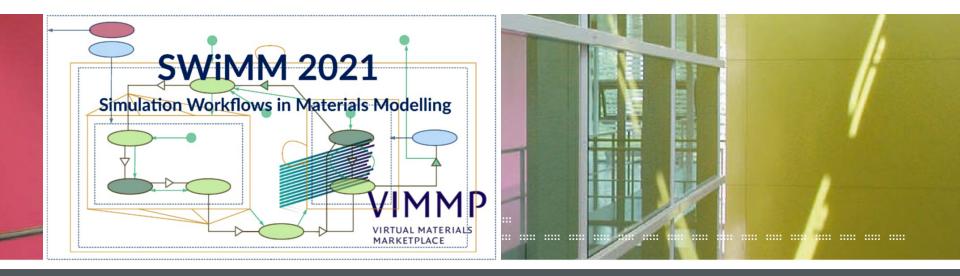


Semantic interoperability and ontology-driven technology for simulation workflows

Silvia Chiacchiera and Martin Thomas Horsch



:: S. Chiacchiera, M. T. Horsch ::





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?]

:::::



S. Chiacchiera, M. T. Horsch 💠

16th March 2021

•••







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.



S. Chiacchiera, M. T. Horsch

16th March 2021

•••







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).







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).

1) Syntax/syntactics: formal relations between signs

2) Semantics: meaning, relations between signs and what they refer to

3) Pragmatics: relation between signs and their users/enviroment

# Mode	el 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
# Mode		
# A	sigma	sigma_err
40.0	1.25022	0.1238
60.0	2.75247	0.2723
	4.05209	0.2691
100.0	4.05401	0.2726

The file format (ASCII text file).

Info about the content: e.g., what each column and block means, the data provenance, etc.

E.g.: we type "rm sigma.dat" in a terminal. Depending on our rights on the file, it will be removed or not.

16th March 2021

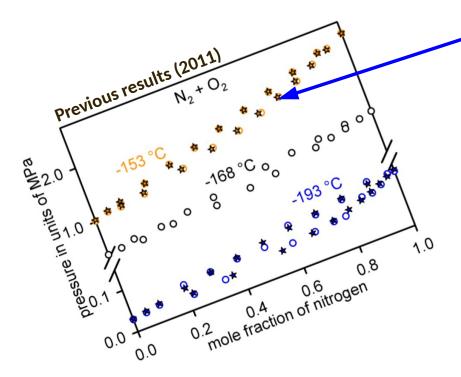
•••







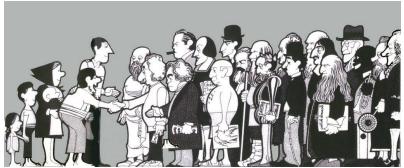
Why do we need semantic technology?



• 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?

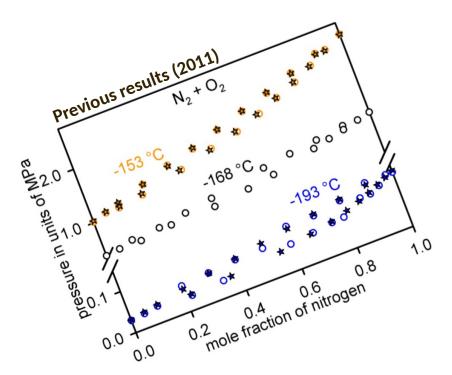


S. Chiacchiera, M. T. Horsch

16th March 2021



Why do we need semantic technology?



Good practice in managing research data:

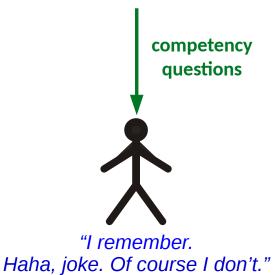
Make all data findable, accessible, interoperable, and reusable (FAIR).

What values did x and p have?

How was the data point obtained?

Н

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



•••

S. Chiacchiera, M. T. Horsch 💠

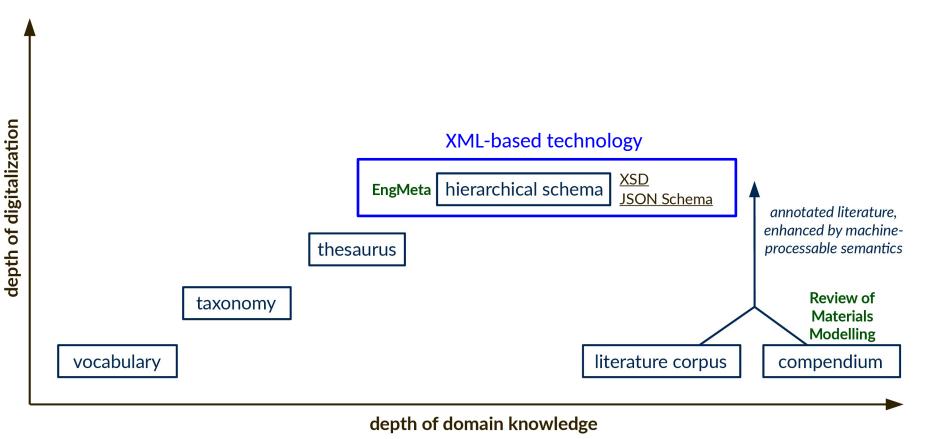
16th March 2021

S



Metadata standardization

Hierarchy of semantic assets

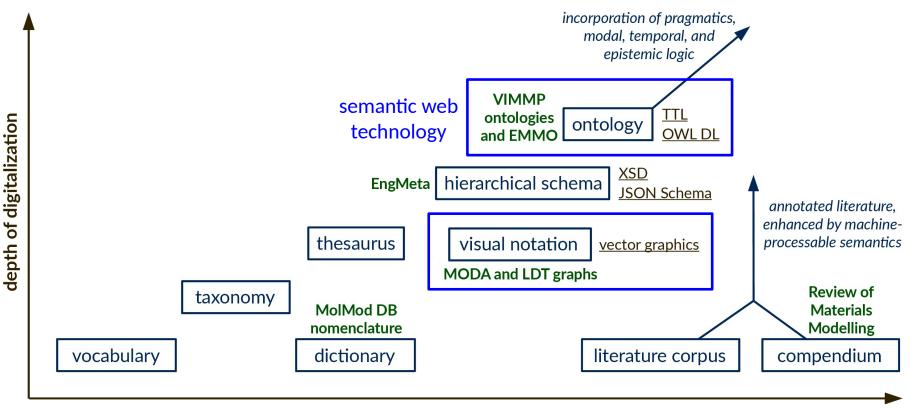


16th March 2021



Metadata standardization

Hierarchy of semantic assets



depth of domain knowledge

S. Chiacchiera, M. T. Horsch 💠





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 has_property other_property class_of_subject; first_object, second_object; 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.

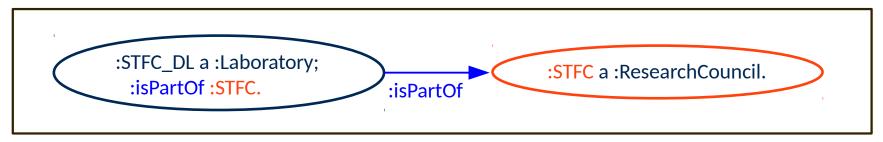
osmo:has_resource a owl:ObjectProperty;

rdfs:domain osmo:logical_access;

rdfs:range osmo:logical_resource.



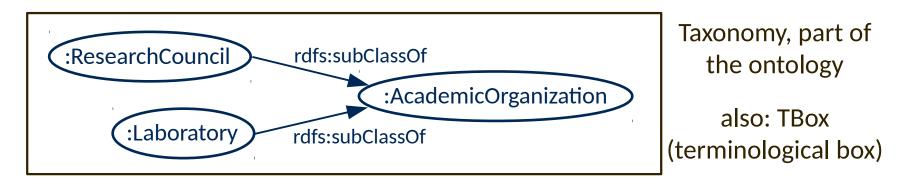
Ontologies (in a nutshell)



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

:STFC_DL :isPartOf :STFC.

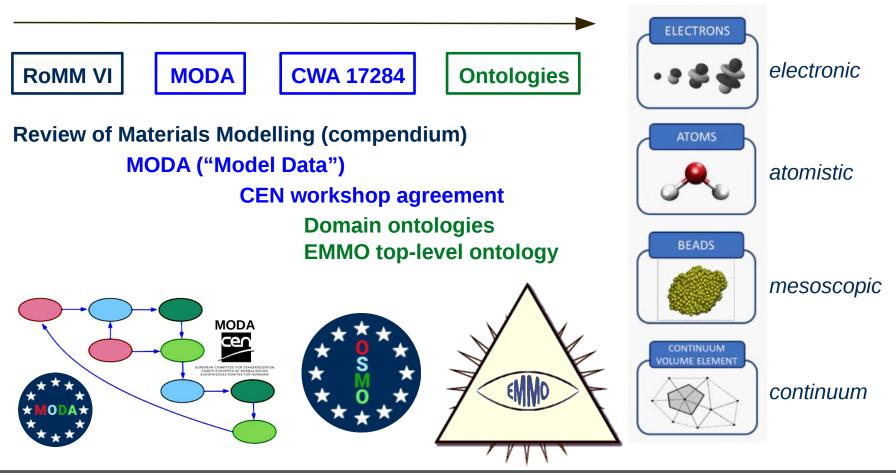
RDF triple, consisting of subject, predicate, and object





Ontology development in materials modelling

Community-governed development of metadata standards



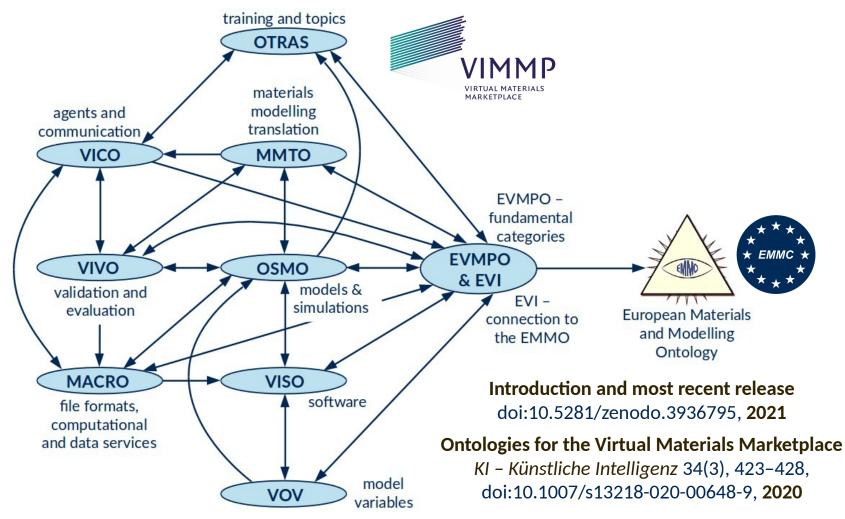
S. Chiacchiera, M. T. Horsch

16th March 2021

•••



Ontology development in materials modelling



S. Chiacchiera, M. T. Horsch 💠

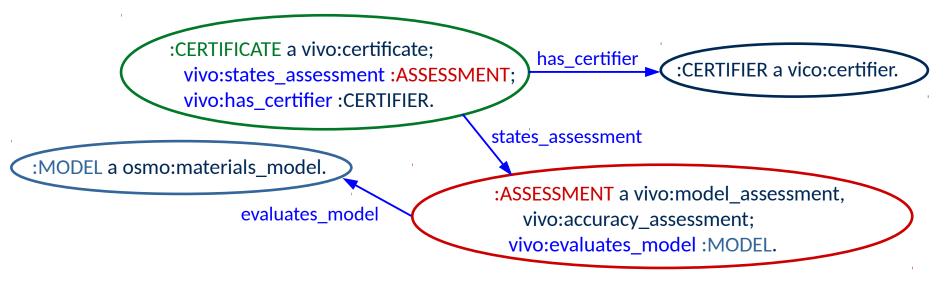
16th March 2021







Knowledge graphs



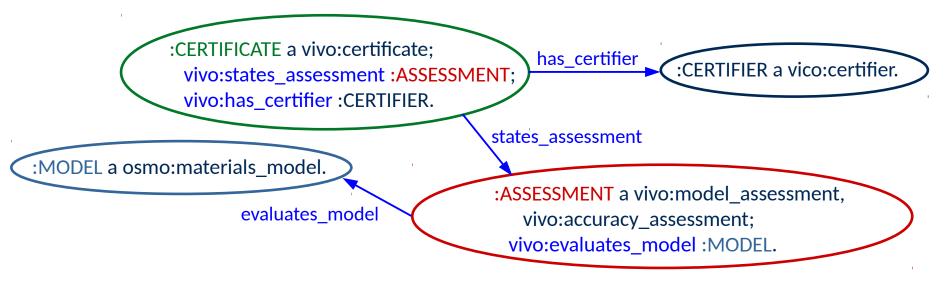
The certifier :CERTIFIER has issued a certificate (the IRI of which is :CERTIFICATE) stating a model accuracy assessment (the IRI of which is :ASSESSMENT) that evaluates the materials model :MODEL.







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.

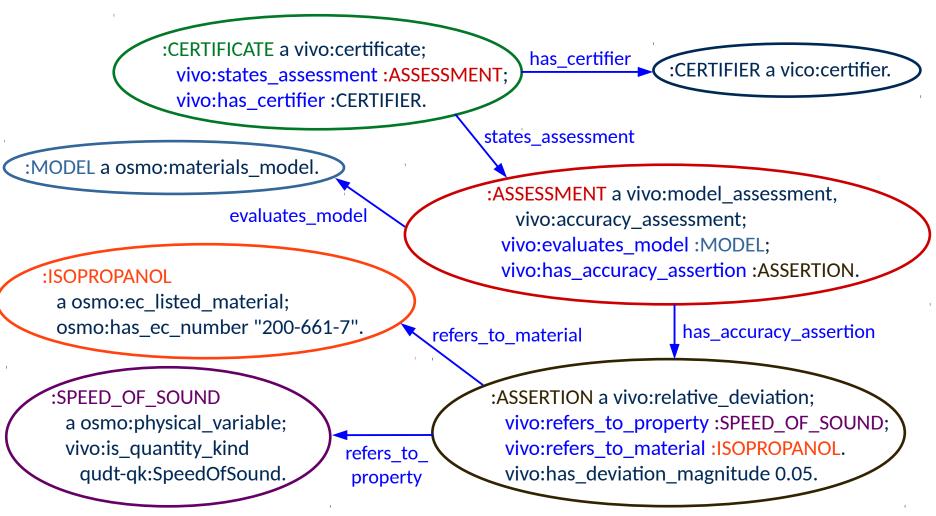
In the model accuracy assessment :ASSESSMENT, it is asserted that the materials model :MODEL has a relative error of 5% for the speed of sound of isopropanol.







Knowledge graphs



S. Chiacchiera, M. T. Horsch

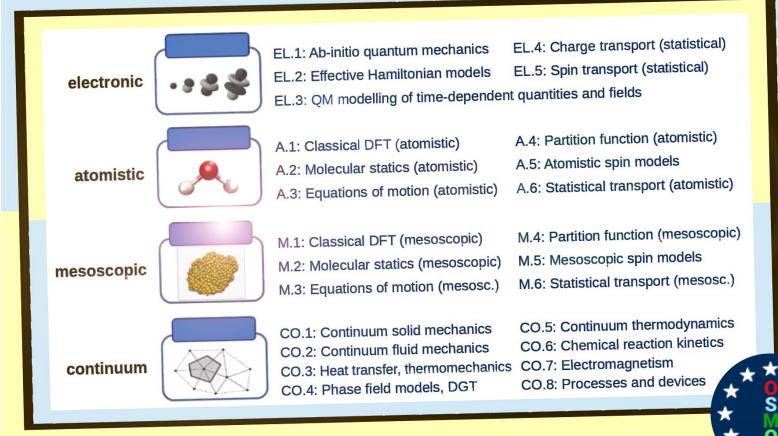
16th March 2021







Simulation data provenance



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

S. Chiacchiera, M. T. Horsch

16th March 2021

•••





:::::

:::::



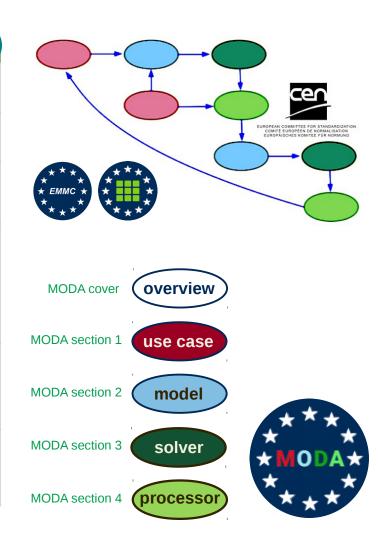
Simulation data provenance

:::::

:::::

:::::

3	SOLVER AND COMPUT	ATIONAL TRANSLATIC			
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		ise give the time step used in the solving operations. Ical time step and this is not the same as the time lapse of the case to 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	SpecificCut-offs	re internal numerical solver details (if applicable), like tolerances c, convergence criteria tor options		



:: S. Chiacchiera, M. T. Horsch ::

16th March 2021





.

....

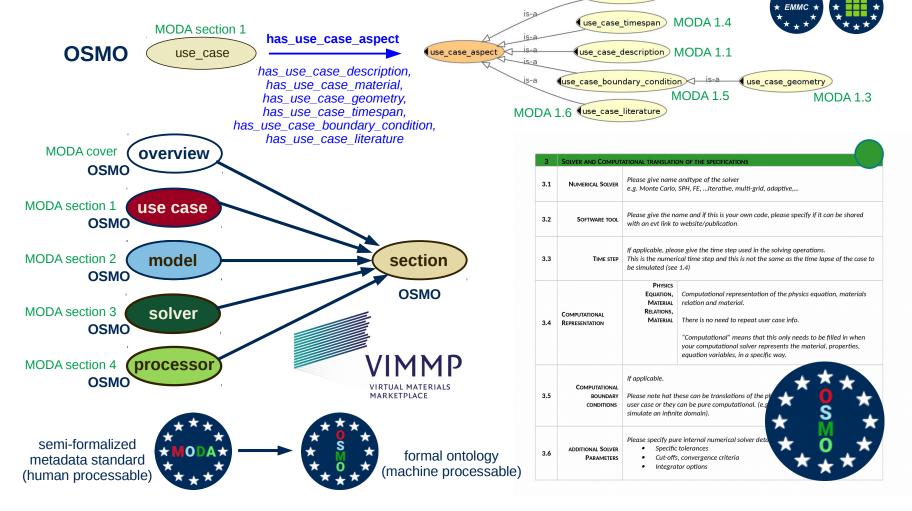
::::

use case material

MODA 1.2

HLRS

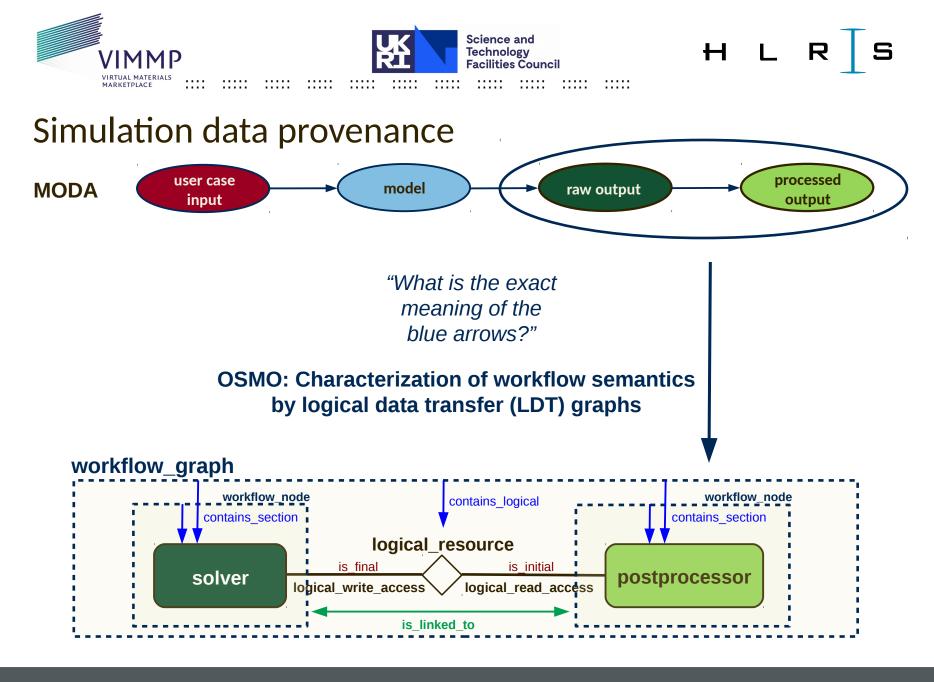
Simulation data provenance



S. Chiacchiera, M. T. Horsch 💠

16th March 2021

•••

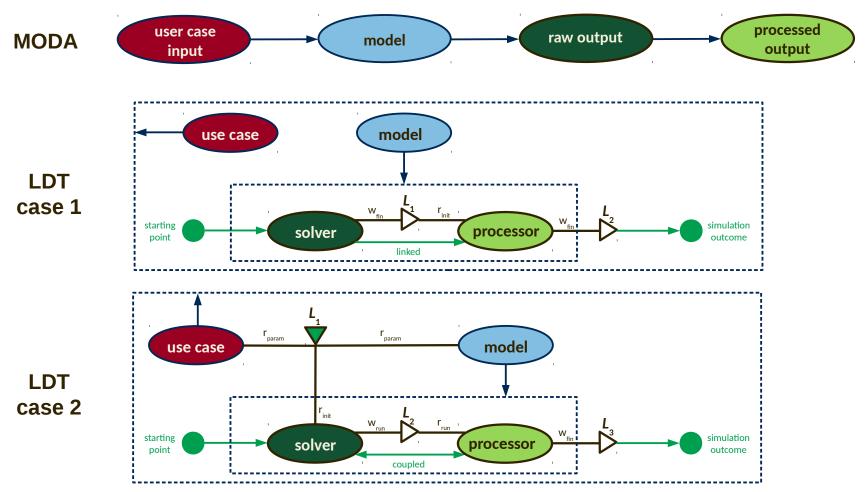


S. Chiacchiera, M. T. Horsch 💠

16th March 2021



Simulation data provenance

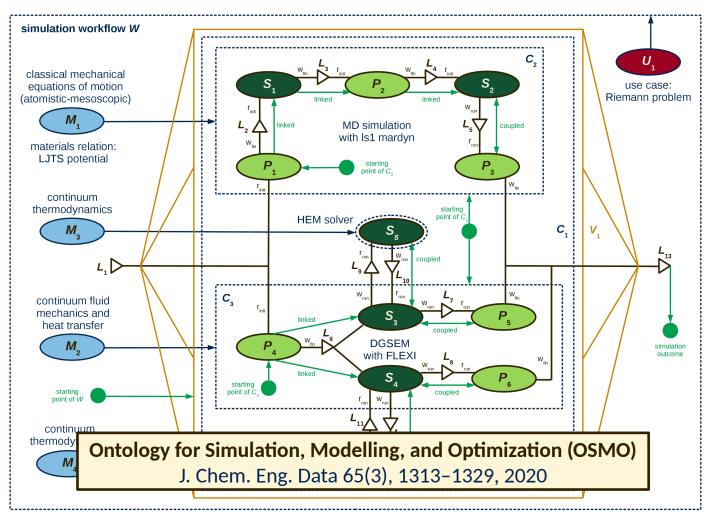


S. Chiacchiera, M. T. Horsch 💠



H L R S

Simulation data provenance



OSMO-based provenance description as an extension of the MODA workflow metadata standard:

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



:: S. Chiacchiera, M. T. Horsch ::







Annotating molecular models

Geometry

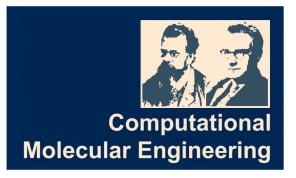
Types and positions of interaction sites

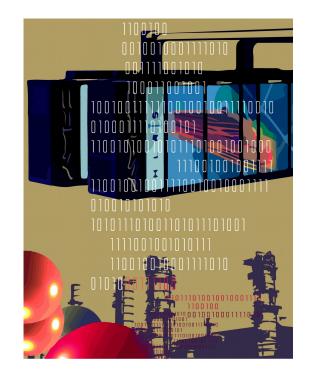
Dispersion and repulsion

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

•••

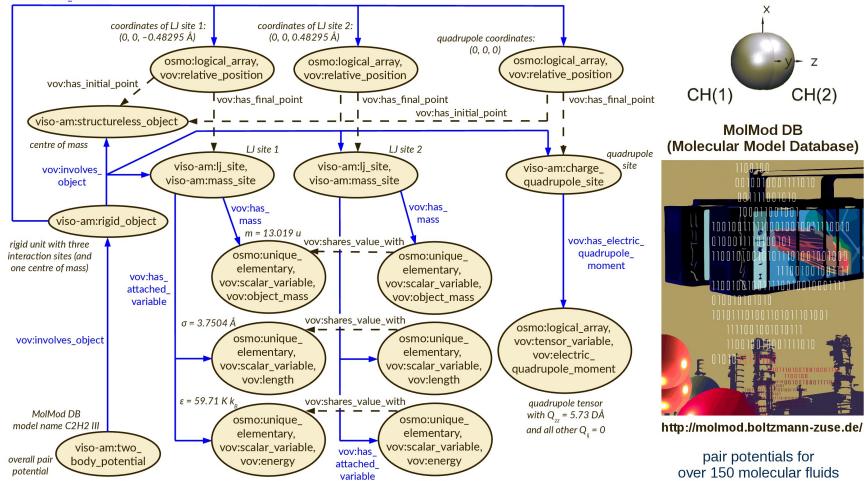
S. Chiacchiera, M. T. Horsch





Annotating molecular models

vov:involves_variable





Annotating molecular models

vov:involves_variable coordinates of LJ site 1: coordinates of LJ site 2: (0, 0, -0.48295 Å) (0, 0, 0.48295 Å) quadrupole coordinates: (0, 0, 0)osmo:logical_array, osmo:logical_array, osmo:logical array, vov:relative_position vov:relative_position vov:relative_position vov:has_initial_point CH(2) CH(1) vov:has_final_point vov:has_final_point vov:has_final_point vov:has initial point viso-am:structureless object MolMod DB centre of mass (Molecular Model Database) V LJ site 1 LJ site 2 quadrupole viso-am:lj site, viso-am:lj site, viso-am:charge vov:involves viso-am:mass site viso-am:mass_site quadrupole site object vov:has vov:has rigid JSON-LD file containing a hierarchical representation of the ABox (scenario): one material-D/acetylene-model-cecam.json VO elementary, elementary, vov:electric_ vov:scalar_variable, vov:scalar variable, quadrupole_moment vov:length vov:length ε = 59.71 K k vov:shares_value_with quadrupole tensor MolMod DB with Q₇₇ = 5.73 DÅ osmo:unique_ osmo:unique model name C2H2 III http://molmod.boltzmann-zuse.de/ and all other $Q_{\mu} = 0$ elementary, elementary, vov:scalar_variable, viso-am:two vov:scalar_variable, vov:has pair potentials for overall pair body potential vov:energy vov:energy attached over 150 molecular fluids potential variable

S. Chiacchiera, M. T. Horsch

16th March 2021

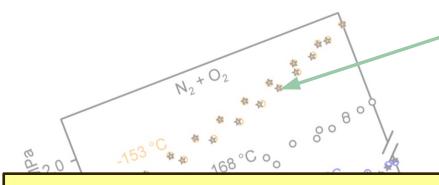
••







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

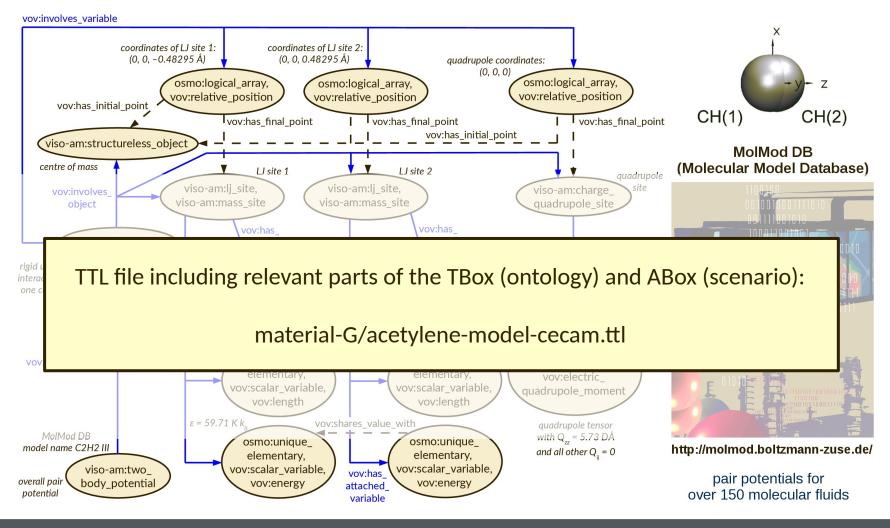
Access https://www.menti.com/ and type the code 8878 5997.



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



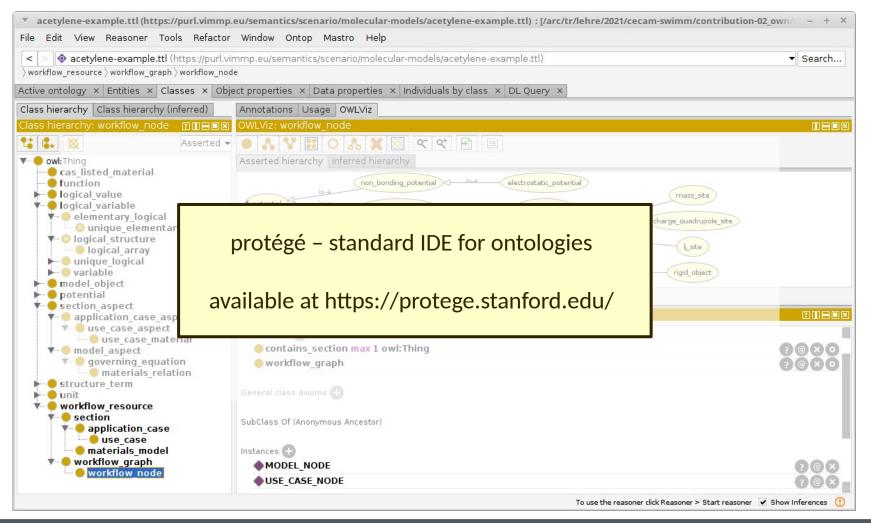
Tools for working with ontologies



S. Chiacchiera, M. T. Horsch



Tools for working with ontologies



S. Chiacchiera, M. T. Horsch



HLRS

Tools for working with ontologies

▼ acetylene-example.ttl (https://purl.vimmp. File Edit View Reasoner Tools Refactor	eu/semantics/scenario/molecular-models/acetylene-example.ttl) : [/arc/tr/lehre/2021/cecam-swimm/contribution-02_ Window Ontop Mastro Help	own/ – + ×
Acetylene-example.ttl (https://purl.vir) workflow_resource) workflow_graph) workflow_node	nmp.eu/semantics/scenario/molecular-models/acetylene-example.ttl) e	▼ Search
Active ontology × Entities × Classes × Obje	ect properties 🗙 Data properties 🗙 Individuals by class 🗙 DL Query 🗙	
Class hierarchy Class hierarchy (inferred)	Annotations Usage OWLViz	
Class hierarchy: workflow_node	OWLViz: workflow_node	
🐮 🔩 🐹 Asserted 🗸		
 owl:Thing cas_listed_material function logical_value logical_variable elementary_logical gical_structure logical_array variable variable svariable svariable section aspect 	Asserted hierarchy Inferred hierarchy non_bonding_potential is-a electrostatic_potential mass_site potential is-a two_body_potential is-a mie_potential is-a charge_quadrupole_site is-a two_body_potential is-a mie_potential is-a charge_quadrupole_site is-a tructureless_object is-a interaction_site is-a is-a is-a is-a is-a is-a is-a is-a	
<pre></pre>	Description: workflow_node SubClass Of + contains_section max 1 owl:Thing	
 governing_equation materials_relation structure_term unit workflow_resource section governing_equation governing_equation 	• workflow_graph General class axioms 🕂 SubClass Of (Anonymous Ancestor)	0000
 materials_model workflow_graph workflow_node 	Instances MODEL_NODE USE_CASE_NODE To use the reasoner click Reasoner > Start reasoner	

:: S. Chiacchiera, M. T. Horsch ::





Tools for working with ontologies

CONTENT TYPE (SELECT)	CONTENT TYPE (GRAPH)	
JSON	Turtle	•
l.vimmp.eu/semantics/osmo/osmo.ttl#>		< 13
fuseki SPARQL serve	r	
available at ps://jena.apache.org/documen	tation/fuseki2/	
	Search: Sho	w 50 v ent
∂ unit		
<https: purl.vimn<="" td=""><td>p.eu/semantics/vivo/vivo.ttl#UNIT_KELVIN_B</td><td>OLTZMANN></td></https:>	p.eu/semantics/vivo/vivo.ttl#UNIT_KELVIN_B	OLTZMANN>
	JSON 3.org/1999/02/22-rdf-syntax-ns#> 1.vimmp.eu/semantics/osmo/osmo.ttl#> .vimmp.eu/semantics/vov/vov.ttl#> fuseki SPARQL serve available at ps://jena.apache.org/document	JSON Turtle 3.org/1999/02/22-rdf-syntax-ns#> 1.vimmp.eu/semantics/osmo/osmo.ttl#> vimmp.eu/semantics/vov/vov.ttl#> fuseki SPARQL server available at ps://jena.apache.org/documentation/fuseki2/

:: S. Chiacchiera, M. T. Horsch ::

16th March 2021

••





:::::



Tools for working with ontologies

:::::

	CONTENT TYPE (SELECT)	CONTENT TYPE (GRAPH)	
/MOLMOD/query	JSON	Turtle	·
<pre>1 * prefix rdf: <http: *="" 10="" 11="" 12="" 199="" 2="" 3="" 4="" 5="" 6="" 7="" 8="" 9="" <https:="" ?elval="" ?iri="" ?magnitu="" ?magnitude="" ?unit="" ?ur="" osmo:="" osmo:has_elementary_value="" osmo:has_variable_unit="" osmo:is_decimal="" ouery="" pre="" prefix="" purl.vimmp.eu="" rdf:type="" results<="" select="" vov:="" vov:pair_variable="" where="" www.w3.org="" {="" }=""></http:></pre>	u/semantics/osmo/osmo.ttl#> /semantics/vov/vov.ttl#> ?elval. de. it.	– query k _B , 3.7504 Å	< 23
Table Raw Response 生			
Table Raw Response Showing 1 to 2 of 2 entries		Search:	Show 50 ∽ entries
	∂ unit	Search:	Show 50 ∨ entries ⇔
Showing 1 to 2 of 2 entries		Search:	ð

:: S. Chiacchiera, M. T. Horsch ::

16th March 2021

••





History: SQL (Structured Query Language) for relational databases category to

model_parameters

be selected

id	name	value	unit	category	reference
		Value		- category	
1	"sigma"	3.7504	Å	pair_variable	
2	"epsilon"	59.71	К·k _в	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

pointers to IDs from other tables





History: SQL (Structured Query Language) for relational databases

id	name	value	unit	category	reference
1	"sigma"	3.7504	Å	pair_variable	
2	"epsilon"	59.71	К· <i>k</i> _в	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

model_parameters

SELECT name, value, unit **FROM** model_parameters , sigma 3.7504 Å **WHERE** category = pair_variable; $V \cdot k_{B}$

S. Chiacchiera, M. T. Horsch 💠

16th March 2021

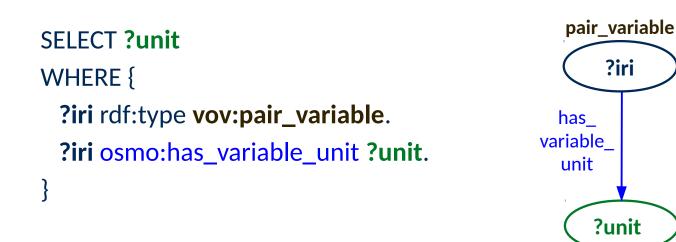
•••

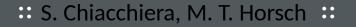




SPARQL ("SPARQL Protocol and RDF Query Language") for the semantic web

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









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

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

SELECT ?magnitude ?unit

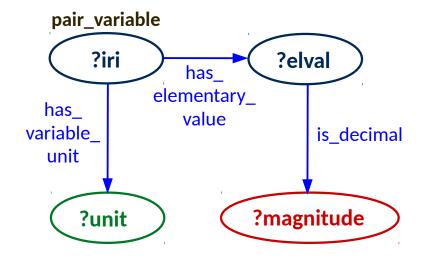
WHERE {

?iri rdf:type **vov:pair_variable**.

?iri osmo:has_elementary_value ?elval.

?elval osmo:is_decimal **?magnitude**.

?iri osmo:has_variable_unit ?unit.



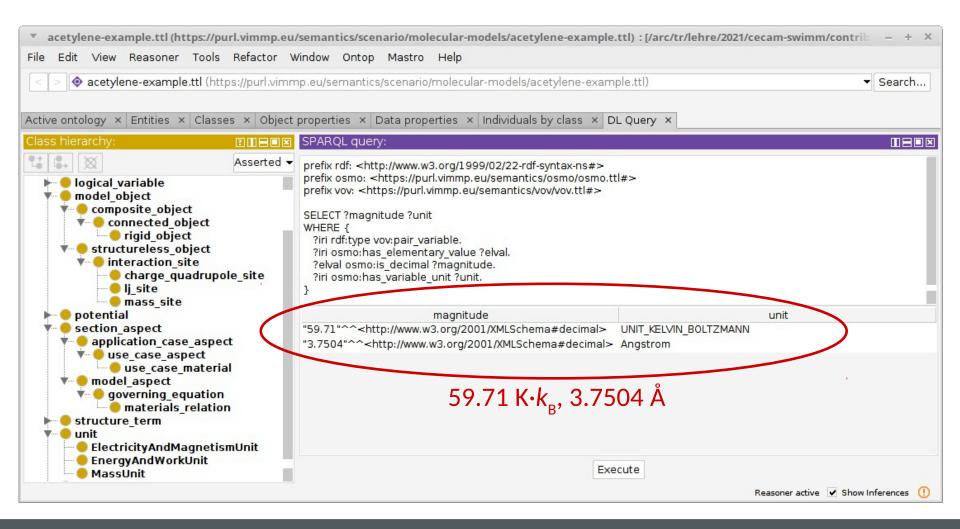
16th March 2021







Query languages (SPARQL in protégé)



S. Chiacchiera, M. T. Horsch

16th March 2021

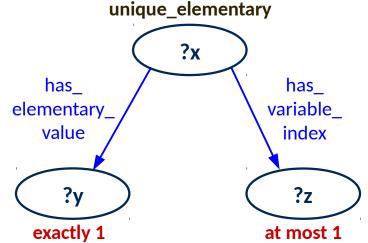




Constraint language (SHACL)

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)

:::::

:::::

:::::

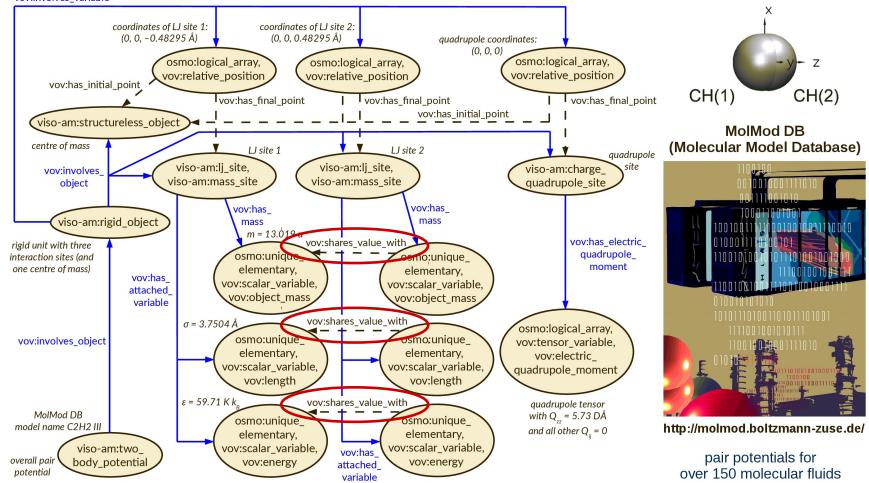
File Edit View Reasoner Tools Refac	tor Window Ontop Mastro Help		
	l.vimmp.eu/semantics/scenario/molecular-models/acetylene-example.ttl)	▼ Searc	ch
$ angle$ logical_variable $ angle$ unique_logical $ angle$ unique_elemer	tary		
Active ontology × Entities × Classes × 0	Object properties × Data properties × Individuals by class × DL Query ×	SHACL Editor ×	
Class hierarchy: unique_elemer 🛙 🛙 🗖 🗖 🗷	SHACL editor:		- • >
🕻 🕻 🕅 Asserted 🕇	Open Save Validate		
🔻 😑 logical_variable	<pre>@prefix osmo: <https: osmo="" osmo.ttl#="" purl.vimmp.eu="" semantics="">.</https:></pre>		
elementary_logical	@prefix owl: ">http://www.w3.org/2002/07/owl#>">http://www.w3.org/2002/02/02/02/02/02/02/02/02/02/02/02/0		
Iogical_structure ↓	<pre>@prefix rdf: <http: 02="" 1999="" 22-rdf-syntax-ns#="" www.w3.org="">. @prefix sh: <http: ns="" shacl#="" www.w3.org="">.</http:></http:></pre>		1
logical_array	epreiix an. <nrtp. .<="" anact#="" na="" td="" www.wo.org=""><td></td><td></td></nrtp.>		
lunique elementary	acetylene-example:UNIQUE_ELEMENTARY_SHAPE a sh:Shape;		
	sh:targetClass osmo:unique_elementary;		
Direct instances:			
●* ※	<pre>sh:path osmo:has_elementary_value; sh:minCount 1:</pre>		
or: 😑 unique_elementary	sh:maxCount 1		
LJ A EPS			
LJ A MASS	<pre>sh:path osmo:has_variable_index;</pre>		
LJ A POS X	sh:maxCount 1		
LJ A POS Y			_
LJ A POS Z	SHACL constraint violations: 3	Ш	
LJ A SIG	Severity SourceShape Message	FocusNode Path V	Valu
LJ B EPS	http:// afdfa4478c Property needs to have at least 1 values, but found 0	https://purl.v http	
LJ B MASS	http:// afdfa4478c Property needs to have at least 1 values, but found 0	https://purl.v http	
LJ_B_POS_X	http:// afdfa4478c Property needs to have at least 1 values, but found 0	https://purl.v http	
	0-1	oner active 🗸 Show Inferences	

:: S. Chiacchiera, M. T. Horsch ::



Constraint language (SHACL)

vov:involves_variable



S. Chiacchiera, M. T. Horsch 💠





HLRS

RDF and OWL libraries

Jena (Java)

https://jena.apache.org/

Jena

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+)

S. Chiacchiera, M. T. Horsch

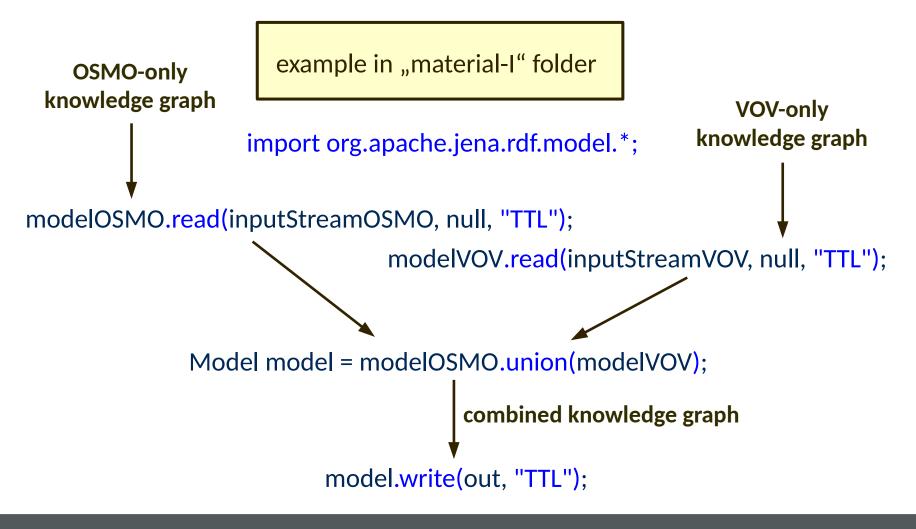
16th March 2021







RDF and OWL libraries: Apache Jena



😳 S. Chiacchiera, M. T. Horsch

16th March 2021







RDF and OWL libraries: Owlready2

example in "material-L" folder

python3 owlready2_example.py

Α	toy	examp	е	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")

16th March 2021







Time for coffee



A sign and the object it (supposedly) refers to

:: S. Chiacchiera, M. T. Horsch ::



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.

This document and all information contained herein is the sole property of the VIMMP Consortium (unless specified otherwise or clear by context). Information presented herein may be subject to intellectual property rights. No intellectual property rights are granted by the delivery of this document or the disclosure of its content. Reproduction or circulation of this document to any third party is prohibited without the consent of the authors.

The statements made herein do not necessarily have the consent or agreement of the VIMMP Consortium. They represent the opinion and findings of the authors.



:: S. Chiacchiera, M. T. Horsch ::