Achieving semantic interoperability in materials science data and simulation workflows

Abril Azocar Guzman

Institute for Advanced Simulations – Materials Data Science and Informatics (IAS-9), Forschungszentrum Jülich

NFDI4Ing CC43 Community Meeting

16.01.2025

NFDI MatWerk

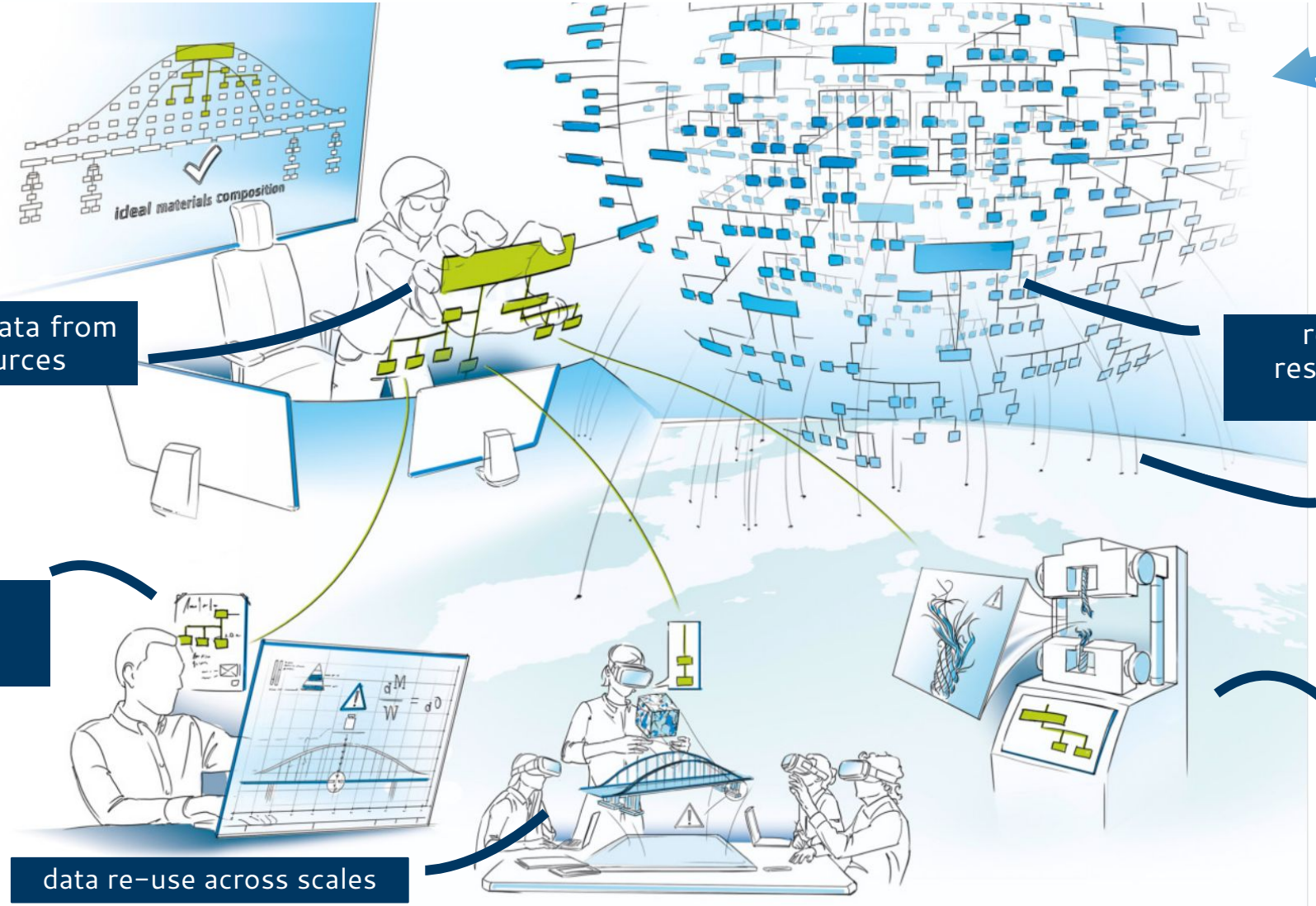
NATIONAL RESEARCH DATA INFRASTRUCTURE FOR MATERIALS SCIENCE & ENGINEERING

Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation

Project number 460247524

JÜLICH Forschungszentrum

NFDI-MatWerk Vision



exchange with other knowledge systems

integration of (meta)data from heterogeneous sources

representation of resources in the MSE domain

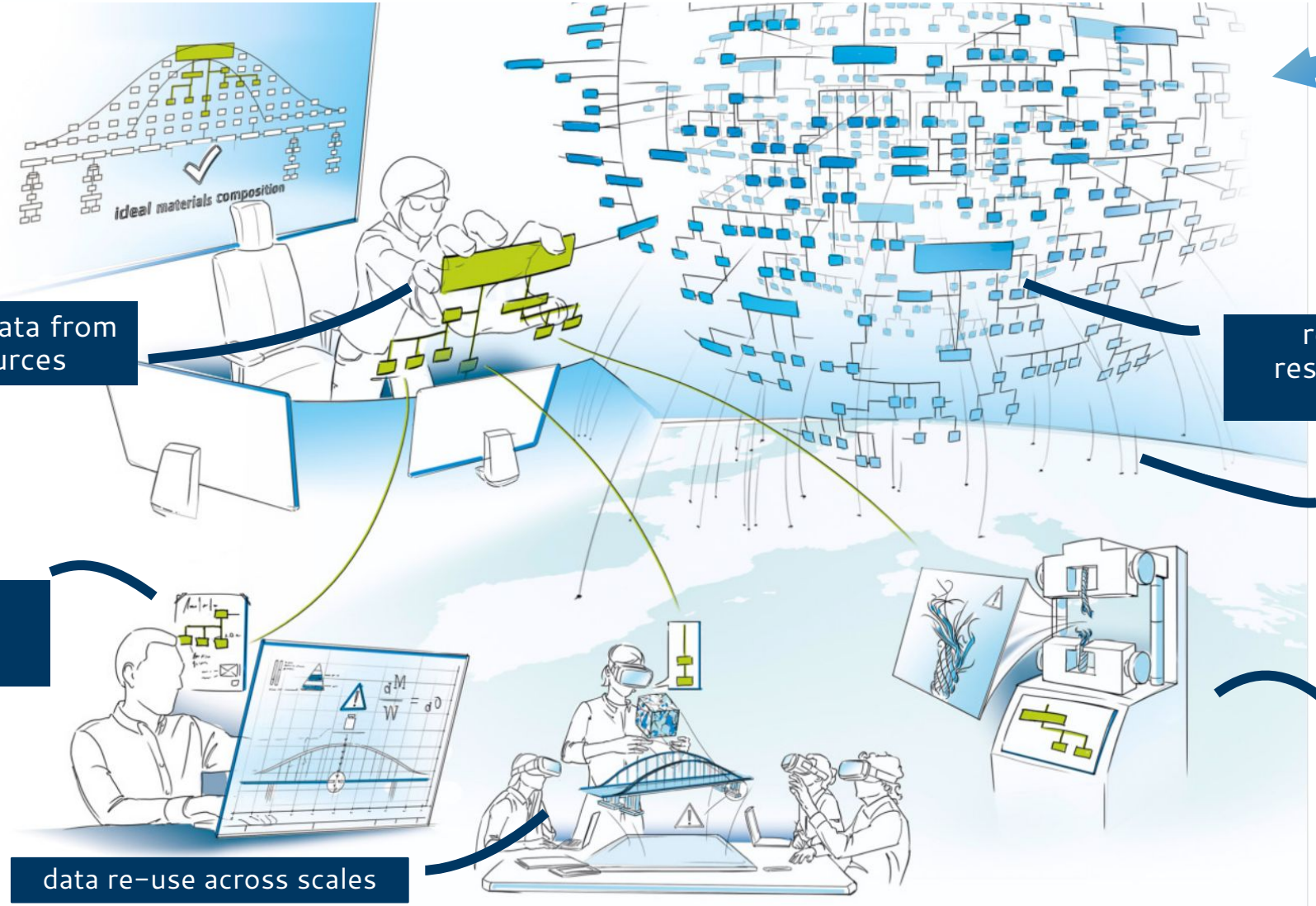
findability & accessibility of federated resources

structured documentation of simulations

structured documentation of experiments

data re-use across scales

WHERE and WHY do we need ontologies?



integration of (meta)data from heterogeneous sources

exchange with other knowledge systems

representation of resources in the MSE domain

findability & accessibility of federated resources

structured documentation of simulations

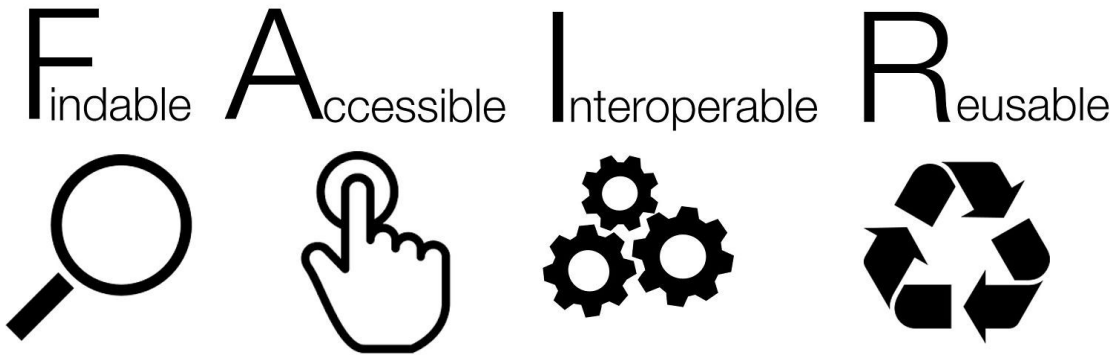
structured documentation of experiments

data re-use across scales

FAIRification of research data

FAIR data ensures
interoperability & re-usability:

- how is data organized?
- what (class) is in what column?
- what is the unit/dimension?
- how was it recorded?
- who does this data belong to?
- …

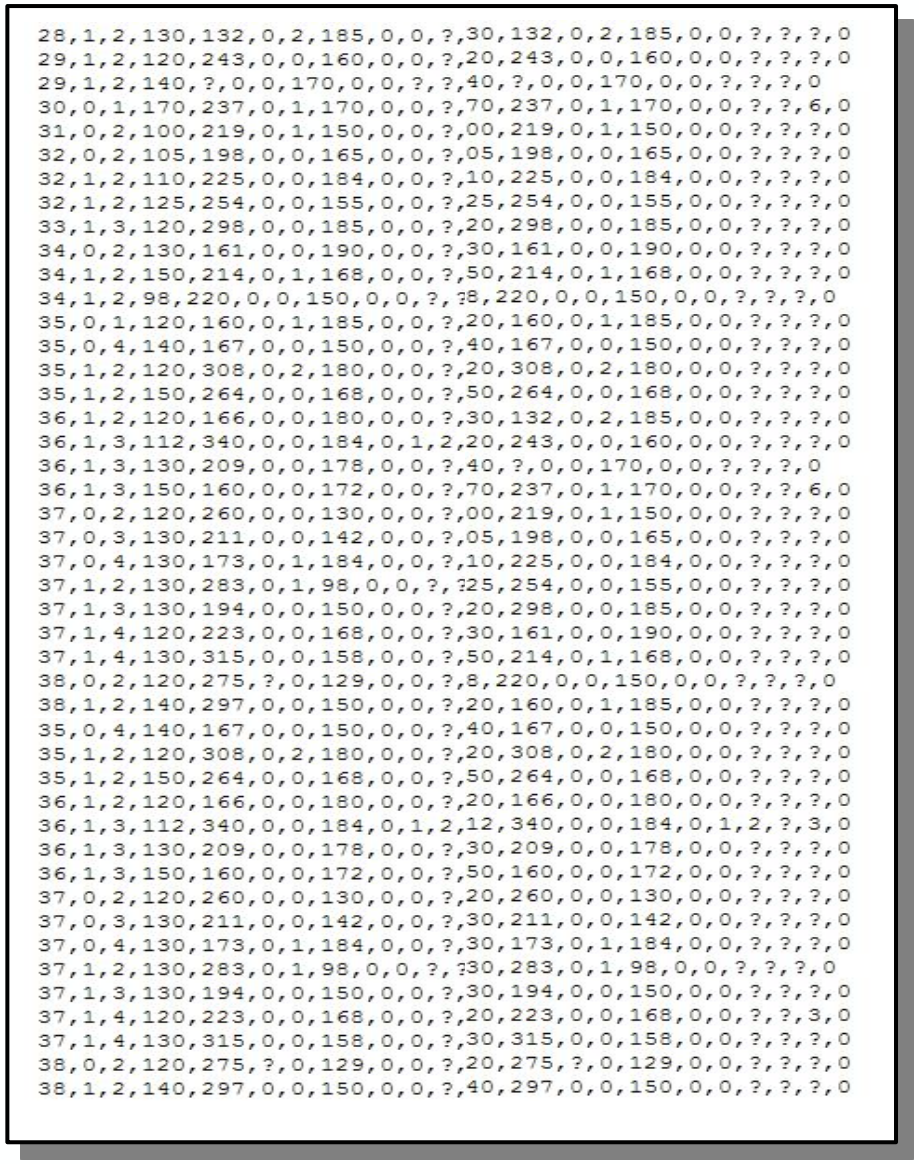


Wilkinson, M., Dumontier, M., Aalbersberg, I. et al., Sci Data, 2016, 3, 160018.
(image source: Sangya Pundir, Wikipedia license CC BY SA)

FAIRification of research data

FAIR data ensures interoperability & re-usability:

- how is data organized?
- what (class) is in what column?
- what is the unit/dimension?
- how was it recorded?
- who does this data belong to?
- …

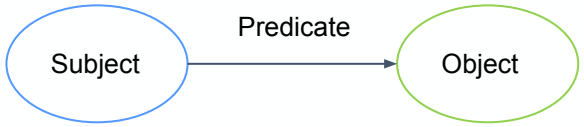


```
28,1,2,130,132,0,2,185,0,0,?,30,132,0,2,185,0,0,?,?,?,0
29,1,2,120,243,0,0,160,0,0,?,20,243,0,0,160,0,0,?,?,?,0
29,1,2,140,?,0,0,170,0,0,?,?,40,?,0,0,170,0,0,?,?,?,0
30,0,1,170,237,0,1,170,0,0,?,70,237,0,1,170,0,0,?,?,6,0
31,0,2,100,219,0,1,150,0,0,?,00,219,0,1,150,0,0,?,?,?,0
32,0,2,105,198,0,0,165,0,0,?,05,198,0,0,165,0,0,?,?,?,0
32,1,2,110,225,0,0,184,0,0,?,10,225,0,0,184,0,0,?,?,?,0
32,1,2,125,254,0,0,155,0,0,?,25,254,0,0,155,0,0,?,?,?,0
33,1,3,120,298,0,0,185,0,0,?,20,298,0,0,185,0,0,?,?,?,0
34,0,2,130,161,0,0,190,0,0,?,30,161,0,0,190,0,0,?,?,?,0
34,1,2,150,214,0,1,168,0,0,?,50,214,0,1,168,0,0,?,?,?,0
34,1,2,98,220,0,0,150,0,0,?,?8,220,0,0,150,0,0,?,?,?,0
35,0,1,120,160,0,1,185,0,0,?,20,160,0,1,185,0,0,?,?,?,0
35,0,4,140,167,0,0,150,0,0,?,40,167,0,0,150,0,0,?,?,?,0
35,1,2,120,308,0,2,180,0,0,?,20,308,0,2,180,0,0,?,?,?,0
35,1,2,150,264,0,0,168,0,0,?,50,264,0,0,168,0,0,?,?,?,0
36,1,2,120,166,0,0,180,0,0,?,30,132,0,2,185,0,0,?,?,?,0
36,1,3,112,340,0,0,184,0,1,2,20,243,0,0,160,0,0,?,?,?,0
36,1,3,130,209,0,0,178,0,0,?,40,?,0,0,170,0,0,?,?,?,0
36,1,3,150,160,0,0,172,0,0,?,70,237,0,1,170,0,0,?,?,6,0
37,0,2,120,260,0,0,130,0,0,?,00,219,0,1,150,0,0,?,?,?,0
37,0,3,130,211,0,0,142,0,0,?,05,198,0,0,165,0,0,?,?,?,0
37,0,4,130,173,0,1,184,0,0,?,10,225,0,0,184,0,0,?,?,?,0
37,1,2,130,283,0,1,98,0,0,?,?25,254,0,0,155,0,0,?,?,?,0
37,1,3,130,194,0,0,150,0,0,?,20,298,0,0,185,0,0,?,?,?,0
37,1,4,120,223,0,0,168,0,0,?,30,161,0,0,190,0,0,?,?,?,0
37,1,4,130,315,0,0,158,0,0,?,50,214,0,1,168,0,0,?,?,?,0
38,0,2,120,275,?,0,129,0,0,?,8,220,0,0,150,0,0,?,?,?,0
38,1,2,140,297,0,0,150,0,0,?,20,160,0,1,185,0,0,?,?,?,0
35,0,4,140,167,0,0,150,0,0,?,40,167,0,0,150,0,0,?,?,?,0
35,1,2,120,308,0,2,180,0,0,?,20,308,0,2,180,0,0,?,?,?,0
35,1,2,150,264,0,0,168,0,0,?,50,264,0,0,168,0,0,?,?,?,0
36,1,2,120,166,0,0,180,0,0,?,20,166,0,0,180,0,0,?,?,?,0
36,1,3,112,340,0,0,184,0,1,2,12,340,0,0,184,0,1,2,?,3,0
36,1,3,130,209,0,0,178,0,0,?,30,209,0,0,178,0,0,?,?,?,0
36,1,3,150,160,0,0,172,0,0,?,50,160,0,0,172,0,0,?,?,?,0
37,0,2,120,260,0,0,130,0,0,?,20,260,0,0,130,0,0,?,?,?,0
37,0,3,130,211,0,0,142,0,0,?,30,211,0,0,142,0,0,?,?,?,0
37,0,4,130,173,0,1,184,0,0,?,30,173,0,1,184,0,0,?,?,?,0
37,1,2,130,283,0,1,98,0,0,?,?30,283,0,1,98,0,0,?,?,?,0
37,1,3,130,194,0,0,150,0,0,?,30,194,0,0,150,0,0,?,?,?,0
37,1,4,120,223,0,0,168,0,0,?,20,223,0,0,168,0,0,?,?,3,0
37,1,4,130,315,0,0,158,0,0,?,30,315,0,0,158,0,0,?,?,?,0
38,0,2,120,275,?,0,129,0,0,?,20,275,?,0,129,0,0,?,?,?,0
38,1,2,140,297,0,0,150,0,0,?,40,297,0,0,150,0,0,?,?,?,0
```

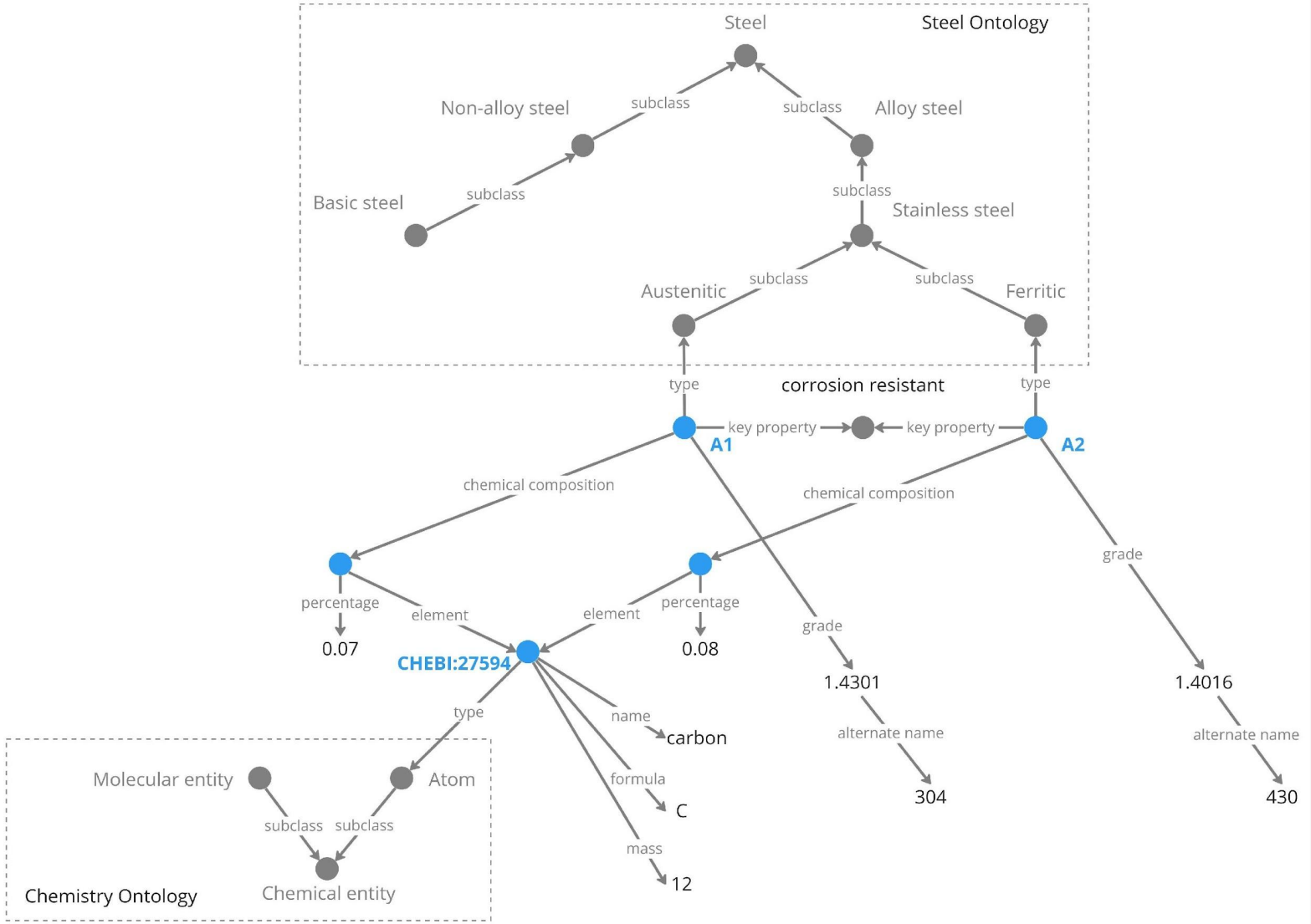
Ontological data models

What is an ontology?

An ontology formalises concepts and their relations to other concepts.



Steel has Iron

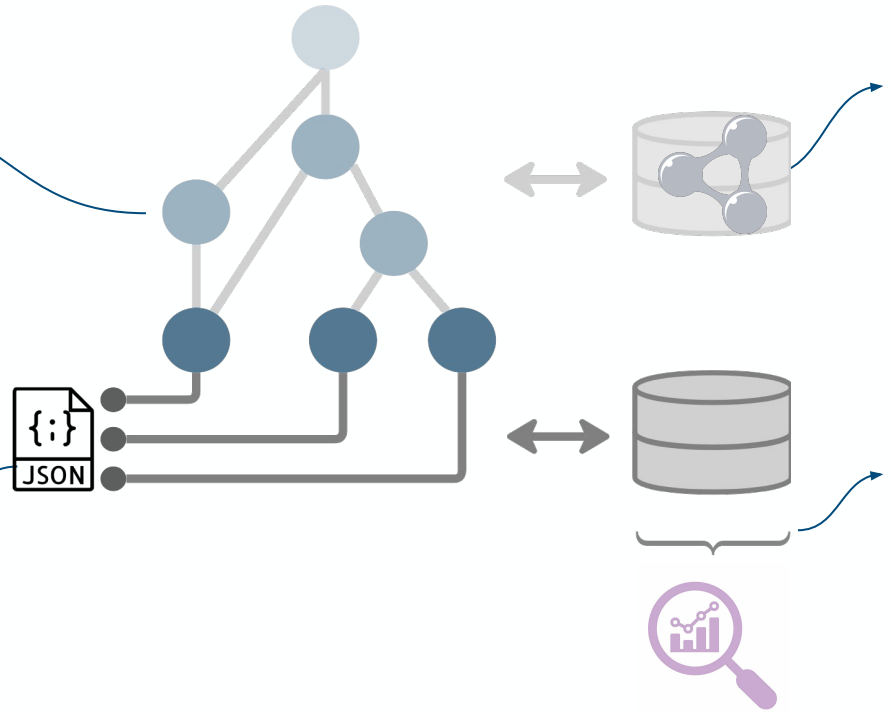


Using ontologies

What are the requirements?

- **Ontology**
  Provides common vocabulary and constraints that hold across different applications
  community agreement

- **Conceptual model**
  Focus on stored data
  e.g., metadata schemas
  FAIR, interoperability
  metadata harmonization



- **Implementation**
  - knowledge graphs
  - databases

  data integration

- **Ontology is part of other tasks**
  - data mining
  - semantic AI
  - reasoning and decision making

Levels of abstraction

top-level

top level alignment

mid-level

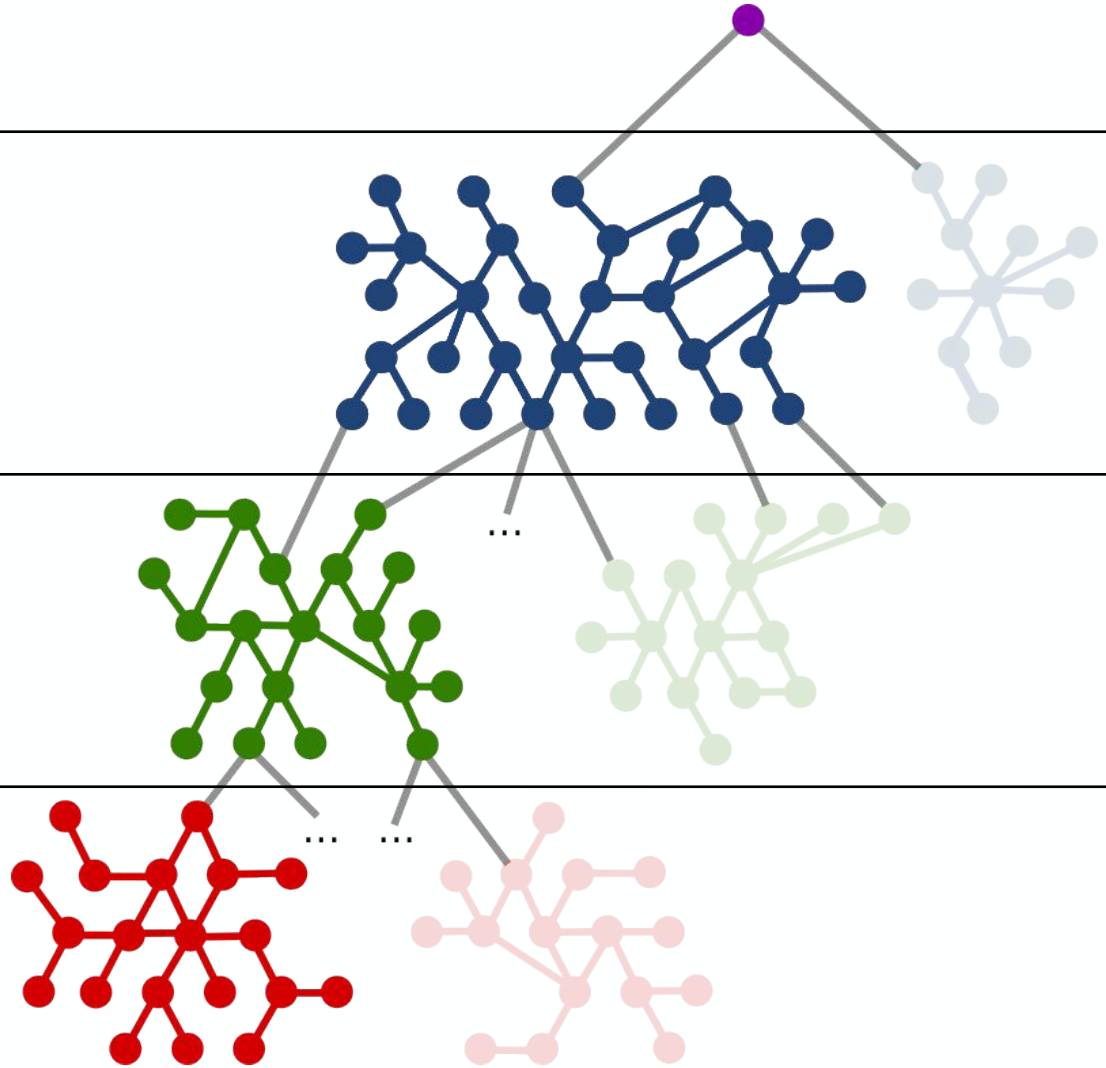
lateral harmonization
bridging domains

domain level

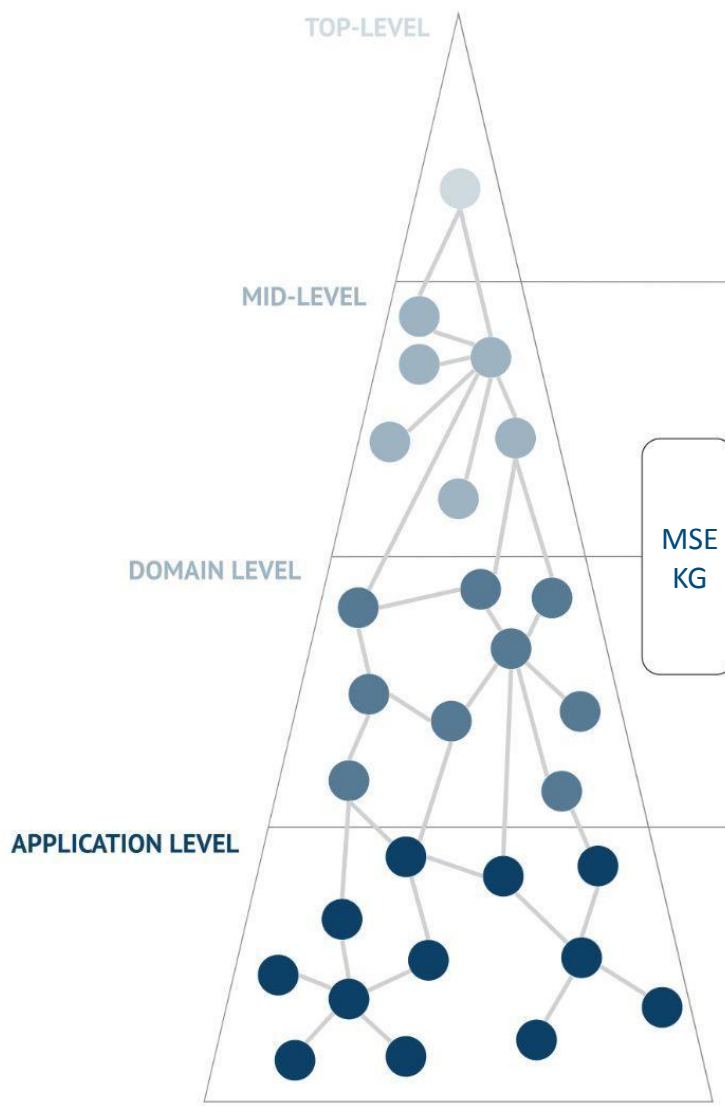
organizing knowledge about data and methods within a domain

application level

describing a specific dataset or process



Ontologies in NFDI-MatWerk



External

Basic Formal Ontology - BFO

European Material Modeling Onto - EMMO

MSE KG

Platform Material Digital Core Ontology - PMDco

NFDI core Ontology - NFDIco

Provenance in Mat. Sci. Ontology - PRIMA

MatWerk Ontology - MWO

Materials Modelling and Simulation Suite - MMSS

Material Design Ontology - MDO

Crystallographic Defect Ontology Suite - CDOS

Dislocation Ontology Suite - DISOS

EM Glossary - EMG

Computational Material Sample Ontology - CMSO

Creep Reference Dataset Schema

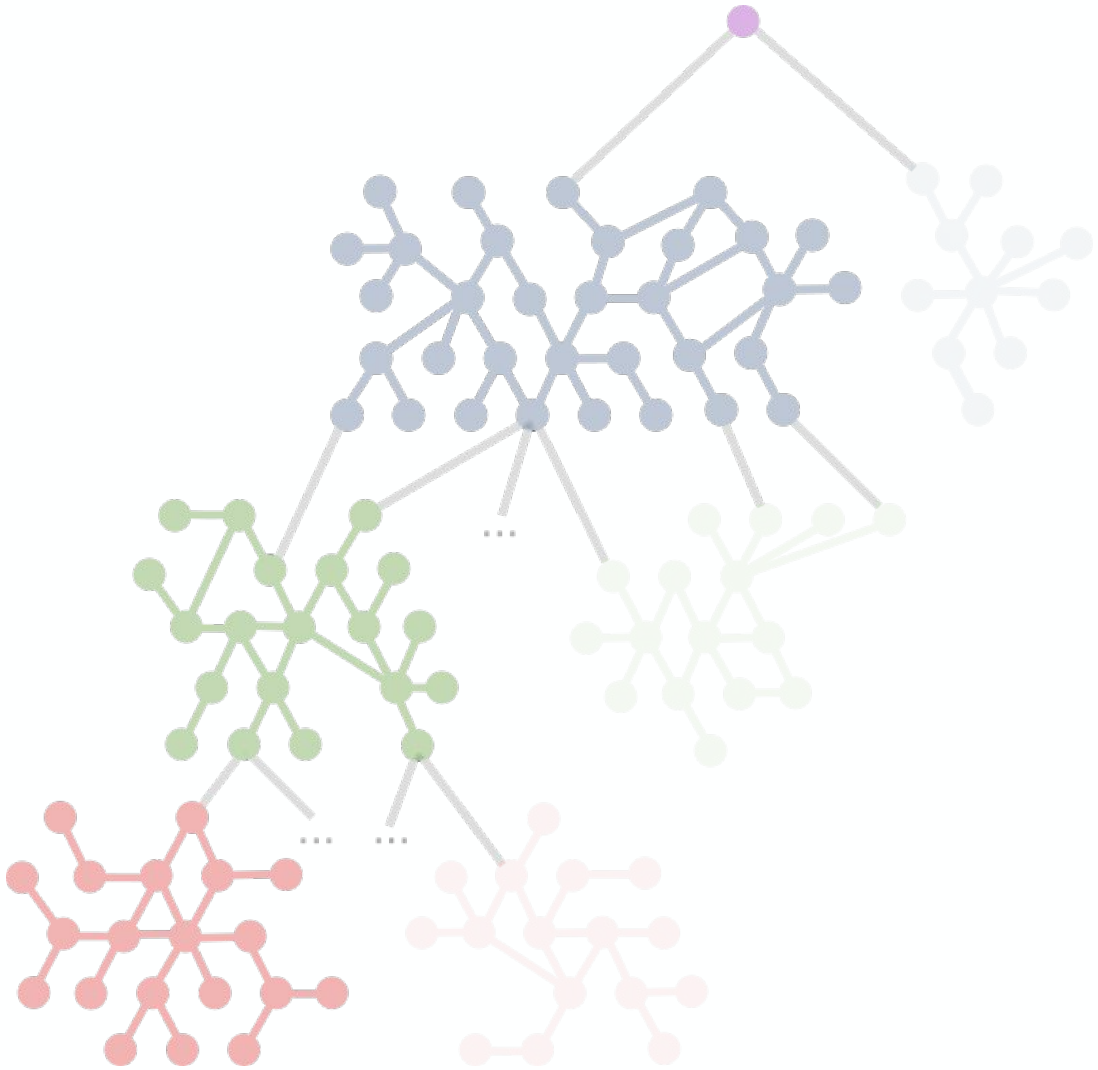
MatWerk Computational Workflow Schema

Reference Dataset Ontology - Creep - RDOC

MatWerk Dataset Schema

MatWerk Experimental Workflow Schema

Outline



Top- and mid- level alignment:

Description of resources in NFDI

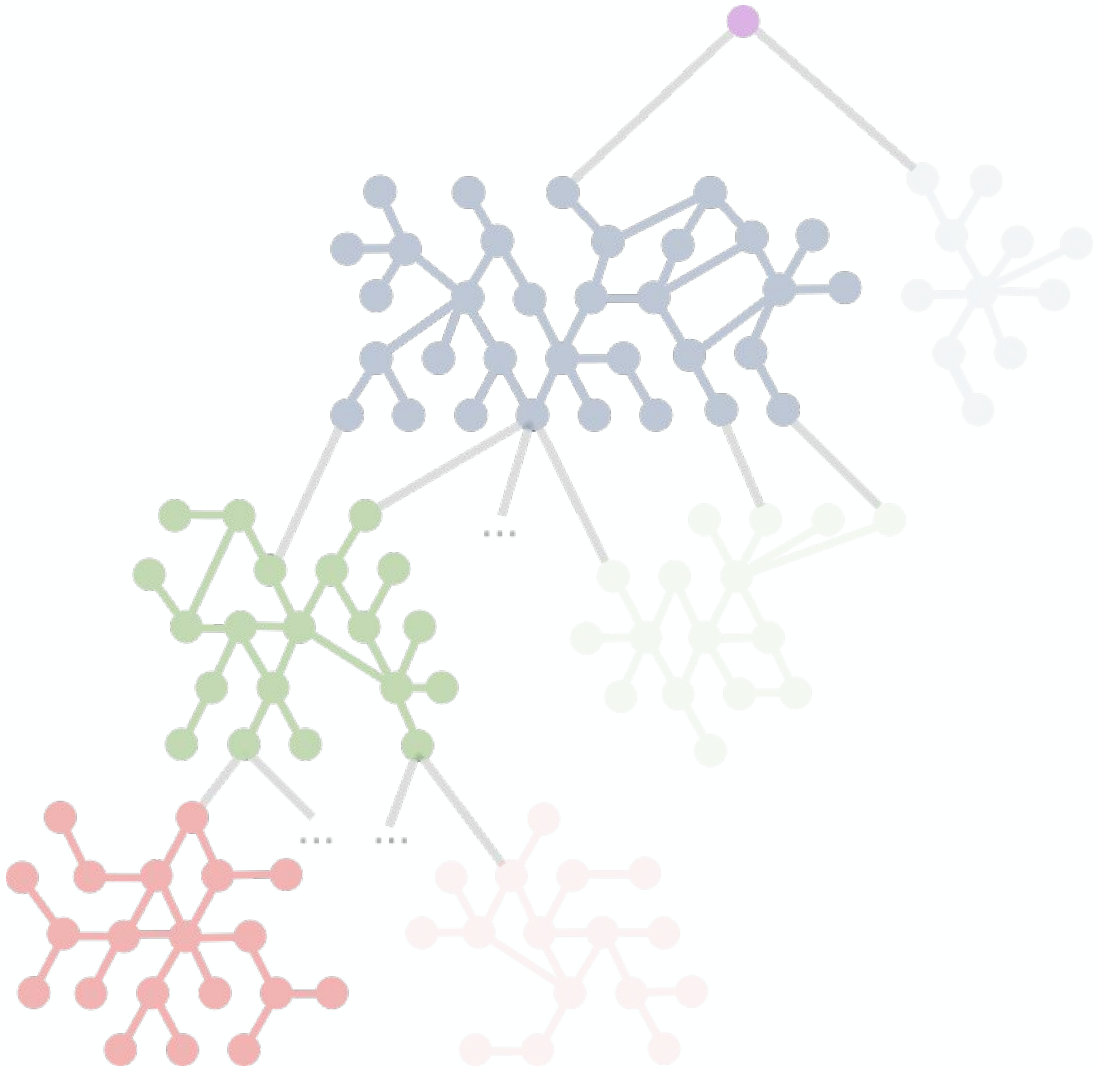
Crystallographic defects

ontologies

Infrastructure Use Case:

Atomistic simulations

Outline



Top- and mid- level alignment:

Description of resources in NFDI

Crystallographic defects

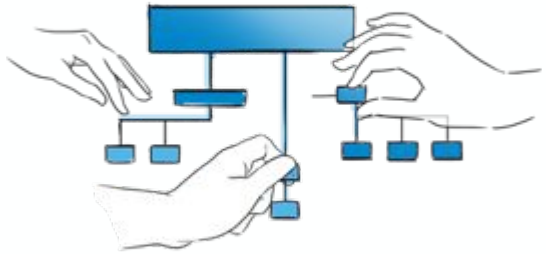
ontologies

Infrastructure Use Case:

Atomistic simulations

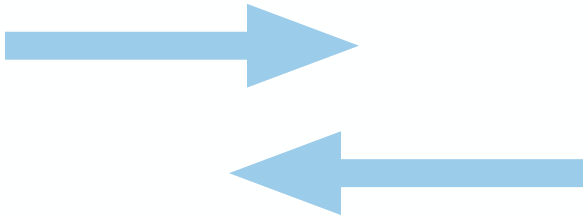
Metadata of NFDI-MatWerk resources

perspective from Task Area Ontologies for Materials Science**



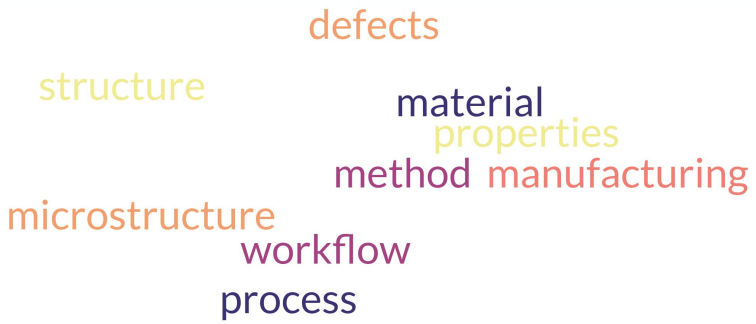
MSE domain

- Development of domain and application level ontologies and metadata schemas



Infrastructure

- Alignment of top- and mid-level
- MSE Knowledge Graph
- Connection with other NFDIs and consortia (RDA,PMD,etc)



FAIRmat MATERIALDIGITAL NFDi4ing RDA RESEARCH DATA ALLIANCE

MatWerk Ontology

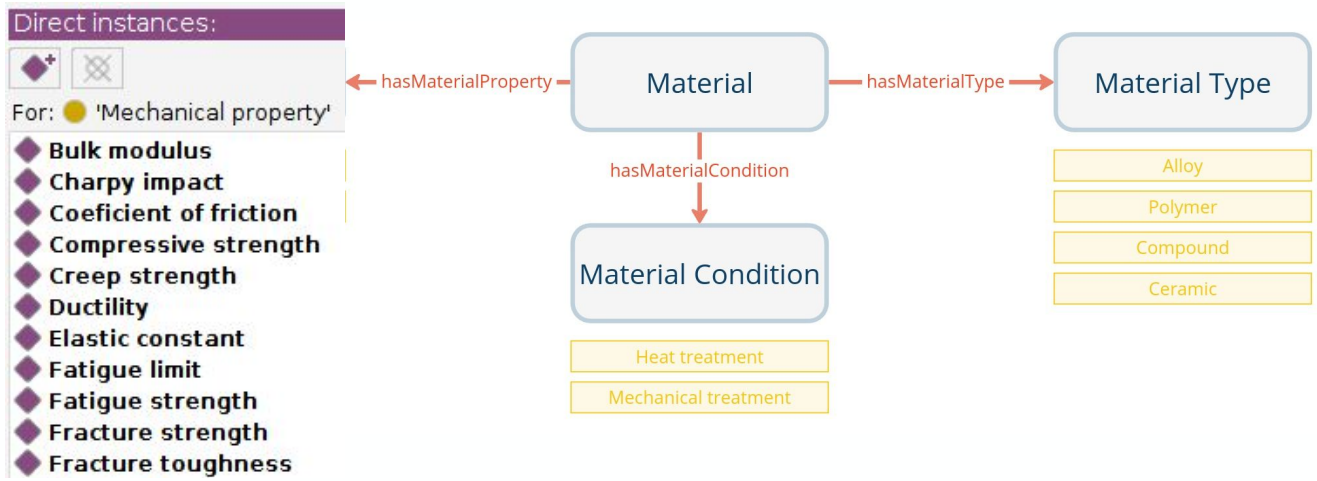
MWO ontology - release: v2.0.0 (2024-03)

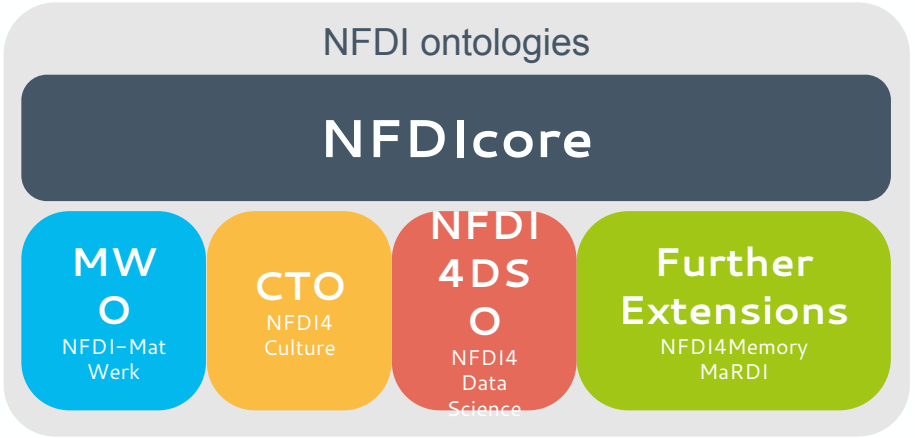
represents resources of NFDI-MatWerk

- community structure
- infrastructure
- data
- material
- computational workflows
- experimental workflows

IRI: http://purls.helmholtz-metadaten.de/mwo/

Checkout MWO at www.matportal.org

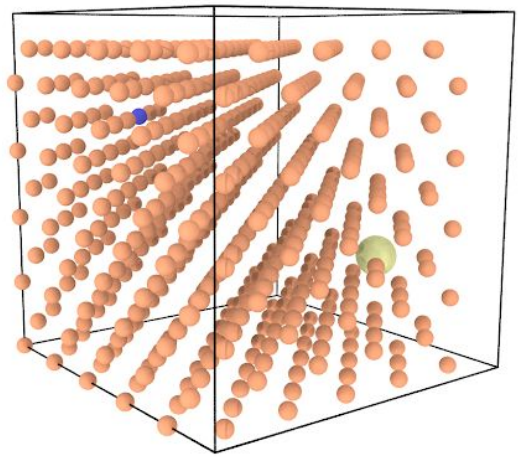




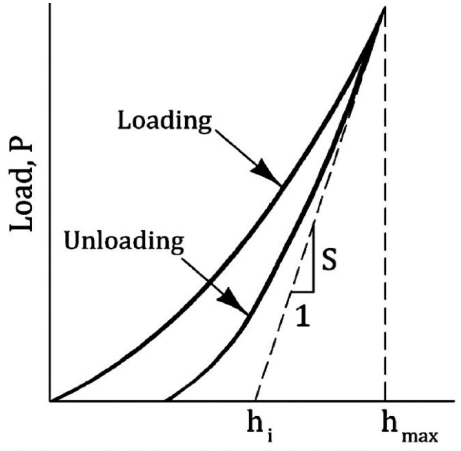
User Journey: calculation of mechanical properties

Simulation ← **Experiment** → **Data science**

- Elastic constant calculation of Al and Al + X (X=Si,Fe,Mg,Cu) impurities in LAMMPS

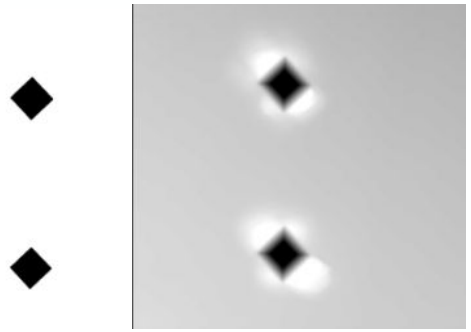


- EN AW 1050A (Al alloy) sample is indented, load-displacement curve is recorded, and the Young's modulus is calculated



Gdoutos, E., Konsta-Gdoutos, M. (2024). Indentation Testing. In: Mechanical Testing of Materials, vol 275. Springer, Cham.

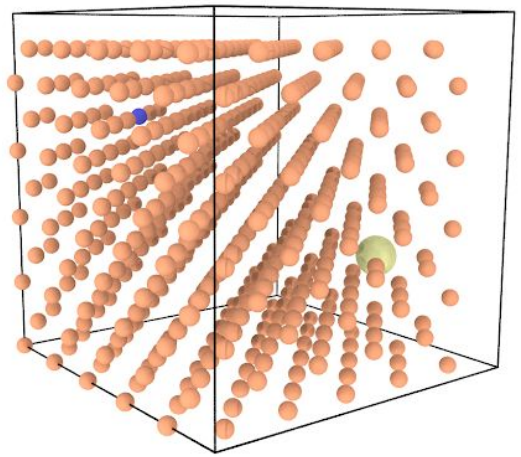
- Measured the physical size of imprints in Al indented sample microscopy images



User Journey: calculation of mechanical properties

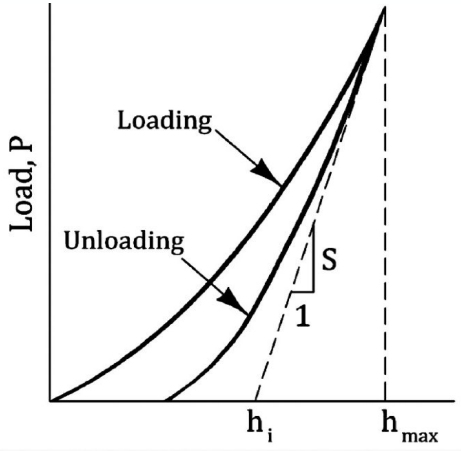
Simulation
pyiron

- Elastic constant calculation of Al and Al + X (X=Si,Fe,Mg,Cu) impurities in LAMMPS



Experiment
PASTA-ELN

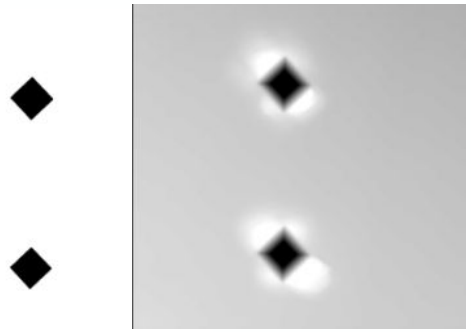
- EN AW 1050A (Al alloy) sample is indented, load-displacement curve is recorded, and the Young's modulus is calculated



Gdoutos, E., Konsta-Gdoutos, M. (2024). Indentation Testing. In: Mechanical Testing of Materials, vol 275. Springer, Cham.

Data science
scikit-image

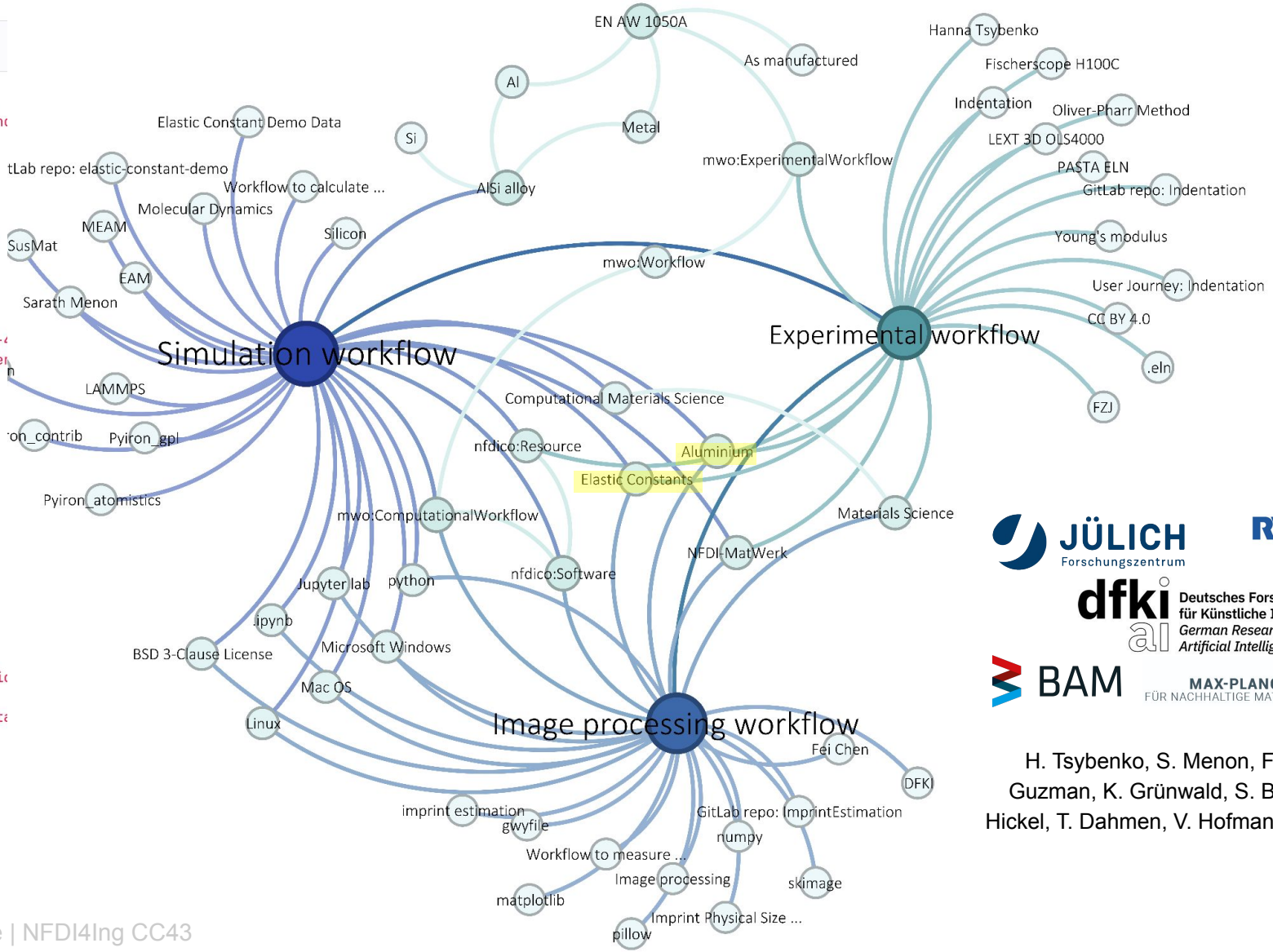
- Measured the physical size of imprints in Al indented sample microscopy images



Metadata of NFDI-MatWerk resources

metadata_exp_workflow.yml 2.22 KiB

```yaml
resourceType: Experimental Workflow
title: Indentation
description: an EN AW 1050A sample is polished and inc
author:
  - firstName: Hanna
    surname: Tsybenko
    title: Dr.-Ing.
    emailAddress: h.tsybenko@fz-juelich.de
    ORCID: "https://orcid.org/0000-0001-7691-2856"
    affiliation:
    - organization: Forschungszentrum Juelich
      RORID: "https://ror.org/02nv7yv05"
license: CC BY 4.0
repository: "http://hdl.handle.net/21.11102/7f78d0ed-4
gitRepository: "https://git.rwth-aachen.de/nfdi-matwer
documentation:
dateCreated: '2023-03-31T15:29:36.232Z'
datePublished: '2023-04-20T00:00:00.000Z'
fundingProject:
  - title: NFDI-MatWerk
    projectIdentifier: "460247524"
discipline: Materials Science
hasFormat:
  - fileFormat: ".json"
relatedResource:
  - resourceType: Software
    title: PASTA-ELN
    version: v1.3.5
    url: https://github.com/PASTA-ELN/
  - resourceType: Computational Workflow
    url: "https://gitlab.mpcdf.mpg.de/smenon/elastic
  - resourceType: Computational Workflow
    url: "https://git.rwth-aachen.de/nfdi-matwerk/ta
method: Non-destructive testing
keywords:
  - Elastic Constants
  - Aluminium
```



JÜLICH Forschungszentrum

RWTH AACHEN UNIVERSITY

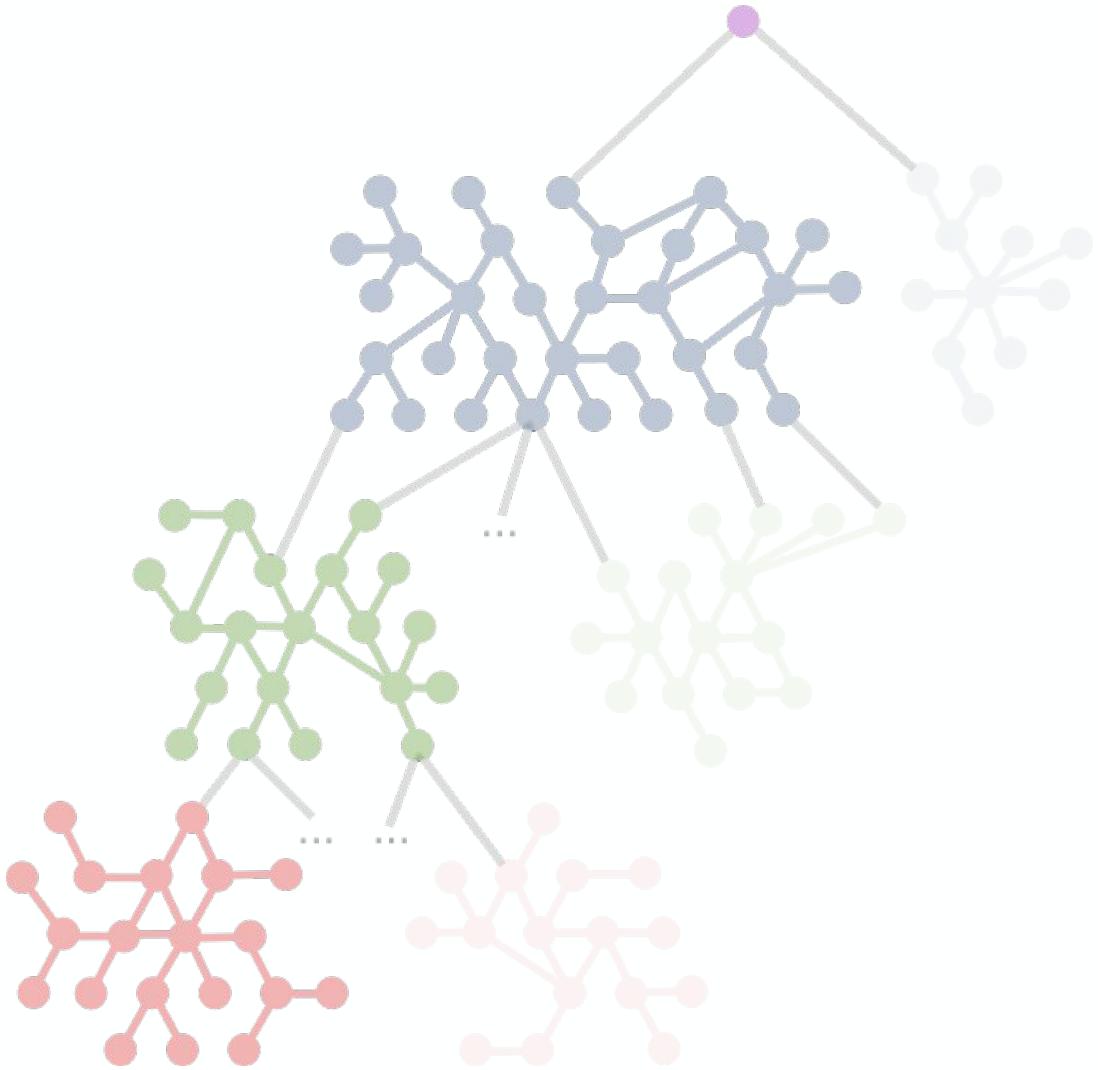
dfki Deutsches Forschungszentrum für Künstliche Intelligenz
German Research Center for Artificial Intelligence

BAM

MAX-PLANCK-INSTITUT FÜR NACHHALTIGE MATERIALIEN GMBH

H. Tsybenko, S. Menon, F. Chen, A.A. Guzman, K. Grünwald, S. Brinckmann, T. Hickel, T. Dahmen, V. Hofmann, R. Schwaiger

Outline



Top- and mid- level alignment:

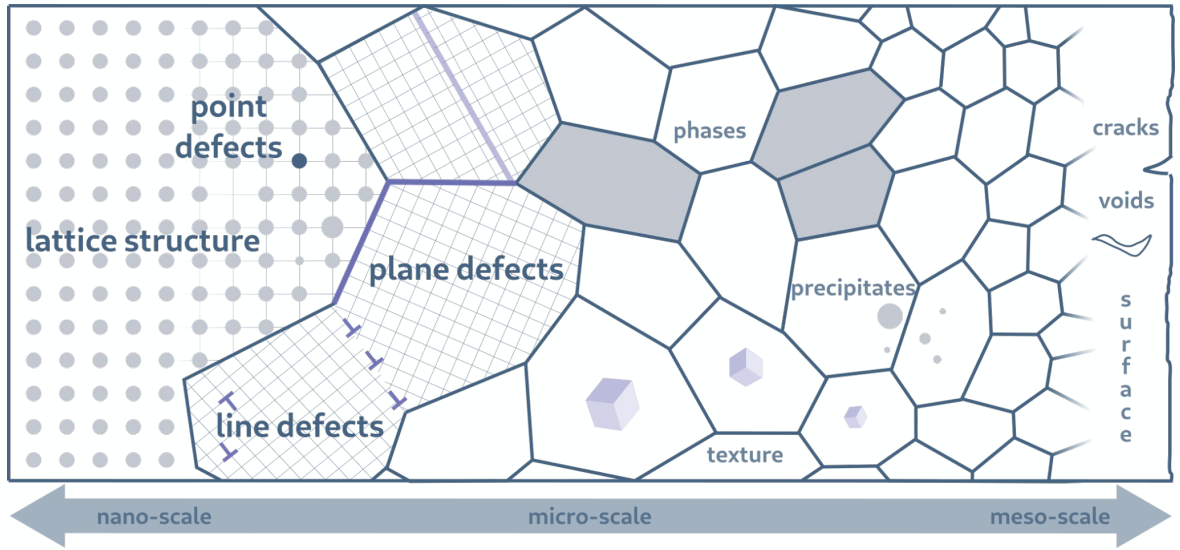
Description of resources in NFDI

Crystallographic defects ontologies

Infrastructure Use Case:

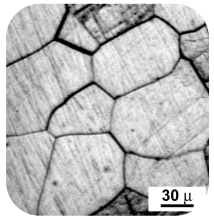
Atomistic simulations

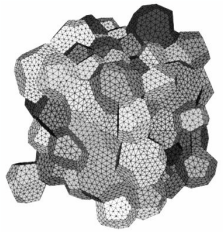
Crystal defects in materials science and engineering



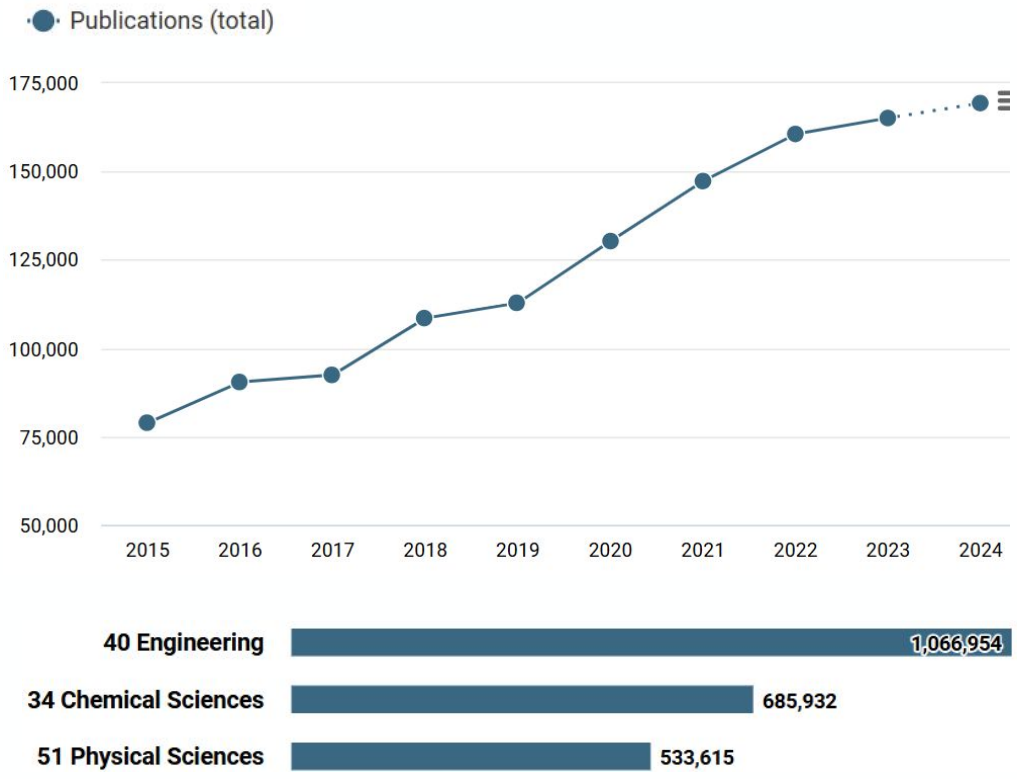
across scales

experiments + simulations





"crystal defects"



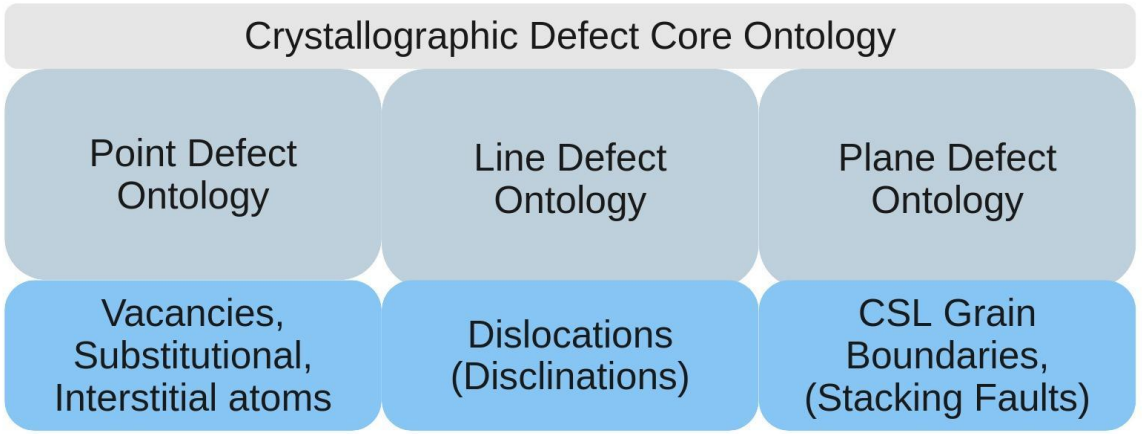
multidisciplinary

Crystal Defect Ontologies

OCDO
Open Crystallographic Defects Ontologies

https://github.com/OCDO

CDOS - Crystallographic Defect Ontology Suite



✓ align to existing ontologies, e.g. EMMO, MDO, QUDT

Open community

Application in atomistic simulations
A. Guzmán, S. Menon, V. Hofmann, T. Hickel, J. Neugebauer, S. Sandfeld

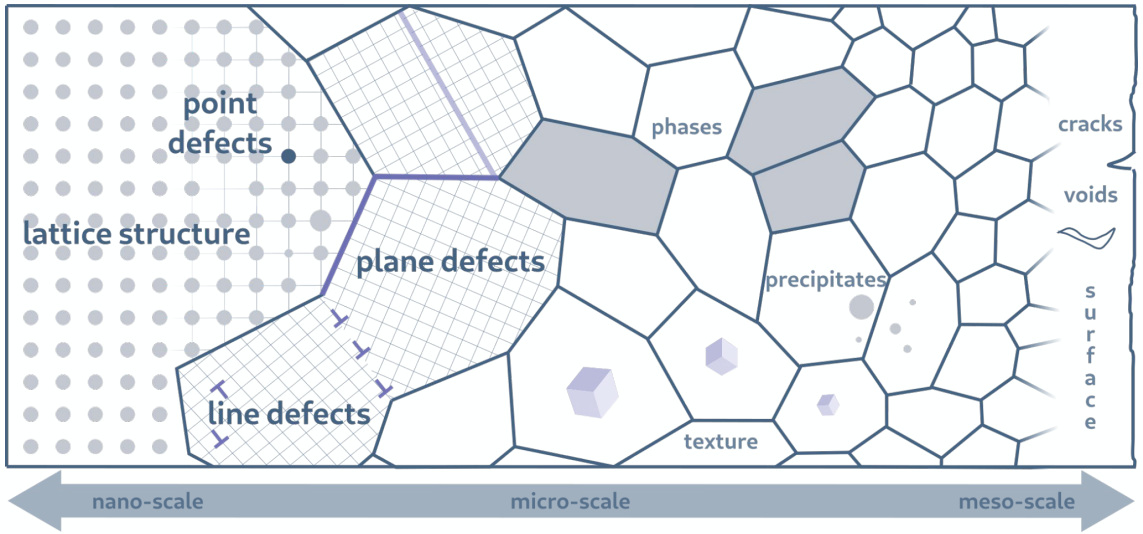
Stacking fault energy calculation workflows
S. B. Ravari, T. Hickel, M. Stricker

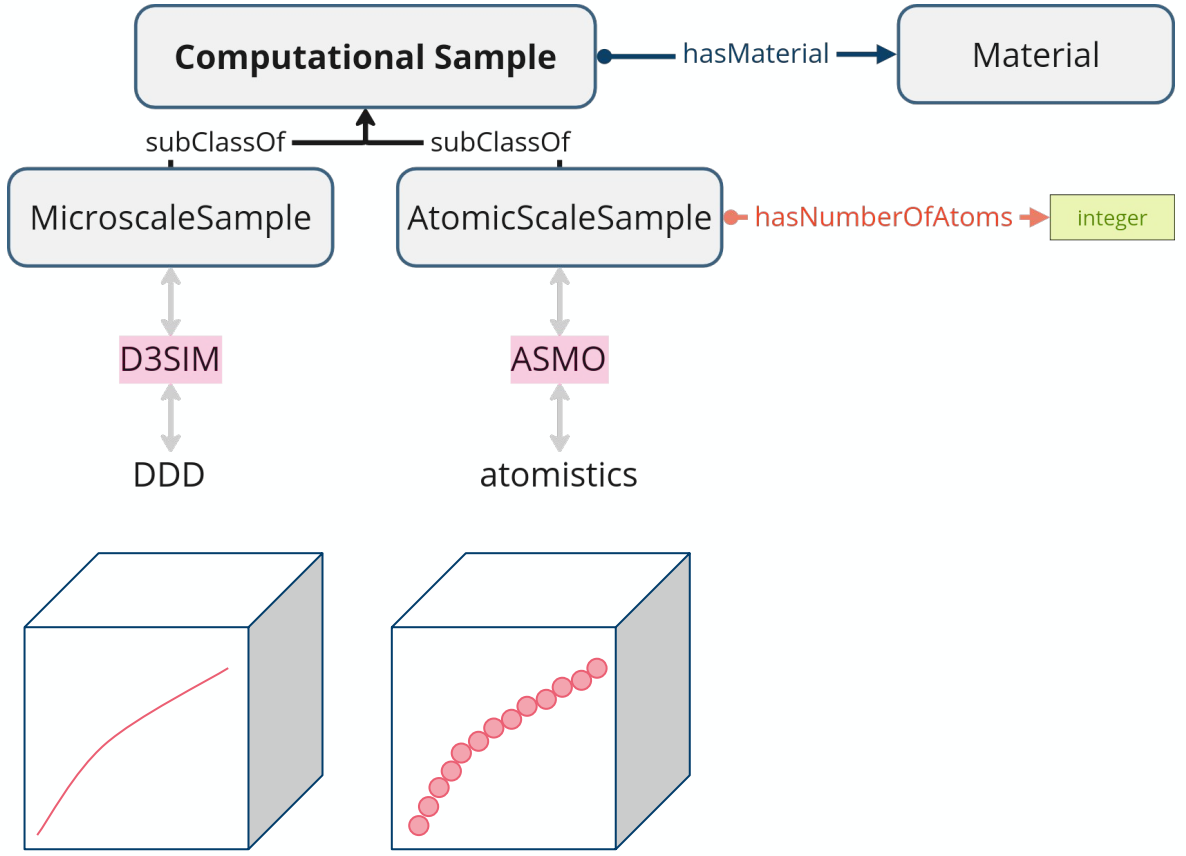
FAIR grain boundaries
E. Bitzek, S. Divinski, T. Hammerschmidt, R. Janisch, J. Mayer, H. Rösner, M. Stricker

Hydrogen-microstructure interaction
T. Hickel, P. von Hartrott, F. Schäfer, et al.

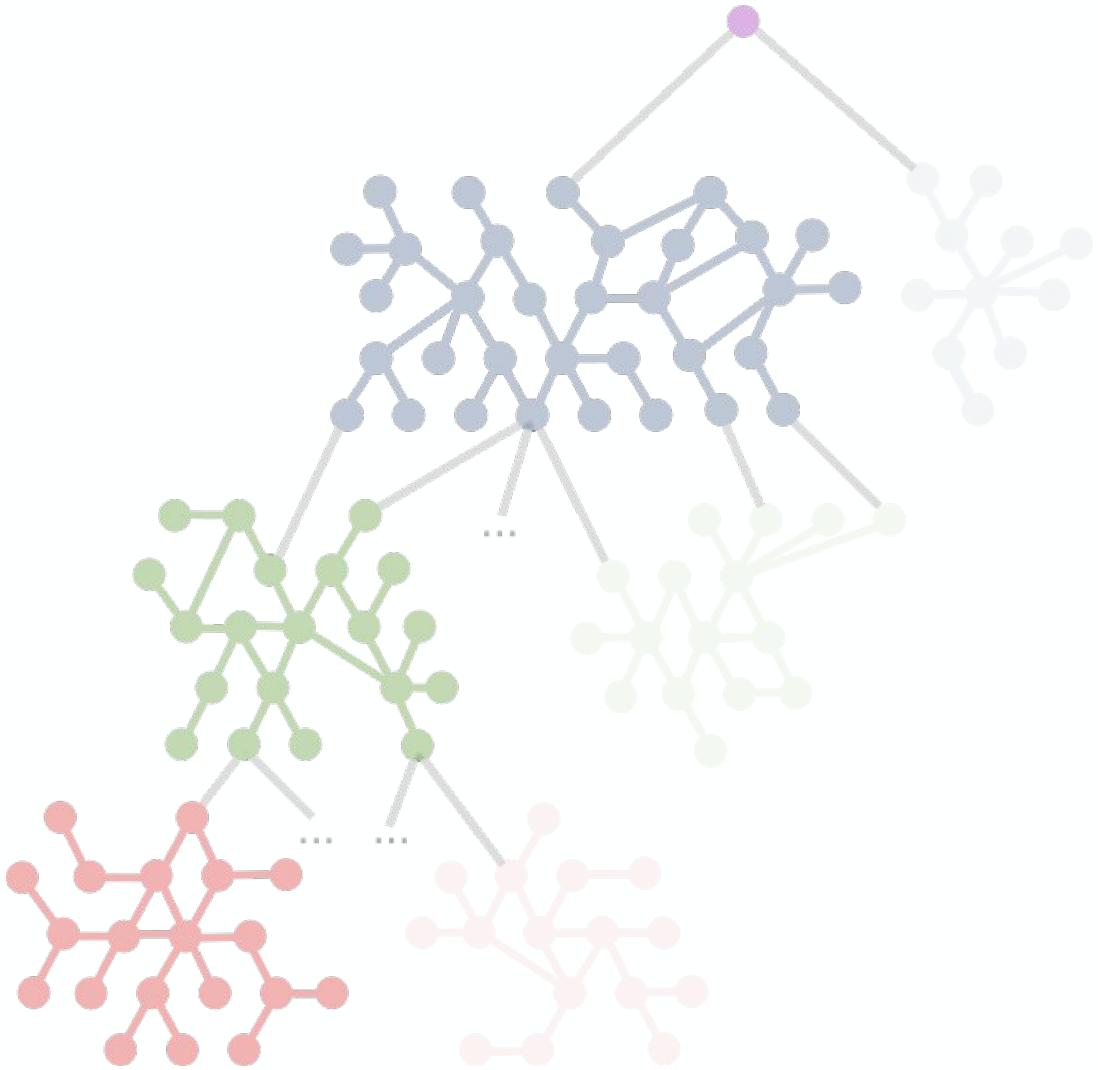
Description of materials

Representation across scales





Outline



Top- and mid- level alignment:

Description of resources in NFDI

Crystallographic defects

ontologies

Infrastructure Use Case:

Atomistic simulations

Computational materials science struggle with reproducibility

Feature | Published: 03 January 2024

Open computational materials science

Aron Walsh

Nature Materials **23**, 16–17 (2024) | Cite this article

"Even something as seemingly straightforward as the calculation of a ***point defect in a crystal*** *can require the combination of* ***dozens of files and several codes****."*

Walsh, A. *Nat. Mater.* 23, 16–17 (2024). https://doi.org/10.1038/s41563-023-01699-7

Is there a grain boundary plane?

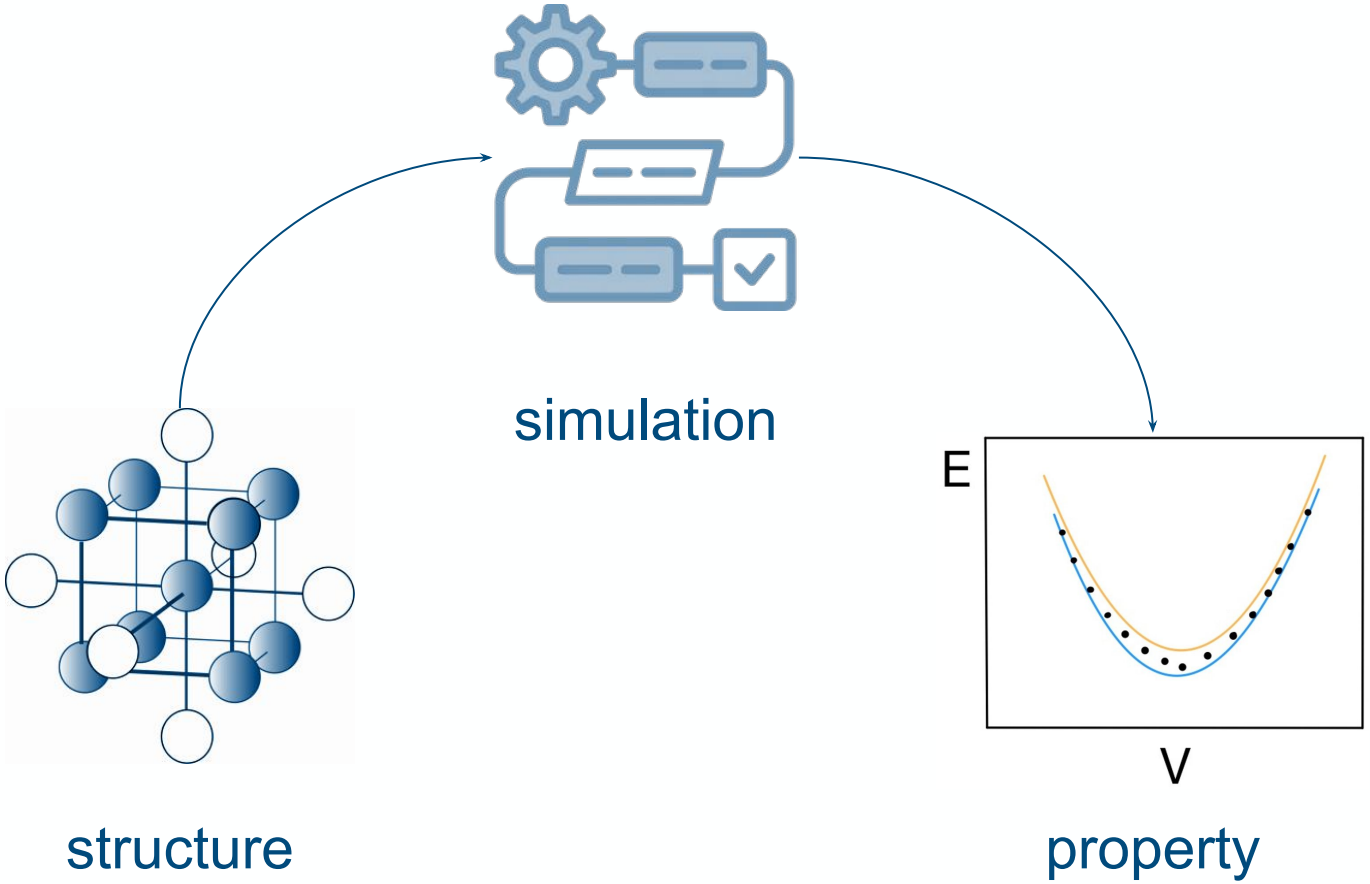
❌

```
data_image0
_chemical_formula_structural       Fe36
_chemical_formula_sum              "Fe36"
_cell_length_a       8.61
_cell_length_b       8.61
_cell_length_c       5.74
_cell_angle_alpha    90
_cell_angle_beta     90
_cell_angle_gamma    90

_space_group_name_H-M_alt    "P 1"
_space_group_IT_number       1

loop_
  _space_group_symop_operation_xyz
  'x, y, z'

loop_
  _atom_site_type_symbol
  _atom_site_label
  _atom_site_symmetry_multiplicity
  _atom_site_fract_x
  _atom_site_fract_y
  _atom_site_fract_z
  _atom_site_occupancy
  Fe  Fe1       1.0  0.00000  0.00000  0.00000  1.0000
  Fe  Fe2       1.0  0.16667  0.16667  0.25000  1.0000
  Fe  Fe3       1.0  0.33333  0.00000  0.00000  1.0000
```

CIF, POSCAR, lammps-data, XYZ, CJSON, struct, …
~70 file formats

Representation of crystal defect data through KGs

- Establish **semantic standards** for representing material structures, especially defect structures, simulation workflows and calculated properties
- Build knowledge graphs, including **structure-property relationships** of the material



Ontology development at the application level

Atomistic simulations

- Software used in the community

community agreement through code compatibility

- Data resources in the field

The Materials Project

NOMAD

OQMD
The Open Quantum Materials Database

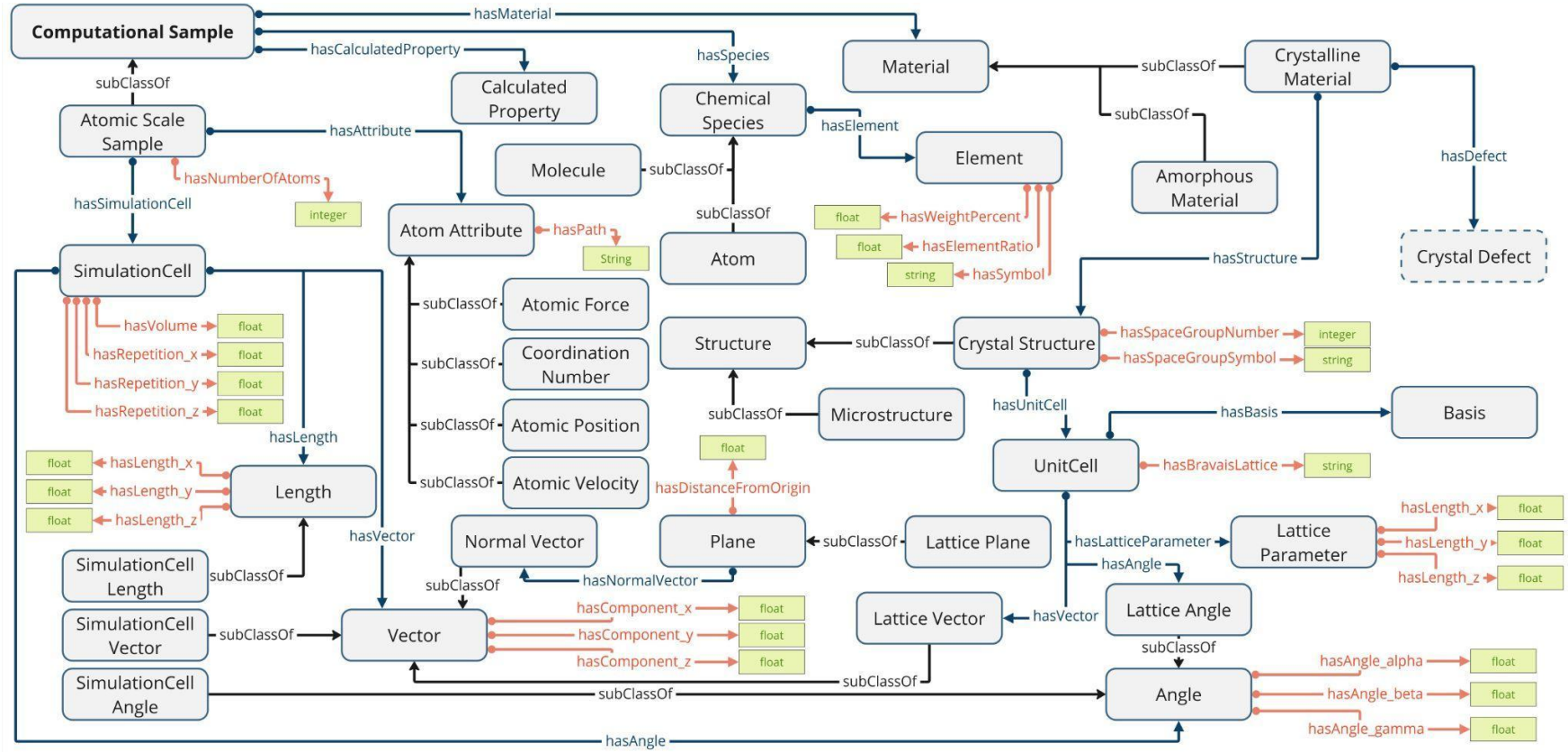
AFLOW
Automatic-FLOW for Materials Discovery

MATERIALSCLOUD

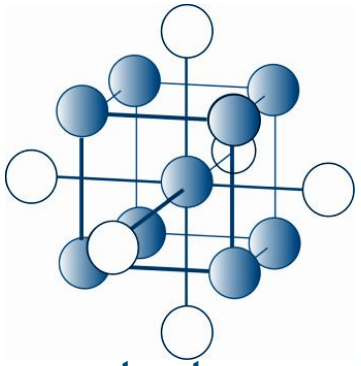
ICSD
FIZ Karlsruhe
Inorganic Crystal Structure Database

Ontologies for simulations

Computational Materials Sample Ontology - CMSO



http://purls.helmholtz-metadaten.de/cmso/

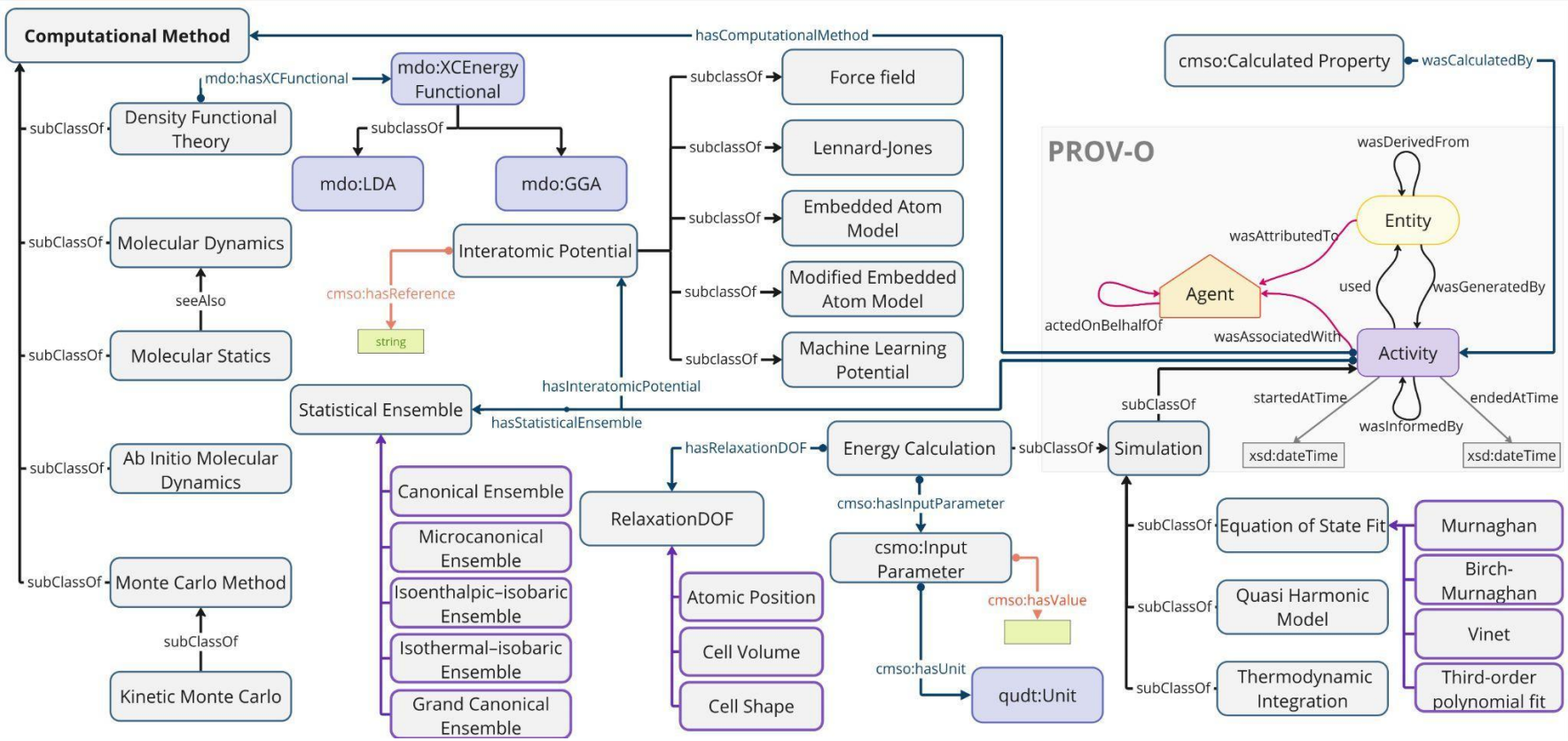


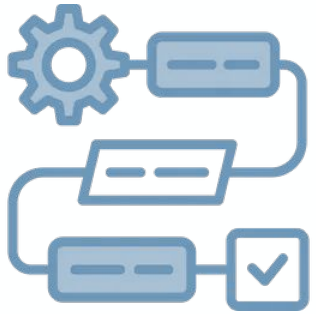
structure

- ➔ pyscal
- ➔ atomic simulation environment - ASE
- ➔ pymatgen
- ➔ Materials Project
- ➔ aimsgb
- ➔ atomman
- ➔ …

Ontologies for simulations

Atomistic Simulation Methods Ontology - ASMO





simulation

- ➔ pyiron_atomistics
- ➔ VASP
- ➔ Quantum ESPRESSO
- ➔ LAMMPS
- ➔ …

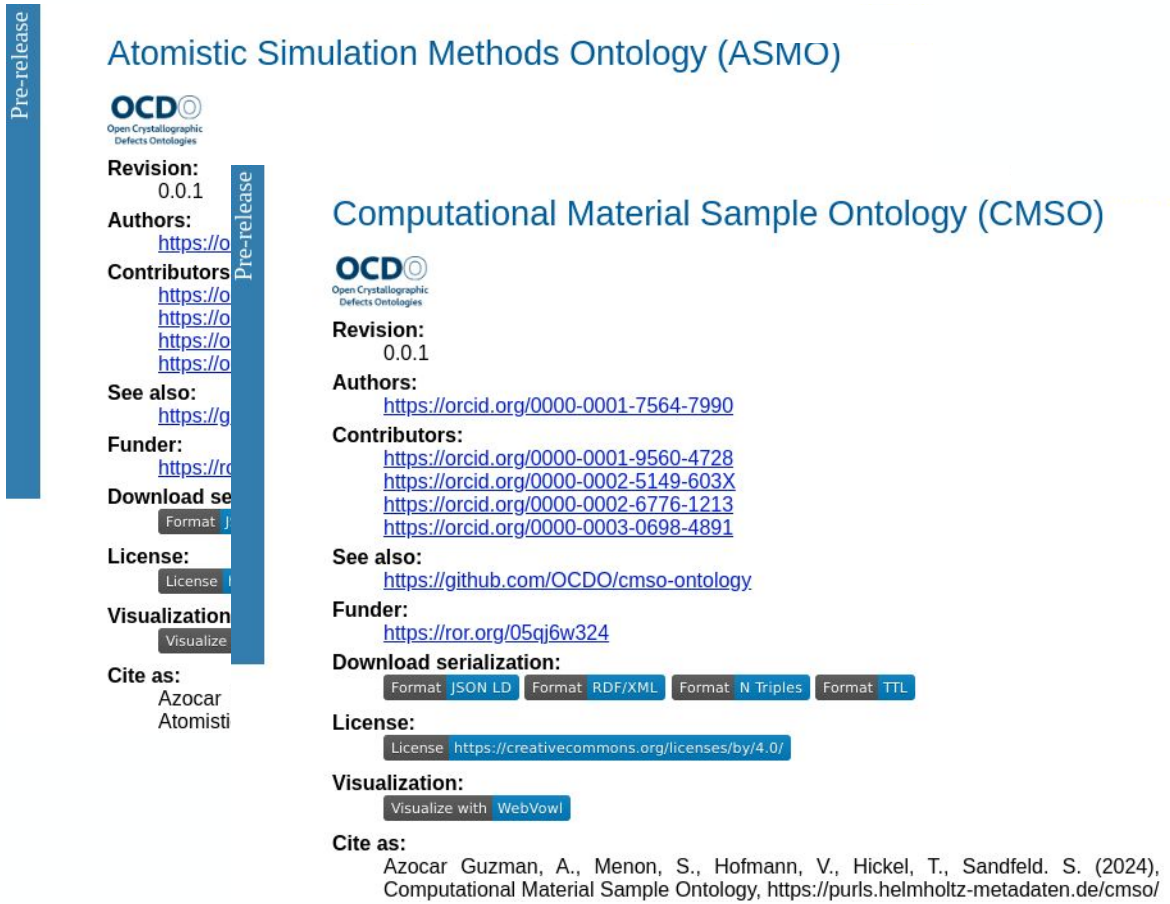
http://purls.helmholtz-metadaten.de/asmo

Ontologies for simulations

CMSO
Computational Material Sample Ontology
http://purls.helmholtz-metadaten.de/cmso/

ASMO
Atomistic Simulation Methods Ontology
http://purls.helmholtz-metadaten.de/asmo/

OCDO
Open Crystallographic Defects Ontologies
https://github.com/OCDO



Automatic instantiation

Different file formats
Varying levels of information

Annotate structures with CMSO
Common description
RDF serialization

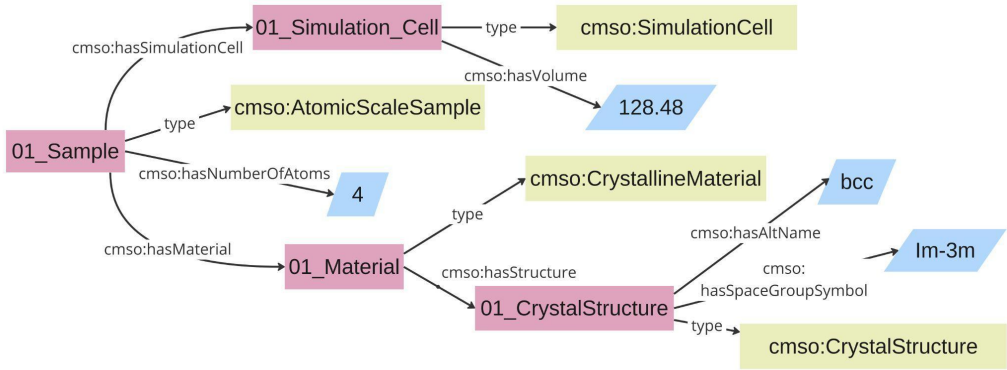
Integration with workflows
Query for structures directly in workflows
Publish W3C compliant datasets

atomRDF

Knowledge graph

Ontology-based representation of structures and workflows

- Interoperability: compatible with existing tools ASE, pyiron …
- Allows for **interoperability** on the semantic level.



User does not need SWT knowledge

Creating knowledge graphs

Dr. Tilmann Hickel, Dr. Sarath Menon
Task Area Workflows and Software Development

BAM MAX-PLANCK-INSTITUT FÜR NACHHALTIGE MATERIALIEN GMBH

- Installable through pip and conda

```
conda install -c conda-forge atomrdf
```

```
pip install atomrdf
```

atomRDF

Ontology-based representation

of structures and workflows

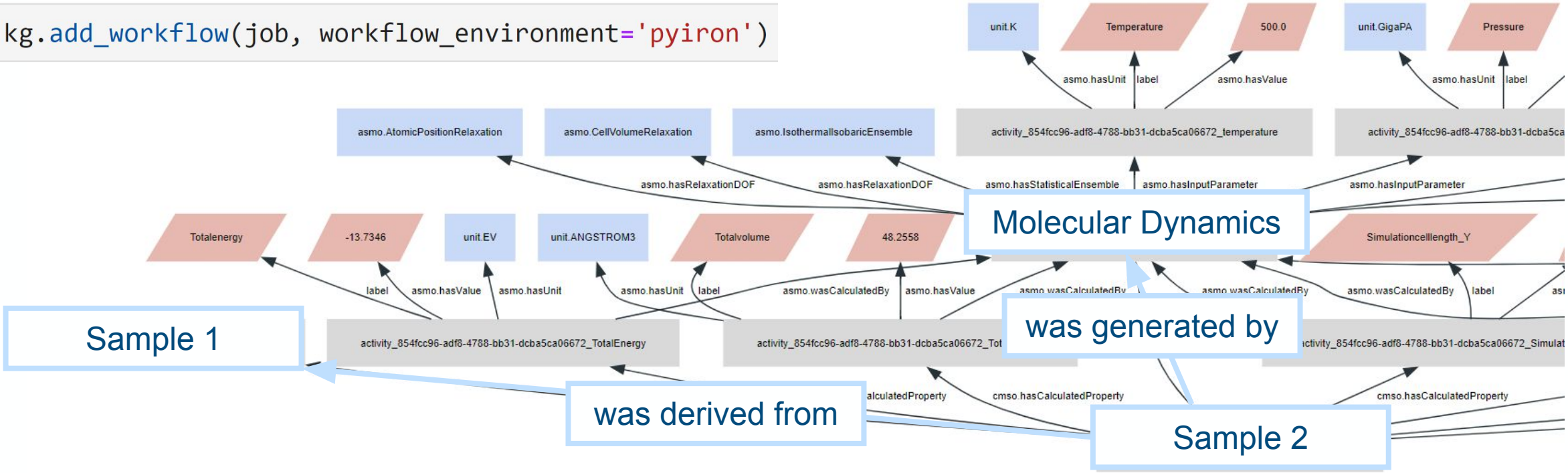
https://atomrdf.pyscal.org/

https://github.com/pyscal/atomRDF

Molecular Dynamics workflow

```
job = pr.create.job.Lammps('j1')
job.structure = structure
job.potential = '2001--Mishin-Y--Cu-1--LAMMPS--ipr1'
job.calc_md(pressure=0, temperature=500)
job.run()
```

```
kg.add_workflow(job, workflow_environment='pyiron')
```



Extracting thermodynamic properties from existing data
Application Use Case I

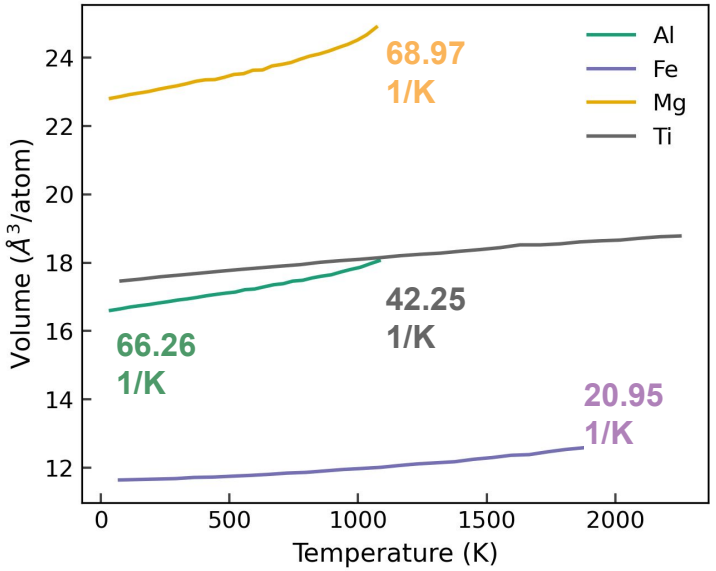
- Using DC3 dataset of MD simulations:
  - atomic configurations of different timesteps and temperatures, for eight elemental systems

Coefficient of thermal expansion

$$\alpha = \frac{1}{V}\frac{\partial V}{\partial T}$$

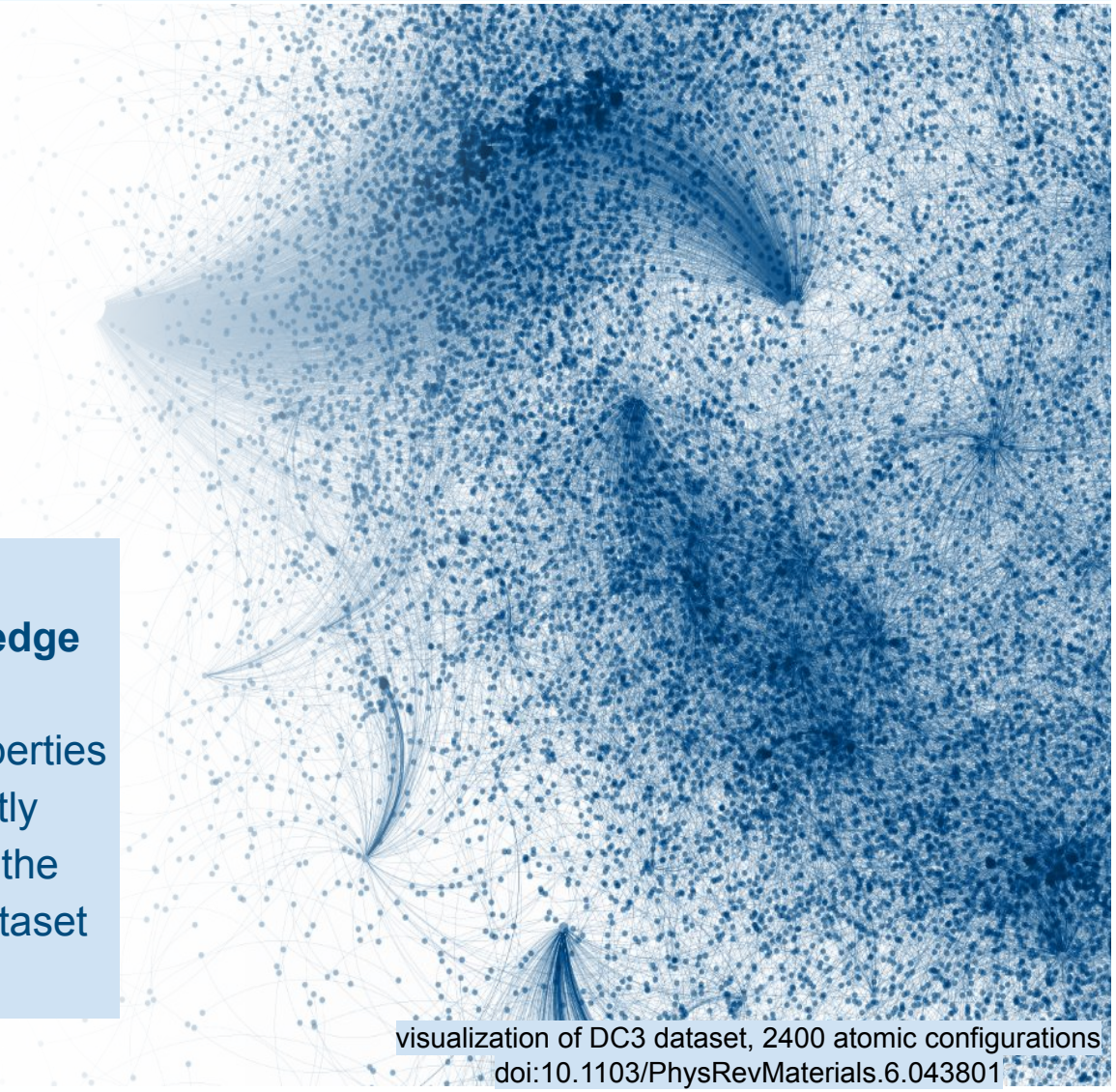
Physics-based constraints

same interatomic potential, same ensemble



New knowledge

materials properties not explicitly included in the published dataset



visualization of DC3 dataset, 2400 atomic configurations
doi:10.1103/PhysRevMaterials.6.043801

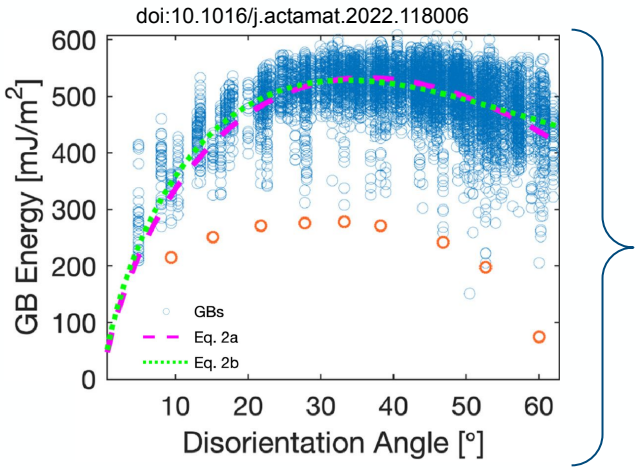
Grain boundary energy data
Application Use Case II

- Aggregate data from heterogeneous sources in one harmonized semantic framework
- Intuitive querying using autocompletion

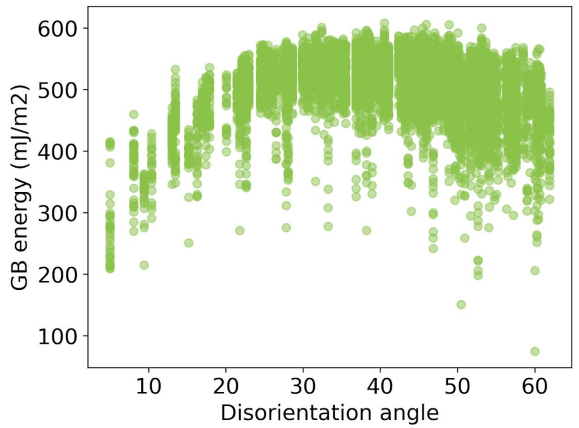
```
condition = (kg.terms.rdfs.label@kg.terms.cmso.CalculatedProperty=="GrainBoundaryEnergy")
kg.auto_query(kg.terms.cmso.AtomicScaleSample,
                    [condition,
                     kg.terms.pldo.hasMisorientationAngle,
                     kg.terms.rdfs.label@kg.terms.cmso.CalculatedProperty,
                     kg.terms.asmo.hasValue@kg.terms.cmso.CalculatedProperty
                    ],
              )
```

- Physics concepts embedded into data
- Easy querying

Published dataset: 7304 unique Aluminium GBs

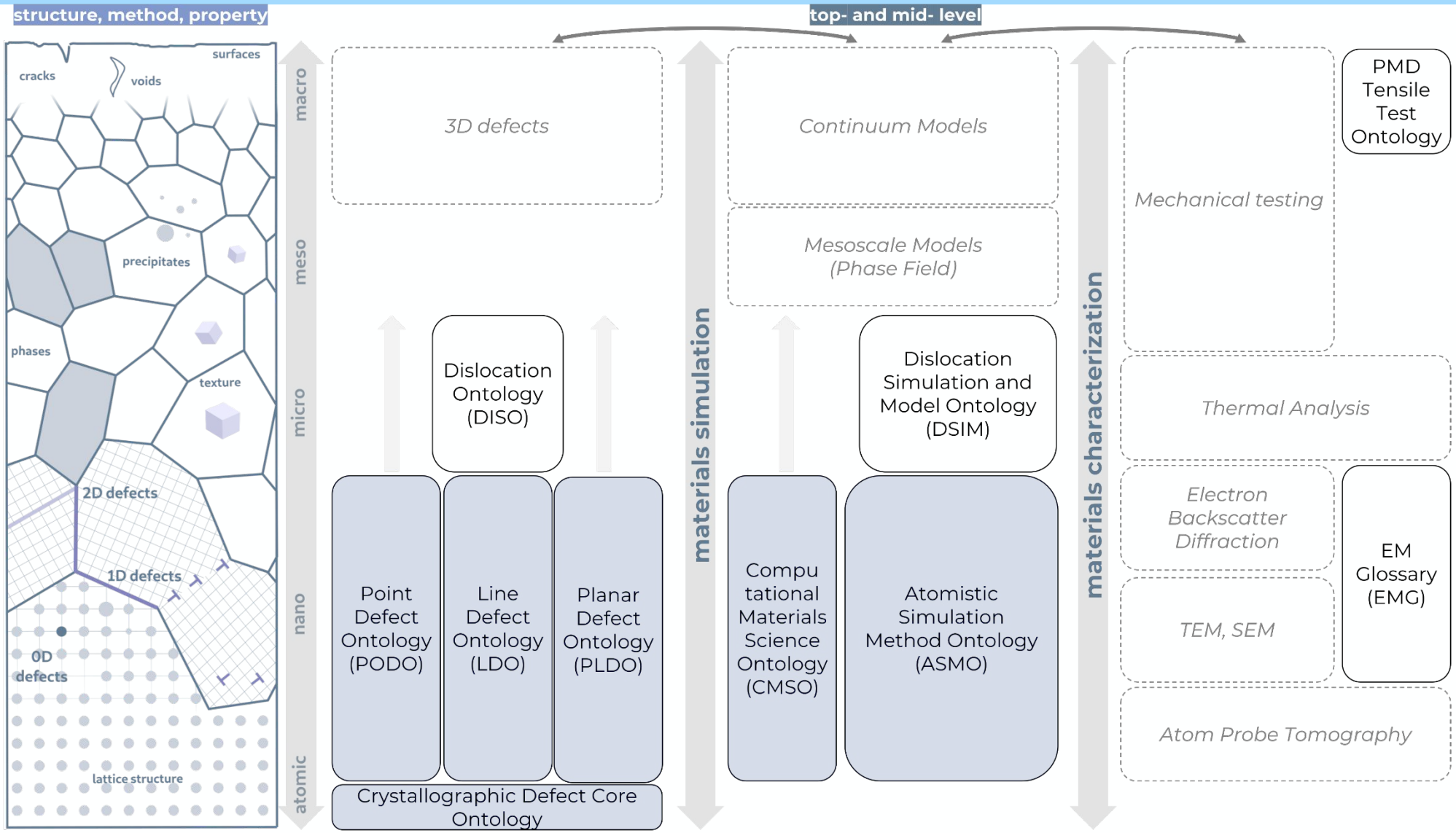


parsed with atomRDF into a knowledge graph



Domain scientist does not need semantic technologies knowledge

Outlook


structure, method, property
top- and mid- level
surfaces
cracks
voids
precipitates
phases
texture
2D defects
1D defects
0D defects
lattice structure
macro
meso
micro
nano
atomic
3D defects
Continuum Models
Mesoscale Models (Phase Field)
Dislocation Ontology (DISO)
Dislocation Simulation and Model Ontology (DSIM)
Point Defect Ontology (PODO)
Line Defect Ontology (LDO)
Planar Defect Ontology (PLDO)
Crystallographic Defect Core Ontology
Compu tational Materials Science Ontology (CMSO)
Atomistic Simulation Method Ontology (ASMO)
materials simulation
materials characterization
Mechanical testing
PMD Tensile Test Ontology
Thermal Analysis
Electron Backscatter Diffraction
TEM, SEM
EM Glossary (EMG)
Atom Probe Tomography

Acknowledgments

Task Area Ontologies for Materials Science @ FZJ

Prof. Stefan Sandfeld, Volker Hofmann, Said Fathalla, Ahmad Zainul Ihsan

Collaborators:

Sarath Menon, Tilmann Hickel, Prof. Joerg Neugebauer
Hanna Tsybenko, Steffen Brinckmann, Prof. Ruth Schwaiger

Summary

- **Ontologies for defects and simulations** are developed with a focus on application to specific research questions.

- Automated annotation allows creation of application-level **knowledge graphs of simulation data**, integration with workflows.

www.nfdi-matwerk.de

https://github.com/OCDO

a.azocar.guzman@fz-juelich.de

Nationale Forschungsdaten Infrastruktur

Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure – NFDI 38/1 – project number 460247524