

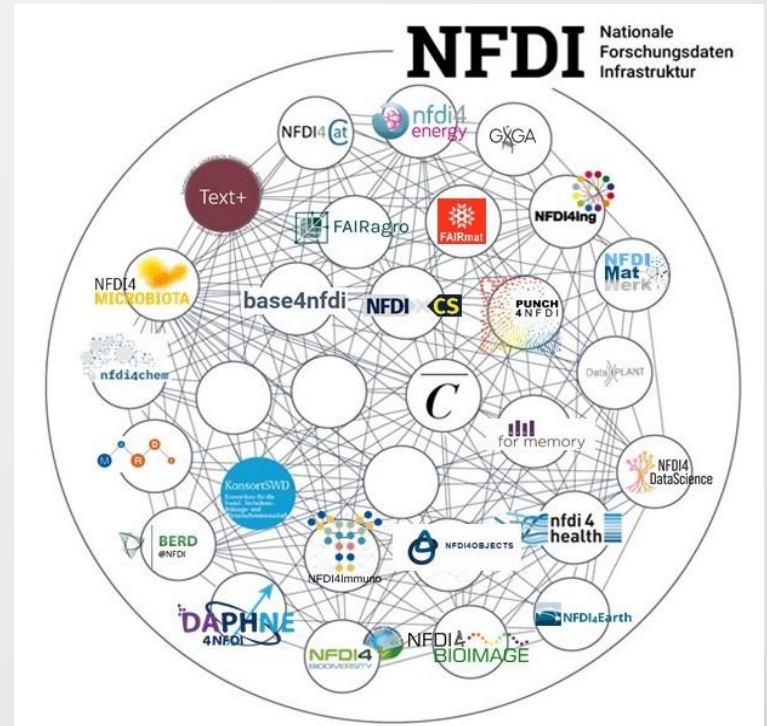
# MSE Knowledge Graph

**Ebrahim Norouzi**, Jörg Waitelonis, Heike Fliegl, Abril Azócar Guzmán, Said Fathalla, Ahmad Zainul Ihsan, Volker Hofmann, Stefan Sandfeld, Felix Fritzen, Amir Laadhar, Harald Sack

Patents4Science Workshop, 05th Oct 2023

## Facilitated access to digital data resources:

- Promote the exchange between different research units and disciplines
- Simplify research data management at the international level
- Exploit potentials across disciplines
- Take the needs of different communities into account



(Most) NFDI consortia have rather similar **information needs**

1. Provision of information **about the consortium** itself, as e.g.,
  - People, organisations, services, resources, etc.
2. Provision of information about the consortium related **research data**, i.e.
  - (Distributed) **data resources** and (data) **services**

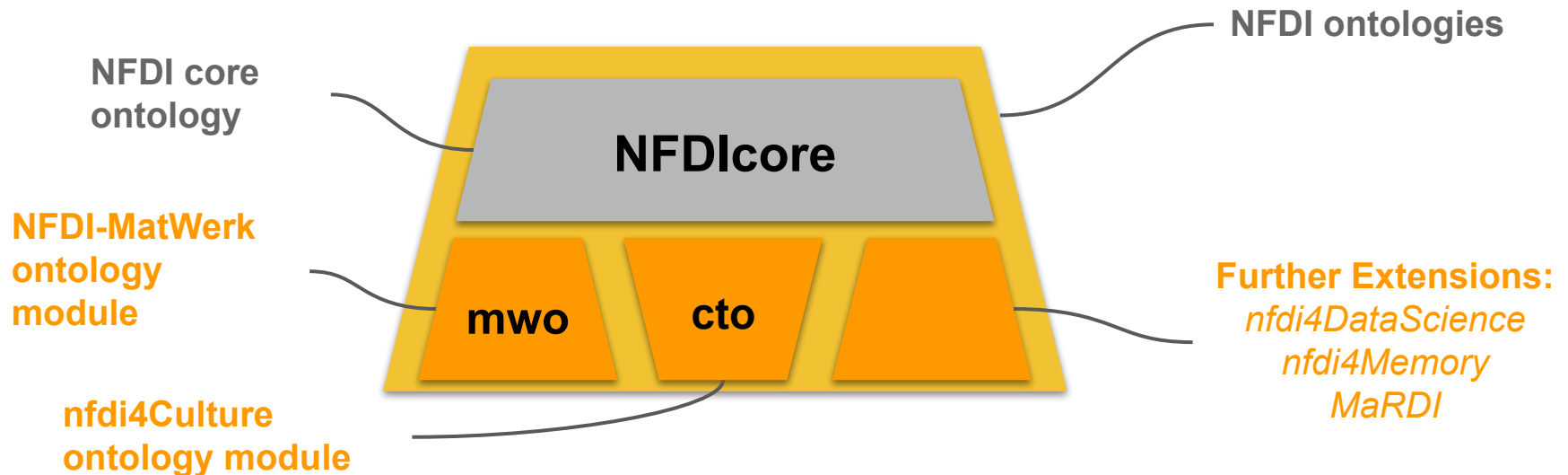
## Team:

- Prof. Harald Sack
- Prof. Stefan Sandfeld
- Prof. Felix Fritzen
- Dr. Volker Hofmann
- Dr. Abril Azocar-Guzman
- Dr. Said Fathalla
- Dr. Jörg Waitelonis
- Dr. Heike Fliegl
- Dr. Amir Laadhar
- Dr. Angelika Gedsun
- Ebrahim Norouzi
- Ahmad Ihsan
- Mirza-Mohtashim Alam



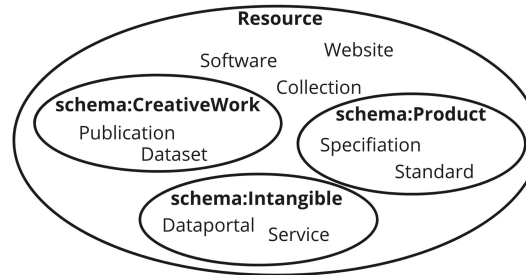
- **NFDI Ontologies Architecture**
  - NFDI Core Ontology (NFDI Core)
  - MatWerk Ontology - MWO
- MSE Knowledge Graph v1.0
- MSE Knowledge Graph v2.0
  - NFDI MatWerk LOD Working Group

- **NFDI ontologies** follow **modular approach**
- **NFDIcore v1.1.0** represents **NFDI consortia in general**
- **NFDI modules** represent **specific consortia**, as e.g. NFDI-MatWerk
- Specific requirements of consortia lead to **ontology extension**



- Represents NFDI consortia and their contributions (services, data sets, guidelines, etc)
- Contains 34 classes and 58 object and data properties
- Links to 20+ external vocabularies
- Available online:  
<https://github.com/ISE-FIZKarlsruhe/nfdicore>
- Documentation:  
[https://nfdi.fiz-karlsruhe.de/ontology\\_v1.1.0](https://nfdi.fiz-karlsruhe.de/ontology_v1.1.0)

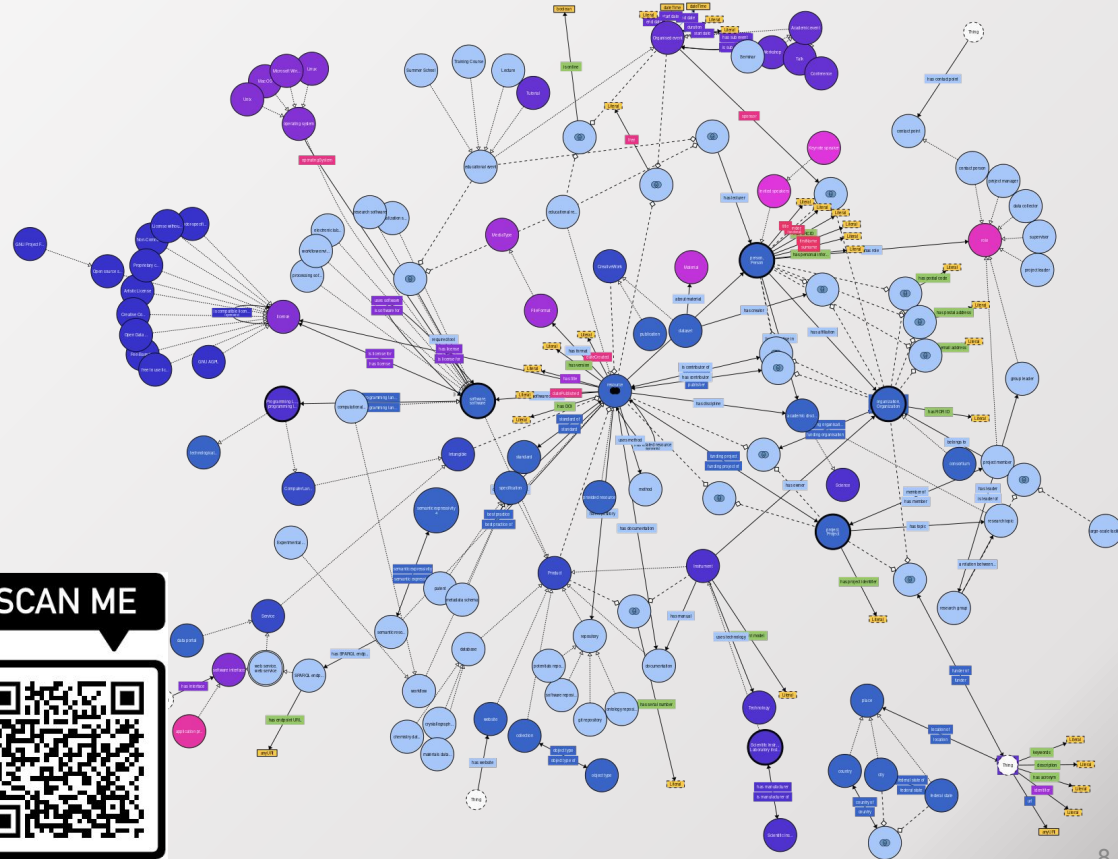
Temporary namespace:  
<https://nfdi.fiz-karlsruhe.de/ontology>



- owl:Thing
  - :AcademicDiscipline
  - :DevelopmentWork
  - :License
  - :MediaType
  - :ObjectType
  - :Organization
    - :Consortium
  - :Person
  - :Place
  - :Project
  - :Resource
    - :Collection
    - :ProvidedResource
    - :Software
    - :Website
    - schema:CreativeWork
    - schema:Intangible
    - schema:Product
  - :SemanticExpressivity
  - :ServiceType
  - :TechnologicalMeans



- **MWO ontology - release: v1.0.0**
- Based on NFDICORE v1.0.2
- Contains **115 classes** and **103 object and data properties**
- Mapped to 11 external vocabularies
- Content negotiation (HTML/OWL) enabled
- Documentation can be found at: <http://purls.helmholtz-metadaten.de/mwo>
- Development version: <https://git.rwth-aachen.de/nfdi-matwerk/ta-oms/mwo>
- Modular design based on NFDI Core + Domain extension (MWO) e.g. Material, Method, Instrument and Workflows.





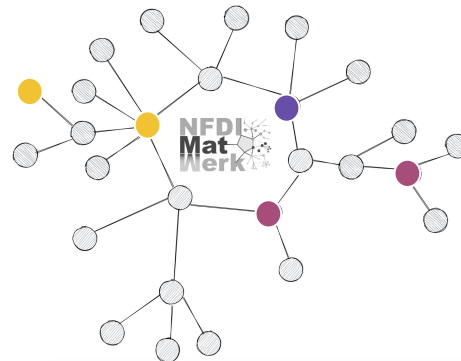
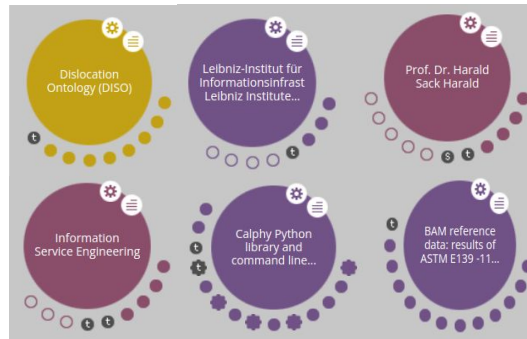
- NFDI Ontologies Architecture
  - NFDI Core Ontology (NFDI Core)
  - MatWerk Ontology - MWO
- **MSE Knowledge Graph v1.0**
- MSE Knowledge Graph v2.0
  - NFDI MatWerk LOD Working Group

## CONTENT:

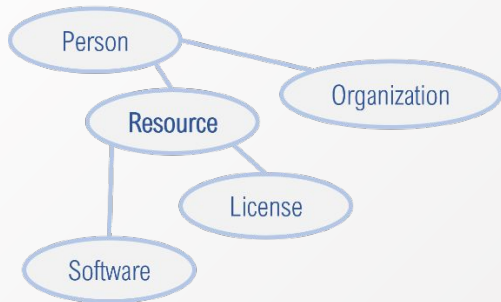
- Index and Metadata for **MatWerk** Project Resources
  - **Community:** persons, institutions, etc.
  - **Infrastructure:** software, workflows, instruments, etc.
  - **Data:** data repositories, publications, etc.
  - **Educational Events:** Lecture, Summer School, Educational Resource, etc.

## PURPOSE:

- Backend (Meta-)Data Resource for the **MatWerk** Portal

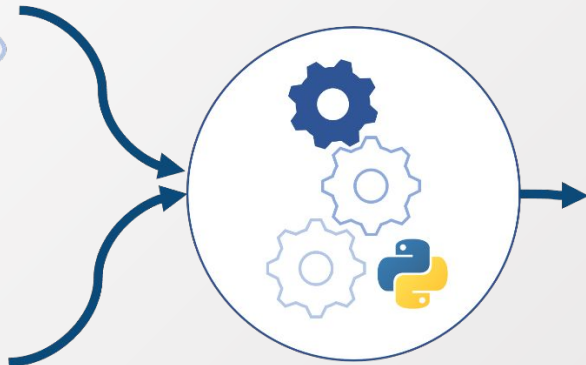


## MatWerk Ontology

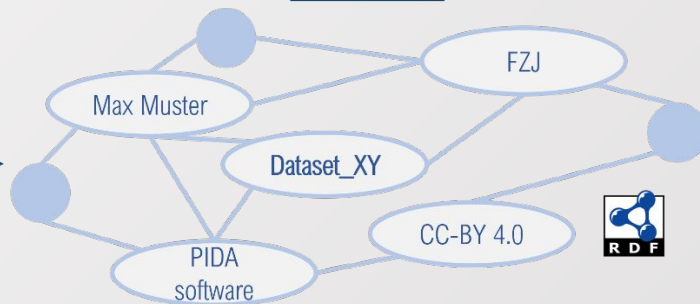


## Collected Data

First name	Last name	Affiliation
Max	Muster	FZJ
.	.	.
..	..	..



## MSE KG



[https://git.rwth-aachen.de/nfdi-matwerk/ta-oms/kg\\_demonstrator](https://git.rwth-aachen.de/nfdi-matwerk/ta-oms/kg_demonstrator)

1. created survey spreadsheet which we sent out to PPs and other TAs
2. Manual gathering of data related to NFDI-MatWerk and MSE community
3. Total of 381 entries in 9 categories

## Currently collected data

Category	Total
People	51
Software	199
Data Portals	67
Metadata & Standardization	33
Published Datasets	7
Large Scale Facilities	8
Instruments	2
Educational Events & Resources	14
Patents	0

### currently collected data on participants:

Personnel info		People
Name		Institution / Institute / Association

### currently collected data on data portals:

Basic info		Data formats	Data portals: Database / Repository	Contact
Name		Interoperability		

### currently collected data on metadata standards:

Basic info		Metadata & Standardization: Schemas/ Ontologies/ Glossaries / Standards for Metadata/ Workflow description
Name		Contact

### currently collected data on software:

Software: Tools, Frameworks, Services, Simulation codes															
Basic info			Contact				Usage								
#	Name	Acronym	Short description	Type	Field of research	Year released	Link (URL / PID) to website	Link (URL / PID) to repository	Name	Email	License	Runtime environment	Link (URL / PID) to documentation	Related infrastructure	Programming language
14	Cafly		Python library and associated test tool for calculation of thermodynamics, molecular dynamics, electronic simulations, thermodynamics, molecular dynamics, electronic simulations, computational materials	Research software	thermodynamics, molecular dynamics, electronic simulations, computational materials	1.1.4	<a href="https://cafly.org/">https://cafly.org/</a>	<a href="https://github.com/CAMSLab/cafly">https://github.com/CAMSLab/cafly</a>	Sarah Monson	s.monson@mpie.de	ASU BSD-3	Linux, OSX	<a href="https://cafly.org/">https://cafly.org/</a>	Calypso users LAMMPS, pymatgen, pyiron	python
16	Melting temperature computational workflow		A repository illustrating the calculation of melting temperature through ab initio simulations using generic Python library written in C++ for calculation of local atomic structure environments	Computational Workflow	thermodynamics, molecular dynamics, ab initio simulations, computational materials	2.10.15	<a href="https://melt.org/">https://melt.org/</a>	<a href="https://github.com/compflow/melting_workflow">https://github.com/compflow/melting_workflow</a>	Sarah Monson	s.monson@mpie.de	BSD-3	Linux, OSX, Windows	<a href="https://melt.org/">https://melt.org/</a>	Pyiron users ASE, Pyiron, pymatgen, pyiron	python, C++
19	Pycol		A python script to implement the generation of disordered atomic structures with external z-axis constraint.	Research software	Computational materials science		<a href="https://pycol.org/">https://pycol.org/</a>	<a href="https://github.com/monson/pycol">https://github.com/monson/pycol</a>	Neucom Fernandes	fernandes@cpm.ford.umd.edu	BSD-3	Linux, OSX, Windows	<a href="https://pycol.org/">https://pycol.org/</a>		python
20	EQPC2		An algorithm implementing the maximum entropy method to reconstruct the distribution of non-hydrogen atoms in a crystal structure from experimental diffraction data.	Research software	Computational mathematics, computational materials science		<a href="https://github.com/EMMA-CPD/eqpc2">https://github.com/EMMA-CPD/eqpc2</a>	<a href="https://github.com/EMMA-CPD/eqpc2">https://github.com/EMMA-CPD/eqpc2</a>	Oliver Kunc, Felix Fritzer	oliver.kunc@mpie.de, felix.fritzer@mpie.de	BSD-3	Linux, OSX, Windows	<a href="https://github.com/EMMA-CPD/eqpc2">https://github.com/EMMA-CPD/eqpc2</a>		python
21	MinimalEnergyPoints		Interpolate and differentiate data in arbitrary dimensions efficiently in a concise manner	Research software	Computational mathematics, computational materials science		<a href="https://github.com/EMMA-CPD/minimal-energy-points">https://github.com/EMMA-CPD/minimal-energy-points</a>	<a href="https://github.com/EMMA-CPD/minimal-energy-points">https://github.com/EMMA-CPD/minimal-energy-points</a>	Oliver Kunc, Felix Fritzer	oliver.kunc@mpie.de, felix.fritzer@mpie.de	BSD-3	Linux, OSX, Windows	<a href="https://github.com/EMMA-CPD/minimal-energy-points">https://github.com/EMMA-CPD/minimal-energy-points</a>		C++
22	ConcorcInterp		An open-source code providing a graphical user interface to predict the effective heat conductivity of nanostructures	Research software	Machine learning, heat conductivity, computational materials science		<a href="https://github.com/EMMA-CPD/concorc-interp">https://github.com/EMMA-CPD/concorc-interp</a>	<a href="https://github.com/EMMA-CPD/concorc-interp">https://github.com/EMMA-CPD/concorc-interp</a>	Oliver Kunc, Felix Fritzer	oliver.kunc@mpie.de, felix.fritzer@mpie.de	Proprietary	Windows, Mac/Linux	<a href="https://github.com/EMMA-CPD/concorc-interp">https://github.com/EMMA-CPD/concorc-interp</a>		python
23	Finite Element Analysis Program	FEAP	A general purpose finite element analysis program which is designed for research in educational use	Research software	Finite Element Method, computational materials science	8.6	<a href="http://ftp.ce.berkeley.edu/~feap/">http://ftp.ce.berkeley.edu/~feap/</a>	<a href="https://github.com/monson/feap">https://github.com/monson/feap</a>	Julian Lüthger, Felix Fritzer	luej@berkeley.edu, ffritzer@berkeley.edu	Proprietary	Windows, Mac/Linux	<a href="http://ftp.ce.berkeley.edu/~feap/">http://ftp.ce.berkeley.edu/~feap/</a>		C++
24	Paraview		Paraview is an open-source, multi-platform data analysis and visualization application based on Visualization	Research software	scientific visualization, data analysis	5.10.1	<a href="https://www.paraview.org/">https://www.paraview.org/</a>	<a href="https://github.com/paraview/paraview">https://github.com/paraview/paraview</a>	Isopcon@kitware.com	isopcon@kitware.com	BSD license	Windows, Mac/Linux	<a href="https://www.paraview.org/">https://www.paraview.org/</a>	VTK	C, C++, Fortran, Python
25	The Visualization Toolkit	VTK	The Visualization Toolkit (VTK) is open source software for image processing and displaying scientific data. It consists	Visualization software	scientific visualization, data analysis	9.2.0	<a href="https://vtk.org/">https://vtk.org/</a>	<a href="https://github.com/paraview/vtk">https://github.com/paraview/vtk</a>	Isopcon@kitware.com	isopcon@kitware.com	BSD license	Windows, Mac/Linux	<a href="https://www.vtk.org/">https://www.vtk.org/</a>	Paraview	python
26	pyXRD		Volume Reconstruction Software for X-ray computed tomography data: reconstruction of 3D or 2D projections	Research software	X-ray imaging, imaging data processing, reconstruction	1.1.5	<a href="https://www.pyxrd.org/">https://www.pyxrd.org/</a>	<a href="https://github.com/pyxrd/pyxrd">https://github.com/pyxrd/pyxrd</a>	Isopcon@kitware.com	isopcon@kitware.com	BSD	Windows, Mac/Linux	<a href="https://www.pyxrd.org/">https://www.pyxrd.org/</a>		python
27	Auto Software		Image data processing, visualization and analysis especially for 3D data sets like X-ray tomography	Research software	imaging data processing, scientific visualization	2019.3	<a href="https://www.hermeslab.com/">https://www.hermeslab.com/</a>	<a href="https://github.com/monson/autosoft">https://github.com/monson/autosoft</a>	Isopcon@kitware.com	isopcon@kitware.com	Proprietary license	Windows, Mac/Linux	<a href="https://www.hermeslab.com/">https://www.hermeslab.com/</a>		python
28	Fi		Fi is an image processing package—a "tomographic" distribution of ImageJ2, handling a lot	Research software	imaging data processing, scientific visualization, data analysis	2.5.0	<a href="https://www.fimg.net/">https://www.fimg.net/</a>	<a href="https://github.com/monson/fi">https://github.com/monson/fi</a>	Isopcon@kitware.com	isopcon@kitware.com	Proprietary license	Windows, Mac/Linux	<a href="https://www.fimg.net/">https://www.fimg.net/</a>		python
29	ImageJ		ImageJ is a public domain software for processing and analyzing scientific images	Research software	imaging data processing, scientific visualization, data analysis	1.53h	<a href="https://imagej.net/imagej2/multiplatform">https://imagej.net/imagej2/multiplatform</a>	<a href="https://github.com/imagej/imagej2">https://github.com/imagej/imagej2</a>	Isopcon@kitware.com	isopcon@kitware.com	public domain	Windows, Mac/Linux	<a href="https://imagej.net/imagej2/multiplatform">https://imagej.net/imagej2/multiplatform</a>	Fi, ImageJ2	Java
30	ImageJ2		ImageJ2 is a new version of ImageJ for multidimensional image data, with a focus on scientific imaging. It includes	Research software	imaging data processing, scientific visualization, data analysis	2.5.0	<a href="https://imagej.net/imagej2/multiplatform">https://imagej.net/imagej2/multiplatform</a>	<a href="https://github.com/imagej/imagej2">https://github.com/imagej/imagej2</a>	Isopcon@kitware.com	isopcon@kitware.com	BSD-2	Windows, Mac/Linux	<a href="https://imagej.net/imagej2/multiplatform">https://imagej.net/imagej2/multiplatform</a>	Fi, ImageJ2	Java, Perl, Shell
31	Veritas into Simulation Package	VASP	The Veritas into Simulation Package (VASP) is a computer program for ab initio, quantum mechanical modeling	Research software	DFT periodic code, plane wave and related methods, computational materials science	6.X	<a href="http://www.vasp.at/">http://www.vasp.at/</a>	<a href="https://github.com/monson/vasp">https://github.com/monson/vasp</a>	support@vasp.at	support@vasp.at	Proprietary academic license	Linux, OSX, Windows	<a href="http://www.vasp.at/">http://www.vasp.at/</a>		Fortran
32	CASTEP		CASTEP is a first-principles materials modeling code based on a plane wave pseudopotential method	Research software	DFT periodic code, plane wave and related methods, computational materials science	21.1.1	<a href="https://www.castep.org/">https://www.castep.org/</a>	<a href="https://github.com/monson/castep">https://github.com/monson/castep</a>	Isopcon@kitware.com	isopcon@kitware.com	Proprietary academic license	Linux, OSX, Windows	<a href="https://www.castep.org/">https://www.castep.org/</a>		Fortran
33	Car	Car	The Car code is a parallelized plane wave / pseudopotential implementation of Density Functional Theory (DFT) for ab initio calculations	Research software	DFT periodic code, plane wave and related methods, computational materials science	9.6	<a href="http://www.cmtl.org/">http://www.cmtl.org/</a>	<a href="https://github.com/monson/car">https://github.com/monson/car</a>	info@abinit.org	info@abinit.org	GNU General Public License	Linux	<a href="http://www.cmtl.org/">http://www.cmtl.org/</a>		Fortran
34	ABINIT		ABINIT is software suite to calculate the optical, mechanical, vibrational, and other observable properties	Research software	DFT periodic code, plane wave and related methods, computational materials science	9.6	<a href="http://www.abinit.org/">http://www.abinit.org/</a>	<a href="https://github.com/monson/abinit">https://github.com/monson/abinit</a>	info@abinit.org	info@abinit.org	GNU General Public License	Linux, OSX, Windows	<a href="http://www.abinit.org/">http://www.abinit.org/</a>		Fortran, python
35	BigFT		A fast, parallel, multi-processor FFT code for ab-initio calculations	Research software	DFT periodic code, plane wave and related methods, computational materials science	1.8.3	<a href="https://github.com/monson/bigft">https://github.com/monson/bigft</a>	<a href="https://github.com/monson/bigft">https://github.com/monson/bigft</a>	info@abinit.org	info@abinit.org	GNU General Public License	Linux, OSX, Windows	<a href="https://github.com/monson/bigft">https://github.com/monson/bigft</a>		Fortran
36	Quantum Espresso		Quantum ESPRESSO is the reference code for ab-initio calculations on periodic systems	Research software	DFT periodic code, plane wave and related methods, computational materials science	7.1	<a href="http://www.quantum-espresso.org/">http://www.quantum-espresso.org/</a>	<a href="https://github.com/monson/quantum-espresso">https://github.com/monson/quantum-espresso</a>	info@quantum-espresso.org	info@quantum-espresso.org	GNU General Public License	Linux, MacOS	<a href="http://www.quantum-espresso.org/">http://www.quantum-espresso.org/</a>		Fortran, python
37	Parallel total energy	Flex	Open-Source computer codes for electronic-structure calculations using plane waves and pseudopotentials	Research software	DFT periodic code, plane wave and related methods, computational materials science	3	<a href="http://www.flexcode.org/">http://www.flexcode.org/</a>	<a href="https://github.com/monson/flex">https://github.com/monson/flex</a>	info@flexcode.org	info@flexcode.org	GNU General Public License	Linux, MacOS	<a href="http://www.flexcode.org/">http://www.flexcode.org/</a>		Fortran
38	JDFTx		JDFTx is a parallel plane wave pseudopotential code designed to be as easy to develop with as it is easy to use	Research software	DFT periodic code, plane wave and related methods, computational materials science	1.7.0	<a href="http://www.jdftx.org/">http://www.jdftx.org/</a>	<a href="https://github.com/monson/jdftx">https://github.com/monson/jdftx</a>	info@jdftx.org	info@jdftx.org	GNU General Public License	Linux, MacOS	<a href="http://www.jdftx.org/">http://www.jdftx.org/</a>		Fortran

## LodView, in conjunction with a SPARQL endpoint

[SPARQL endpoint](#)

[The MatWerk ontology \(mwo\)](#)

[The NFDI Core Ontology \(nfdicore\)](#)

[Data collection spreadsheets](#)

[NFDI-MatWerk LOD Working Group](#)

[About NFDI-MatWerk consortium](#)

[Contact Us](#)

# MSE Knowledge Graph v1.0

## Purpose

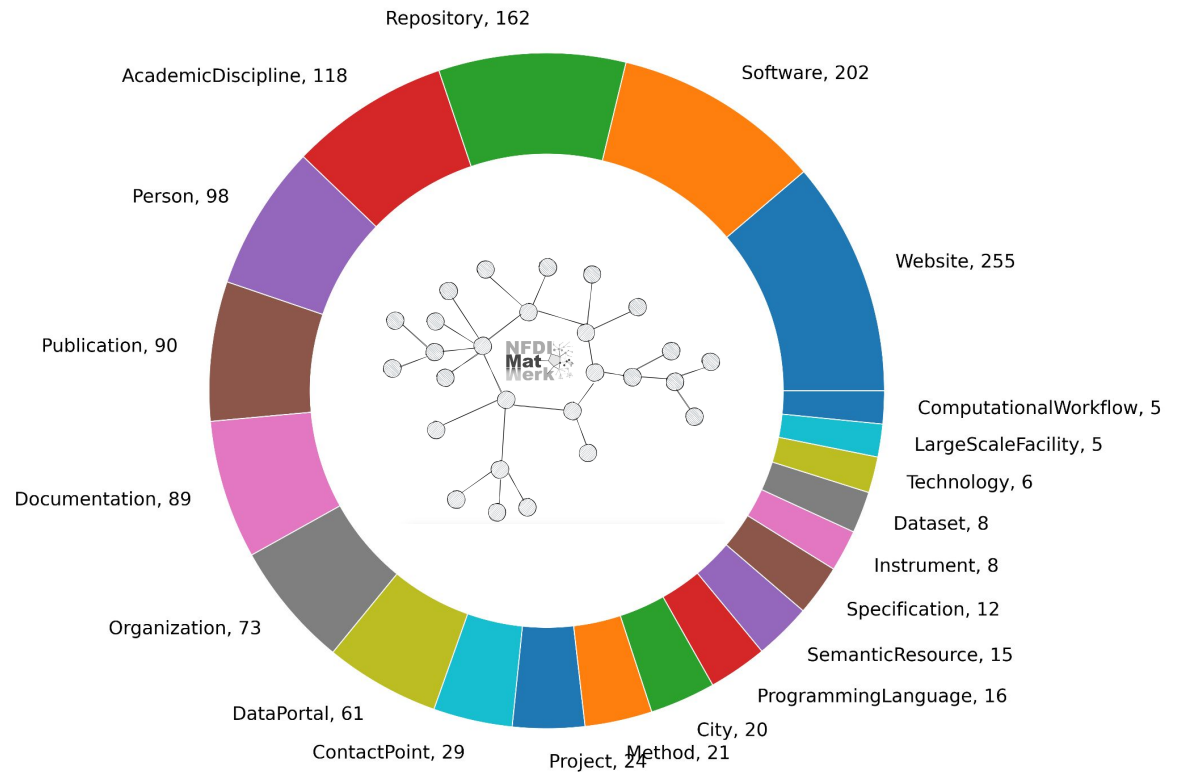
The purpose of creating the MSE KG v1.0 is to represent the consortium resources to enable integrative access to distributed heterogeneous research data within the institutions participating in NFDI-MatWerk, the MSE community, as well as across consortiums within the NFDI network and beyond. It will be the backend data resource for the NFDI MatWerk portal. The knowledge graph will be continuously updated based on new information from the consortium.



<https://demo.fiz-karlsruhe.de/matwerk>

- Number of triples: 8203
- Number of entities: 1893

## Number of entities for each Classes



# The MSE KG v1.0 - An Instance of a Person

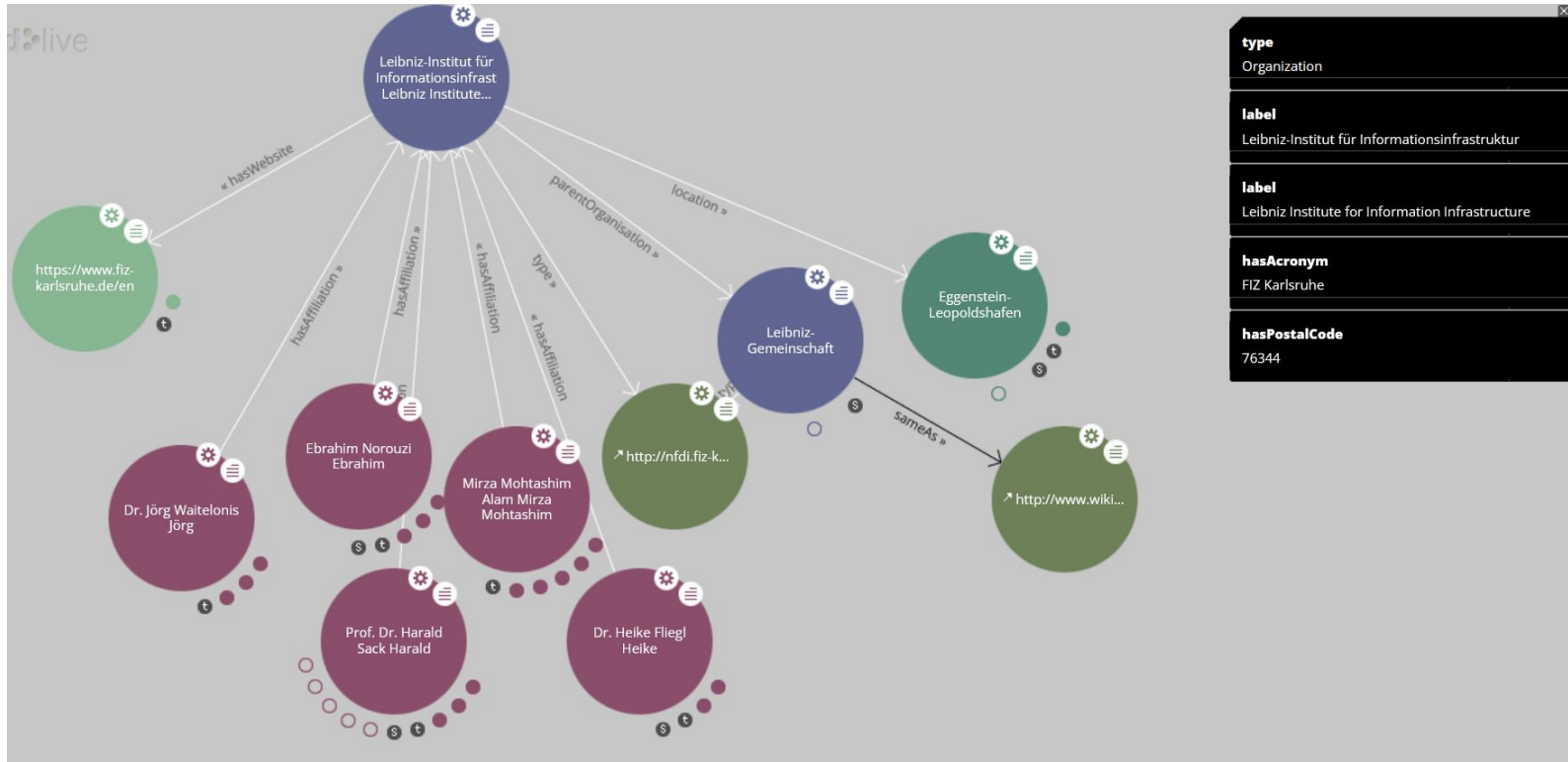


<b>type</b>	Person
<b>label</b>	Prof. Dr. Harald Sack
<b>surname</b>	Sack
<b>emailAddress</b>	harald.sack@fiz-karlsruhe.de
<b>firstName</b>	Harald
<b>title</b>	Prof. Dr.

<https://demo.fiz-karlsruhe.de/matwerk/E14862>



# The MSE KG v1.0 - An Instance of an Organization



<b>type</b>	Organization
<b>label</b>	Leibniz-Institut für Informationsinfrastruktur
<b>label</b>	Leibniz Institute for Information Infrastructure
<b>hasAcronym</b>	FIZ Karlsruhe
<b>hasPostalCode</b>	76344

<https://demo.fiz-karlsruhe.de/matwerk/E1016>

# The MSE KG v1.0 - An Instance of a Dataset



<https://demo.fiz-karlsruhe.de/matwerk/E1173747>

## LodView, in conjunction with a SPARQL endpoint

Here are the list of competency questions with the corresponding SPARQL query:

- Who is working with Researcher "Ebrahim Norouzi" in the same group? Return the ORCID IDs? [SPARQL](#)
- Who are the contributors of the data "BAM reference data"? List the contributors with their affiliations and its license, standard, repository, and which material it is about? [SPARQL](#)
- What is the email address of the contactpoint of "NOMAD" DataPortal? [SPARQL](#)
- What are "Molecular Dynamics" Software? List the programming language, documentation page, repository, and license information? [SPARQL](#)
- What are the ontologies in nanomaterials domain? [SPARQL](#)
- What are the software used to produce the data in the Materials Cloud repository? [SPARQL](#)
- What is the field of research and type of the "BAM reference data"? [SPARQL](#)
- What are the organizations in the KG that are categorized as a Public University in Wikidata? [SPARQL](#)
- Give me the contact point of Elemental Multiperspective Material Ontology (EMMO) and the related projects. [SPARQL](#)
- List all ontologies with the Creative Commons Attribution 4.0 license. [SPARQL](#)
- List people who have expertise in Information Service Engineering and the lecture they give. [SPARQL](#)
- List software written in Python with a GNU General Public License. [SPARQL](#)

Browse MatWerk Knowledge graph resources interactively via LodLive

Sample resources

- **Person:** Prof. Dr. Harald Sack (resource [E14862](#))
- **Organization:** Leibniz-Institut für Informationsinfrastruktur (resource [E1016](#))
- **Resources:** Dislocation Ontology (DISO) (resource [E1129754](#))
- **Dataset:** BAM reference data (resource [E1173747](#))

<https://demo.fiz-karlsruhe.de/matwerk>

[Who is working with Researcher "Ebrahim Norouzi" in the same group? Return the ORCID IDs.](#)

SPARQL Query Editor
About Tables ▾
Conductor
Permalink

```

SELECT ?person ?personlabel ?orcidid
WHERE {

  # Find the researcher with name "Ebrahim Norouzi"
  ?researcher rdfs:label ?researcherlabel FILTER regex(?researcherlabel , "brahim norouzi", "i")

  # Get the affiliation of the researcher
  ?researcher mwo:hasAffiliation ?Affiliation .

  # Find persons with the same affiliation but different from the researcher
  ?person mwo:hasAffiliation ?Affiliation FILTER (?person != ?researcher) .

```

Results Format HTML

Execute Query
Reset

<https://demo.fiz-karlsruhe.de/sparql>

[Who is working with Researcher "Ebrahim Norouzi" in the same group? Return the ORCID IDs.](#)

SPARQL | HTML5 table

person	personlabel	orcidid
<a href="http://demo.fiz-karlsruhe.de/matwerk/E9779">http://demo.fiz-karlsruhe.de/matwerk/E9779</a>	"Dr. Heike Fliegl"	<a href="https://orcid.org/0000-0002-7541-115X">https://orcid.org/0000-0002-7541-115X</a>
<a href="http://demo.fiz-karlsruhe.de/matwerk/E1245566">http://demo.fiz-karlsruhe.de/matwerk/E1245566</a>	" Mirza Mohtashim Alam"	<a href="https://orcid.org/0000-0002-7047-0791">https://orcid.org/0000-0002-7047-0791</a>
<a href="http://demo.fiz-karlsruhe.de/matwerk/E14862">http://demo.fiz-karlsruhe.de/matwerk/E14862</a>	"Prof. Dr. Harald Sack"	<a href="https://orcid.org/0000-0001-7069-9804">https://orcid.org/0000-0001-7069-9804</a>
<a href="http://demo.fiz-karlsruhe.de/matwerk/E16052">http://demo.fiz-karlsruhe.de/matwerk/E16052</a>	"Dr. Jörg Waitelonis"	<a href="https://orcid.org/0000-0001-7192-7143">https://orcid.org/0000-0001-7192-7143</a>

<https://demo.fiz-karlsruhe.de/sparql>

## LodView, in conjunction with a SPARQL endpoint

[SPARQL endpoint](#)

[The MatWerk ontology \(mwo\)](#)

[The NFDI Core Ontology \(nfdicore\)](#)

[Data collection spreadsheets](#)

[NFDI-MatWerk LOD Working Group](#)

[About NFDI-MatWerk consortium](#)

[Contact Us](#)

### Infrastructure use cases (IUCs)-related questions with the corresponding SPARQL queries:

- **(IUC02)** What are datasets produced by the BAM organization? List the title, standard, license, hosting repository and which material it [SPARQL](#) is about.
- **(IUC04)** What are the resources related to the SFB1394 Project? [SPARQL](#)
- **(IUC09)** What are Computational Workflows associated with the Atom Probe Tomography method? List the funding project(s), [SPARQL](#) license and the repository URI.
- **(IUC17)** What are ontologies which describe "crystalline defects"? [SPARQL](#) List the repositories and related project.
- **(Indentation)** What are workflows related to keywords: Aluminium and Elastic Constants? List the type, URL, funding project and [SPARQL](#) authors.

<https://demo.fiz-karlsruhe.de/matwerk>

- NFDI Ontologies Architecture
  - NFDI Core Ontology (NFDI Core)
  - MatWerk Ontology - MWO
- MSE Knowledge Graph v1.0
- **MSE Knowledge Graph v2.0**
  - NFDI MatWerk LOD Working Group

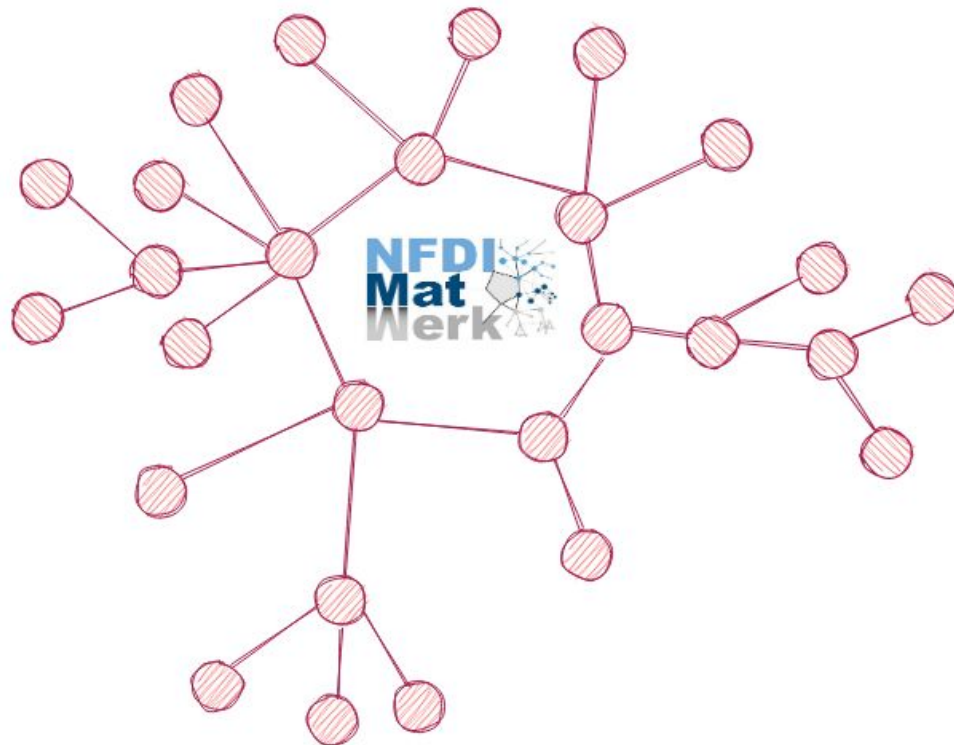


## CONTENT:

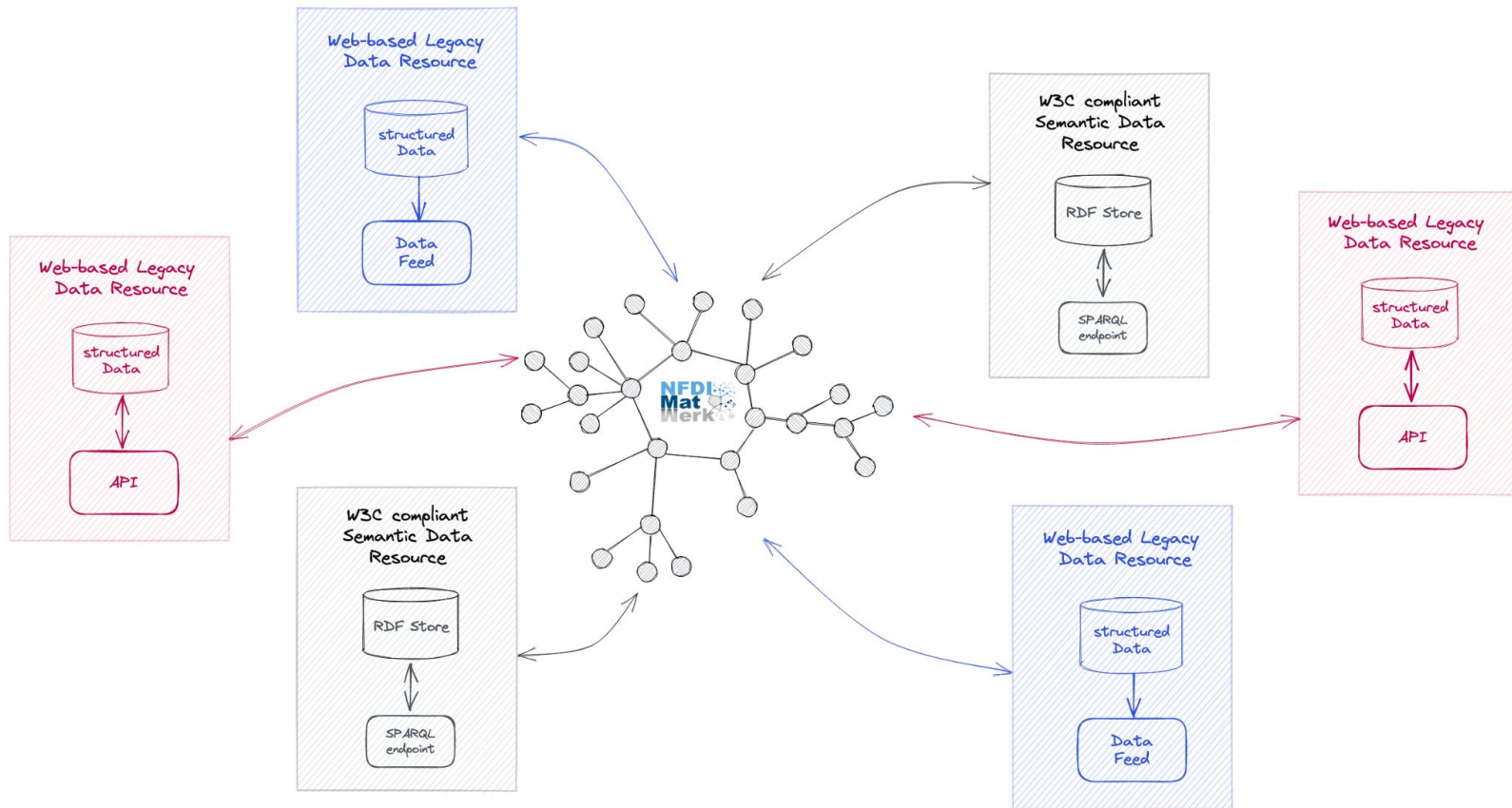
- Index and Metadata for **MatWerk** Project Resources (v1.0)
- **plus Metadata for MatWerk content** as e.g., from
  - (1) Fully fledged Semantic Data resources
  - (2) Structured Data resources (Legacy Data resources)
  - (3) Wikibase-based resources

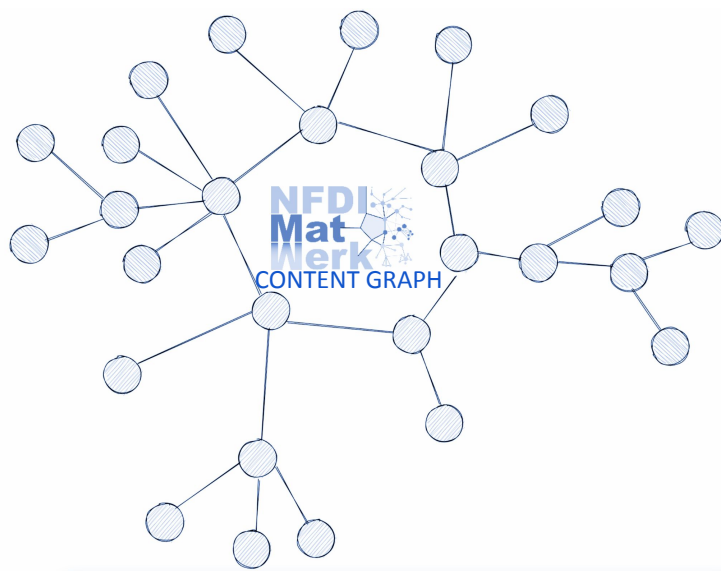
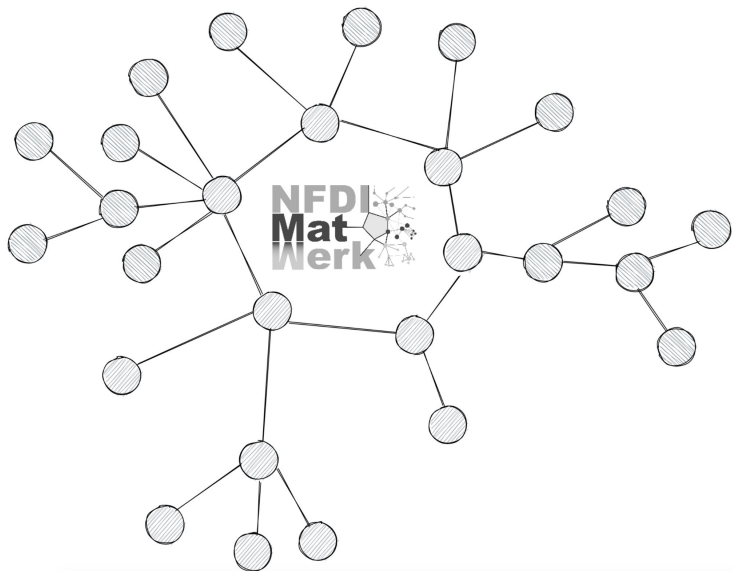
## PURPOSE:

- Metadata for retrieval, exploration, and analysis of **MatWerk** research data resources



# MSE KG and MatWerk Data Resources





MSE KG v1.0

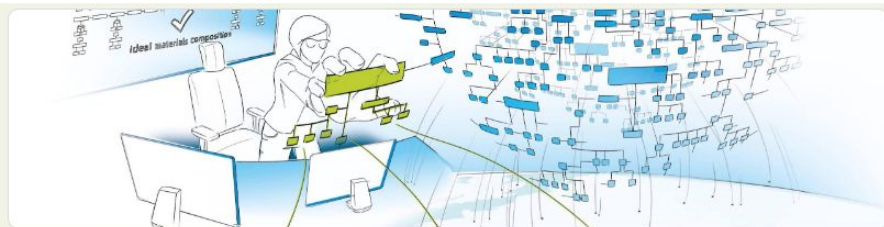
+

Content Graph

=

MSE KG v2.0

- NFDI Ontologies Architecture
  - NFDI Core Ontology (NFDI Core)
  - MatWerk Ontology - MWO
- MSE Knowledge Graph v1.0
- MSE Knowledge Graph v2.0
  - **NFDI MatWerk LOD Working Group**



## NFDI-MatWerk LOD Working Group

Linked Open Data defines a vision of globally accessible and linked data on the internet based on the [RDF](#) standards of the [semantic web](#). Tim Berners-Lee presented a set of four basic rules for publishing, connecting and consuming the structured data on the Web ([LOD rules](#)).

The goal of this working group is to address the problems that participant projects face when working with or towards LOD. The information collected in LOD working group will be integrated into the knowledge graph. This is a small survey to collect people interested to join the NFDI-MatWerk LOD Working Group as well as potential topics to discuss.

Useful links:

[Linked Data Engineering course](#)

[Linked Data Platform Best Practices and Guidelines](#)

[The Linked Open Data Cloud](#)

[awesome-semantic-web](#)

Which topics you would like to see covered in the Working Group? \*

- Brief Seminars on the Basics of LOD and related technologies
- Presentations of NFDI-MatWerk projects related to LOD
- Presentations of new ideas, techniques, methodologies to be applied in the NFDI-MatWerk LOD context
- HandsOn tutorials on NFDI-MatWerk LOD related technologies
- Hackathons on NFDI-MatWerk LOD related technologies
- Presentations from other NFDI consortia on NFDI LOD related solutions

Please suggest further potential topics

Your answer

What is the LOD status of your data?

- 1 star: data is available on the web, in whatever format, but with an open licence, so as to be Open Data.
- 2 star: data is open + machine-readable structured, e.g. excel spreadsheet
- 3 star: data is open, machine-readable and does not require proprietary software package in order to analyze it, e.g. CSV format
- 4 star: data uses open standards from W3C, e.g. RDF and SPARQL and URIs.
- 5 star: data is open, uses open standards from W3C and is linked to other data on the web.
- Not applicable (if it is e.g. a service or a tool)

The goal of this working group is to address the problems that participant projects face when working with or towards LOD.

The information collected in LOD working group will be integrated into the knowledge graph.

**Interested to join the  
LOD Working group?**



### Motivation

- Complexity in MSE: Quick advancements make data management tools essential.
- Role of Ontologies in MSE: Structure domain-specific knowledge.
- Competency Questions (CQs): Validate if ontology meets its intended requirements.
- Potential of LLMs: Simplify CQ answers from Knowledge Graphs without SPARQL expertise.
- LLM Evaluation: Crucial for avoiding errors and verifying efficacy against SPARQL.

### Goals and Objective

- Streamlined Pipeline: Auto-answer MatWerk KG-specific CQs.
- LLM Evaluation: Assess LLM performance metrics.

### System Demo

Natural Language to Answer from KG

Within the interface, users input natural language competency questions, and the system generates responses from the MatWerk KG using LLMs. To use the demo please scan

### Methodology

The methodology flowchart illustrates the process: Knowledge Graph and Process in Checks/Books are used to generate a Trained Embedding from LLM. This embedding is processed through a Compression Process to create a Reference Vector (e.g., [0.0163 3.2424 0.2128 0.0172 0.2240]). This vector is compared against a Vector Storage (e.g., [0.0164 3.9840 0.2128]). The process then involves a Prompt Template (Instruction, Content, Query) and a Competency Question to generate an Answer from an LLM. The answer is evaluated against Ground Truths using an Evaluation Process (Assess generated from LLM vs Ground Truths of CQs) to produce an Answer and various Evaluation Metrics (Precision, Recall, F1, Accuracy).

### Classification of Competency Questions

- Direct answers**  
Question: What is the license of the dataset "Elastic Constant Demo Data"?  
Answer: BSD 3 Clause license  
Ground Truth: BSD 3 Clause license
- Descriptive answers**  
Question: What are the different data formats in the "BAM reference data"?  
Answer: csv; pdf  
Ground Truth: csv; pdf
- Direct answers but can have different representations by LLM answers**  
Question: What is the affiliation of Volker Hofmann?  
Answer: http://demo.fiz-karlsruhe.de/matwerk/E13022  
Ground Truth: Forschungszentrum Jülich, Forschungszentrum Jülich@de
- The answer contains multiple different items, Matching with the ground truth is challenging for LLM**  
Question: What is "Molecular Dynamics" Software? List the programming language, documentation page, repository, and license information  
Answer: [Detailed list of software information]  
Ground Truth: [Detailed list of software information]
- Descriptive answers**  
Question: What is the description of the "BAM reference Data"?  
Answer: The description of the "BAM reference data" is "The creep behavior of a certified reference material was determined according to ASTM E139-11 using calibrated equipment of an accredited testing laboratory."  
Ground Truth: BAM reference data is "The creep behavior of a certified reference material was determined according to ASTM E139-11 using calibrated equipment of an accredited testing laboratory."

### Results

- Evaluation Metrics:** Employed Rouge, precision, recall, accuracy, F1 scores.
- Answer Complexity:** Metrics adapt for LLM-generated answer types and diverse CQs.
- Simple vs Descriptive:** Standard metrics for direct, ROGUE for complex answers.
- Initial Evaluation:** Tested on 37 MatWerk KG CQs.

**Categorized the existing competency questions**

- Direct answers (28)
- Direct answers but can have different representations by LLM answers (1)
- The answer contains multiple different items, Matching with the ground truth is challenging for LLM (2)
- Descriptive answers (2)

**Evaluation on direct answers**

Metric	Average Value
$F_1$	0.88
Precision	0.83
Recall	0.84
Accuracy	0.83

**Evaluation on all CQs (verified vs non-verified)**

Metric	Report 1 (Non-Verified)	Report 2 (Verified)	Report 3 (Non-Verified)	Report 4 (Verified)	Report 5 (Non-Verified)	Report 6 (Verified)
$F_1$	0.83	0.86	0.83	0.83	0.87	0.85
Precision	0.82	0.80	0.80	0.80	0.87	0.82
Recall	0.89	0.90	0.87	0.84	0.82	0.87

### Findings

- Usability LLM:** effective with limitations.
- Question Creation:** Simplified CQ generation/validation, boosted by structured content.
- Performance:** Accelerated & Insights: Enhances data-driven understanding.
- Malfunctionals:** LLM risks, need reduction.

### Takeaways

- Expert & Newcomer:** LLM use in KG answer retrieval is challenging.
- Question Creation:** Simplified CQ generation/validation, boosted by structured content.
- Performance:** Accelerated & Insights: Enhances data-driven understanding.

### Challenges

- Automation Complexity:** LLM use in KG answer retrieval is challenging.
- Evaluation Hurdles:** Metrics tricky due to LLM & CQ nuances.
- Future Research:** Target better LLM-CQ accuracy.

# Utilising Large Language Models for Ontology Evaluation in the Field of Materials Science Engineering



- MSE KG v1.0
- How to contribute to the MSE KG v1.0 (Project **Resources**)?
- How to contribute to the MSE KG v2.0, Metadata for MatWerk **content** (MatWerk Data Resources)?
- Query the MSE KG using SPARQL endpoint and LLM in **natural text**.

<https://open.hpi.de/courses/knowledgegraphs2023>

## Knowledge Graphs - Foundations and Applications

11. October 2023 - 21. November 2023



OPEN HPI

Enroll now!



Thank you  
very much  
for your  
attention!