Bridging Interoperability Gaps in Manufacturing with a Bottom-Up Approach to Managing Semantics

Yoav Nahshon
Team Materials Informatics, Business Unit Manufacturing Processes

VMAP User Meeting, 19.02.2025

Fraunhofer Institute for Mechanics of Materials IWM

- Focus on traditional research fields in materials science
  - Materials modelling and characterization
  - Assessment of components and manufacturing processes (efficiency, life-time, …)
- Nowadays, digitalization plays a key role in materials science and manufacturing

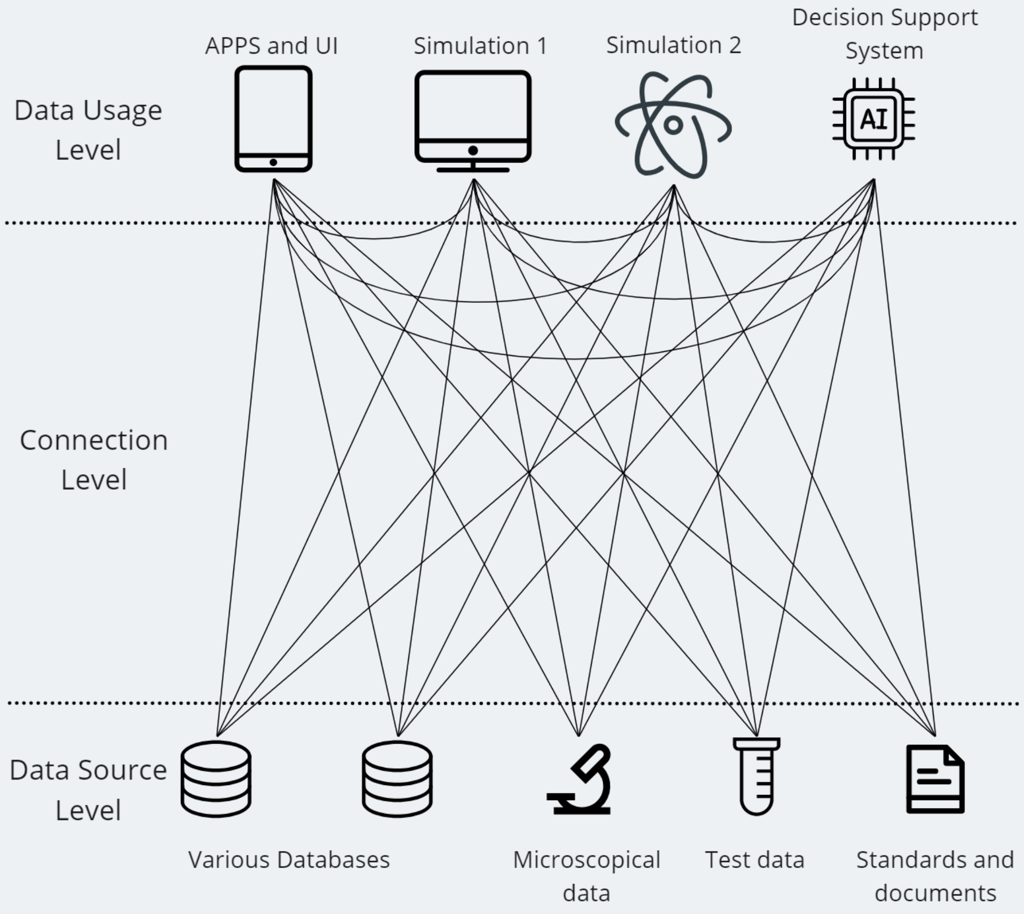
→ Current focus: Digitization of materials and process data
→ Generate value by using data-driven tools

Fraunhofer IWM

Applying data-driven solutions
Obstacles

- **Interoperability gaps hinder effective communication**
- **Limited data access restricts solution potential**
- **Data conversion and pre-processing scripts are required**

➡ Engineers lose valuable time



Fraunhofer IWM

Applying data-driven solutions

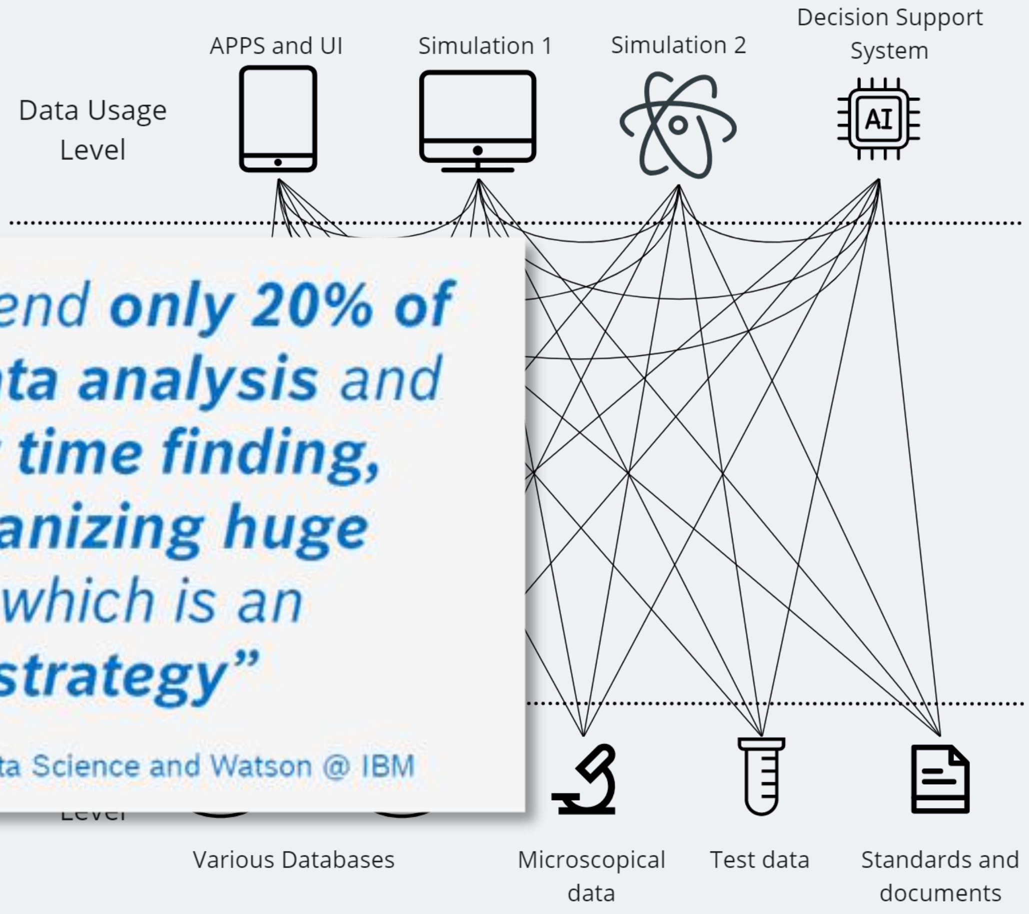
Obstacles

- **Interoperability gaps hinder effective communication**
- **Limited data access restricts solution potential**
- **Data conversion and pre- required**

➡ Engineers los

*"Most data scientists spend **only 20% of their time on actual data analysis** and **80% percent of their time finding, cleaning, and reorganizing huge amounts of data,** which is an **inefficient data strategy"***

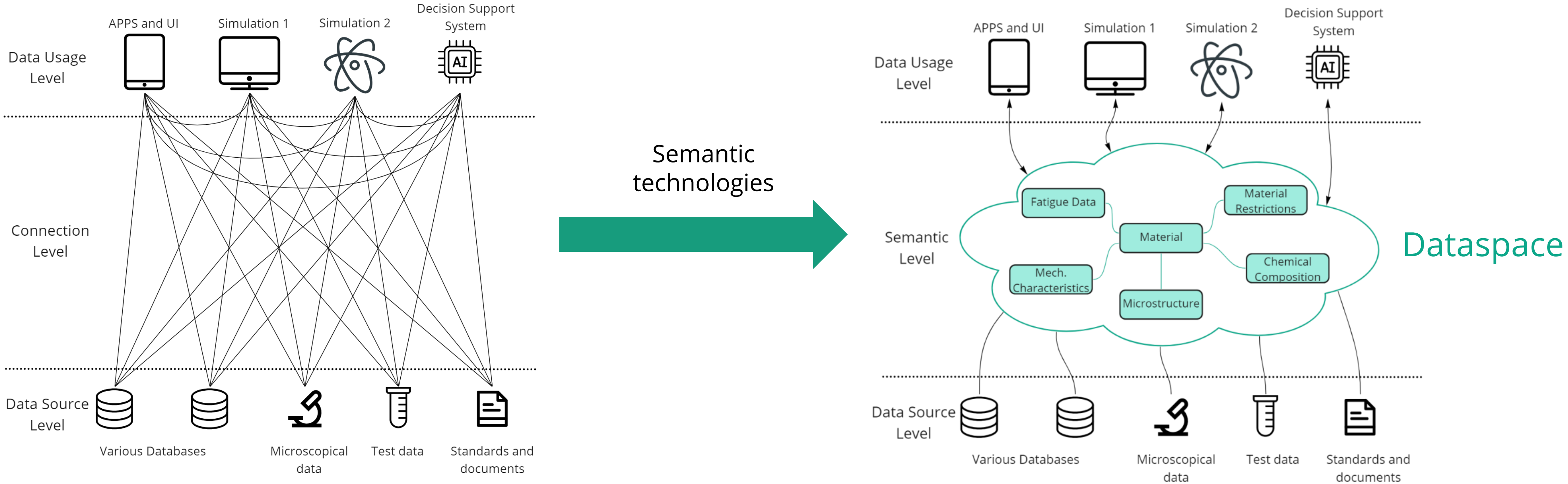
Armand Ruiz, Lead Product Manager IBM Data Science and Watson @ IBM



The importance of FAIR data management for SMEs

What is FAIR data and why it is important

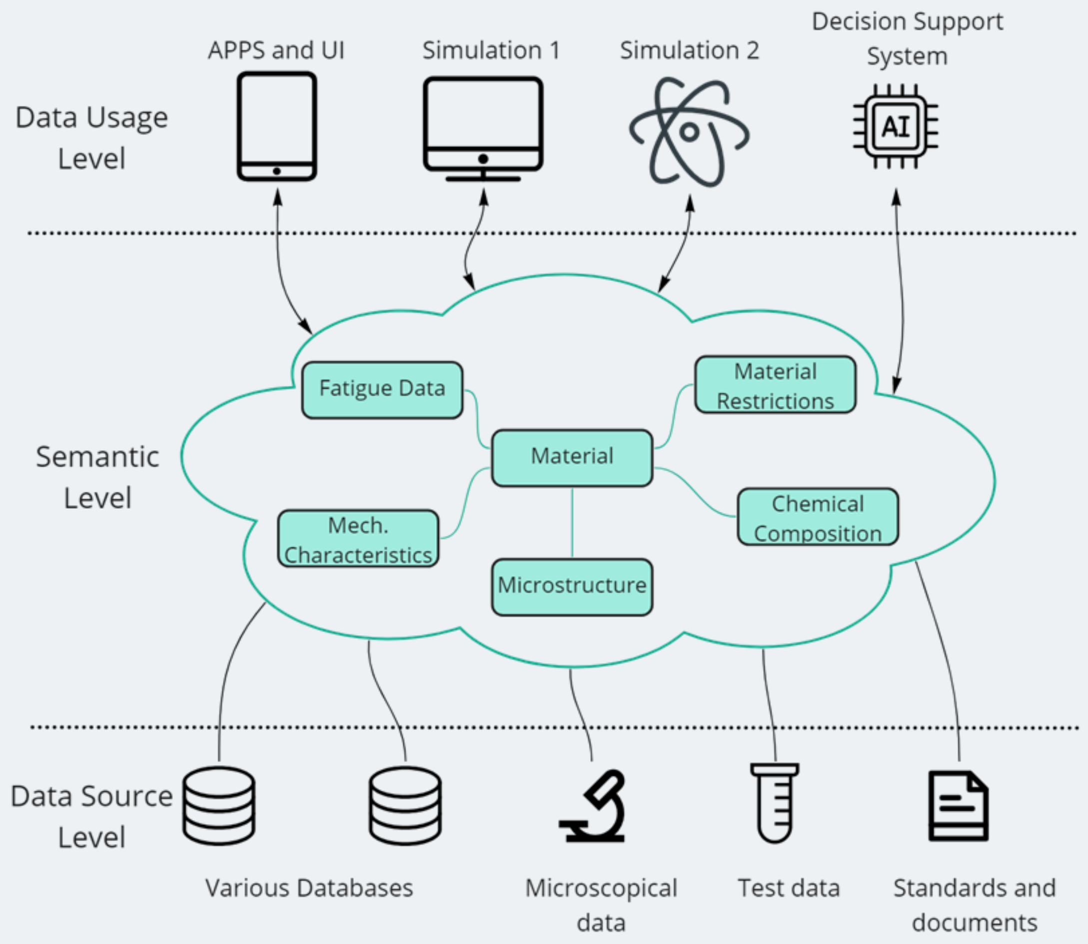
- **FAIR: Findable, Accessible, Interoperable, Reusable**
- **Often data is stored in file structures or relational databases within an organization → not FAIR**



Data-driven solutions with FAIR data

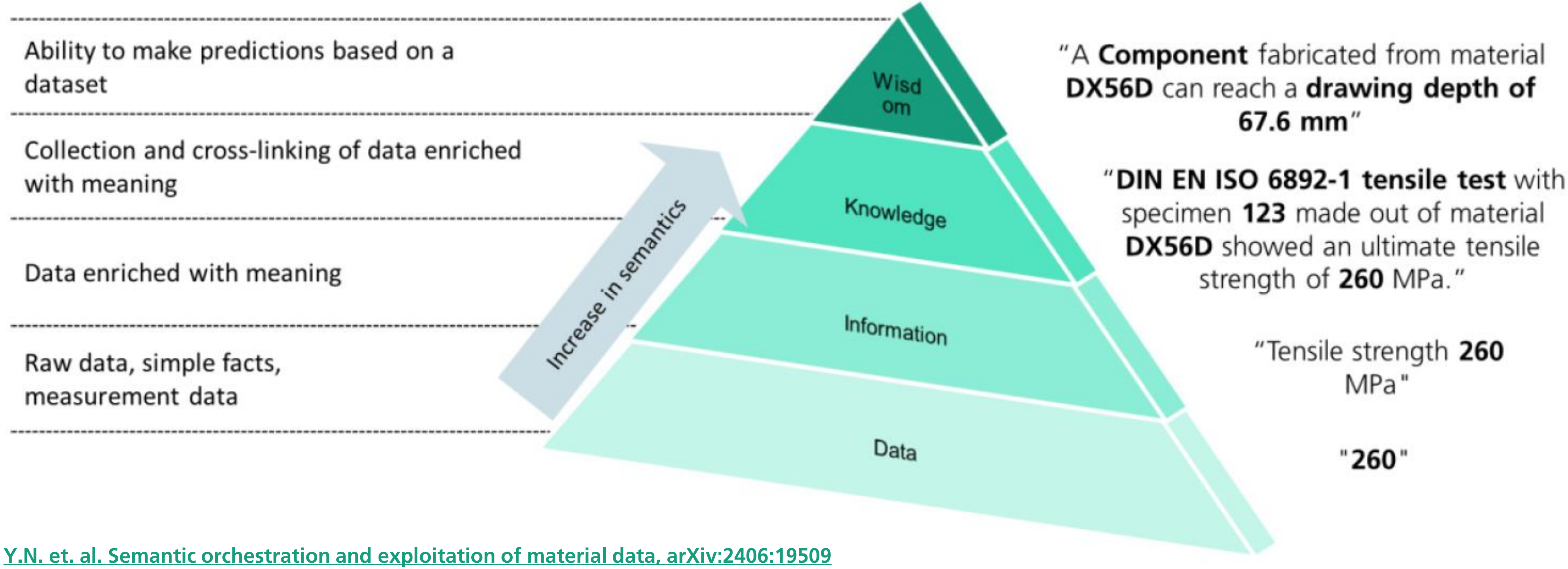
Strengths

- **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

Data-information-knowledge-wisdom (DIKW) hierarchy applied to materials data



Y.N. et. al. Semantic orchestration and exploitation of material data, arXiv:2406:19509

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.

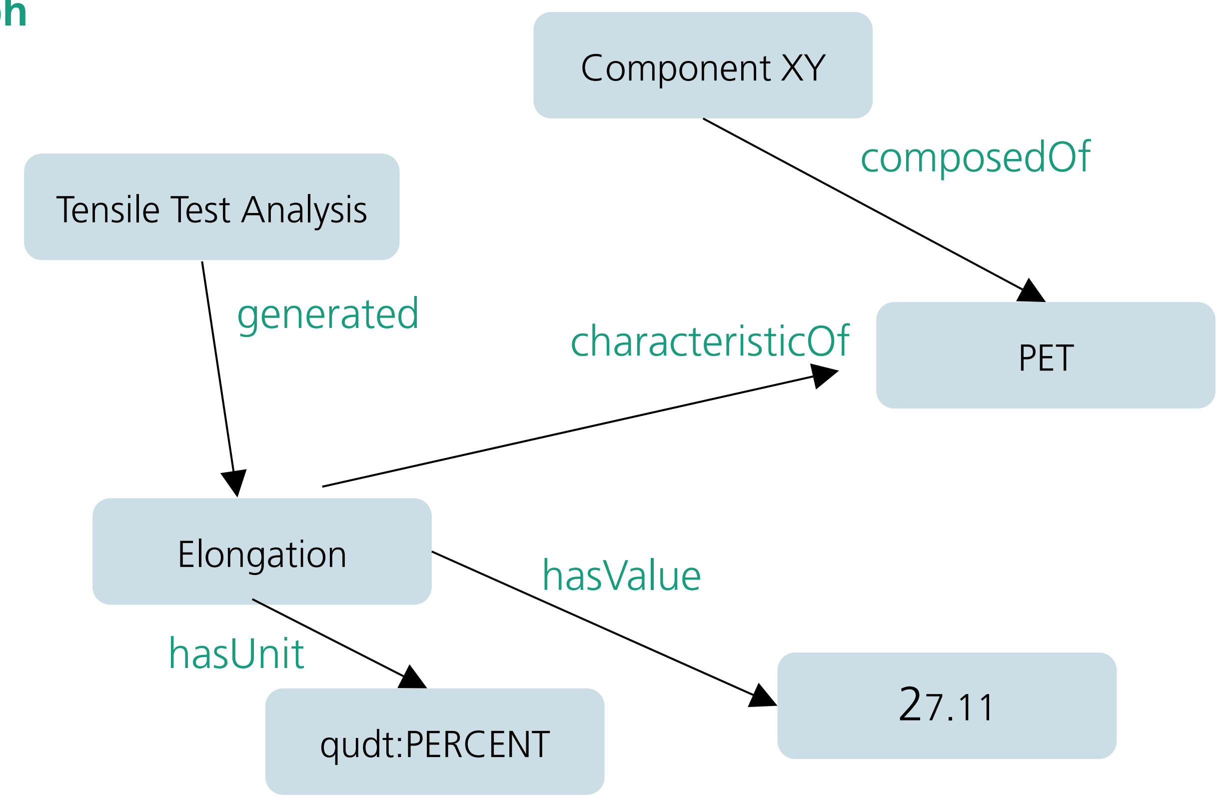
URI: http://qudt.org/vocab/unit/PERCENT

Description

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

Human and machine readable

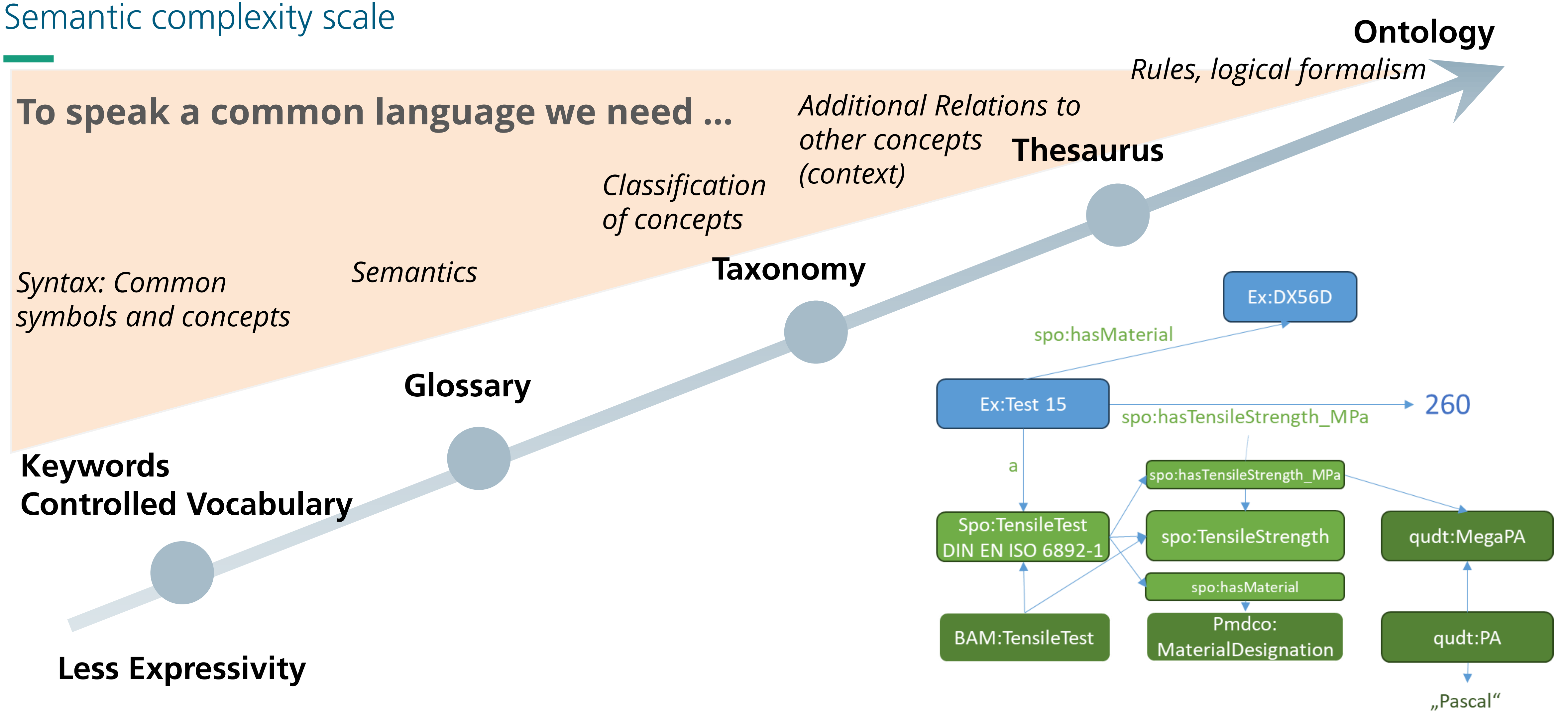
<u>But:</u> Working with semantic technologies is challenging



Fraunhofer IWM

Semantic technologies for FAIR data management

Semantic complexity scale



Data management in materials manufacturing

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"
"Messlä
"Versuc
"Prober
"Prober
"Prüfge
"Vorkra
"Temper
"Bemerk
"Prüfze
"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

⚠ **Structural heterogeneity:** ASCII on top of TSV

⚠ **Semantic heterogeneity:** Terms are specified in a particular language

Fraunhofer IWM

Data management in materials manufacturing

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
```

Data2RDF

```
fileid:TensileTestExperiment a prov:Activity ;
    prov:generated fileid:AbsoluteCrossheadTravel,
        fileid:Extension,
        fileid:Remark,
        fileid:StandardForce,
        fileid:TimeStamp,
        fileid:dataset ;
    prov:hadPlan fileid:TestStandard ;
    prov:used fileid:DisplacementTransducer,
        fileid:ForceMeasuringDevice,
        fileid:TensileTestSpecimen,
        fileid:TensileTestingMachine,
        fileid:TestingFacility ;
    prov:wasAssociatedWith fileid:Tester ;
    prov:wasInfluencedBy fileid:ExperimentPreparation .

fileid:TestingStandard a prov:Plan .

fileid:ExperimentPreparation a prov:Activity ;
    prov:atLocation fileid:TestingLab ;
    prov:generated fileid:OriginalGaugeLength,
        fileid:Preload,
        fileid:TestingRate ;
    prov:wasAssociatedWith fileid:DisplacementTransducer,
        fileid:ForceMeasuringDevice,
```

https://github.com/MI-FraunhoferIWM/data2rdf

Data management in materials manufacturing

Tensile test example

Fraunhofer IWM

Data management in materials manufacturing

Opportunities of using semantic technologies

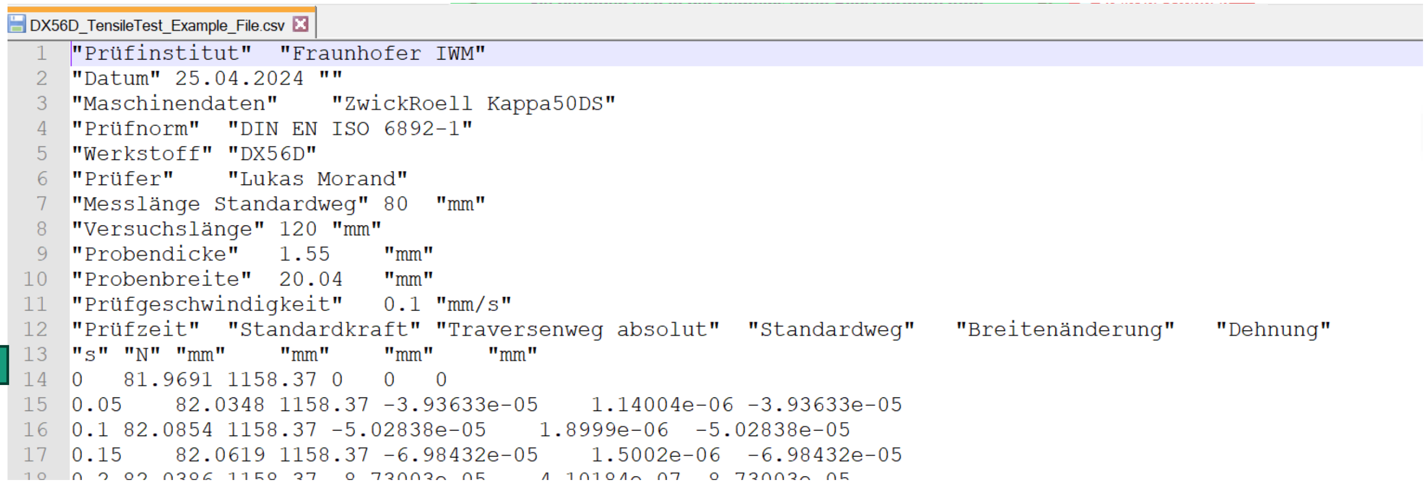
- Integrate all data into one dataspace in a FAIR manner
- Enriching data with meaning and link data
- Make information from (external) data silos available
- Enable information tracking
- Seamlessly connect digital tools (interoperability)

Semantic technologies for FAIR data management

Approaches

Different development philosophies :

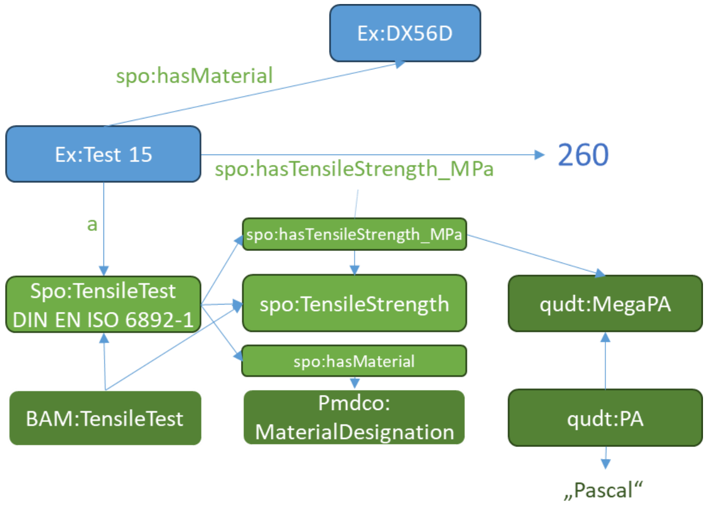
- Top-down: from generic to specific
- Bottom-up: from the practical to the more general
- Middle-Out: start in the middle and add where necessary


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:

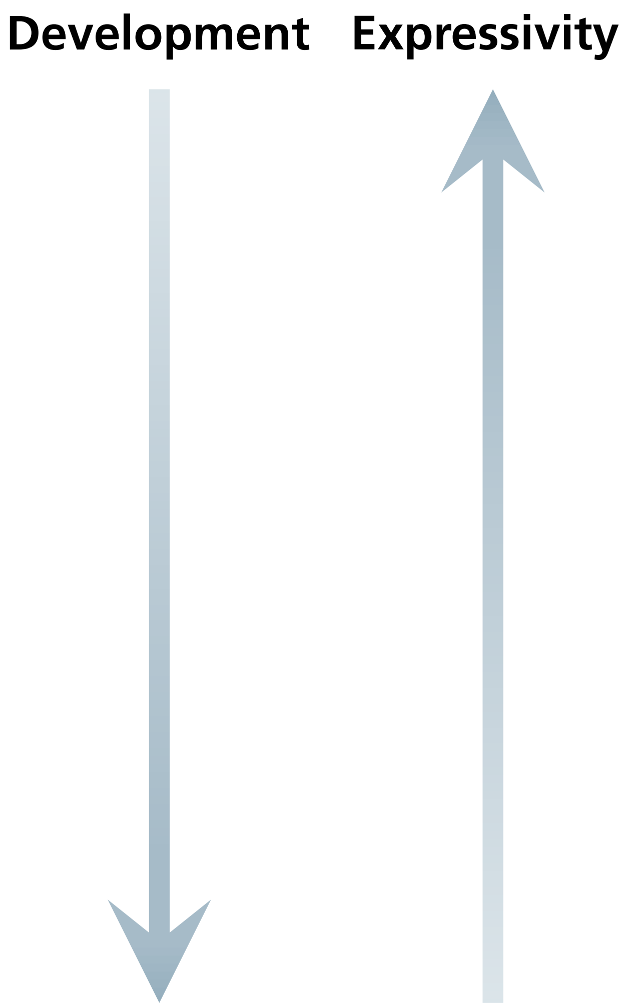
- Protégé / WebProtégé
- Chowlk / RDF Diagram Framework
- Metafactory, Corporate Memory, GraphDB, AllegroGraph, Virtuoso, PoolParty, …
- Xlsx2owl
- Vocabulary manager



Fraunhofer IWM

Ontology development approaches

Top-down



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

Process:

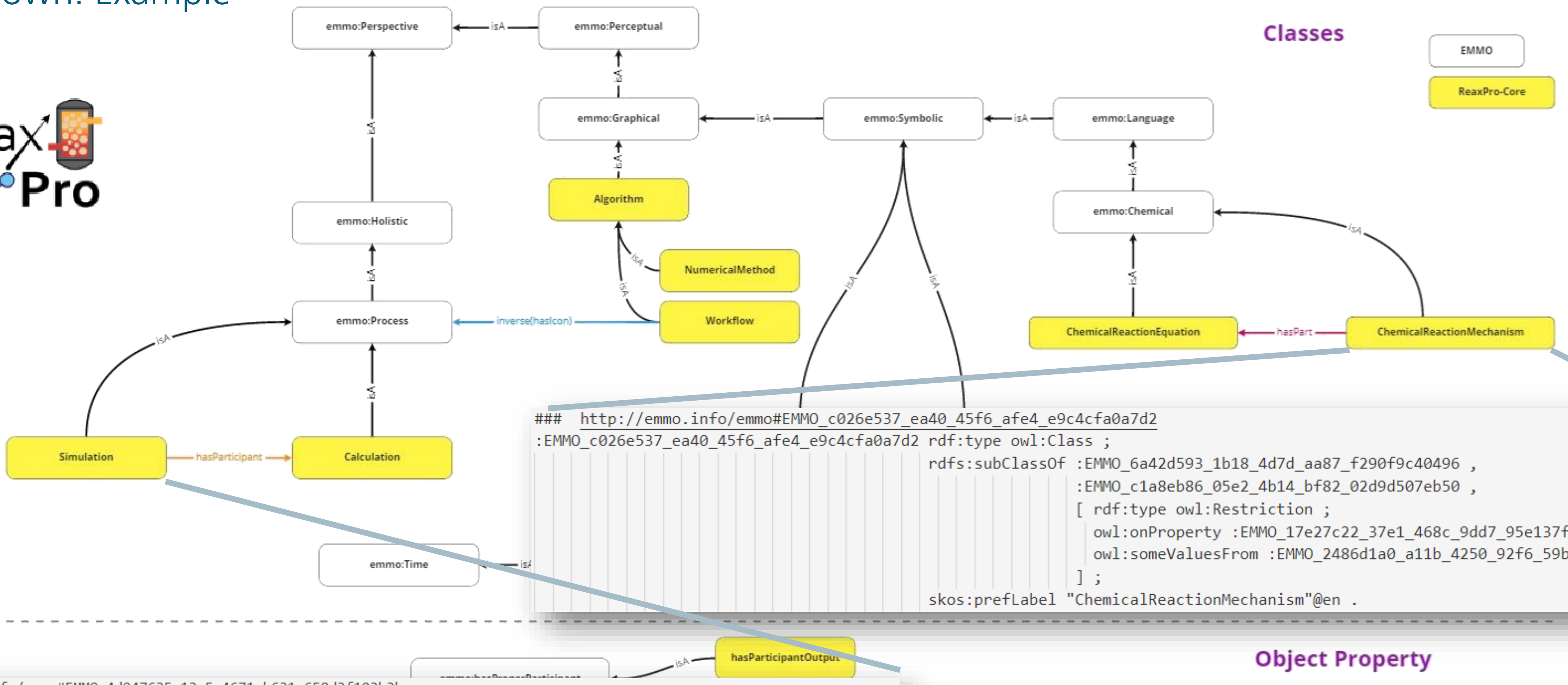
Define the most generic concepts, such as "Entity" or "Object."

Gradually refine these concepts into increasingly detailed subclasses or properties.

Ensure consistency and logical relationships between levels.

Ontology development approaches

Top down: Example

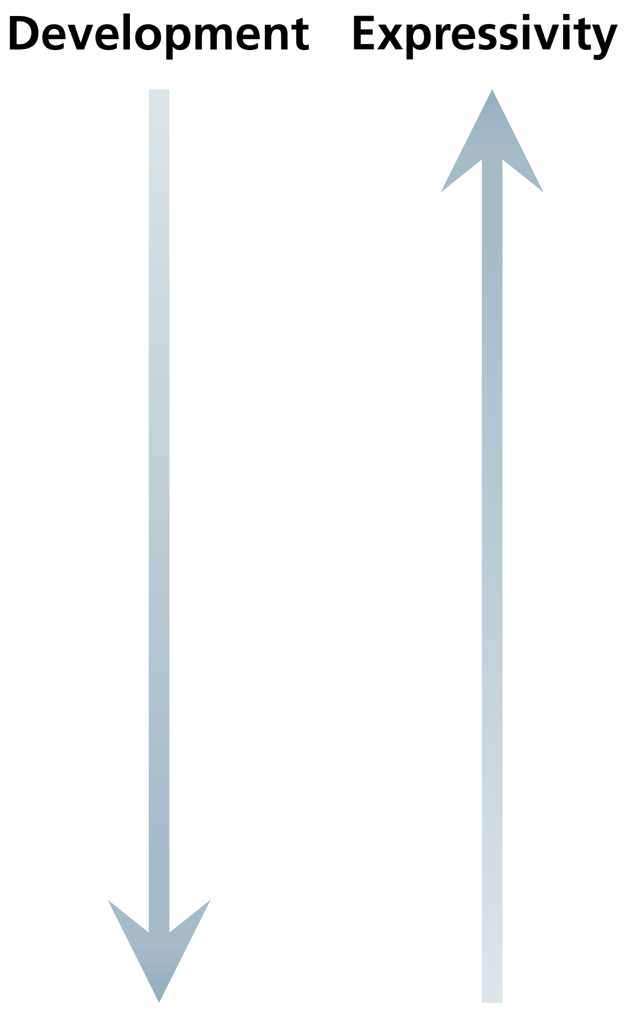



```
### http://emmo.info/emmo#EMMO_c026e537_ea40_45f6_afe4_e9c4cfa0a7d2
:EMMO_c026e537_ea40_45f6_afe4_e9c4cfa0a7d2 rdf:type owl:Class ;
                                           rdfs:subClassOf :EMMO_6a42d593_1b18_4d7d_aa87_f290f9c40496 ,
                                                           :EMMO_c1a8eb86_05e2_4b14_bf82_02d9d507eb50 ,
                                                           [ 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

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	

Ontology development approaches

Bottom-up: Example

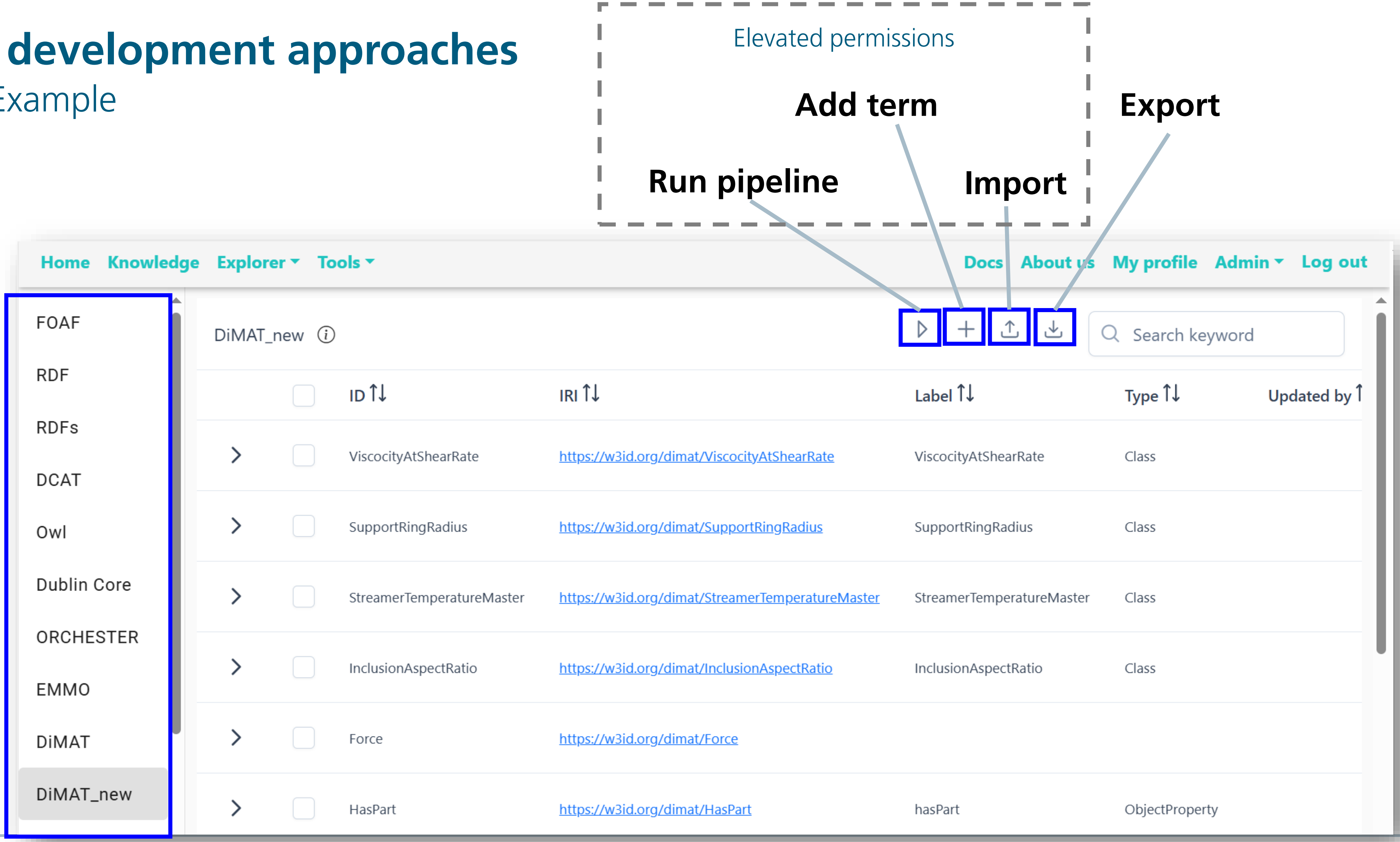
Watch on YouTube

DiMAT

Ontology development approaches

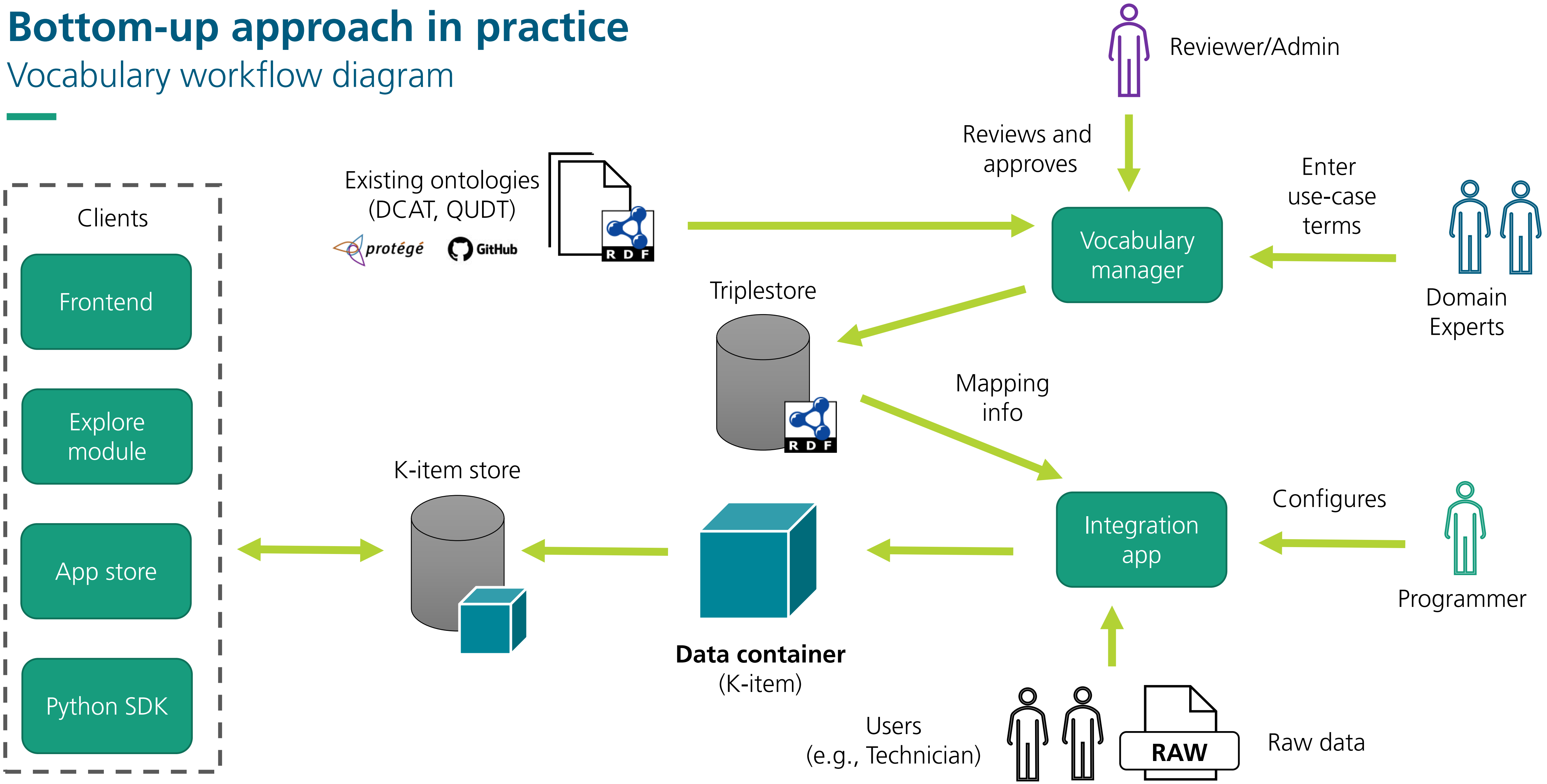
Bottom-up: Example

Registered vocabularies and ontologies



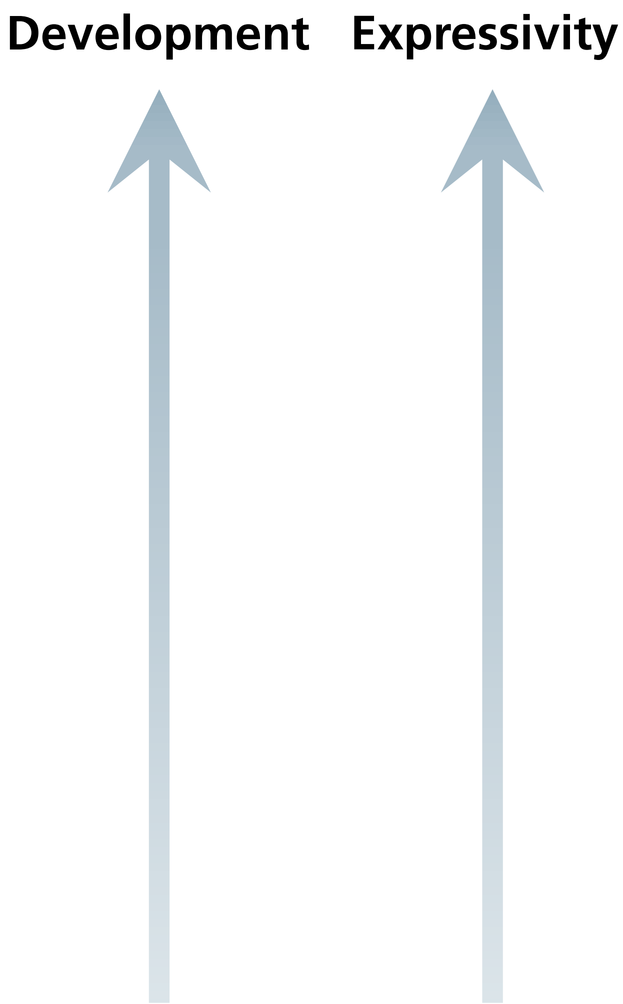
Bottom-up approach in practice

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

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.

Take-home message

Top-down vs. bottom-up

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)

Acknowledgement

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under the Grant Agreement 101091496.

Thank you for your attention!

Yoav Nahshon
Team Materials Informatics, Business Unit Manufacturing Processes
Fraunhofer IWM
yoav.nahshon@iwm.fraunhofer.de