Interoperable CHADA: A Semantic Approach for Managing and Exploiting Characterization Data and Protocols

Yoav Nahshon
On behalf of the AID4GREENTST Consortium

MatCHMaker workshop, 07.04.2025

AID4GREENEST

AID4GREENEST
Revolutionizing Steel Manufacturing with AI & Digitalization

- **AI powereD characterization and modelling for GREEn STeel technology**

- Challenges in traditional steel production:
  - High costs, inefficiencies, and material waste due to trial-and-error methods

- AI-Powered Transformation:
  - AI-driven tools optimize steel design, processing, and performance for faster, smarter, and greener manufacturing.

- Key Impact:
  - Reduced material waste & emissions
  - AI-powered process optimization
  - Standardized & interoperable data management
  - Enhanced sustainability in steel production

Co-funded by the European Union

AID4GREENEST

Task 5.1 - Standardizing Documentation & Workflows

- Challenges:
  - Lack of structured documentation for experiments
  - Inconsistent workflows lead to poor reproducibility
  - Absence of standardized metadata affects knowledge sharing

- Objectives of Task 5.1
  - Collect and unify experimental data across project partners
  - Standardize and assist documentation using CHADA (Characterization) & MODA (Modeling)
  - Ensure metadata completeness for reproducibility
  - Create a shared vocabulary to standardize terminology across domains

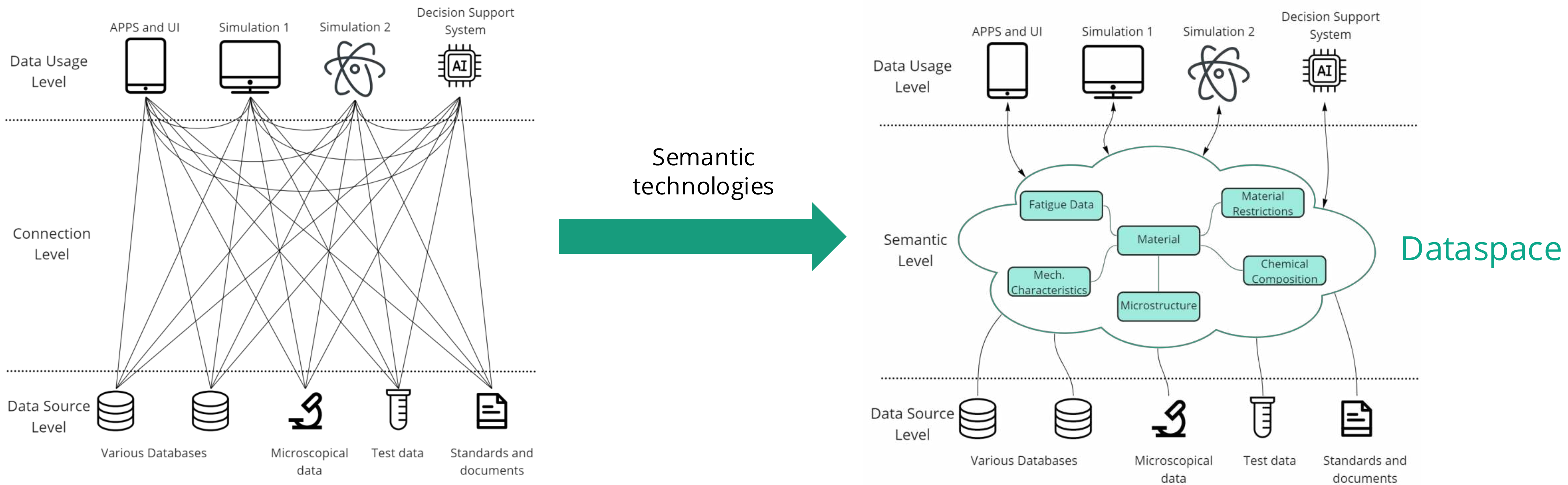
- Key tools developed
  - **Vocabulary manager:** Standardized repository for terminology & metadata attributes
  - **Process editor digital tool:** – Web app for documenting processes in a FAIR manner.
    - Particularly: characterization workflow in accordance with the CHADA guidelines
    - https://www.cencenelec.eu/media/CEN-CENELEC/CWAs/RI/2025/cwa17815_2025.pdf

Co-funded by the European Union

Semantics management

The importance of FAIR data management for SMEs
What is FAIR data and why it is important

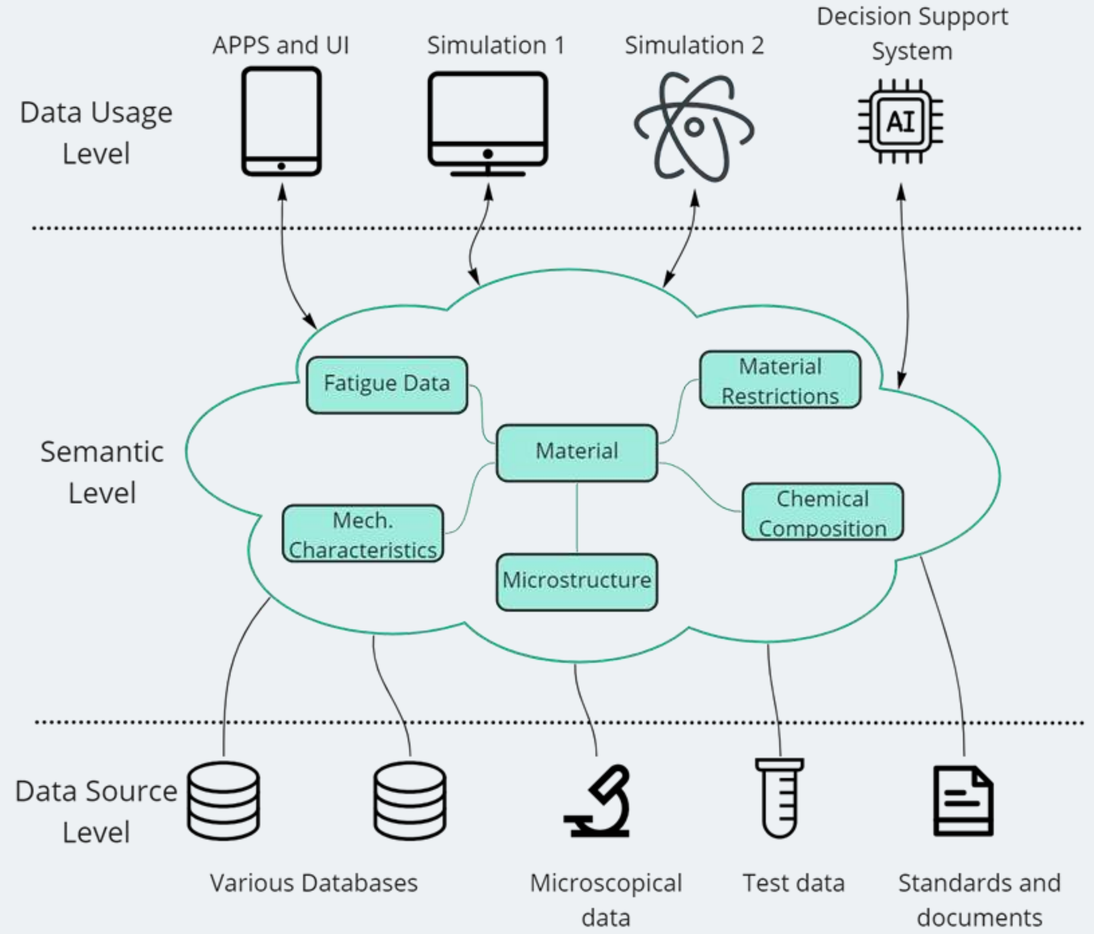
- FAIR: Findable, Accessible, Interoperable, Reusable
- Often data is stored in file structures or relational databases within an organization → not FAIR



Co-funded by the European Union

Data-driven solutions with FAIR data
Strengths

- Interoperability enables seamless communication
- Full data access fosters deeper insights and better decisions
- Less data conversion streamlines processes and boosts efficiency
- FAIR data is AI-ready



Semantic technologies for FAIR data management
Knowledge graphs

Information is linked and stored in a knowledge graph

Every term is unique and defined, e.g.

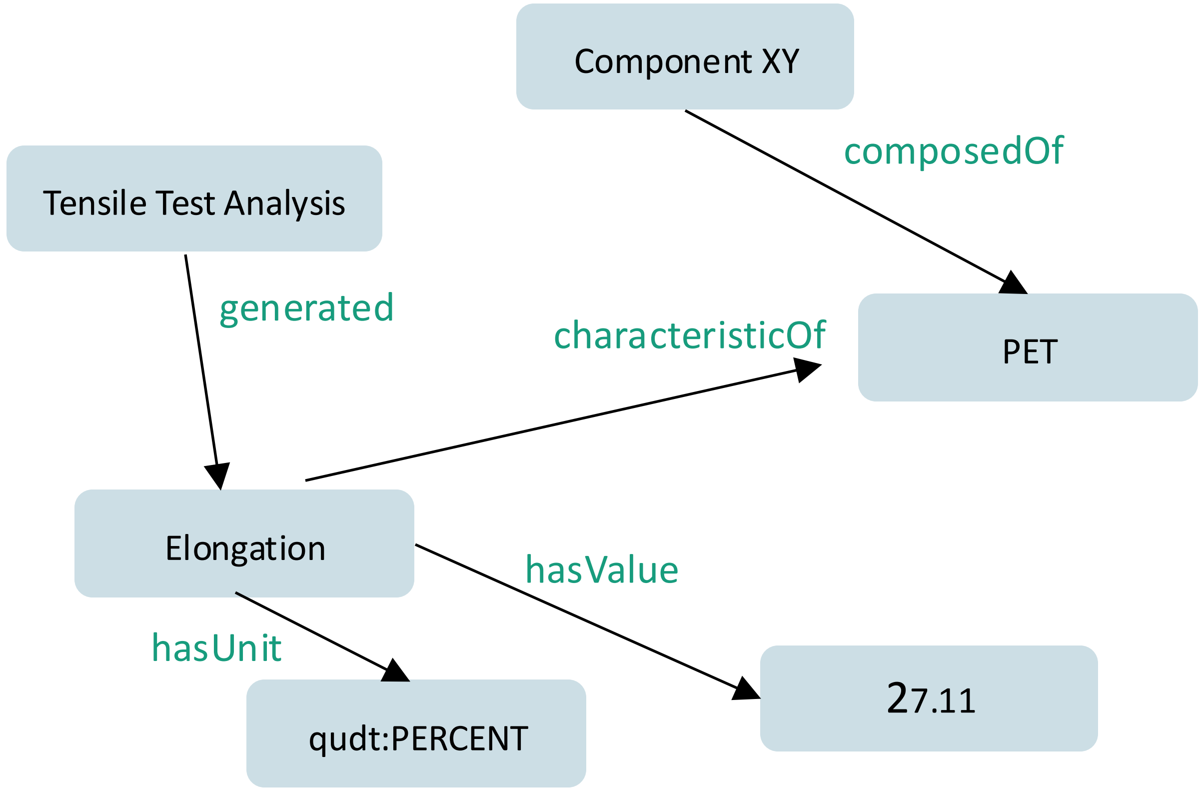
URI: http://qudt.org/vocab/unit/PERCENT

Description

"Percent" is a unit for 'Dimensionless Ratio' expressed as %.

Human and machine readable

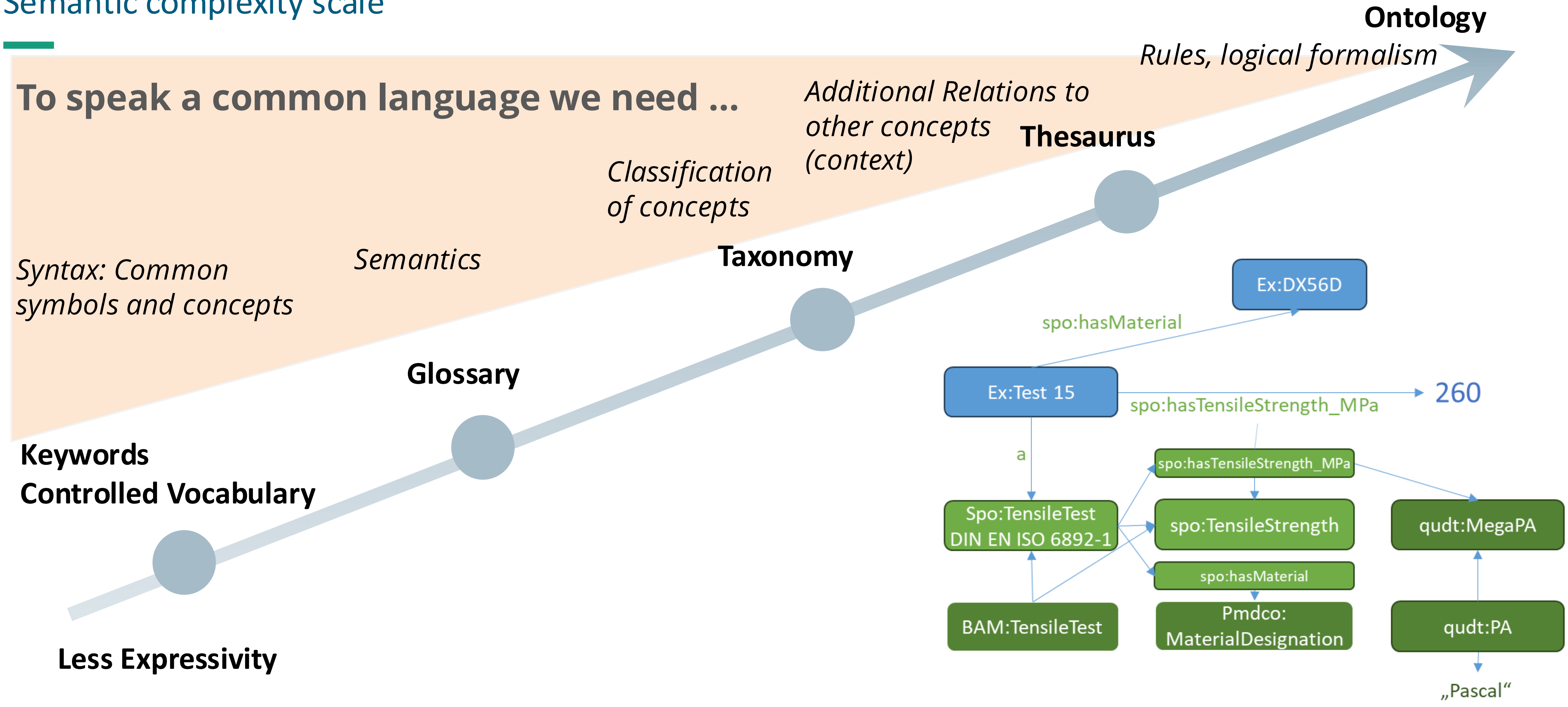
But: Working with semantic technologies is challenging



Co-funded by the European Union

Semantic technologies for FAIR data management

Semantic complexity scale

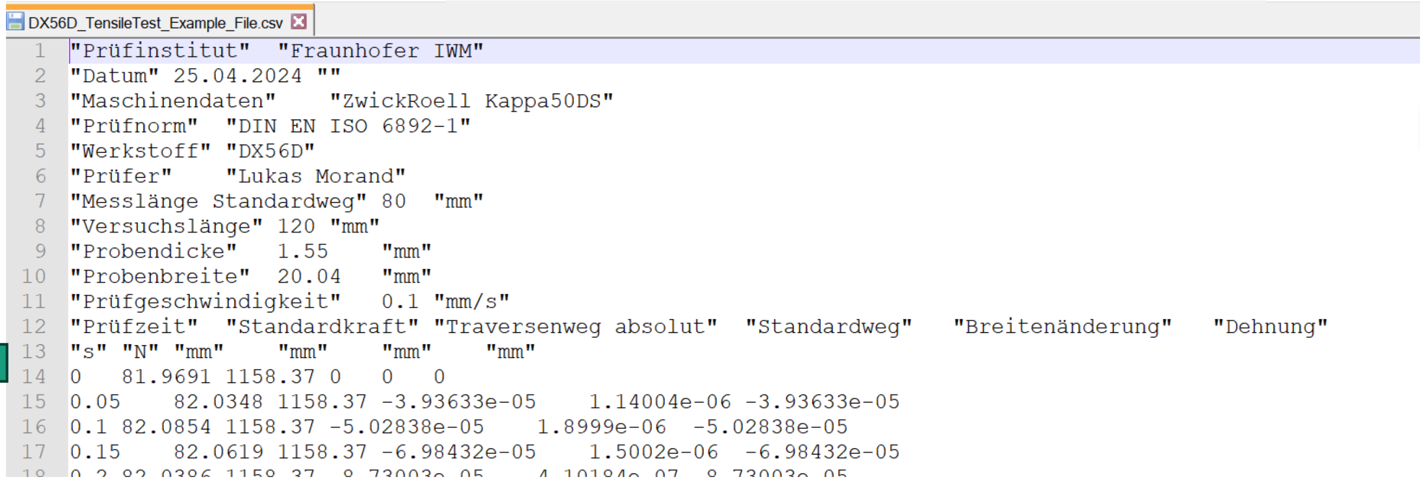


Semantic technologies for FAIR data management

Approaches

Different development philosophies :

- Top-down: from generic to specific
- Bottom-up: from the practical to the more general
- Middle-Out: start in the middle and add where necessary

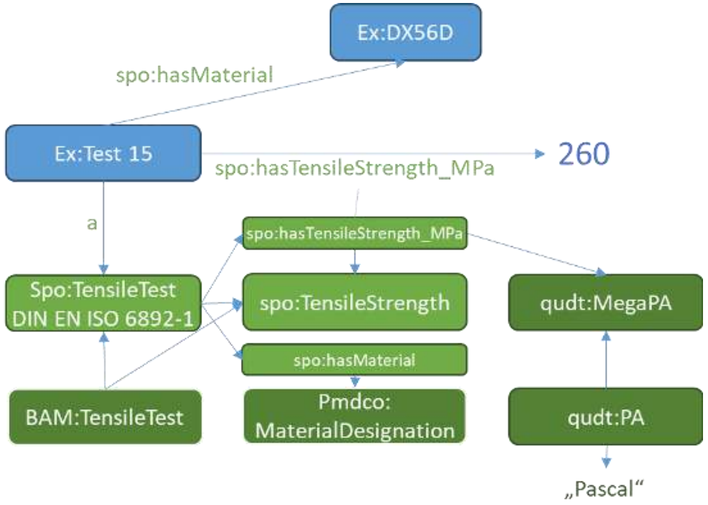


DX56D_TensileTest_Example_File.csv

```
"Prüfinstitut"	"Fraunhofer IWM"
"Datum"	25.04.2024	""
"Maschinendaten"	"ZwickRoell Kappa50DS"
"Prüfnorm"	"DIN EN ISO 6892-1"
"Werkstoff"	"DX56D"
"Prüfer"	"Lukas Morand"
"Messlänge Standardweg"	80	"mm"
"Versuchslänge"	120	"mm"
"Probendicke"	1.55	"mm"
"Probenbreite"	20.04	"mm"
"Prüfgeschwindigkeit"	0.1	"mm/s"
"Prüfzeit"	"Standardkraft"	"Traversenweg absolut"	"Standardweg"	"Breitenänderung"	"Dehnung"
"s"	"N"	"mm"	"mm"	"mm"	"mm"
0	81.9691	1158.37	0	0	0
0.05	82.0348	1158.37	-3.93633e-05	1.14004e-06	-3.93633e-05
0.1	82.0854	1158.37	-5.02838e-05	1.8999e-06	-5.02838e-05
0.15	82.0619	1158.37	-6.98432e-05	1.5002e-06	-6.98432e-05
```

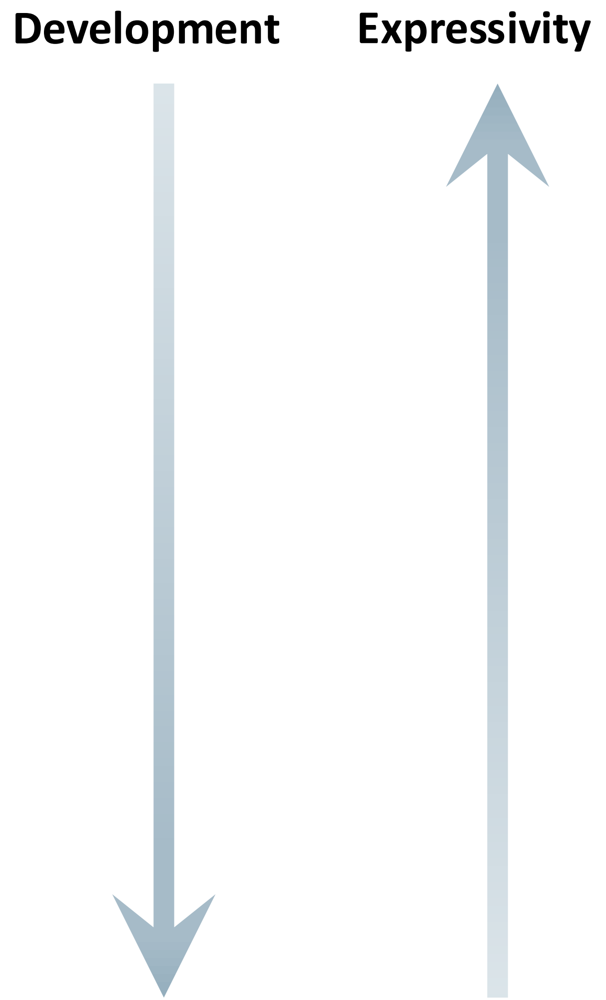
Various tools:

- Protégé / WebProtégé
- Chowlk / RDF Diagram Framework
- Metafactory, Corporate Memory, GraphDB, AllegroGraph, Virtuoso, PoolParty, …
- Xlsx2owl
- Vocabulary manager



Co-funded by the European Union

Ontology development approaches
Top-down



Start from broad, high-level concepts and progress to more specific details

Process:

Define the most generic concepts, such as "Entity" or "Object."

Gradually refine these concepts into increasingly detailed subclasses or properties.

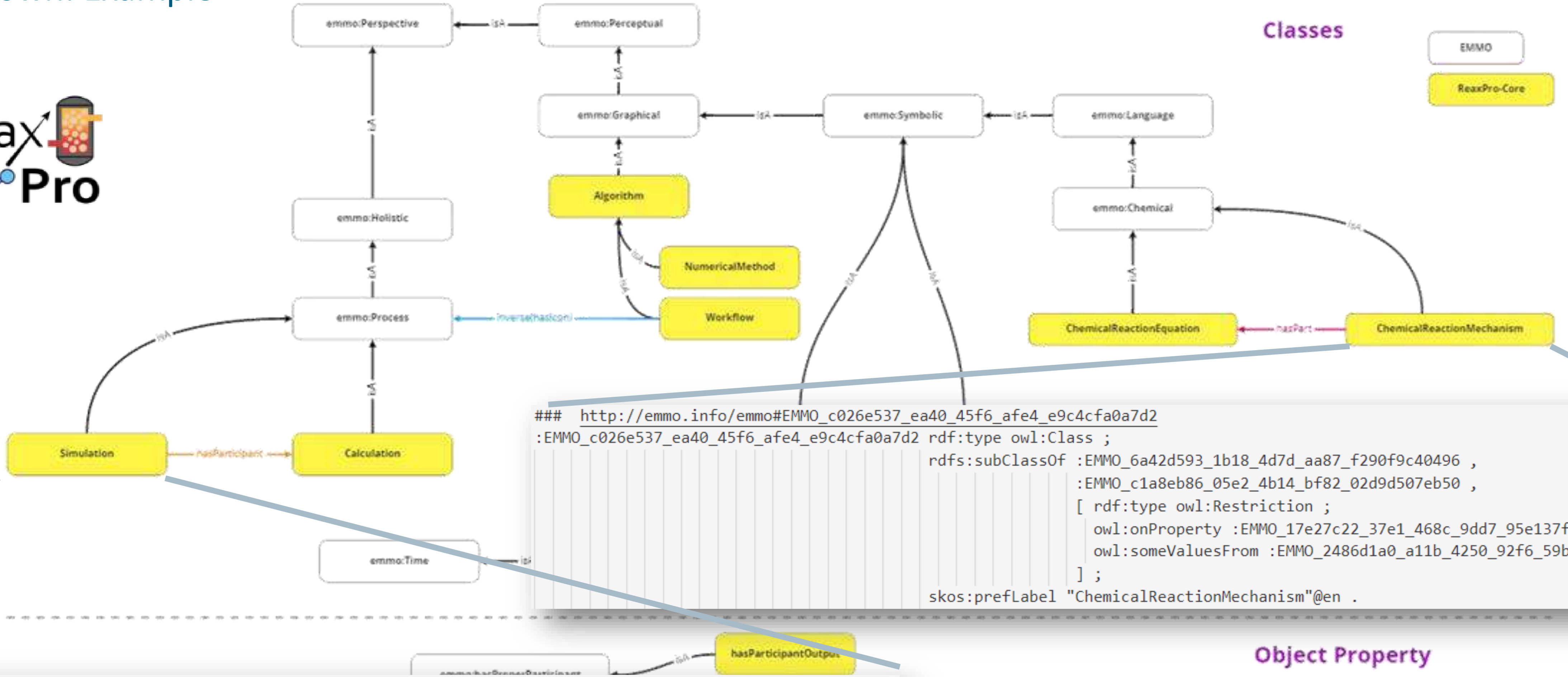
Ensure consistency and logical relationships between levels.

Co-funded by the European Union

Ontology development approaches

Top down: Example

ReaxPro

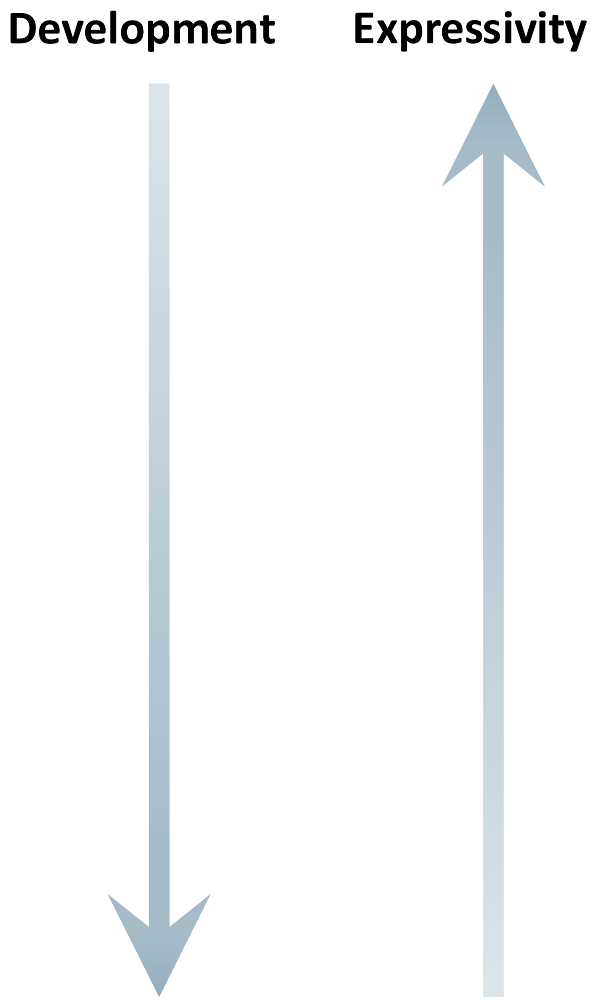



```
###  http://emmo.info/emmo#EMMO_c026e537_ea40_45f6_afe4_e9c4cfa0a7d2
:EMMO_c026e537_ea40_45f6_afe4_e9c4cfa0a7d2 rdf:type owl:Class ;
                                           rdfs:subClassOf :EMMO_6a42d593_1b18_4d7d_aa87_f290f9c40496 ,
                                                           :EMMO_c1a8eb86_05e2_4b14_bf82_02d9d507eb50 ,
                                                           [ rdf:type owl:Restriction ;
                                                             owl:onProperty :EMMO_17e27c22_37e1_468c_9dd7_95e137f73e7f ;
                                                             owl:someValuesFrom :EMMO_2486d1a0_a11b_4250_92f6_59ba8886473f
                                                           ] ;
                                           skos:prefLabel "ChemicalReactionMechanism"@en .
```

```
###  http://emmo.info/emmo#EMMO_4d947635_13a5_4671_b631_658d3f103b3b
:EMMO_4d947635_13a5_4671_b631_658d3f103b3b rdf:type owl:Class ;
                                           rdfs:subClassOf :EMMO_43e9a05d_98af_41b4_92f6_00f79a09bfce ,
                                                           :EMMO_c1a8eb86_05e2_4b14_bf82_02d9d507eb50 ,
                                                           [ rdf:type owl:Restriction ;
                                                             owl:onProperty :EMMO_ae2d1a96_bfa1_409a_a7d2_03d69e8a125a ;
                                                             owl:someValuesFrom :EMMO_790d1d4e_15cc_4e55_b5da_9e2efceecd4f
                                                           ] ;
                                           rdfs:comment "A \"collection\" of calculations with a specific objective (output)." ;
                                           skos:prefLabel "Simulation"@en .
```

Ontology development approaches

Top-down: Trade-off analysis



Challenges:

Time-to-application: Takes typically half a year until practical use of the application is being reached

Technoligcal barrier: Requires ontology experts

Development time: No parallel work; long iteration cycles between ontology experts and domain experts

Advantages:

Ensures a well-structured, hierarchical foundation, consistent with a particular framework

Provides a comprehensive, global perspective

Encourages reusability of ontology for various domains

Co-funded by the European Union

Ontology development approaches

Bottom-up

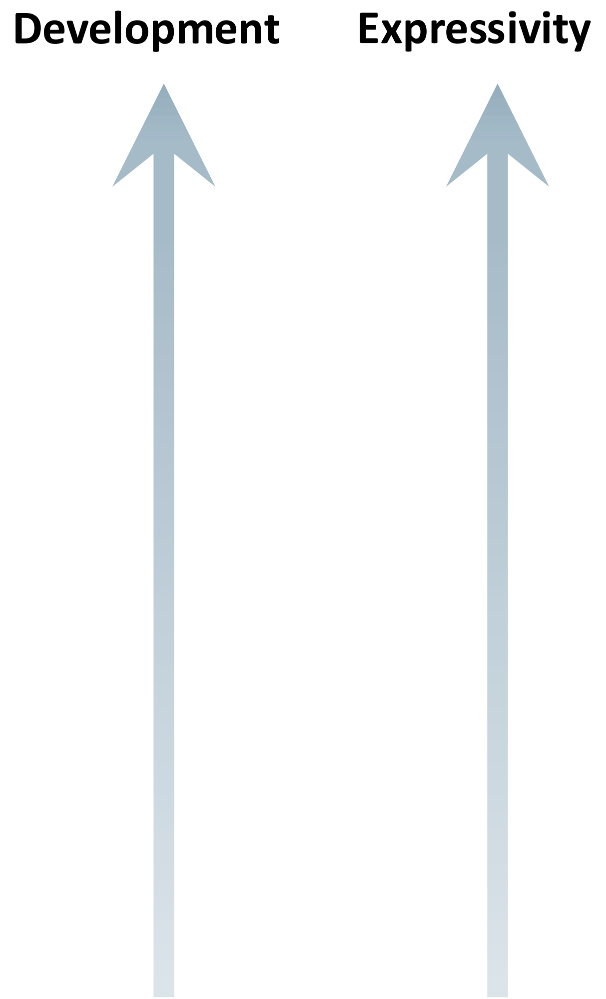
Begin with specific, detailed concepts or instances and generalize to broader categories

Process:

Identify real-world data, examples, or use cases

Organize these instances into classes and relationships

Generalize and abstract broader categories from the specifics.



Ontology development approaches

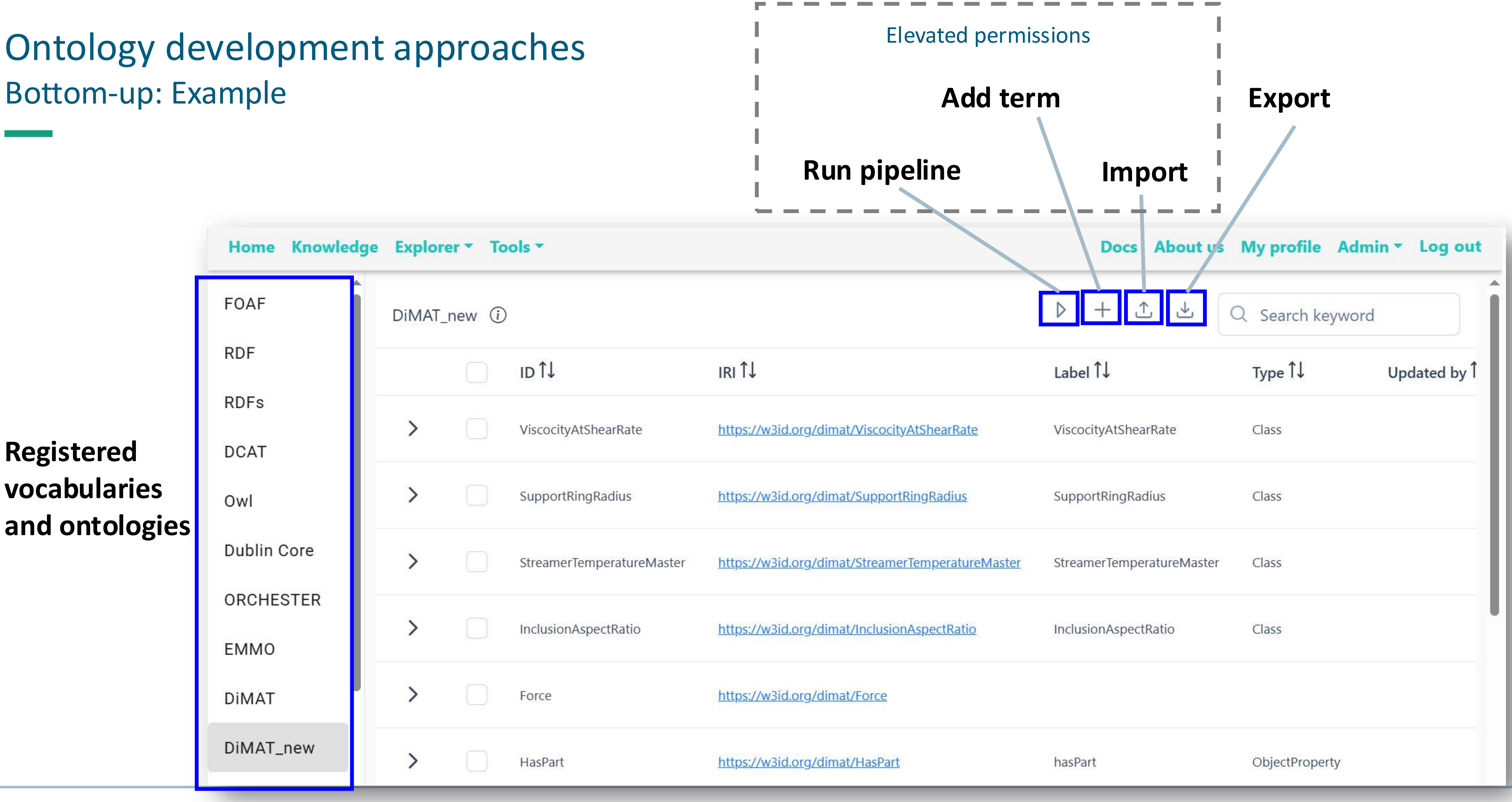
Bottom-up: Example

ID	Label	Type	Description	Comment
SpinneretCapillaries	SpinneretCapillaries	http://www.w3.org/2002/07/owl:	Number of capillaries found in the spinneret, equal to the number of filaments of the yarn obtained	
SpinneretCapillariesDiameter	SpinneretCapillariesDiameter	http://www.w3.org/2002/07/owl:	Diameter of the capillaries found in the spinneret	
SpinneretCapillariesLengthToDiameterRatio	SpinneretCapillariesLengthToDiameterRatio	http://www.w3.org/2002/07/owl:	Ratio between the lenght and the diameter of the capillaries found in the spinneret	L/D
SpinPackMeshSize	SpinPackMeshSize	http://www.w3.org/2002/07/owl:	Pore size of the wire mesh in the filter in micrometers. Indicates the smallest particle that can pass through the filter.	
SpinPackSandBedDepth	SpinPackSandBedDepth	http://www.w3.org/2002/07/owl:	Vertical depth of the sand bed in centimeters. Indicates the height of the sand column used in the filter	
SpinPackSandGrainSize	SpinPackSandGrainSize	http://www.w3.org/2002/07/owl:	Diameter of the individual grains of sand	
SpinPackTemperature	SpinPackTemperature	http://www.w3.org/2002/07/owl:	Temperature of the SpinPack	
PumpSpeed	PumpSpeed	http://www.w3.org/2002/07/owl:	speed of the pump	2,92cm3/rpm
ExtruderSpeed	ExtruderSpeed		speed at which the extruder	

Co-funded by the European Union

Ontology development approaches

Bottom-up: Example

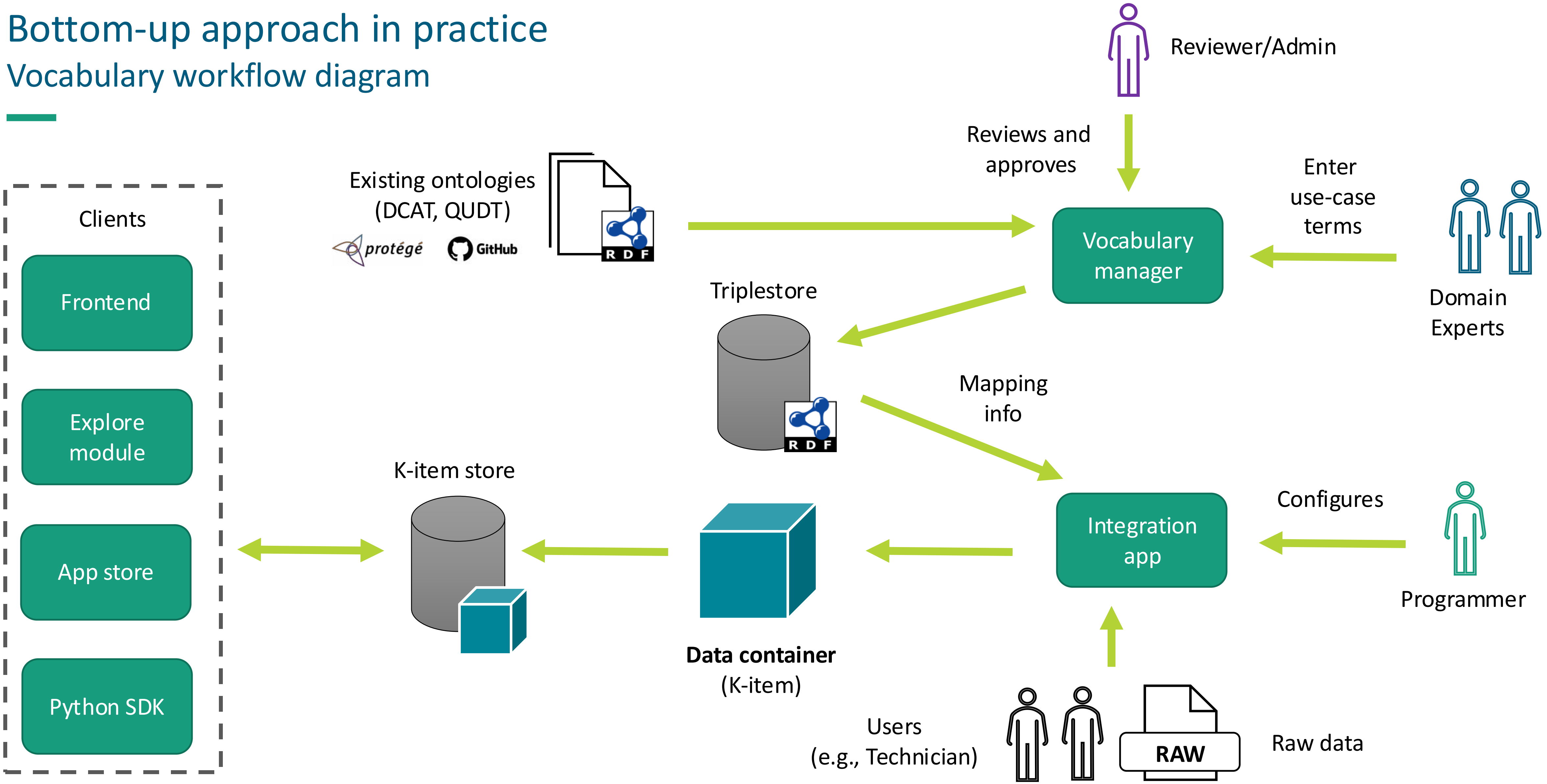


Co-funded by the European Union

Fraunhofer IWM

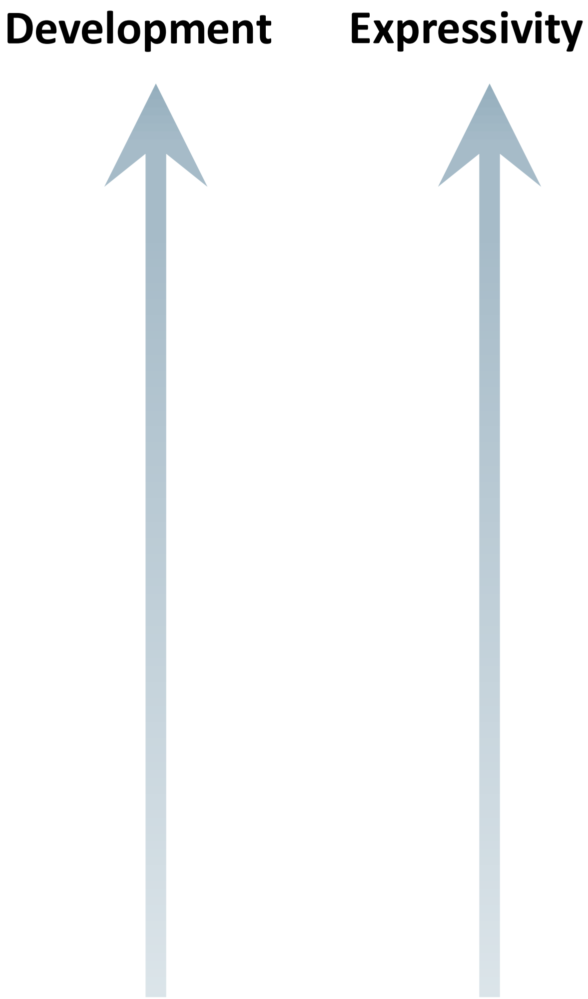
Bottom-up approach in practice

Vocabulary workflow diagram



Ontology development approaches

Bottom-up: Trade-off analysis



Advantages:

Agile development: Application development is reached very fast; updates are done on the fly

Reduces bottlenecks by breaking down complexity into modules → The user is independent

Improves transparency in the system as vocabularies as more easily understandable than ontologies

Challenges:

Results in duplications and inconsistencies → Requires quality control

Requires effort to align with upper level ontologies

Lacking the necessary tools and modules

Delegates vocabulary managements to users, which need to be guided

Co-funded by the European Union

Bottom-up approach in practice
Upcoming CWA

CEN Workshop: Vocabulary definition for domain ontologies in materials science and manufacturing

- Scope: CWA to outline a guideline and workflow for identifying and defining terminology related to materials science applications, aimed at facilitating the creation of a domain ontology
- Motivation: Address the challenge of inaccessible heterogeneous data by promoting FAIR data management.
- Bottom-up ontology creation: Facilitates practical application and relevant term identification.
- Aim of CWA: Establish guidelines for defining terminology in application ontologies within materials science.
- Overall goal: Enable consistent and effective development of application ontologies to enhance interoperability.

Co-funded by the European Union

Take-home message
Semantics management

Top-down:

👎 Usually long iteration cycles between ontology experts and domain experts

→ Often creates bottleneck, can hinder progress and demotivate end uses

👍 Ensures a well-structured, hierarchical foundation, consistent with a particular framework

Bottom up:

👎 Introduces duplicates and inconsistencies → Requires a review process (automation, via NLP / LLMs)

👍 Fast/agile that delegates the tasks to domain experts

→ But experts need guidance (CWA) and tools (vocabulary manager module)

Vocabulary Manager Tool:

- Collects and organizes vocabulary terms across projects/domains
- Serves as a mapping system for metadata attributes, ensuring consistency

Co-funded by the European Union

Semantic Representation of Characterization Processes

Challenges in curating characterization data

Tensile test example

```
"Prüfinstitut"  "Fraunhofer IWM"
"Projektnummer" "142003"
"Projektname"   "3D-Blechmodelle2"
"Datum/Uhrzeit" "16.04.2016 13:53"
"Maschinendaten"        "ZwickRoell Kappa50DS"
"Kraftaufnehmer"        "xForce K"
"Wegaufnehmer"  "makroXtens"
"Prüfnorm"      "DIN EN ISO 6892-1"
"Werkstoff"     "DX56D"
"Probentyp"     "FZ2 (L0=80_b0=20_R20)"
"Prüfer"        "wes"
"Probenkennung 2"       "DX56_D_FZ2_WR00_43"
"Messlänge Standardweg" 80      "mm"
"Versuchslänge" 120     "mm"
"Probendicke"   1.55    "mm"
"Probenbreite"  20.04   "mm"
"Prüfgeschwindigkeit"   0.1     "mm/s"
"Vorkraft"      2       "MPa"
"Temperatur"    22      "°C"
"Bemerkung"     ""
"Prüfzeit"      "Standardkraft" "Traversenweg absolut"  "Standardweg"   "Breitenänderung"       "Dehnung"
"s"     "N"     "mm"    "mm"    "mm"    "%"
0       81.9691 1158.37 0       0       0
0.05    82.0348 1158.37 -3.93633e-05    1.14004e-06     -3.93633e-05
0.1     82.0854 1158.37 -5.02838e-05    1.8999e-06      -5.02838e-05
0.15    82.0619 1158.37 -6.98432e-05    1.5002e-06      -6.98432e-05
0.2     82.0386 1158.37 -8.73003e-05    4.10184e-07     -8.73003e-05
0.25    82.0924 1158.37 -9.95444e-05    9.4993e-07      -9.95444e-05
0.3     82.1014 1158.37 -0.000103205    1.49995e-06     -0.000103205
0.35    82.1089 1158.37 -9.91844e-05    1.75019e-06     -9.91844e-05
0.4     82.1103 1158.37 -0.000109016    9.60408e-07     -0.000109016
```

Metadata (key-value pairs)

Measurement data

Co-funded by the European Union

Challenges in curating characterization data

Tensile test example

```
"Prüfinstitut"   "Fraunhofer IWM"
"Projektnummer"  "142003"
"Projektname"    "3D-Blechmodelle2"
```

⚠ heterogeneity:

- **Structural heterogeneity:** ASCII on top of TSV
- **Semantic heterogeneity:** Terms are specified in a particular language

⚠ Inconsistencies in experimental procedures (e.g., sample preparation)

⚠ Variations in executions (e.g., displacement-controlled vs. load-controlled)

⚠ Reproducibility: Is all necessary information available?

⚠ Collaboration:

- Confidence in shared data
- Non-uniform protocols

```
0.25    82.0924 1158.37 -9.95444e-05     9.4993e-07     -9.95444e-05
0.3     82.1014 1158.37 -0.000103205     1.49995e-06    -0.000103205
0.35    82.1089 1158.37 -9.91844e-05     1.75019e-06    -9.91844e-05
0.4     82.1103 1158.37 -0.000109016     9.60408e-07    -0.000109016
```

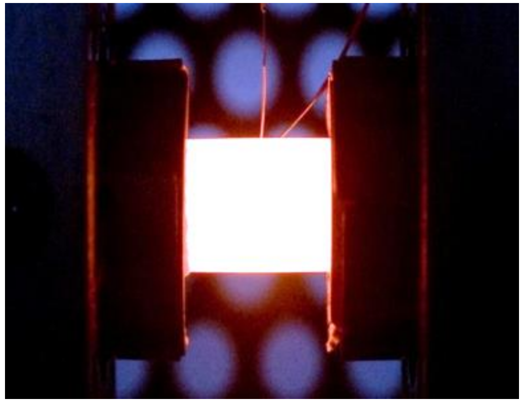
Measurement data

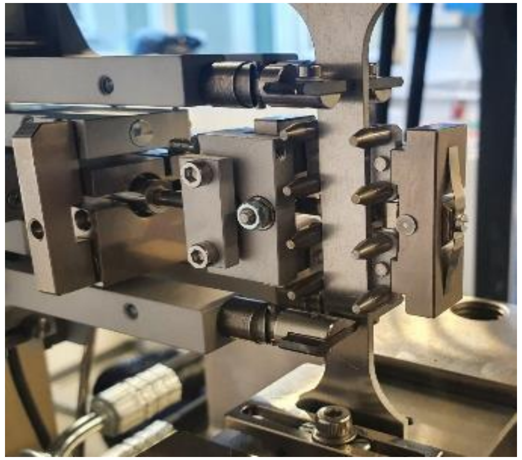
Co-funded by the European Union

Benefits of standardized terminology and documentation

Motivation

- **Enhanced data FAIRness**
  - Making data Findable, Accessible, Interoperable, and Reusable unlocks the value hidden within
  - Standardized documentation ensures transparent reporting of parameters and methods
  - Common frameworks (e.g., CHADA) enable seamless data exchange
- **Support for advanced analytics**
  - Machine-readable documentation aids data mining, AI training, and digital twin development, driving innovation
- **Accelerates technology transfer**
  - Standard protocols bridge academia and industry, speeding material development and commercialization



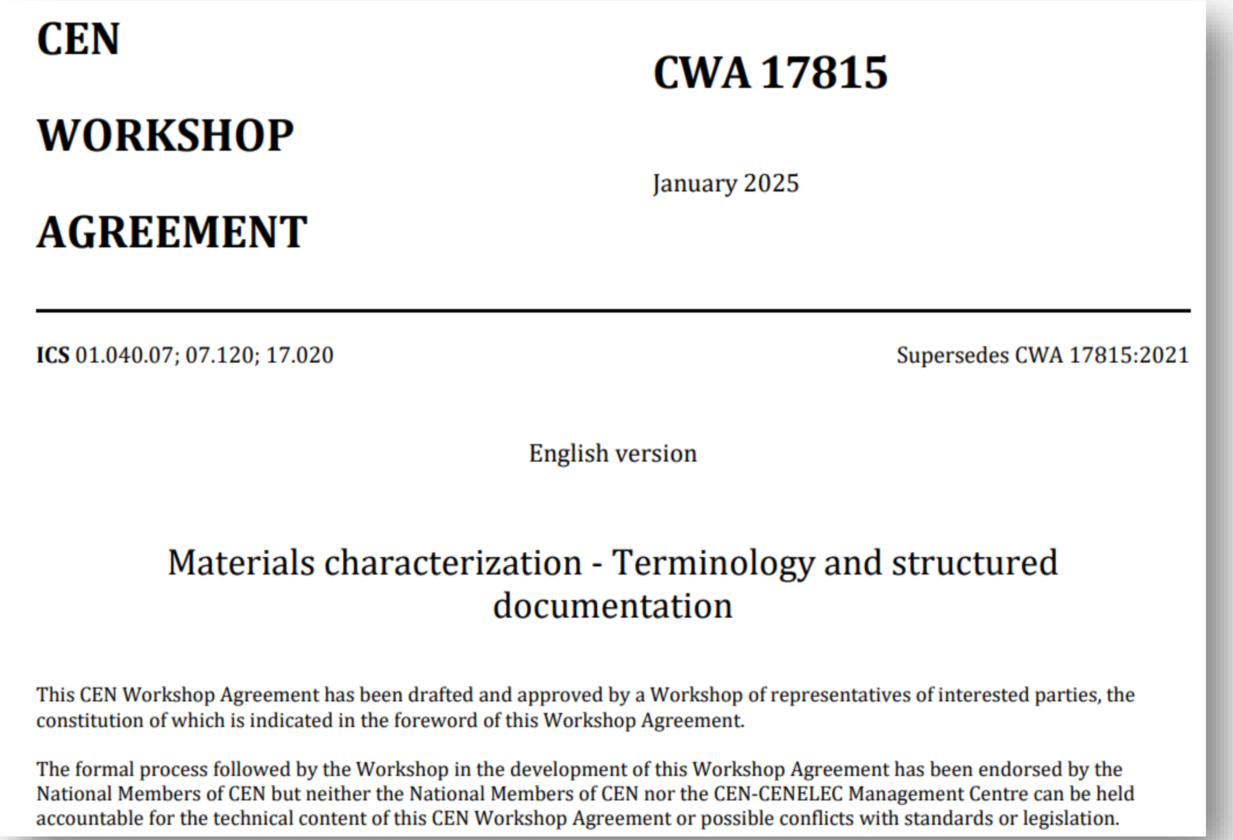


Co-funded by the European Union

CHADA

Introduction

- A framework that provides a standard set of concepts, metadata, and terminology for materials characterization
- Ensures documentation follows a structured, reproducible workflow.
- Covers materials, sample preparation, experimental procedures, and analysis workflows.



CEN

WORKSHOP

AGREEMENT

CWA 17815

January 2025

ICS 01.040.07; 07.120; 17.020

Supersedes CWA 17815:2021

English version

Materials characterization - Terminology and structured documentation

This CEN Workshop Agreement has been drafted and approved by a Workshop of representatives of interested parties, the constitution of which is indicated in the foreword of this Workshop Agreement.

The formal process followed by the Workshop in the development of this Workshop Agreement has been endorsed by the National Members of CEN but neither the National Members of CEN nor the CEN-CENELEC Management Centre can be held accountable for the technical content of this CEN Workshop Agreement or possible conflicts with standards or legislation.

Co-funded by the European Union

CHADA

Introducion

The overall structure of the CHADA v2 form is the following:

1. **User case**: A high-level description of the user case, inc
   there are no details related to the characterisation proc
2. **Characterisation procedure**: An overview on the charac
   - the scientific validation of the approach (if any)
   - the laboratory and operator (if unique for the w
   - a link to the BPMN diagram and/or the actual c
3. **Sample:** With the size, description of the physical/chem
4. **Characterisation procedure steps:** They are processes t
   - **Sampling process**
   - **Sample preparation**
   - **Sample inspection**
   - **Calibration process**
   - **Measurement parameters adjustment**
   - **Measurement process**
   - **Data normalisation**
   - **Data filtering**
   - **Data post-processing**

   (Data normalisation, Data filtering, Data post-processing: Data processing)

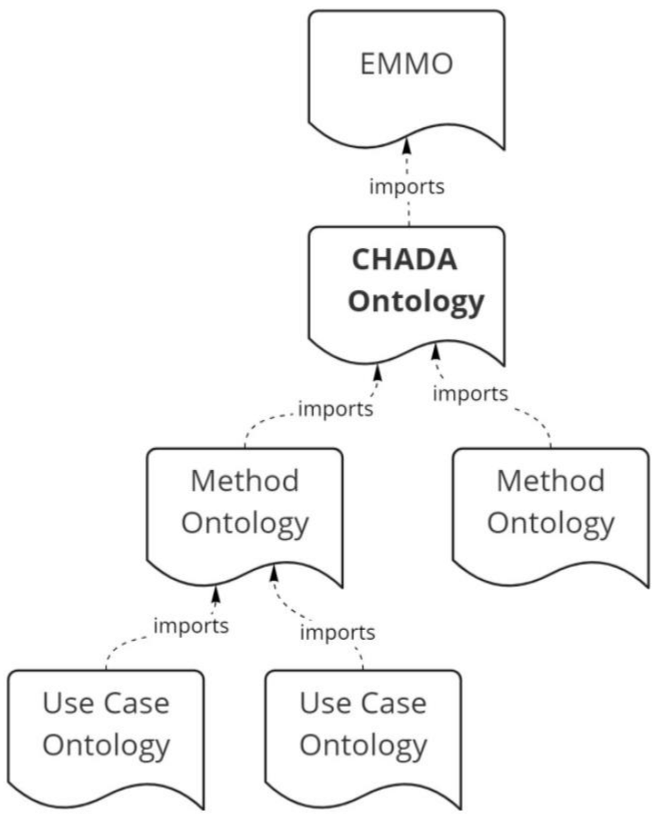
Measurement Process chameo: CharacterisationMeasurementProcess	Characterisation method chameo: CharacterisationMeasurement Process	Nanoindentation
	Name rdfs:label	Nanoscratch
	Description rdfs:comment	Constant load nanoscratch to assess the breakage phenomena of surface features
	Instrument chameo: CharacterisationMeasurementInstrument	KLA G200
	Measurement parameters [2.7] chameo:MeasurementParameter	Target load = 100-200mN Scratch distance = 500um Scratch velocity = 10um/s
	Probe [2.1] chameo:Probe	Spheroconical indenter (50 um radius)

Co-funded by the European Union

CHAMEO

An ontology for CHADA

- A domain ontology designed to model the common aspects across the different characterization methodologies
- Based on a recent CEN Workshop Agreement (CWA 17815) which introduced a standardised terminology and the Characterisation Data (CHADA) documentation scheme
- EMMO based, developed under the EMMC Task group "Materials Characterisation Methodology: Domain Ontology"
- Designed from selected industrial cases of NanoMECommons
- Resources:
  - GitHub: https://github.com/emmo-repo/domain-mechanical-testing/tree/modularization
  - Paper: https://doi.org/10.3233/AO-220271
    Del Nostro, P., Goldbeck, G. and Toti, D., 2022. CHAMEO: An ontology for the harmonisation of materials characterization methodologies. *Applied Ontology.*



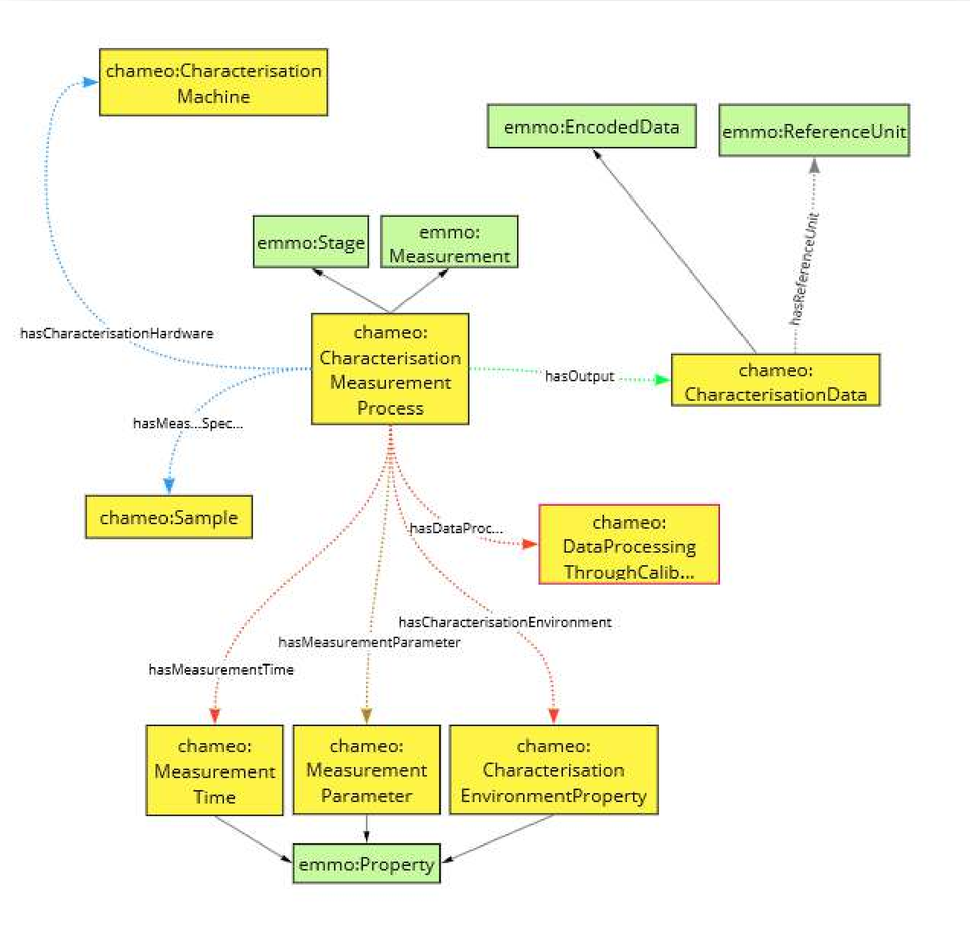
Modular ontology design

Co-funded by the European Union

Fraunhofer IWM

CHAMEO

An ontology for CHADA



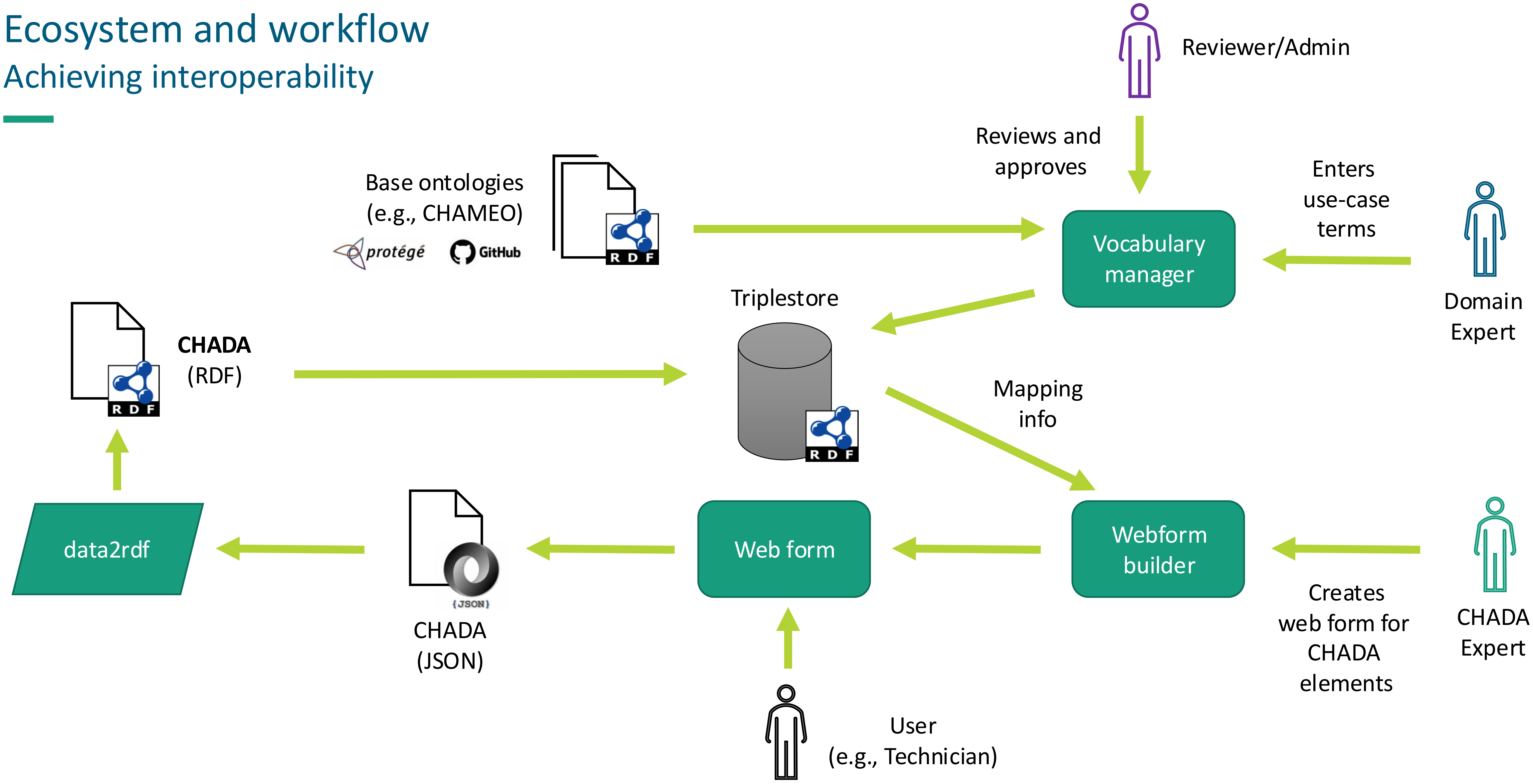
Measurement Process chameo: CharacterisationMeasurementProcess	Characterisation method chameo: CharacterisationMeasurement Process	Nanoindentation
	Name rdfs:label	Nanoscratch
	Description rdfs:comment	Constant load nanoscratch to assess the breakage phenomena of surface features
	Instrument chameo: CharacterisationMeasurementInstrument	KLA G200
	Measurement parameters [2.7] chameo:MeasurementParameter	Target load = 100-200mN Scratch distance = 500um Scratch velocity = 10um/s
	Probe [2.1] chameo:Probe	Spheroconical indenter (50 um radius)

Co-funded by the European Union

Fraunhofer IWM

Ecosystem and workflow

Achieving interoperability



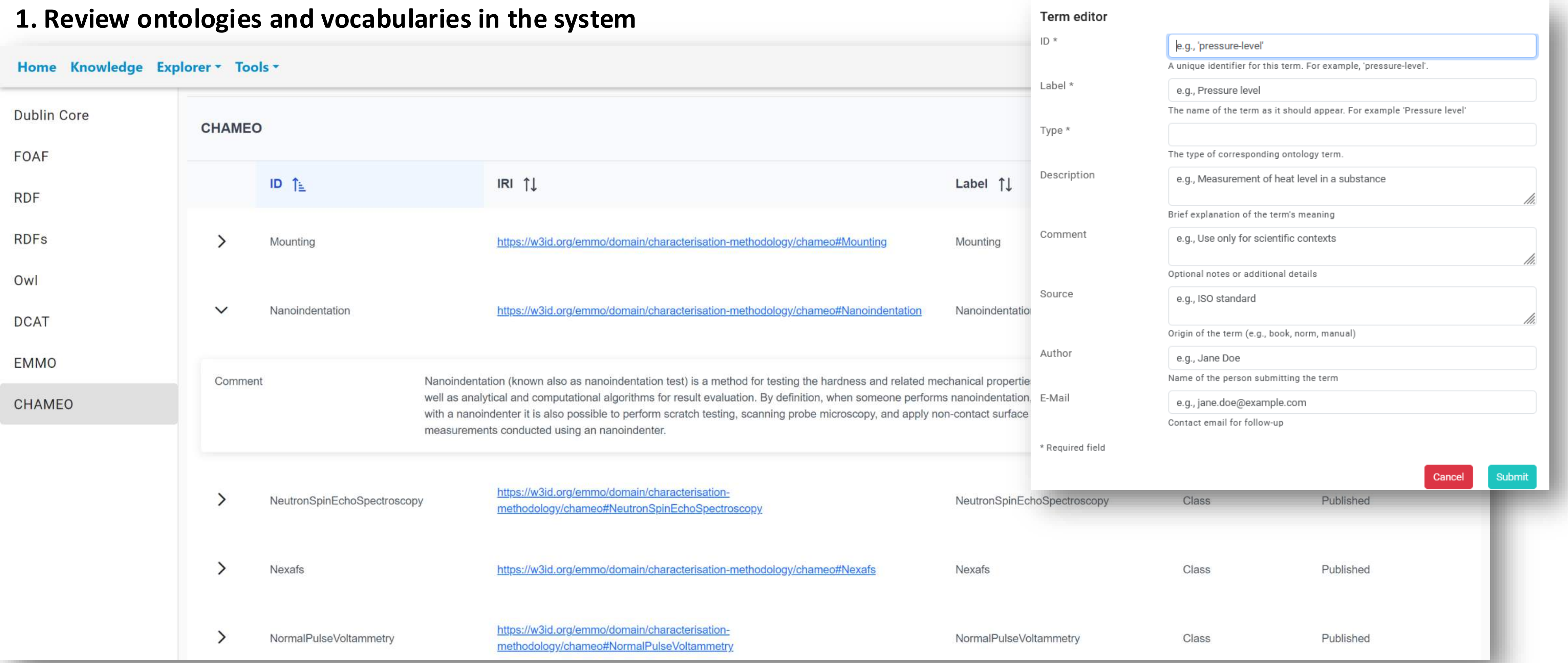
Co-funded by the European Union

Vocabulary/Ontology manager

Achieving interoperability

2. Enter use-case specific terms

1. Review ontologies and vocabularies in the system



Co-funded by the European Union

Fraunhofer IWM

Summary
CHADA digital tool

- CHADA: A framework that provides a standard set of concepts, metadata, and terminology for materials characterization

- CHADA Digital Tool:
  - Researchers document each characterization step using a simple web form
  - The tool automatically structures data in RDF format - No manual semantic input required
  - Enhances compliance with FAIR principles (Findable, Accessible, Interoperable, Reusable)
  - It will be showcased in the AID4GREENEST project

Co-funded by the European Union

Acknowledgement

Co-funded by the European Union

This work has been supported by the European Union under the Horizon Europe Programme

Grant Agreement 101091912

Co-funded by the European Union

Thank you for your attention!

Yoav Nahshon
Team Materials Informatics, Business Unit Manufacturing Processes
Fraunhofer IWM
[redacted]