÅ	tensor	276
5	“site 1 position”		Å	vector	178
6	“site 2 position”		Å	vector	179

SELECT name, value, unit
FROM model_parameters
WHERE category = pair_variable;

→

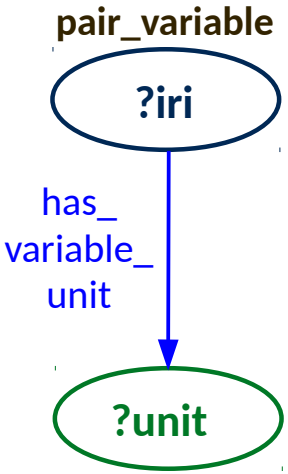
„sigma“	3.7504	Å
„epsilon“	59.71	$K \cdot k_B$

Query languages

SPARQL („SPARQL Protocol and RDF Query Language“) for the semantic web

SELECT ?x ?y ... WHERE {sequence of triples involving ?x, ?y, ...}

```
SELECT ?unit
WHERE {
  ?iri rdf:type vov:pair_variable.
  ?iri osmo:has_variable_unit ?unit.
}
```

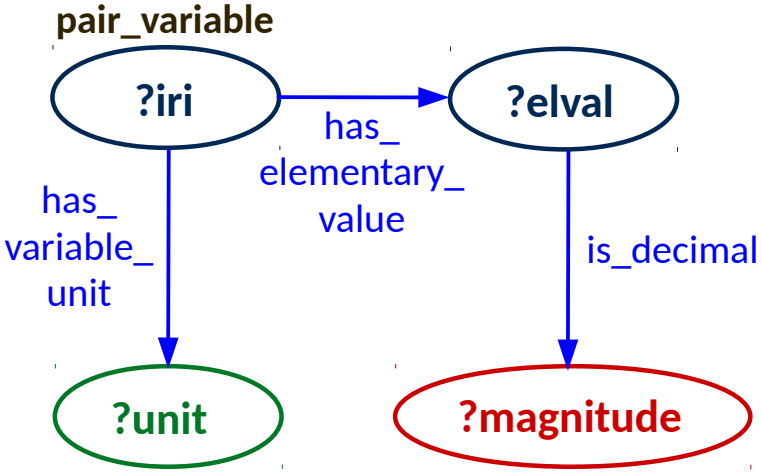


Query languages

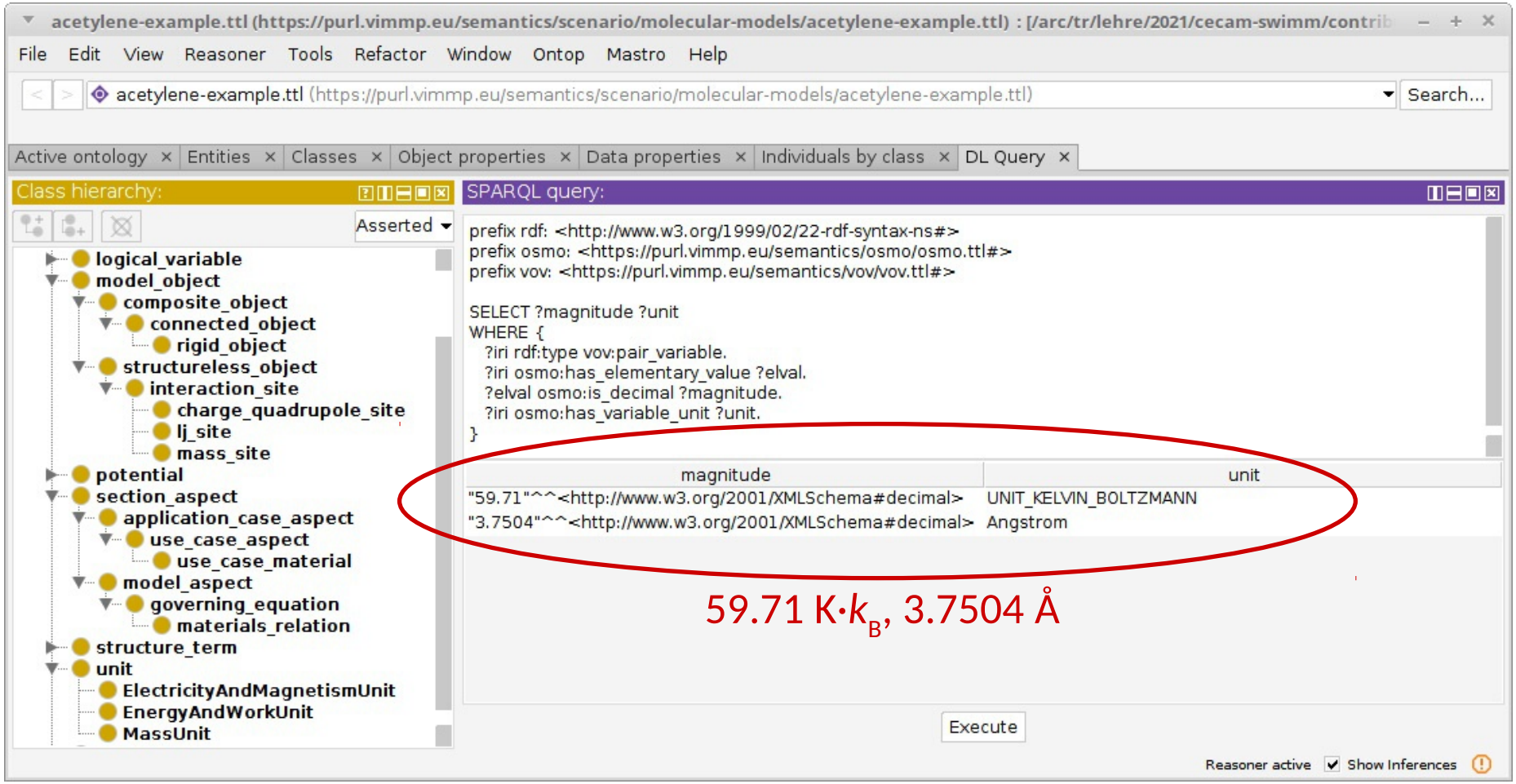
SPARQL („SPARQL Protocol and RDF Query Language“) for the semantic web

SELECT ?x ?y ... WHERE {sequence of triples involving ?x, ?y, ...}

```
SELECT ?magnitude ?unit
WHERE {
  ?iri rdf:type vov:pair_variable.
  ?iri osmo:has_elementary_value ?elval.
  ?elval osmo:is_decimal ?magnitude.
  ?iri osmo:has_variable_unit ?unit.
}
```



Query languages (SPARQL in protégé)

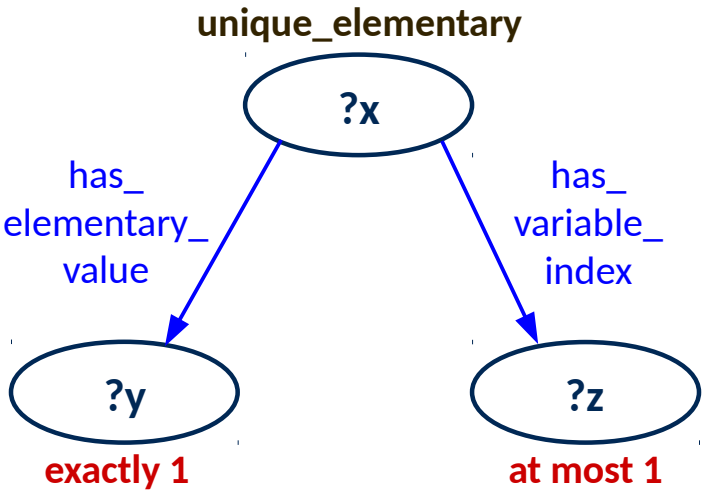


59.71 K·k_B, 3.7504 Å

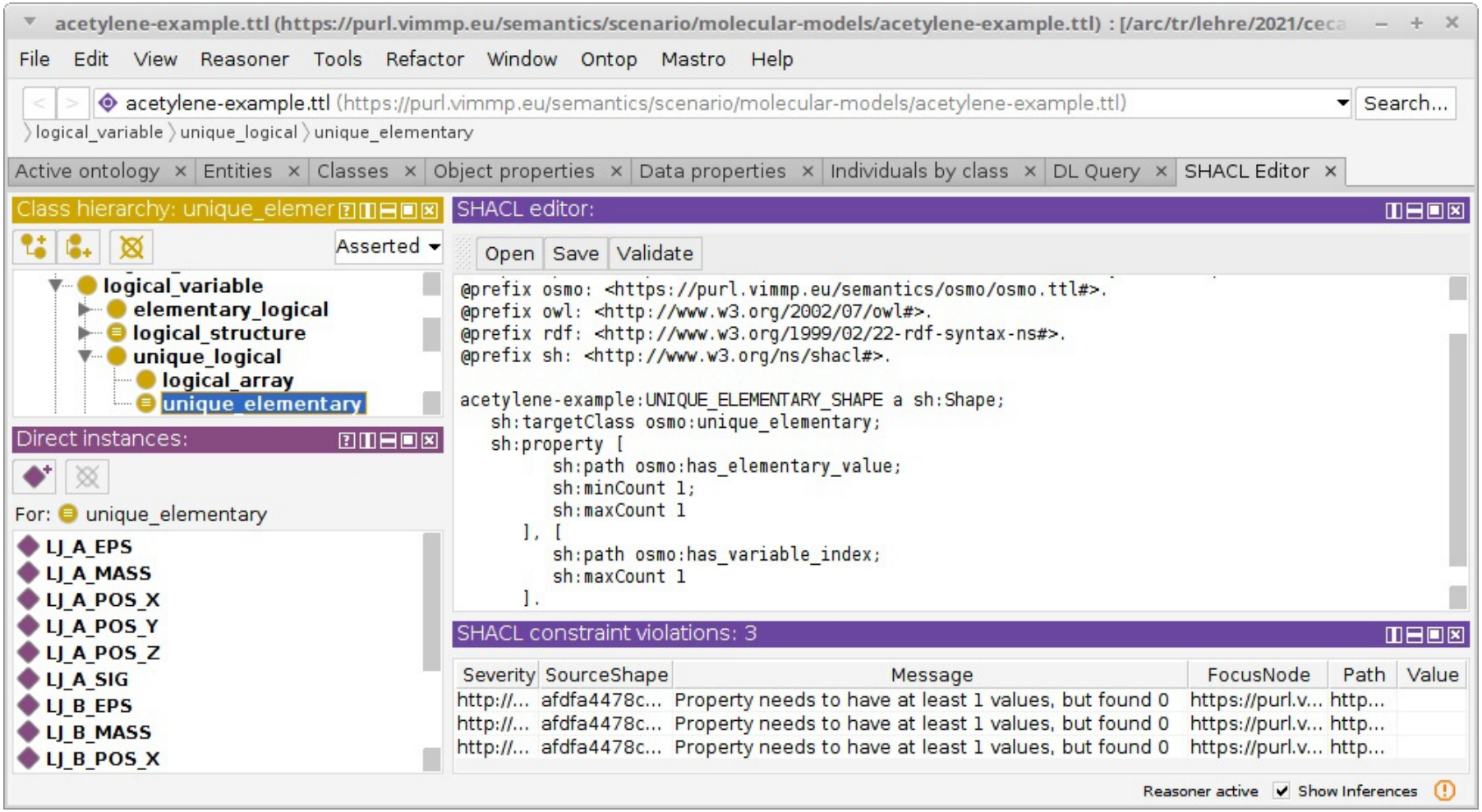
Constraint language (SHACL)

SHACL (Shapes Constraint Language) for validating knowledge graphs

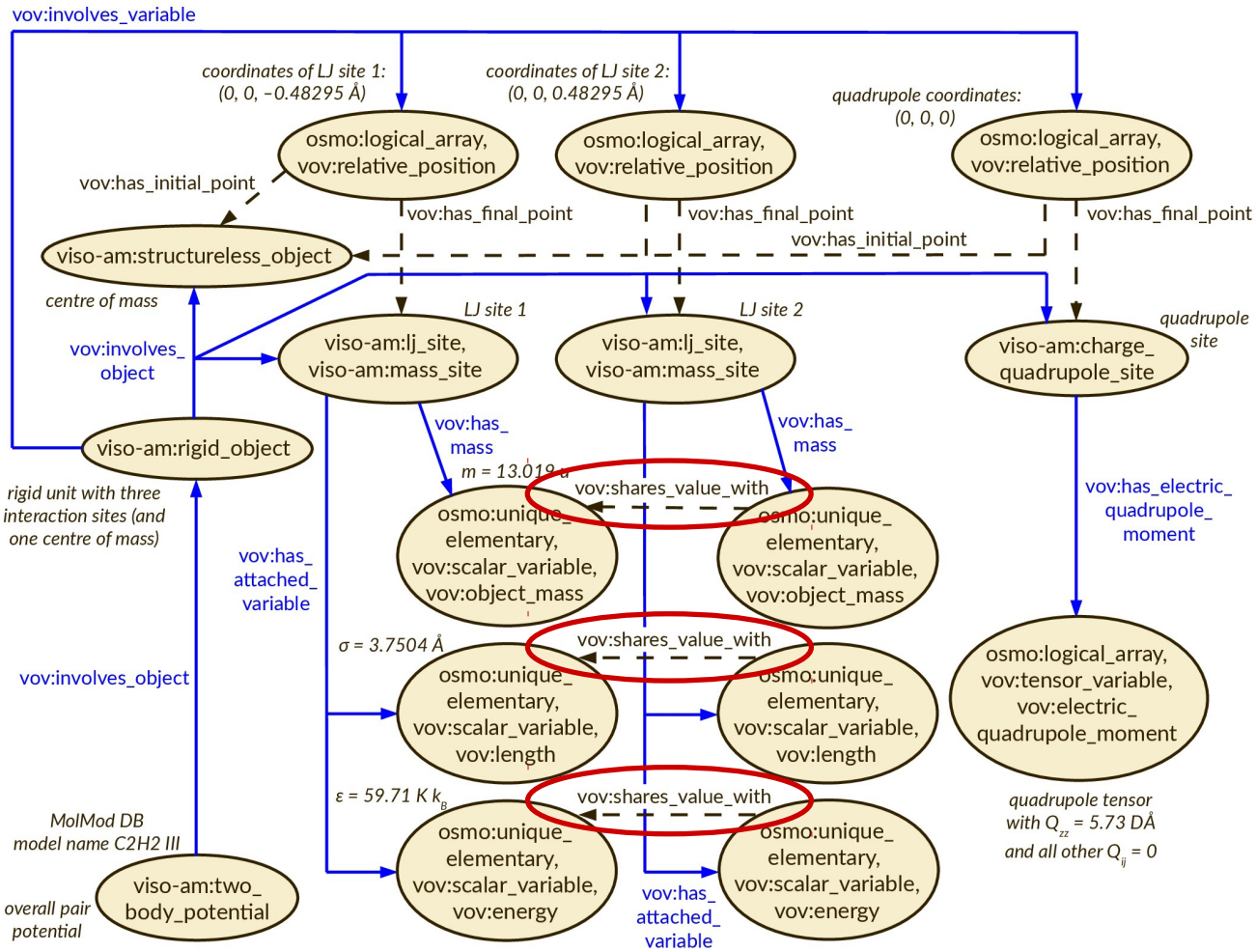
```
:UNIQUE_ELEMENTARY_SHAPE a sh:Shape;
  sh:targetClass osmo:unique_elementary;
  sh:property [
      sh:path osmo:has_elementary_value;
      sh:minCount 1;
      sh:maxCount 1
    ], [
      sh:path osmo:has_variable_index;
      sh:maxCount 1
    ].
```

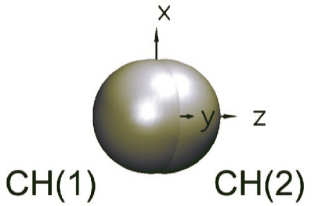


Constraint language (SHACL in protégé)



Constraint language (SHACL)





MolMod DB (Molecular Model Database)

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

pair potentials for over 150 molecular fluids

RDF and OWL libraries

Jena (Java)

https://jena.apache.org/

Developer: Apache Software Foundation

Functionality: TTL and RDF/XML I/O, SPARQL querying (used by fuseki), reasoning („Inference API“)

License: Apache Software License

owlready2 (Python)

https://pypi.org/project/Owlready2/

Functionality: manipulate OWL 2.0 ontologies as Python objects; reasoning; Ntriples and RDF/XML I/O, OWL/XML as I, Reasoning, RDF quadstore

Developer: Jean-Baptiste Lamy

License: GNU Lesser General Public License v3 or later (LGPLv3+)

RDF and OWL libraries: Apache Jena

example in „material-I“ folder

OSMO-only knowledge graph

VOV-only knowledge graph

```
import org.apache.jena.rdf.model.*;

modelOSMO.read(inputStreamOSMO, null, "TTL");
modelVOV.read(inputStreamVOV, null, "TTL");

Model model = modelOSMO.union(modelVOV);
```

combined knowledge graph

```
model.write(out, "TTL");
```

RDF and OWL libraries: Owlready2

example in „material-L“ folder

python3 owlready2_example.py

A toy example to

- **load** an ontology
- **navigate** it (classes and properties)
- **modify it** (create an individual)
- **save** the modified ontology **in a file**
- .. all *via* Python commands

```
myonto = get_ontology("vimmp-ontology-fragment.owl")
myonto.load()

classes_list=list(myonto.classes())
op_list=list(myonto.object_properties())

my_ind = (my_class)("my_individual", label="my_label")

myonto.save(file="my_first_test.owl", format = "rdfxml")
```

Time for coffee

A sign and the object it (supposedly) refers to

The following (only) refers to the contribution to this presentation from the VIMMP project:

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