

Materials knowledge and data representation with a European ontology ecosystem

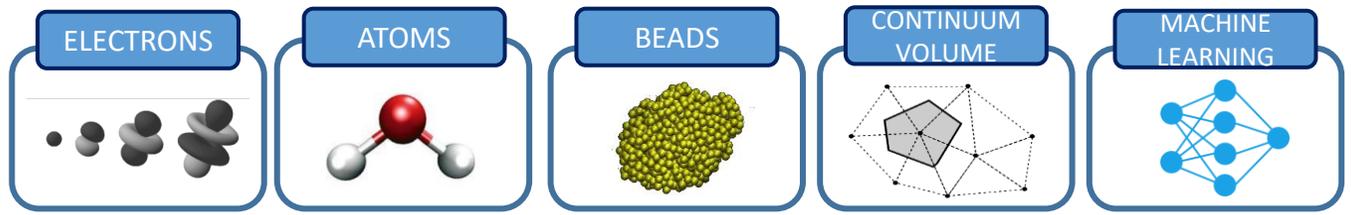
Gerhard Goldbeck

Goldbeck Consulting Ltd
Semantic Materials
EMMC ASBL



EMMC: modelling the materials world

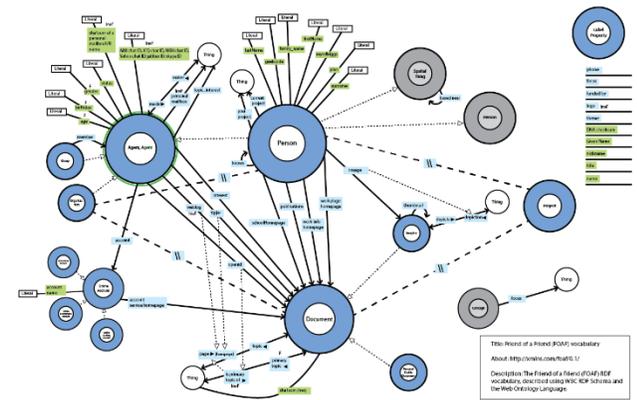
- **ALL types of materials models:** physics and data based, incl AI



- **ALL** chemical/material and application fields

- **Data models and ontologies**, shared conceptualisations based on materials science foundations

Now including Materials Characterisation



- **Harmonisation and standardisation.**



<https://emmc.eu>

The value of knowledge organisation

Your shopping experience....

- Random stuff by car
- No insights about what other sellers have
- Random level of quality

Car Boot Sale



Junk Yard

- Randomly distributed stuff
- Lots of effort to find stuff
- Only the 'owner' knows where stuff might be at best



Moroccan Street Market

- Grouping by product category i.e. spices vs carpets vs etc
- Improved level of quality
- Good intentions but **not really scalable**

Local Market



- Well organised
- Well labelled items
- Limited produce

Hyper Market



- Everything in once place
- Specialist by product category
- Data driven display/groupings

Amazon Prime



- Not physically constrained
- Recommendation engines
- Huge Scale and Variety
- Fuzzy search
- Services ...
- AI ...

Adapted from: B. Gardner;
<https://www.slideshare.net/slideshow/fair-datacentric-information-architecturepptx/265566803>

Credit to Graham Henry, Sandra McGarry, Kerstin Forsberg and Ben Gardner

Semantic data management: what and why?

- Applying semantic web technologies to the management of data
- Benefit is a **flexible, interconnected Data Model**
- **Saves time, cost, and improves maintainability**
 - No fixed schema
 - Open World Assumption (OWA): new facts can be easily included.

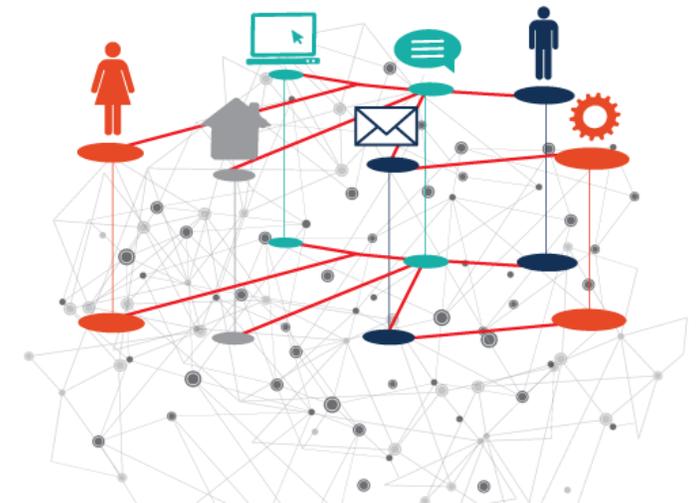
<https://cambridgesemantics.com/blog/semantic-university/semantic-technologies-applied/applying-semantic-web-two-camps/>

<https://devopedia.org/semantic-web>

Web



Semantic Web



Linked Open Data Cloud

Linked data is **structured** data which is **interlinked** with other data so it becomes more useful through [semantic queries](#).

It builds upon standard [Web](#) technologies.

Part of the vision of linked data is for the [Internet](#) to become a global [database](#).^[1]

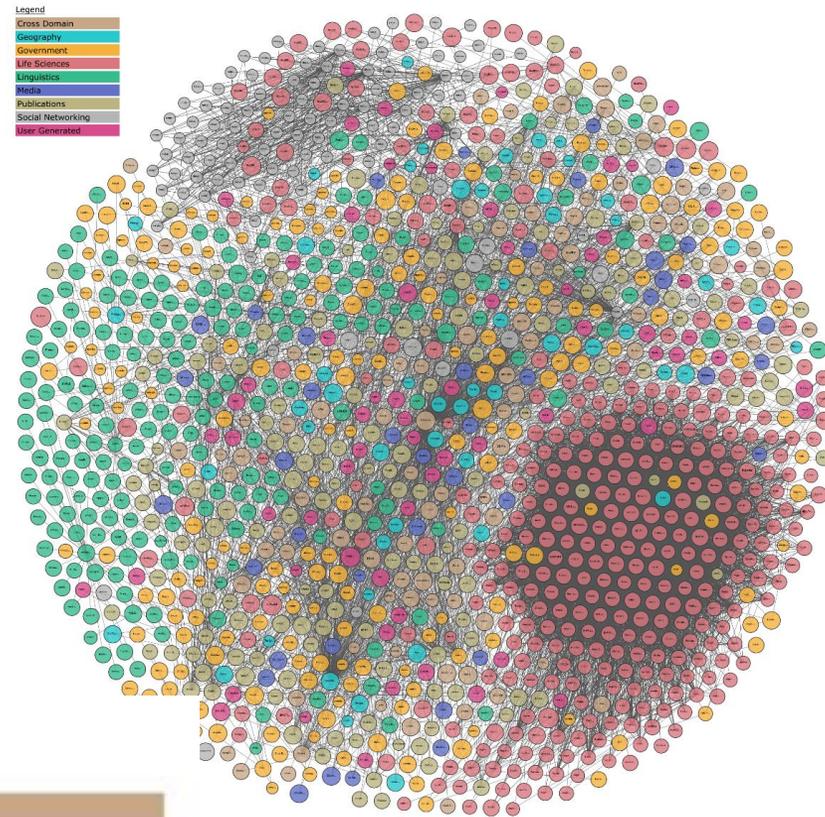
[Tim Berners-Lee](#) coined the term in a 2006 design note about the [Semantic Web](#) project.^[2]

The dataset contained **1356** datasets on 31 May 2025

Materials/Chemistry is not covered!

https://en.wikipedia.org/wiki/Linked_data

<https://www.lod-cloud.net/>



© 0

Schema.org and also Dbpedia.org

About: <http://dbpedia.org/ontology/>

An Entity of Type: [Vocabulary](#), from Named Graph: <http://dbpedia.org/resource/classes#>, within Data Space: [dbpedia.org](#)

This ontology is generated from the manually created specifications in the DBpedia Mappings Wiki. Each release of this ontology corresponds to a new release of the DBpedia data set which contains instance data extracted from the different language versions of Wikipedia. For information regarding changes in this ontology, please refer to the DBpedia Mappings Wiki.

Dbpedia and Schema.org have a Chemical Substance class
but nothing specific to materials

Full Hierarchy

Schema.org is defined as two hierarchies: one for textual property values

This is the main schema.org hierarchy: a collection of types (or "classes") where each type has at least one super-type, here we show each type in one branch of the tree

Types:

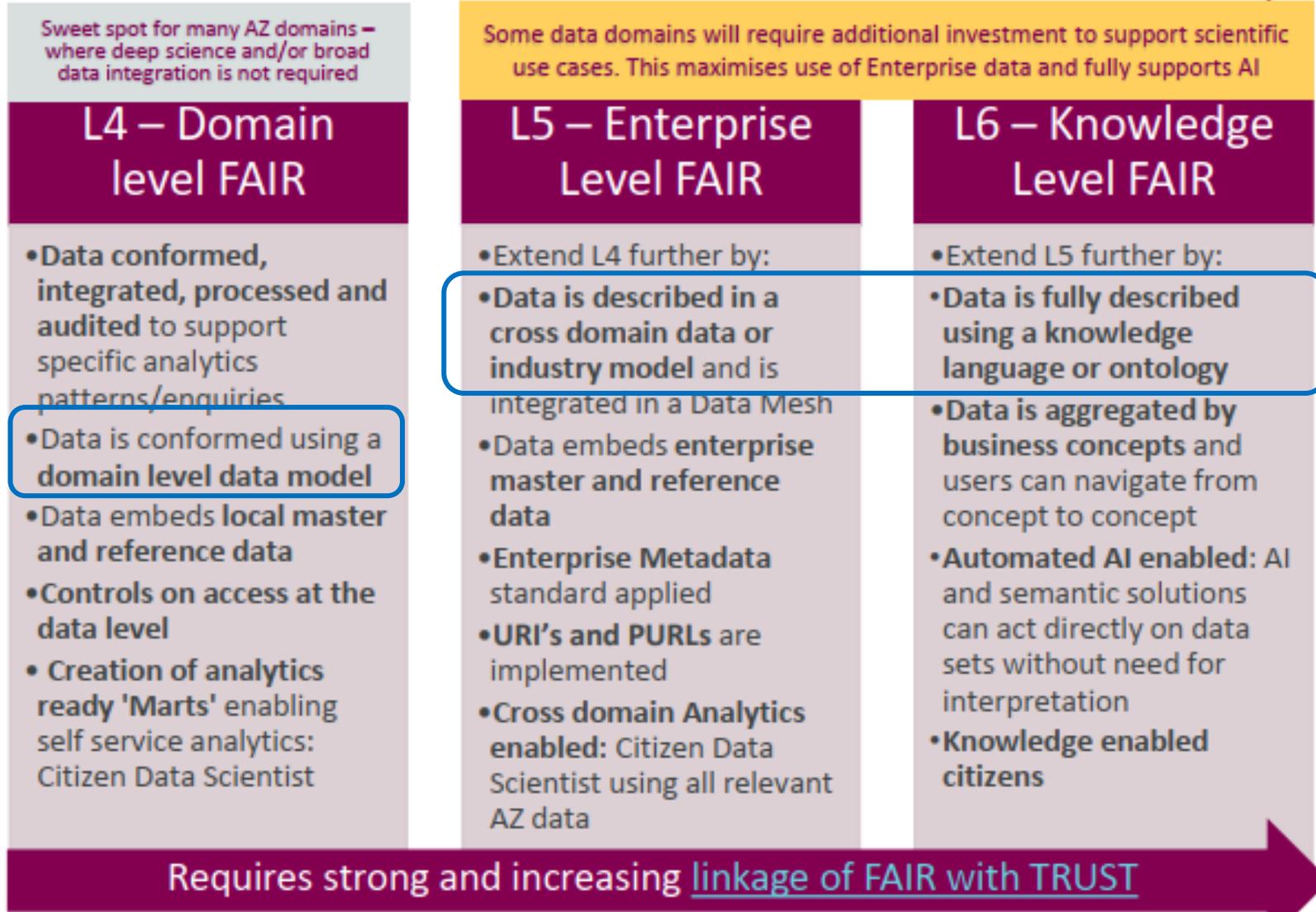
[Close hierarchy](#) / [Open hierarchy](#)

Thing -

- ▶ Action +
- ▼ BioChemEntity -
 - ChemicalSubstance
 - Gene
 - MolecularEntity
 - Protein
- ▶ CreativeWork +
- ▶ Event +
- ▶ Intangible +
- ▶ MedicalEntity +
- ▶ Organization +
- ▶ Person +
- ▶ Place +
- ▼ Product -
 - DietarySupplement
 - Drug
 - IndividualProduct
 - ProductCollection
 - ProductGroup
 - ProductModel
 - SomeProducts
 - ▶ Vehicle +
- Taxon

Towards Knowledge Level FAIR

Evolution of Data Centric Thinking to Knowledge Centric



<https://www.slideshare.net/slideshows/fair-datacentric-information-architecturepptx/265566803>

Knowledge Organisation Challenges in Materials Sciences

- Lack of agreed conceptualisation and terminologies
 - no equivalent to IUPAC Goldbook
 - basic terms (e.g. material vs chemical substance) are not standardised.
- Lack of coverage of materials in widely used schema (Schema.org, DCAT-AP)
- Low FAIR score of many semantic assets.
- Poor and inconsistent coverage of materials terms in ontologies (at least up to 2023*)
- Reflects the EOSC Semantic Interoperability Challenges:
 - *Lack of common explicit definitions about the terms that are used by user communities*
 - *Lacking common semantic artefacts across sub-domains*

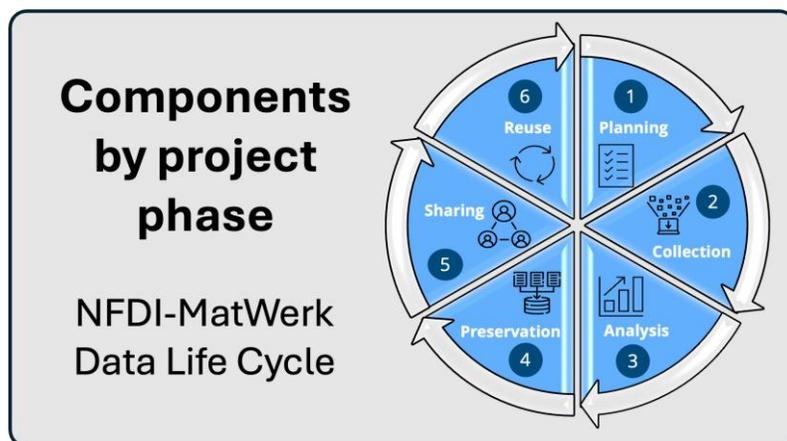
* Review and Alignment of Domain-Level Ontologies for Materials Science

Anne De Baas, Pierluigi Del Nostro, Jesper Friis, Emanuele Ghedini, Gerhard Goldbeck, Ilaria Maria Paponetti, Andrea Pozzi, Lan Yang, Arkopaul Sarkar, Francesco Antonio Zaccarini and Daniele Toti (2023) <https://ieeexplore.ieee.org/document/10296887>

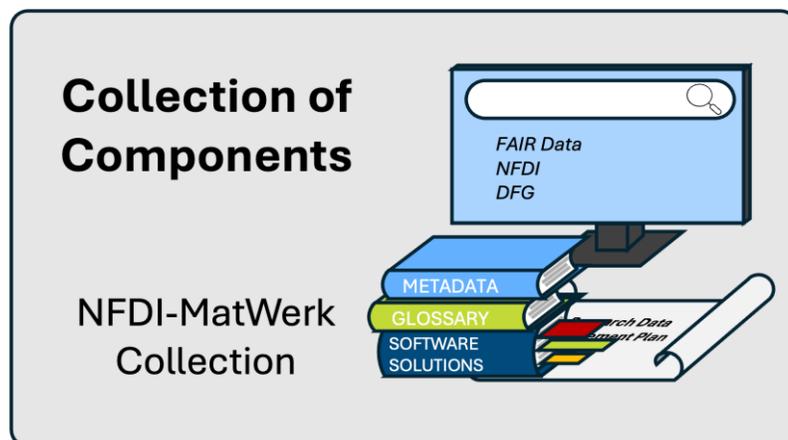


Things are improving ...

<https://nfdi-matwerk.de/solutions/via-data-life-cycle>



<https://nfdi-matwerk.de/solutions/collection/metadata>



...but lack standards

Metadata Schemas

Name

MWO metadata template for datasets

MWO metadata template for experimental workflows

MWO metadata template for computational workflows

Metadata schemas for materials science data

Web Standards and Reference Data for First-Principles Simulations

EngMeta

Essential Source of Schemas and Examples (ESSE)

MGI JSON Schema

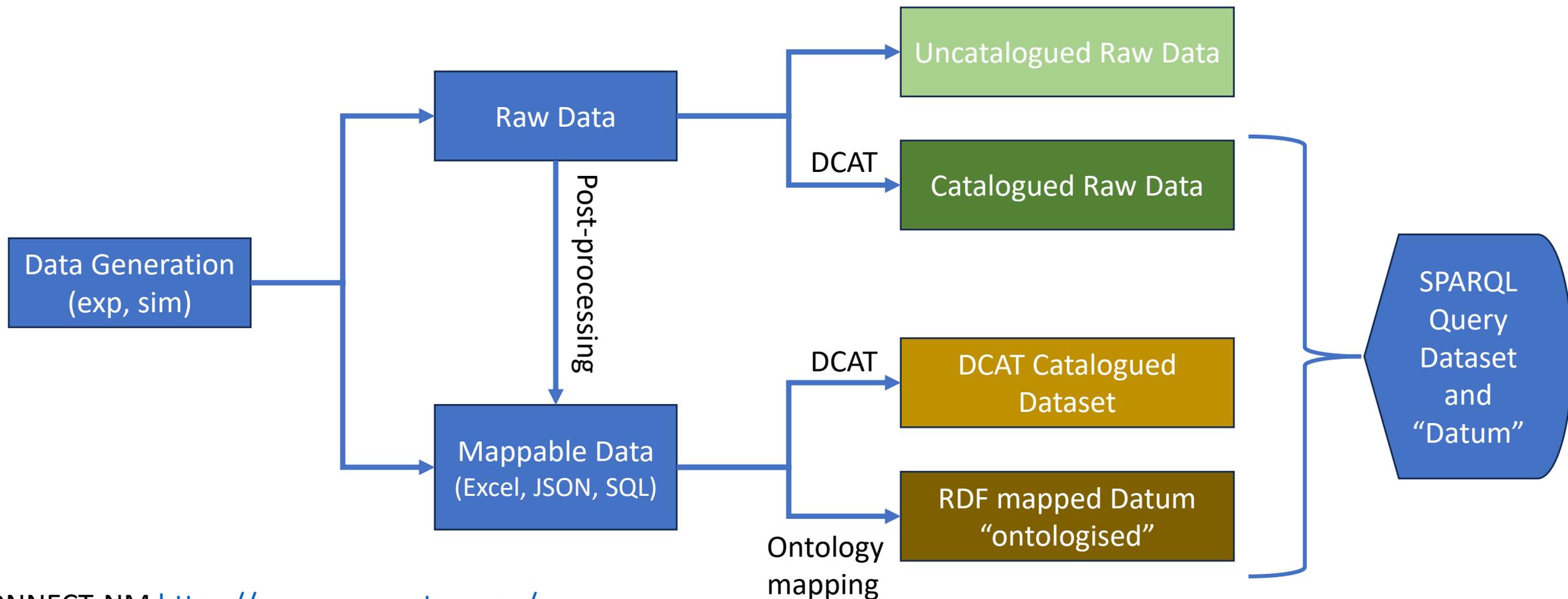
Schema for scanning electron microscopy (SEM)

Schema for transmission electron microscopy (TEM)

Schema for magnetic resonance imaging (MRI)

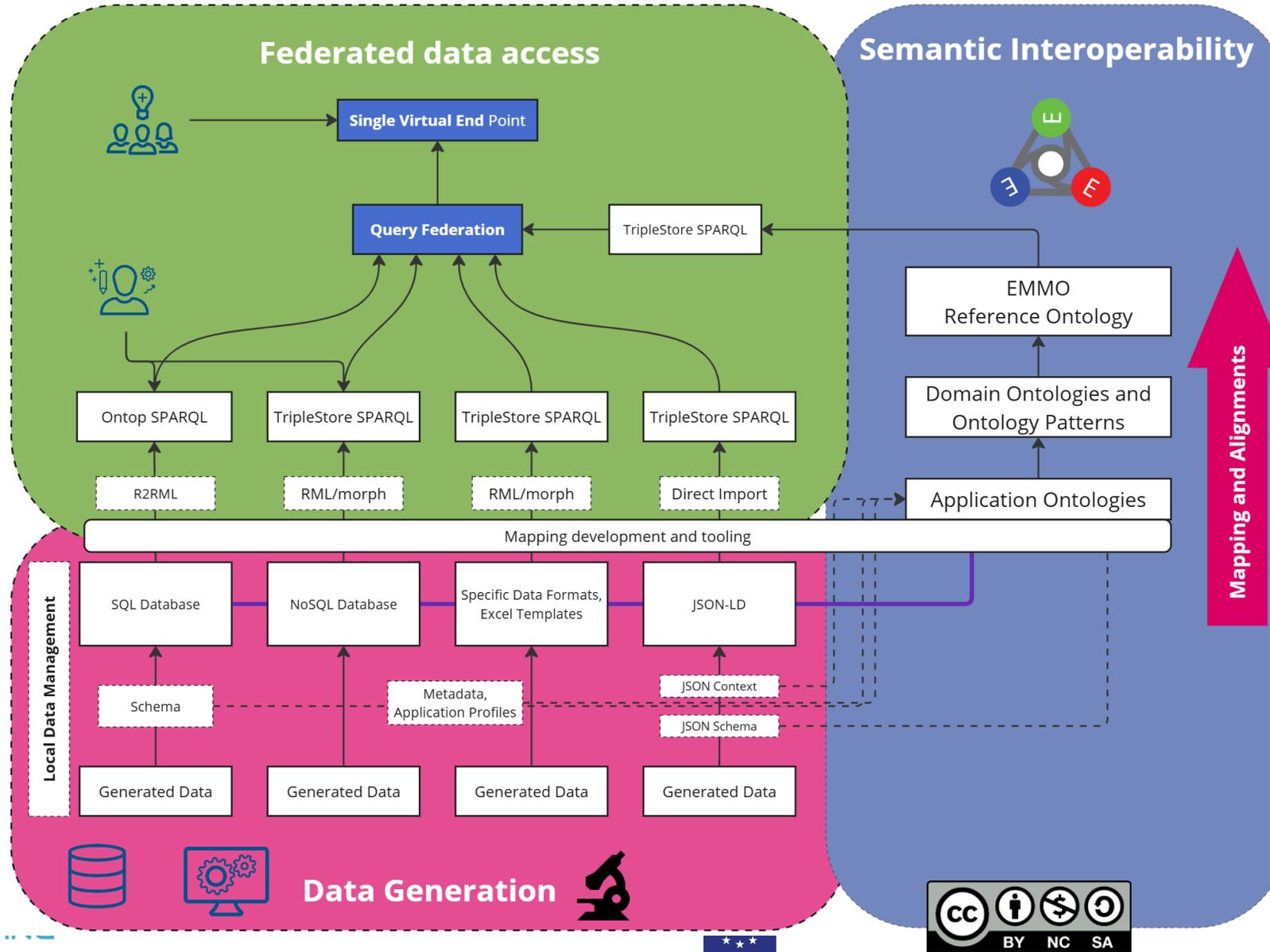
Schemas for FIB/SEM serial sectioning tomography

Need to agree on a standard (?) architecture



CONNECT-NM <https://www.connect-nm.eu/>

Knowledge Organisation System: following industry standards



Credit:
Emanuele Ghedini,
Univ. Bologna,
CONNECT-NM
<https://www.connect-nm.eu/>

Requirements

- For Datasets
 - Require generic and materials specific metadata
 - Support cataloguing of datasets
- For Datum (full data integration)
 - RDF mapping with a coherent set of ontologies
 - Support data federation, federated queries



You are here: [Thermal Material Properties ...](#)

Dataset: Thermal Material Properties of Commercial NMC532 / Graphite Lithium-Ion Battery Cell

[RADAR Metadata](#)

[Content](#)

[Statistics](#)

1

Creator/Author:

Cloos, Lisa  <https://orcid.org/0009-0006-1001-2891> [Institut für Thermische Verfahrenstechnik]

Aluminum 5024-H116 Al-Scandium Alloy

Categories: [Metal](#); [Nonferrous Metal](#); [Aluminum Alloy](#); [5000 Series Aluminum Alloy](#)

Material Notes: Scandium containing alloy is suggested for aerospace applications. Fea
Composition information provided by the Aluminum Association and is r

Vendors: No vendors are listed for this material. Please [click here](#) if you are a sup

 [Printer friendly version](#)  [Download as PDF](#)  [Download to Excel \(requires Excel and Windows\)](#)

 [Export data to your CAD/FEA program](#)

Physical Properties

	Metric
Density	2.65 g/cc

Mechanical Properties

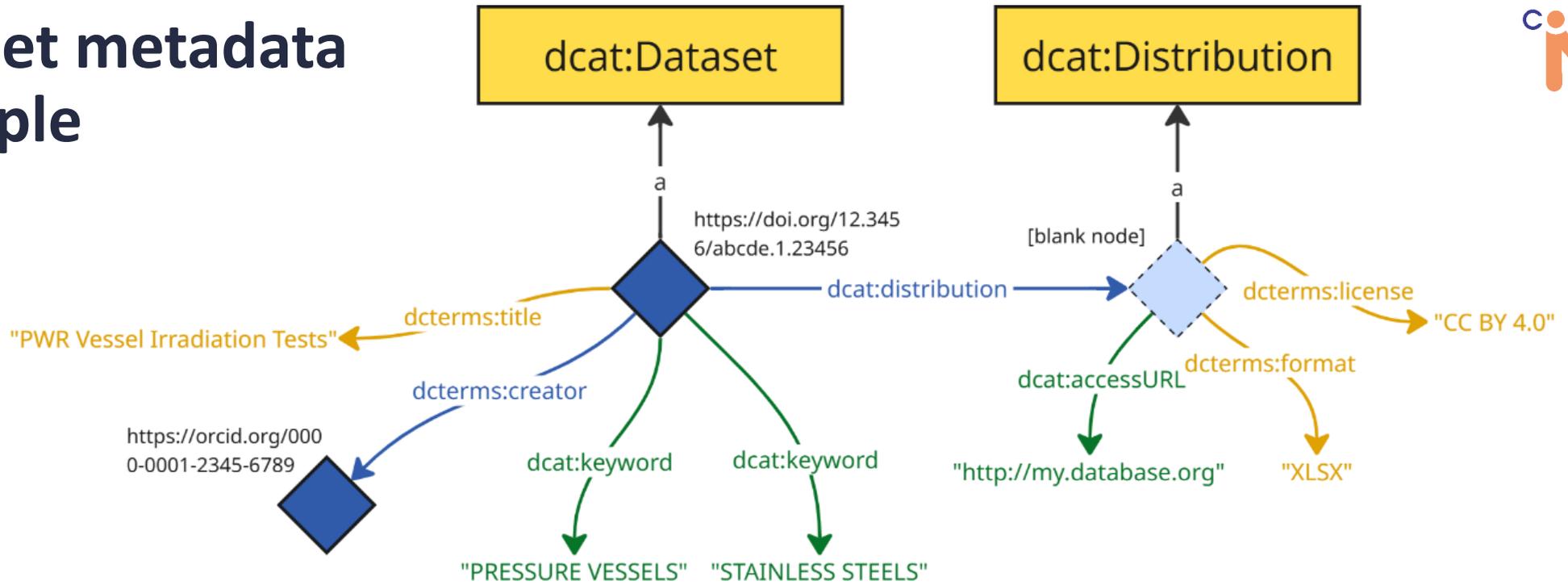
	Metric
Tensile Strength, Ultimate 	315 MPa

@Thickness 1.60 mm

Data Cataloguing Standards

- Schema.org
 - Extensions for bio-chemistry, see also www.bioschemas.org
 - Is it worth creating profiles and extensions for materials?
- DCAT (W3C standard)
 - DCAT-AP (Application Profiles) promoted as standard by EU
 - Also adopted by domains, e.g. GeoDCAT-AP, a spatio-temporal profile of the W3C DCAT Recommendation
 - Need a standard for materials: “MaterialsDCAT-AP”
 - RDA WG looking into this: [Harmonised terminologies and schemas for FAIR data in materials science and related domains](#)

Dataset metadata Example



```

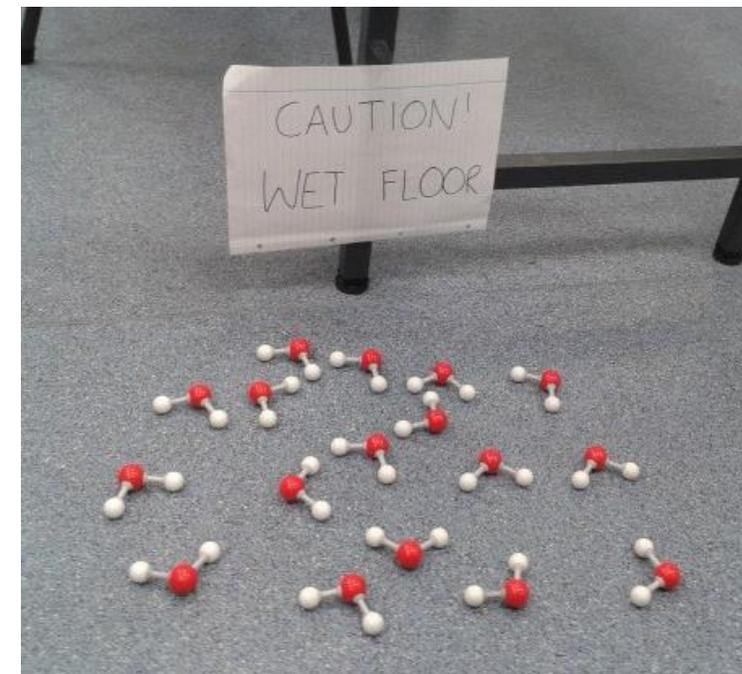
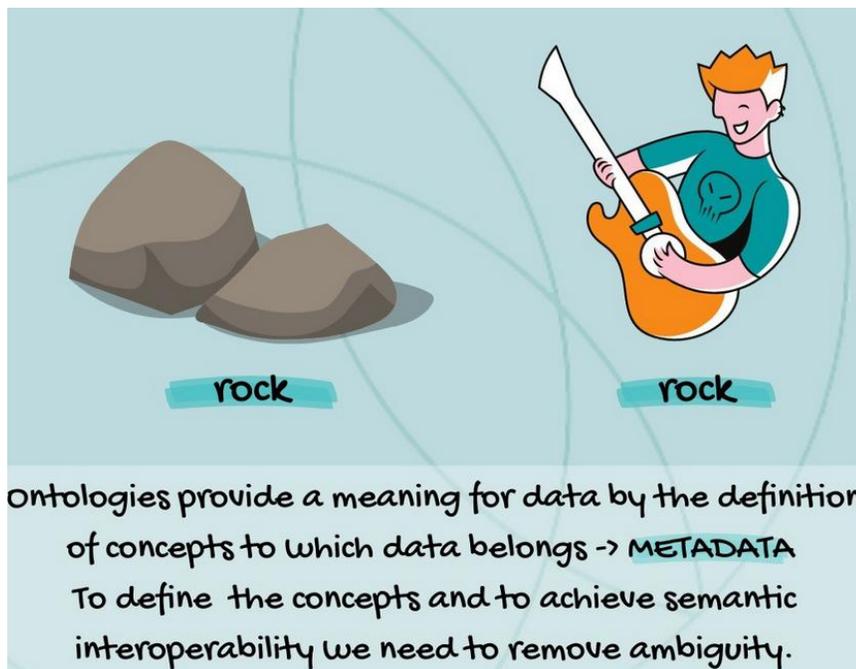
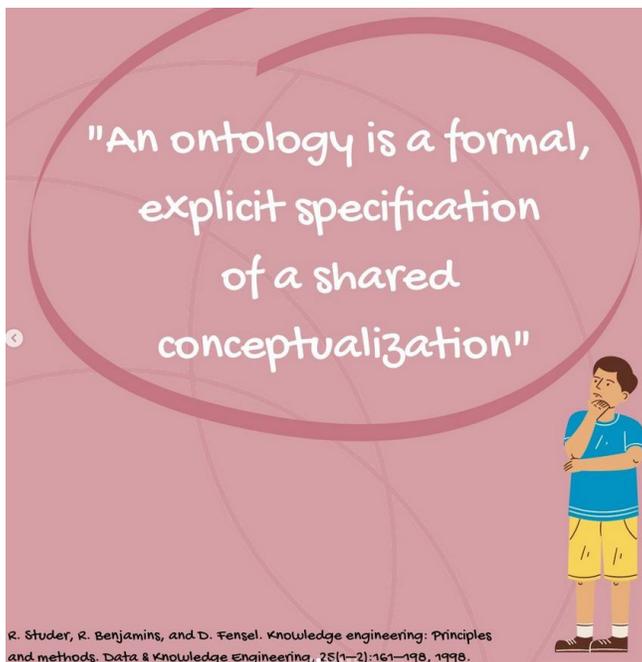
{
  "@context": {
    "dcat": "http://www.w3.org/ns/dcat#",
    "dcterms": "http://purl.org/dc/terms/"
  },
  "@id": "https://doi.org/12.3456/abcde.1.23456",
  "@type": "dcat:Dataset",
  "dcterms:creator": {
    "@id": "https://orcid.org/0000-0001-2345-6789"
  },
  "dcat:keywords": ["PRESSURE VESSELS", "STAINLES STEELS"],
  "dcterms:title": "PWR Vessel Irradiation Tests",
  "dcat:distribution": {
    "@type": "dcat:Distribution",
    "dcat:accessURL": "http://my.database.org",
    "dcterms:license": "CC BY 4.0",
    "dcterms:format": "XLSX"
  }
}

```

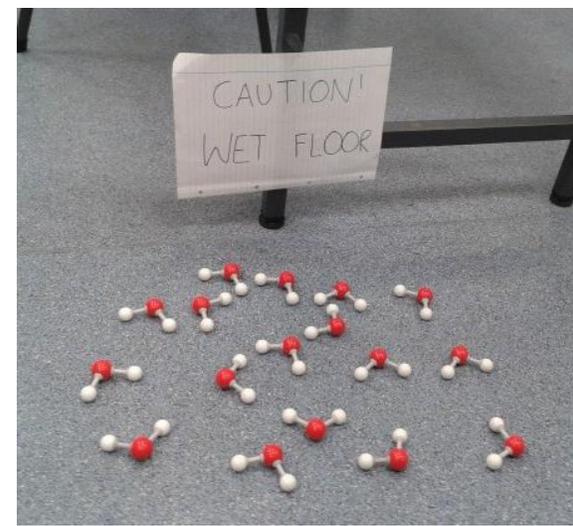
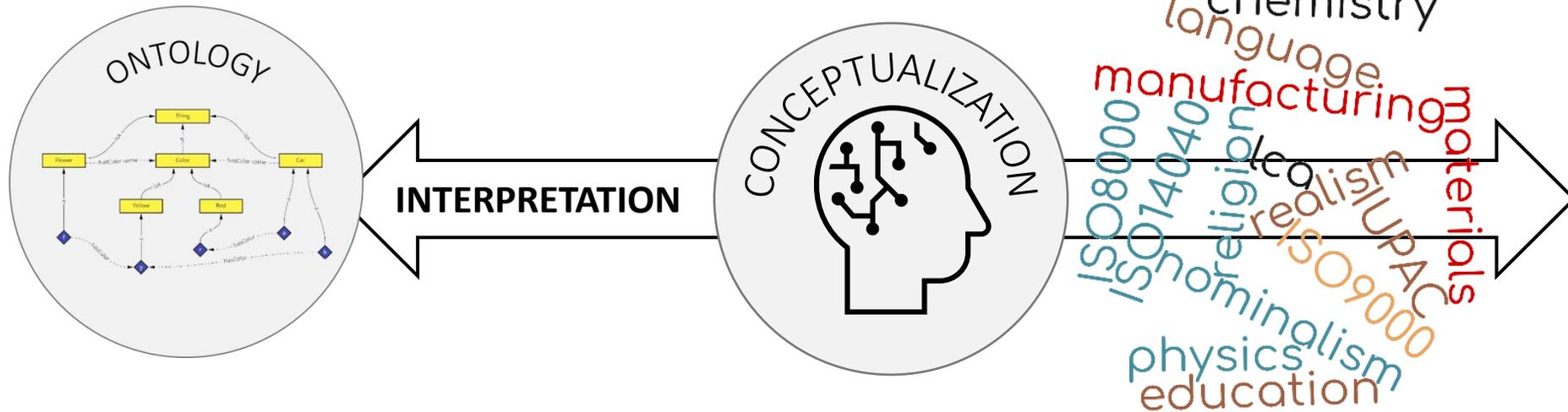
Subject	Predicate	Object
_:b0	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/ns/dcat#Distribution
_:b0	http://purl.org/dc/terms/format	XLSX
_:b0	http://purl.org/dc/terms/license	CC BY 4.0
_:b0	http://www.w3.org/ns/dcat#accessURL	http://my.database.org
https://doi.org/12.3456/abcde.1.23456	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/ns/dcat#Dataset
https://doi.org/12.3456/abcde.1.23456	http://purl.org/dc/terms/creator	https://orcid.org/0000-0001-2345-6789
https://doi.org/12.3456/abcde.1.23456	http://purl.org/dc/terms/title	PWR Vessel Irradiation Tests
https://doi.org/12.3456/abcde.1.23456	http://www.w3.org/ns/dcat#distribution	_:b0
https://doi.org/12.3456/abcde.1.23456	http://www.w3.org/ns/dcat#keywords	PRESSURE VESSELS
https://doi.org/12.3456/abcde.1.23456	http://www.w3.org/ns/dcat#keywords	STAINLES STEELS

“Datum” level integration

- *Structure data around the science* (Ben Gardner, AstraZeneca)
- *Start with meaning* (Dave McComb, Semantic Arts) <https://www.brcommunity.com/articles.php?id=b972>
- *Start with conceptualisation* (Emanuele Ghedini, Univ Bologna, EMMO)



Conceptualisation issues



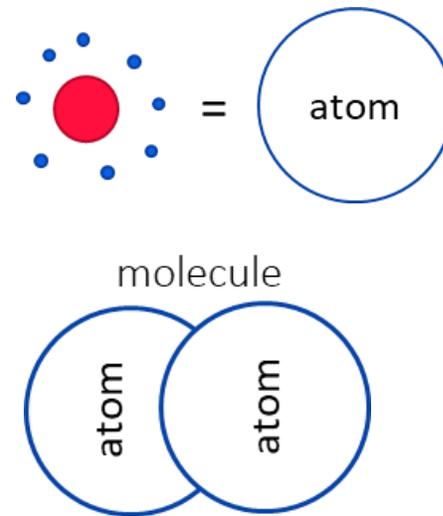
Several conceptualizations for the same things
Different things can have the same label

Non-Univocal Scientific Conceptualisations

IUPAC Goldbook Definitions (Chemistry)

An **Atom** is a nucleus of Z positive charge and Z electrons

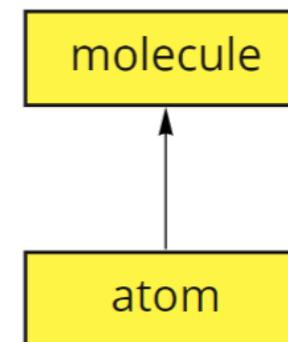
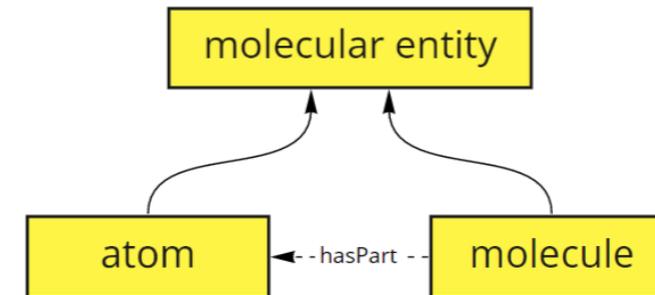
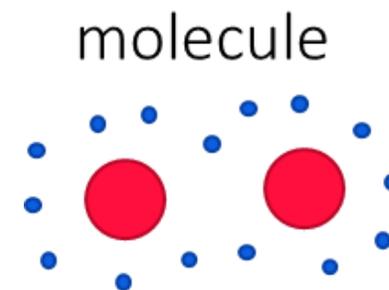
A **Molecule** is an electrically neutral entity consisting of more than one atom



Chemical-Physics Definitions

A **Molecule** is an electrically neutral entity consisting of nuclei and electrons.

An **Atom** is a singular type of molecule with one nucleus only.



Ontology requires conceptualisation

Ontological Framework

Real World



Conceptualisation



Ontology

(a formal theory expressed in
e.g., FOL, HOL, OWL)

$$DP(x, y) \equiv PP(x, y) \wedge$$
$$\exists \phi [\phi(x) \leftrightarrow \neg \phi(y)]$$

Without **conceptualisation** an ontology is simply a symbolic system without meaning.

Why a new foundational ontology for materials science (and generally natural science)?



Generic Ontologist

EMMC Community Member

The world expressed
in philosophical
language



The world expressed
in applied science
language

endurant
continuant
realizable entity
abstract entity
perdurant
generically dependent
continuant

specifically dependent
continuant
fiat object
occurrent
qualia

Physics

Models

Material Science

Chemistry

Characterisation

EMMO Foundations

Everything is made of indivisible **discrete quantum elements** as suggested by SMP (**reductionism, mereology**)

Must be coherent with **special relativity (causality)**

Must be coherent with **quantum mechanics and QFT (mereocausality)**

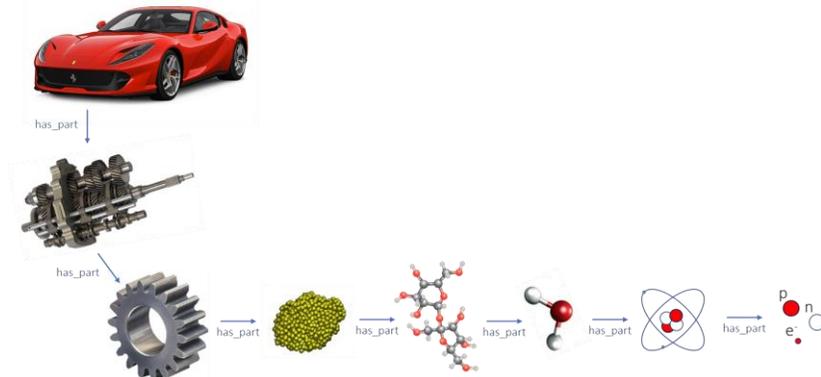
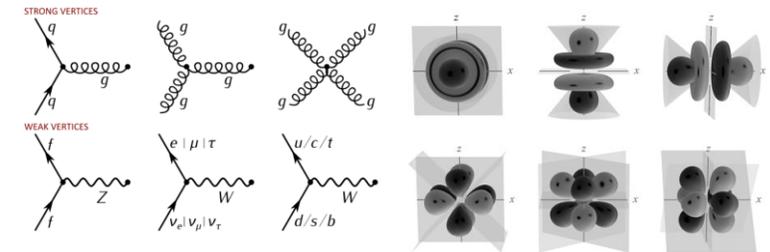
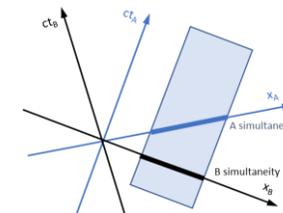
Must be coherent with **conservation laws**, i.e., every change in an entity requires an interaction with at least another entity (**4D**)

Nothing exists if doesn't interact causally with another entity, i.e. **abstract objects do not exist (physicalism, nominalism)**

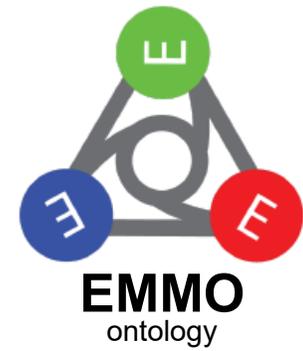
Properties are not inherent to things, but are bounded to a specific observer and observation mechanism (**semiotic approach**)

Standard Model of Elementary Particles

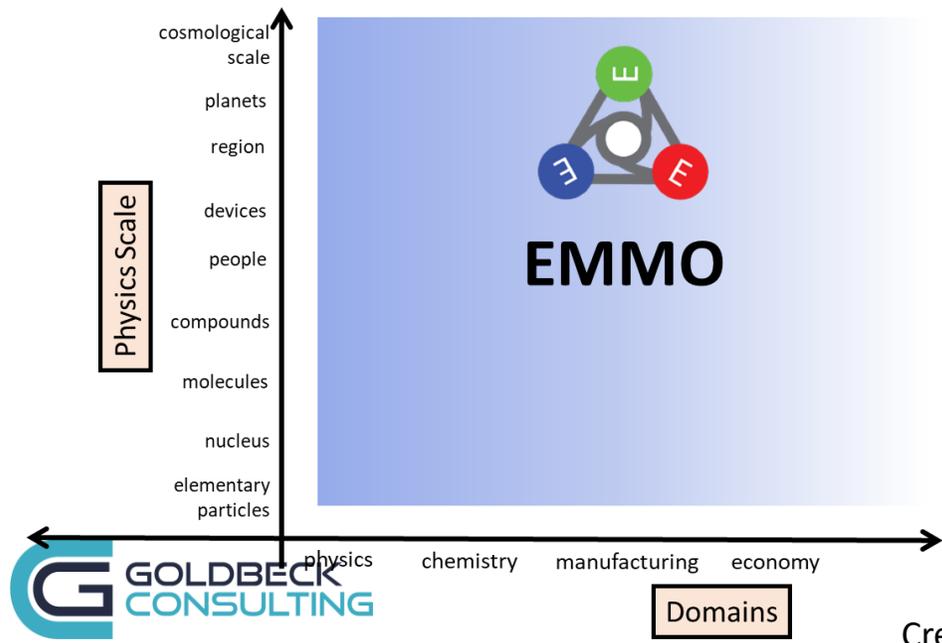
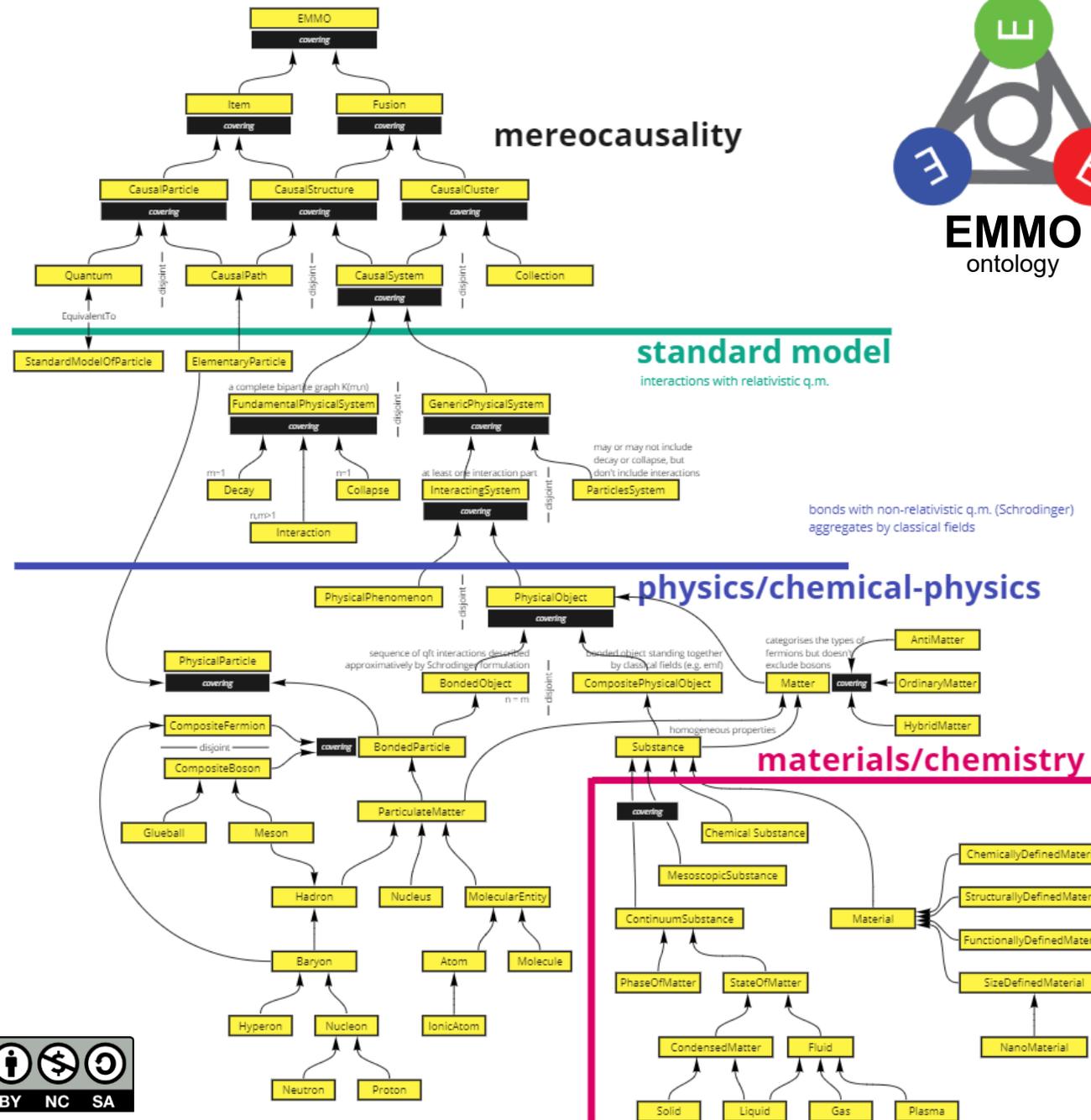
three generations of matter (fermions)			interactions / force carriers (bosons)		
	I	II	III		
mass	=2.2 MeV/c ²	=1.28 GeV/c ²	=173.1 GeV/c ²	0	=124.97 GeV/c ²
charge	2/3	2/3	2/3	0	0
spin	1/2	1/2	1/2	1	0
	u up	c charm	t top	g gluon	H higgs
	d down	s strange	b bottom	γ photon	
	e electron	μ muon	τ tau	Z Z boson	
	ν_e electron neutrino	ν_μ muon neutrino	ν_τ tau neutrino	W W boson	



Physics, Chemistry and Materials Knowledge Organisation



EMMO achieves complete **multiscale** representational capabilities with solid foundations in physics



Credit: Emanuele Ghedini, University of Bologna

EMMO: 3 prime principles for all relations



Mereology: parthood, “spatial”

Causality: incl time sequence

Semiotics: “signs”

“model for”, “property of”, “name of” etc

- EMMORelation
- causal
- mereological
- semiotical

- DOLCERelation
- constantConstituentOf
- constantlyOverlaps
- constantlySpatiallyLocatedAt
- constantQualeOf
- overlaps
- presentAt
- qualeOf
- specificallyDependsOn

- BFORelation
- concretizes at some time
- continuant part of at some time
- environs
- exists at
- first instant of
- generically depends on at some time
- has continuant part at some time
- has first instant
- has history
- has last instant
- has material basis at some time
- has occurrent part
- has participant at some time
- has realization
- history of
- is carrier of at some time
- is concretized by at some time
- last instant of
- located in at some time
- location of at some time
- material basis of at some time
- occupies spatial region at some time
- occupies spatiotemporal region
- occupies temporal region
- occurrent part of
- occurs in
- participates in at some time
- preceded by
- precedes
- realizes
- spatially projects onto at some time
- specifically depended on by
- specifically depends on
- temporally projects onto

They proved to be so **powerful** that every relation (object property) concept can be placed under them.



EMMO complies with ISO/IEC 21838-1



Information technology — Top-level ontologies (TLO) — Part 1: Requirements

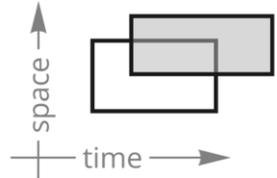
Domain neutrality: Based on scientific principles and analytic philosophy

Logics: EMMO is available at the Top Level in First Order Logic as well

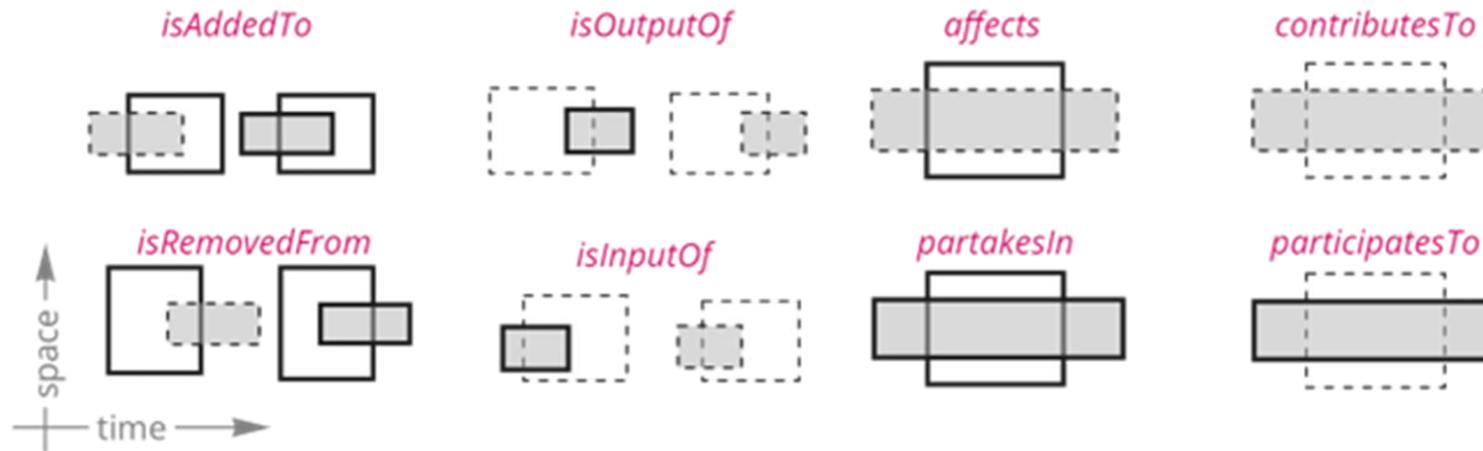
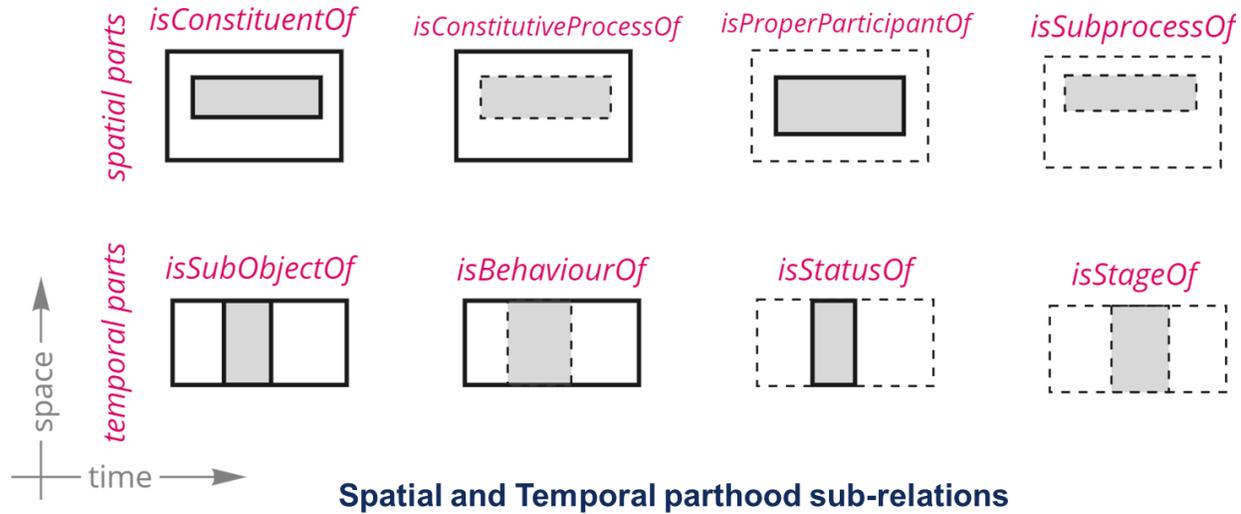
A clear stance on all ISO standard requirements such as

- **Space and time:** practically 4-dimensional
- **Qualities and quantities**
- **Parthood, Location, Constitution, Time and Change, and Events and Processes**
- **Classifications and Mathematical, Informational, Social, & Mental Entities**
- **Causation and scale & granularity** have been individuated as especially important topics a TLO tailored for MSE should focus on
- Further details to be published!

Macro-entities, EMMO toolbox



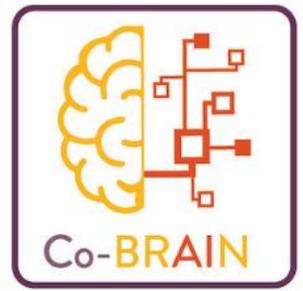
4D approach and space/time graphical representation of entities.



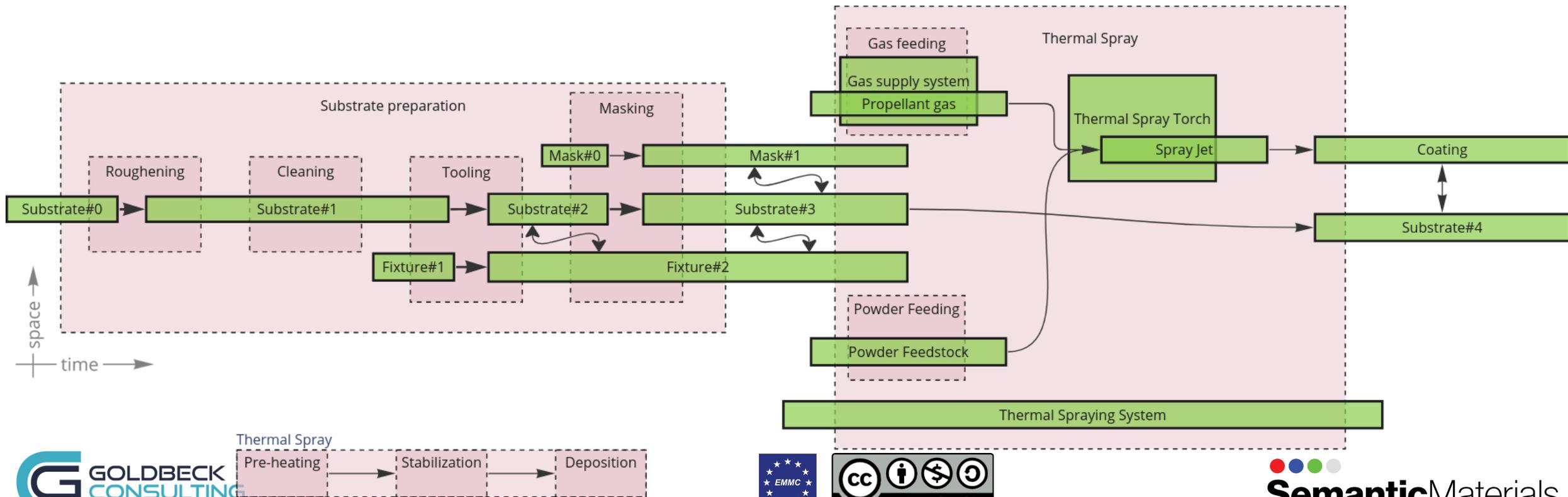
Proper overlap sub-relations (continuous line standing for objects, dashed line for processes, and relations going from the grey to the white boxes)

<https://github.com/emmo-repo/EMMO/wiki>

Representation of processes and objects



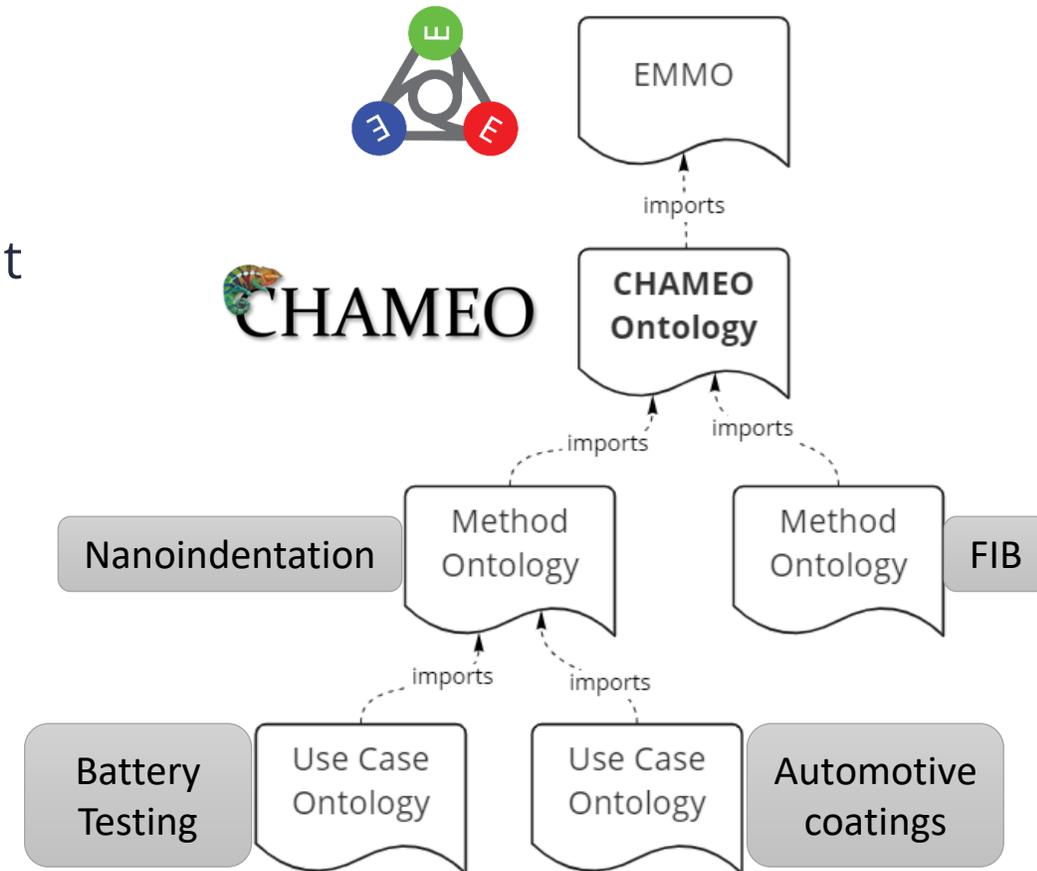
4D mereocausal representation of a **Thermal Spraying process** in **CoBRAIN** project connecting all the relevant entities using mereocausality relations.



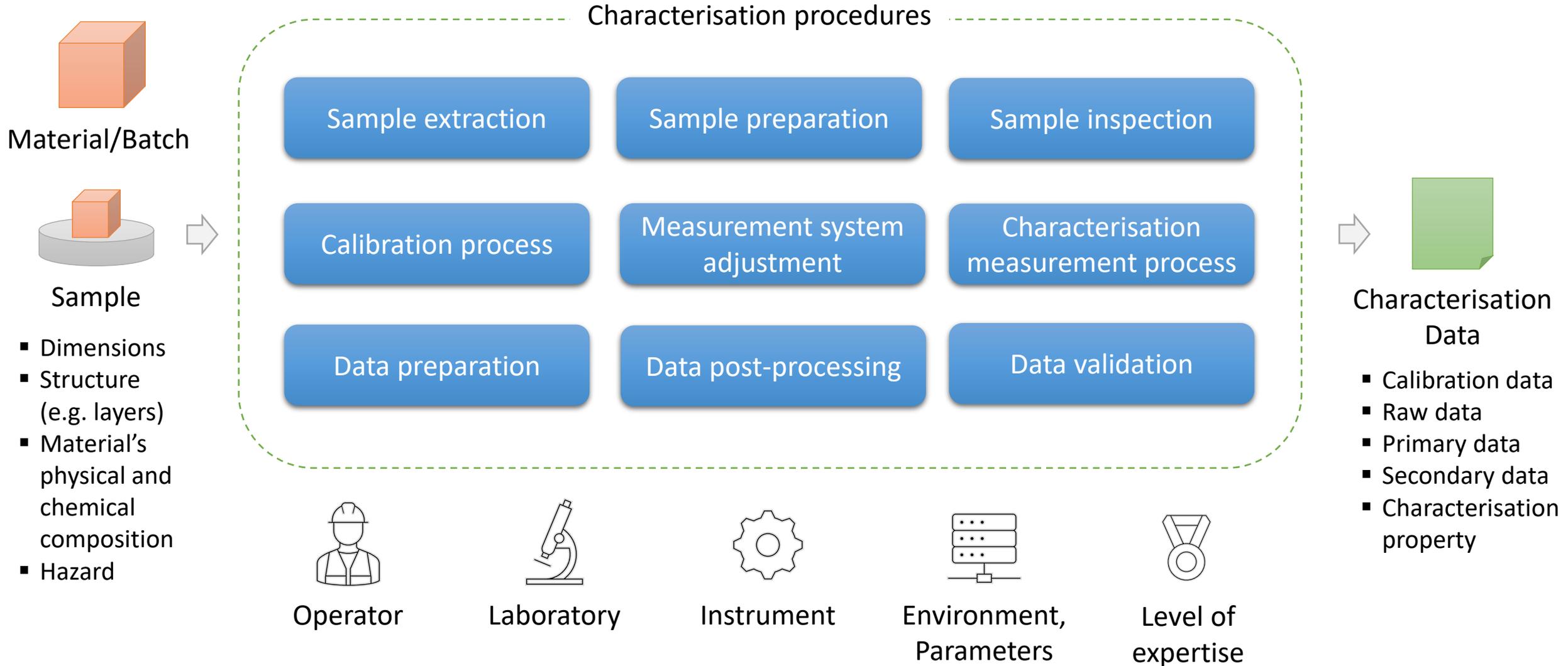
Materials Characterisation

- CHAMEO: ontology of materials characterisation experiments
- Models the generic aspects in common across different characterisation techniques
- Method specific ontologies have been developed
- Different application cases have been demonstrated
- Aligned with a CEN Workshop Agreement: Materials characterizatón - Terminology and structured documentatón

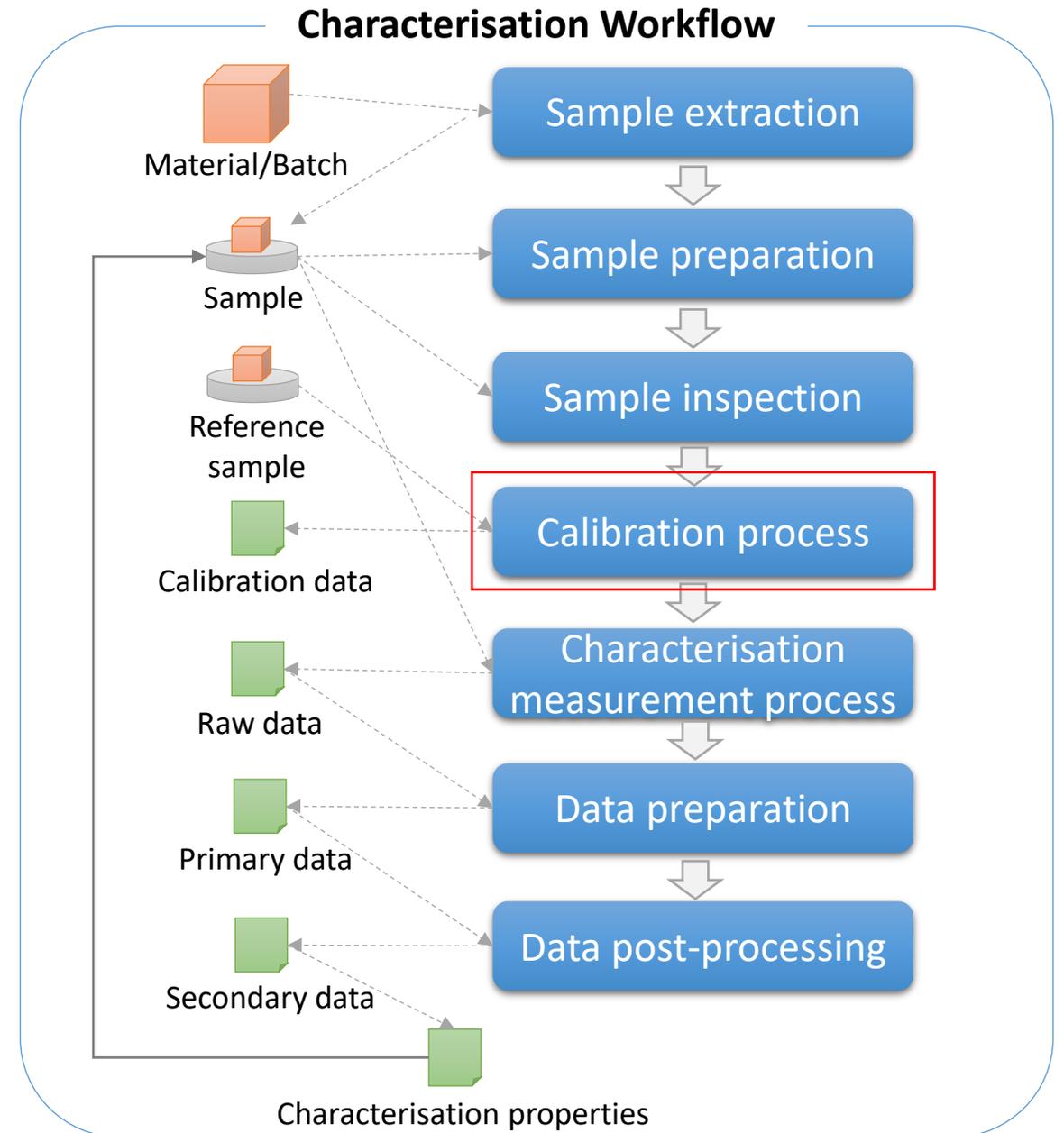
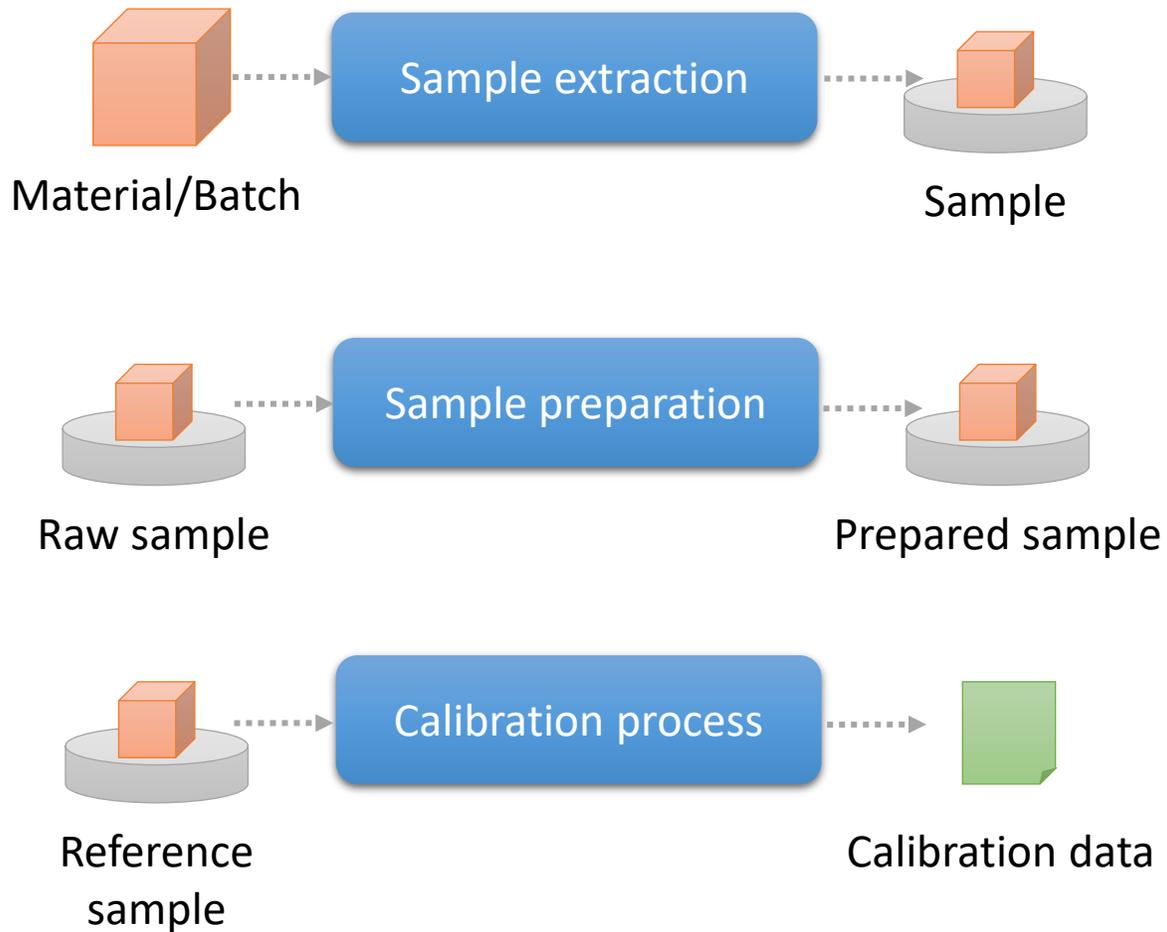
https://www.cencenelec.eu/media/CEN-CENELEC/CWAs/RI/2025/cwa17815_2025.pdf



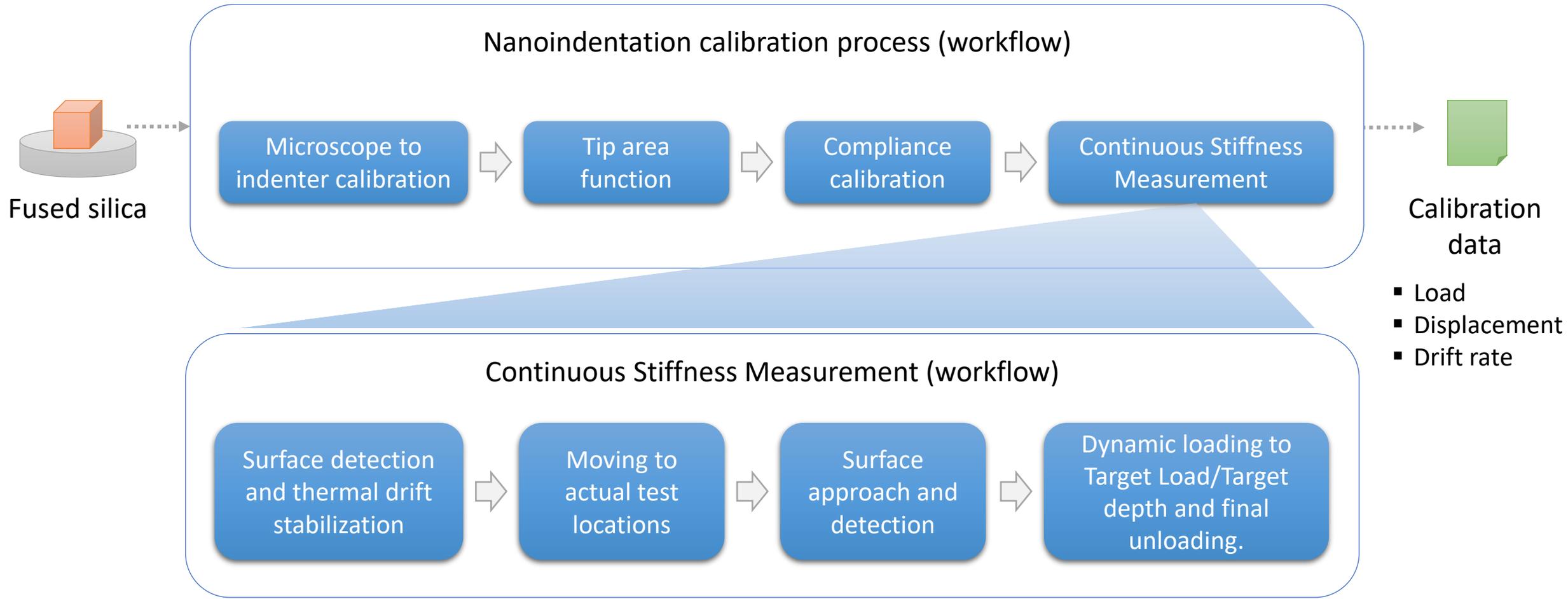
What is in CHAMEO?



CHAMEO is flexible!



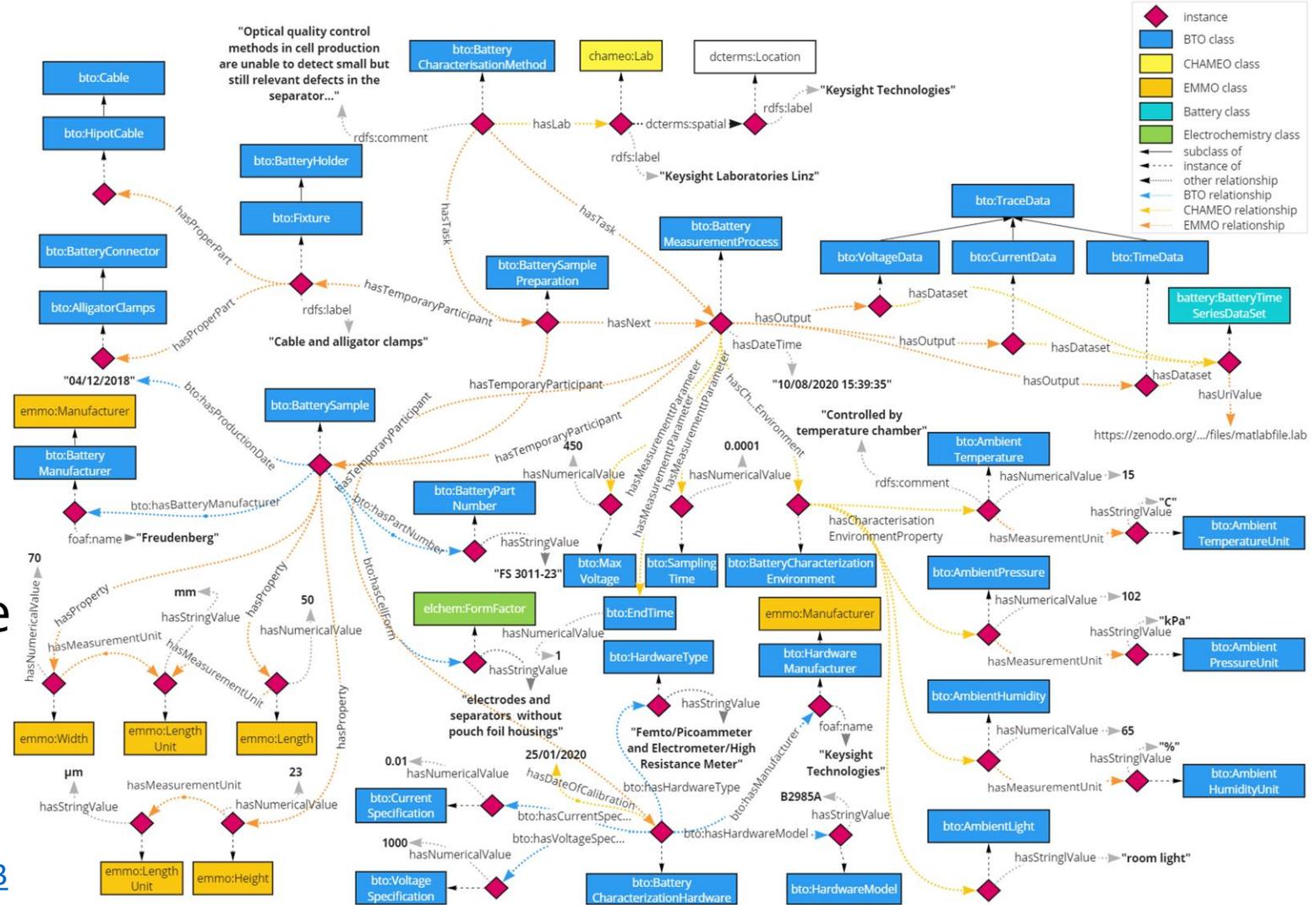
The characterisation process can be broken-down to the desired level of detail



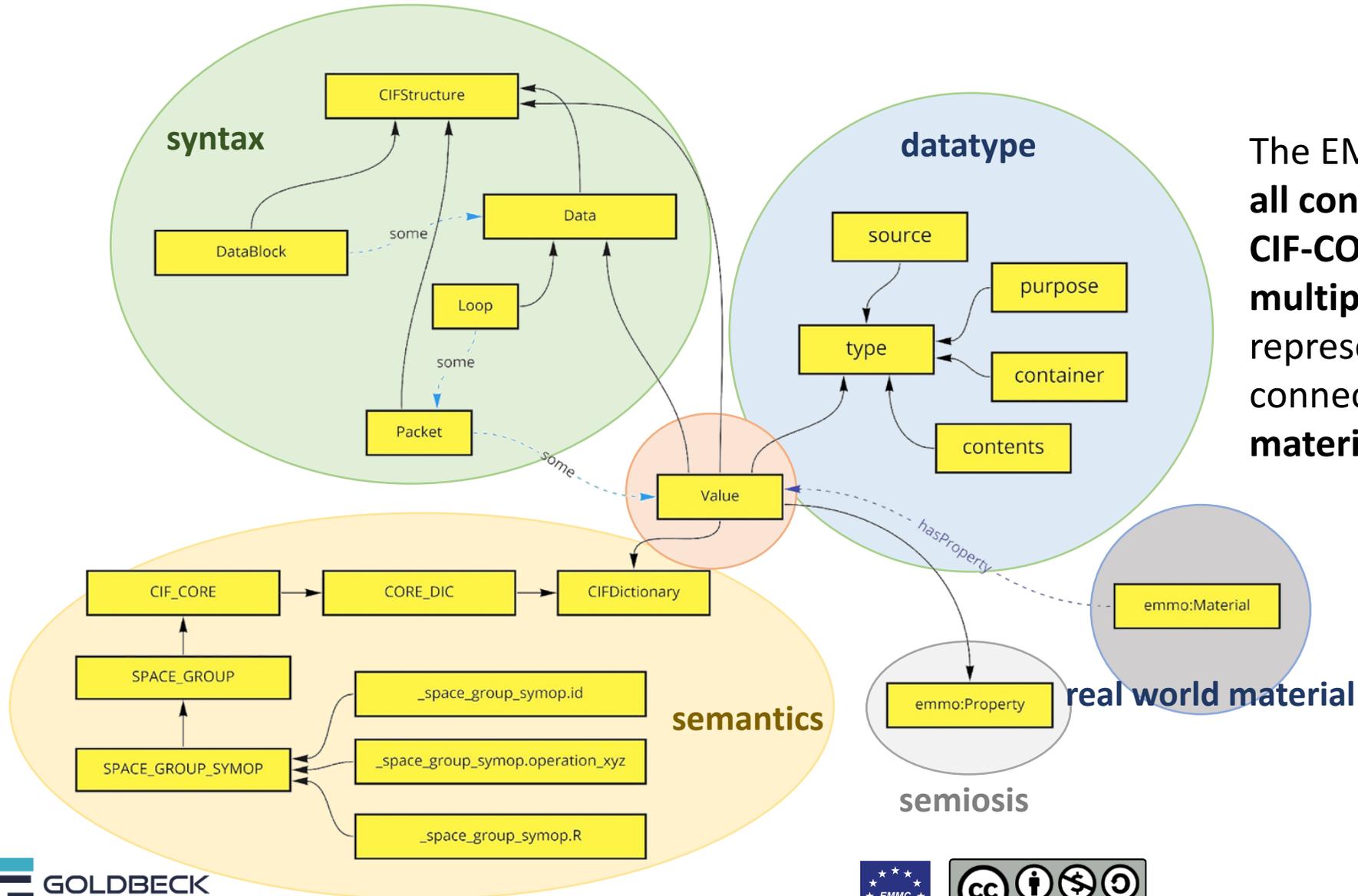
Battery Testing Industrial Use Case from Keysight

- Unified structure for organising knowledge and data in battery testing and battery quality control.
- Supports a range of electrical battery cell tests, including impedence spectroscopy, self-discharge, and high-voltage

Battery testing ontology: An EMMO-based semantic framework for representing knowledge in battery testing and battery quality control Computers in Industry, <https://www.sciencedirect.com/science/article/pii/S0166361524001313>

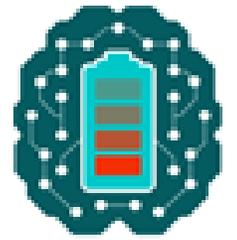


EMMO CIF: Crystallographic Ontology



The EMMO CIF ontology maps **all concepts of CIF-DDL and CIF-CORE** and provides a **multiperspective approach** to represent CIF data and to connect them to **real world materials**.

Clear separation of concerns: multiple perspectives



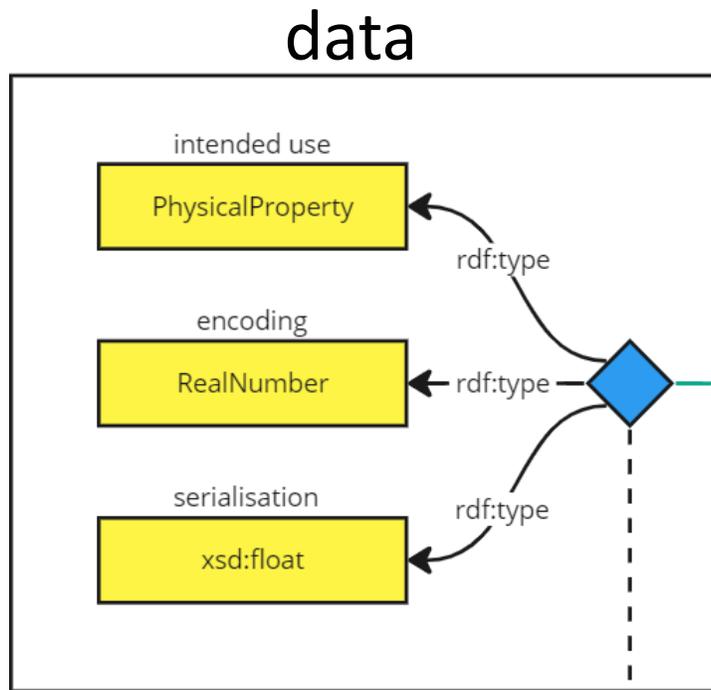
BIG-MAP

information



Battery Ontology

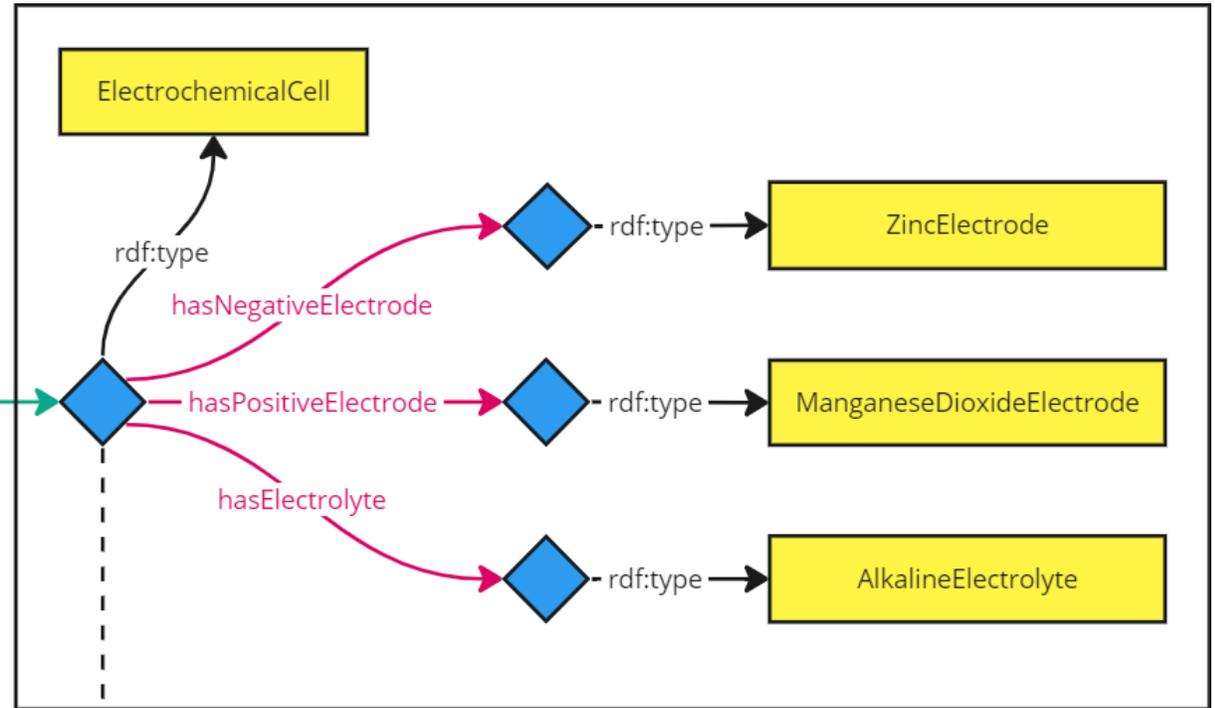
- Based on EMMO
- Representing batteries cells, materials, components, methods and properties



intended interpretation

0.1

physical entity (mereocausality, no data)

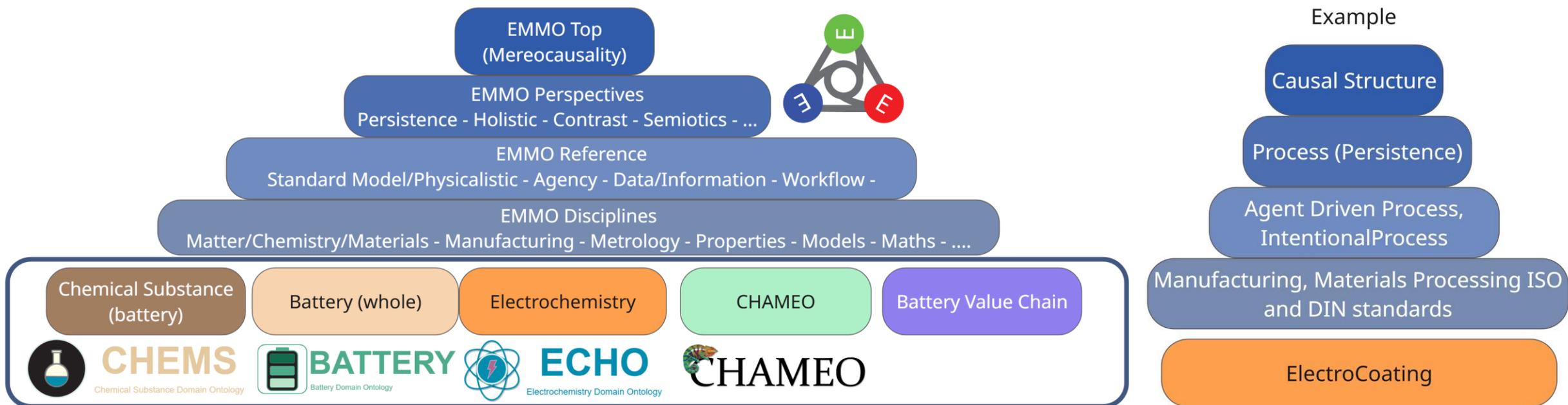


intended interpretation



<https://big-map.github.io/BattINFO/>

Battery Ontologies Ecosystem in EMMO



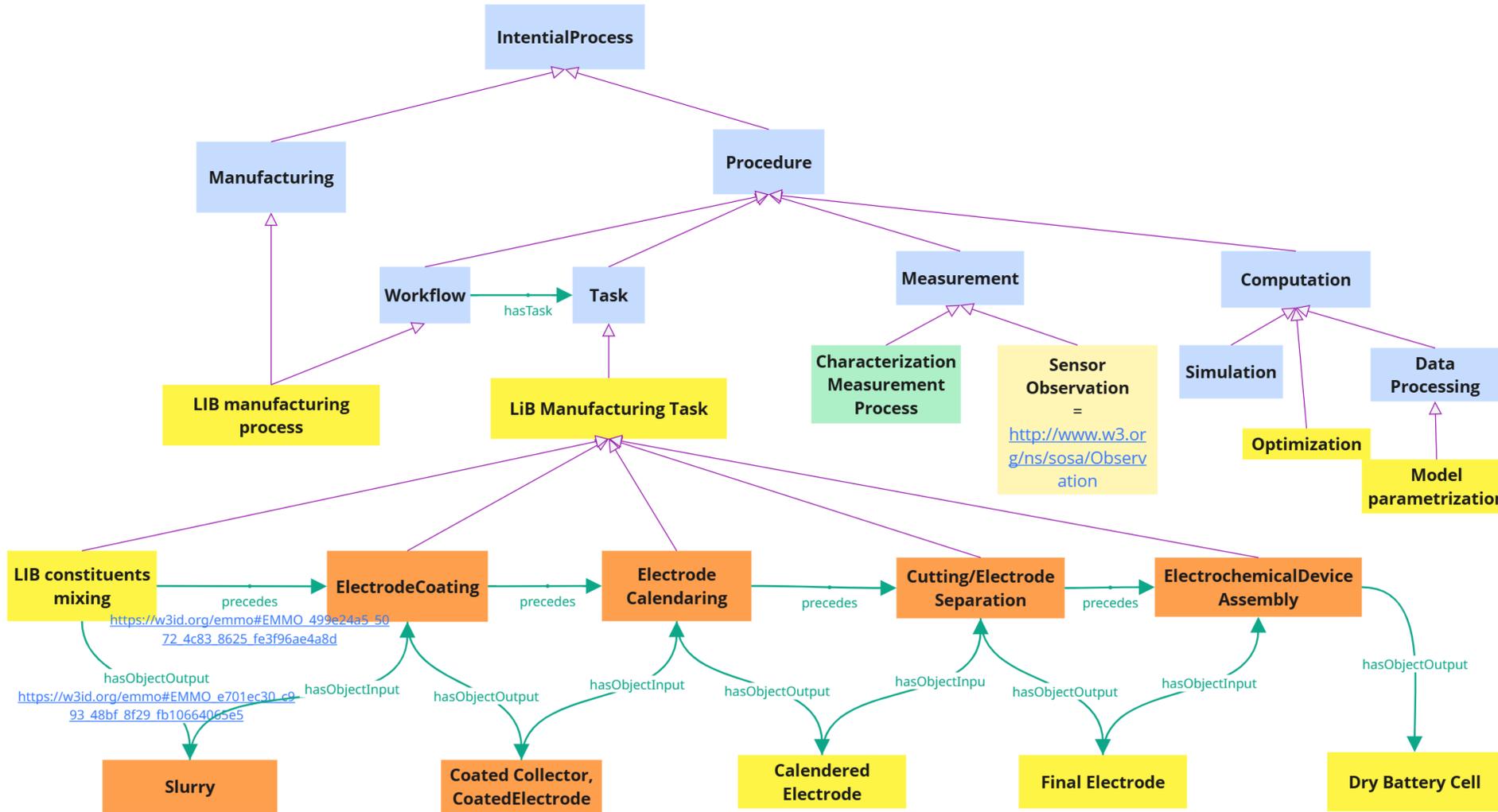
Plus alignment with: **Semantic Sensor Network Ontology**



References

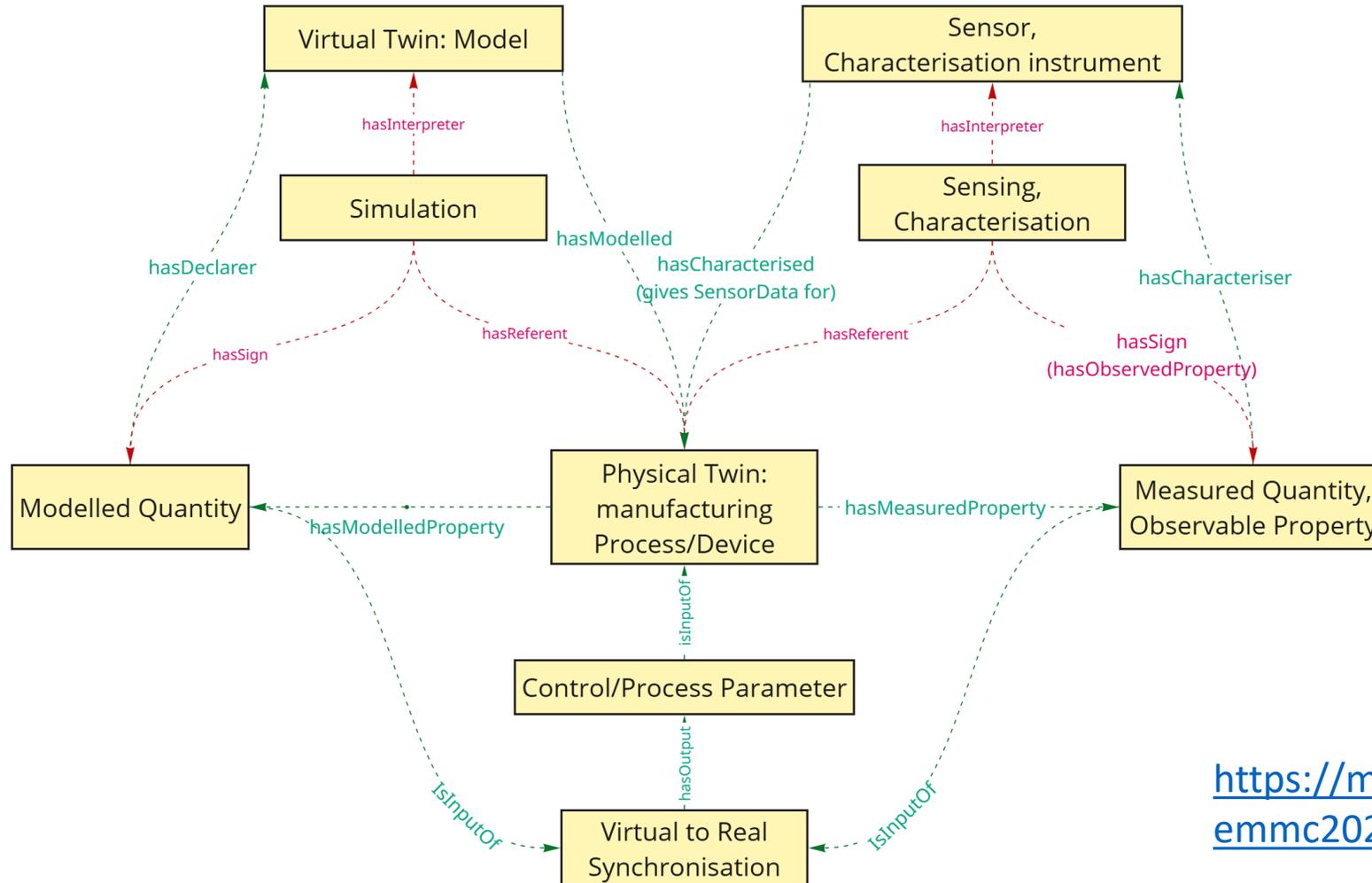
- [1] Battery Cell Assembly Twin (BatCAT) <http://www.batcat.info/>
- [2] Elementary Multiperspective Material Ontology, <https://github.com/emmo-repo/emmo>
- [3] Characterisation Methodology Domain Ontology (CHAMEO), <https://github.com/emmo-repo/domain-characterisation-methodology>
- [4] Electrochemistry Ontology <https://github.com/emmo-repo/domain-electrochemistry>
- [5] Battery Domain Ontology, <https://emmo-repo.github.io/domain-battery/index.html>
- [6] Semantic Sensor Network Ontology, <https://www.w3.org/TR/vocab-ssn/>
- [7] Chemical Substance Ontology, <https://github.com/emmo-repo/domain-chemical-substance>
- [8] Battery Value Chain Ontology, <https://github.com/Battery-Value-Chain-Ontology/ontology>

Battery Cell Assembly Digital Twin (BatCAT)



EMMO-based ontology design for BatCAT, re-using EMMO (blue), CHAMEO (green), Electrochemistry Domain Ontology (orange) to anchor BatCAT concepts (yellow).

Battery Cell Assembly Digital Twin (BatCAT)



<https://materialsmodelling.com/emmc2025-conference-wien/>

EMM Versions and domain ontologies

EMMO complete versions

Name	Link	Comment
emmo	https://w3id.org/emmo	Loads all EMMO modules, excluding the full standard model and specialised units
emmo-tol	https://w3id.org/emmo/tlo	EMMO top level
emmo-reference	https://w3id.org/emmo/reference	EMMO reference level
emmo-mlo	https://w3id.org/emmo/mlo	EMMO middle level
emmo-full	https://w3id.org/emmo/emmo-full	Loads all EMMO modules, including the full standard model and specialised units

EMMO simplified versions

Name	Link	Comment
emmo-for-humans	https://w3id.org/emmo/emmo-for-humans	Version of EMMO middle with IRIs replaced with human readable names. Only intended for examples.
emmo-lite	https://w3id.org/emmo/emmo-lite	Selected leaf classes and properties for rapid development and deployment in graph databases.
emmo-inferred	https://w3id.org/emmo/inferred	Pre-inferred version of EMMO middle level

Domain ontology	Base IRI
Characterisation Methodology Domain Ontology (CHAMEO)	https://w3id.org/emmo/domain/characterisation-methodology/chameo#
Electrochemistry	https://w3id.org/emmo/domain/electrochemistry#
Battery Ontology	https://w3id.org/emmo/domain/battery#
Battery Interface Ontology (BattINFO)	https://w3id.org/battinfo#
Nuclear Energy Ontology	https://w3id.org/emmo/domain/neo#
Nanoindentation Ontology	https://w3id.org/emmo/domain/nanoindentation#
Chemical Substance Domain Ontology	https://w3id.org/emmo/domain/chemical-substance#
Microstructure Domain Ontology	https://w3id.org/emmo/domain/microstructure#
OTE Interface Ontology (OTEIO)	https://w3id.org/emmo/domain/oteio#
Coating Domain Ontology	https://w3id.org/emmo/domain/coating#
Domain Ontology for Concrete	https://w3id.org/emmo/domain/concrete#
Domain ontology for solid oxide fuel cells	https://w3id.org/emmo/domain/sofc#
Domain ontology for equivalent circuit models	https://w3id.org/emmo/domain/equivalent-circuit-model#
Domain Ontology for Additive Manufacturing (DOAM)	https://w3id.org/emmo/domain/doam#
Atomistic and Electronic Modelling	https://w3id.org/emmo/domain/atomistic#
Crystallography	http://emmo.info/domain-crystallography/crystallography#
CIF ontology	http://emmo.info/CIF-ontology/ontology/cif_core#
Magnetic Materials Ontology (MaMMoS)	https://w3id.org/emmo/domain/magnetic_material#
General Process Ontology (GPO)	https://gpo.ontology.link/
Ontology for the Battery Value Chain (BVC)	https://bvco.ontology.link/

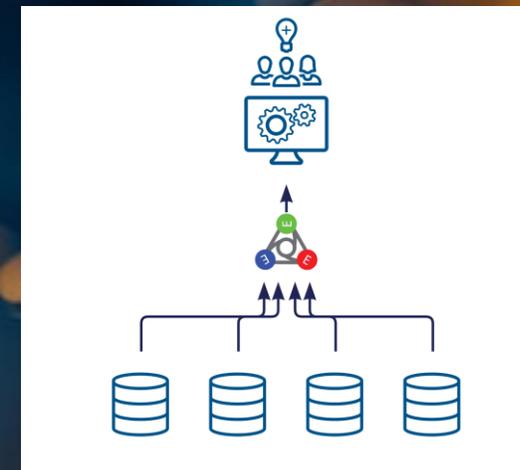
Semantic Materials

Semantic data and knowledge management

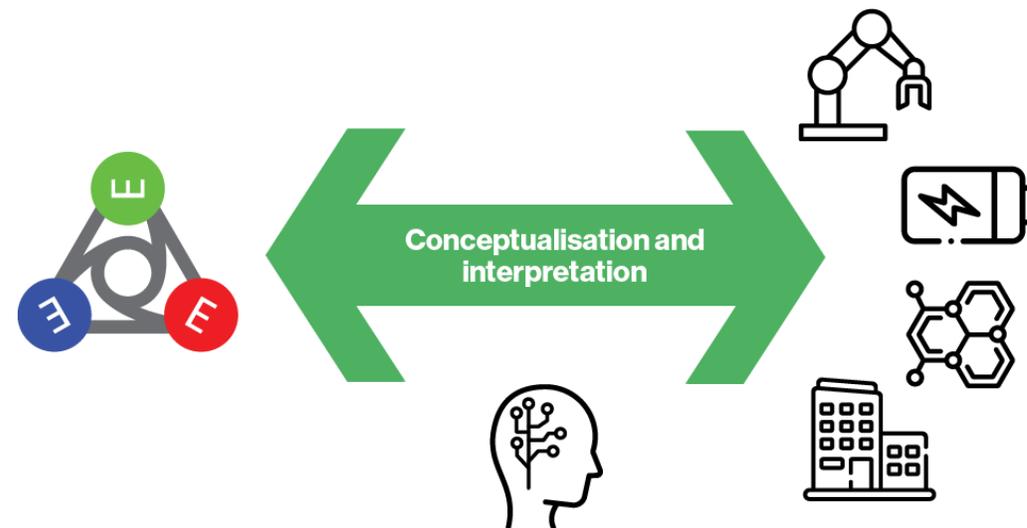
for chemicals and materials-based industries

OUR CONSORTIUM

OUR SOLUTIONS



- A team of EMMO authors and experts
- Technology partners



Outlook: Materials Commons

General information

Programme

Horizon Europe (HORIZON)

Call

INDUSTRY (HORIZON-CL4-2025-01)

Type of action

HORIZON-IA HORIZON Innovation Actions

Type of MGA

HORIZON Action Grant Budget-Bas

Deadline model

single-stage

Opening date

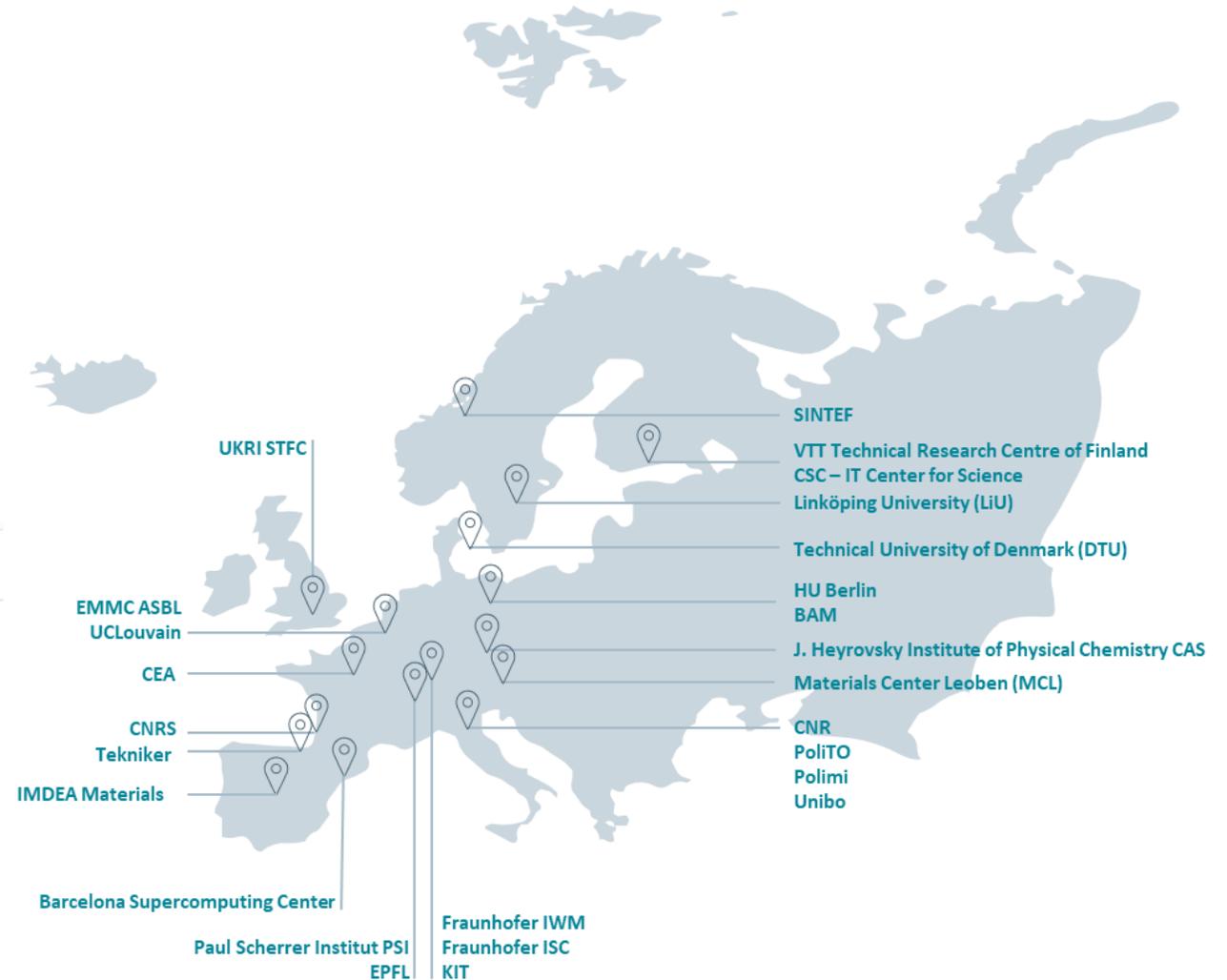
22 May 2025

ased [HORIZON-AG]

 Open For Submission

Deadline date

23 September 2025 17:00:00 Brussels time



<https://materialscommons4.eu/>

Related Initiatives

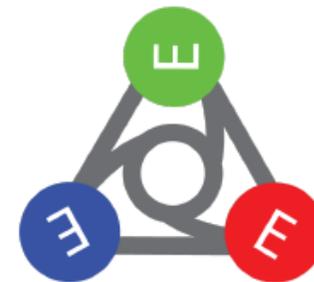
The EMMO core development group:

- Emanuele Ghedini (UNIBO)
- Gerhard Goldbeck (GCL)
- Jesper Friis (SINTEF)
- Adham Hashibon (UCL)
- Georg Schmitz (ACCESS)
- Anne de Baas (GCL)
- Francesco Zaccarini (UNIBO)
- Sebastiano Moruzzi (UNIBO)
- Francesca Lønstad Bleken (SINTEF)
- Simon Clark (SINTEF)
- Otello Roscioni (ex-GCL)

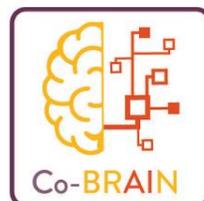
And many more...



Elementary
Multiperspective
Material
Ontology



<https://emmc.eu/emmo/>



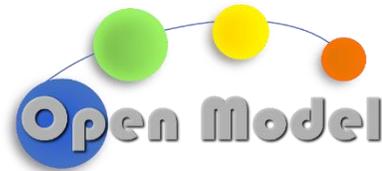
nanomecommons



Acknowledgement

Funding from the European Commission via the Horizon 2020 projects

- OntoCommons (GA n. 958371) www.ontocommons.eu
- OntoTrans (GA n. 952869) www.ontotrans.eu
- OpenModel (GA n. 953167) www.openmodel.eu
- NanoMECommons (GA n. 952869) www.nanoMECommons.eu



Funding by the UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee [GA 10091190]:

- **Battery Cell Assembly Twin (BatCAT) <http://www.batcat.info/>**



- **CONNECT-NM <https://www.connect-nm.eu/>**

