MSE Knowledge Graph

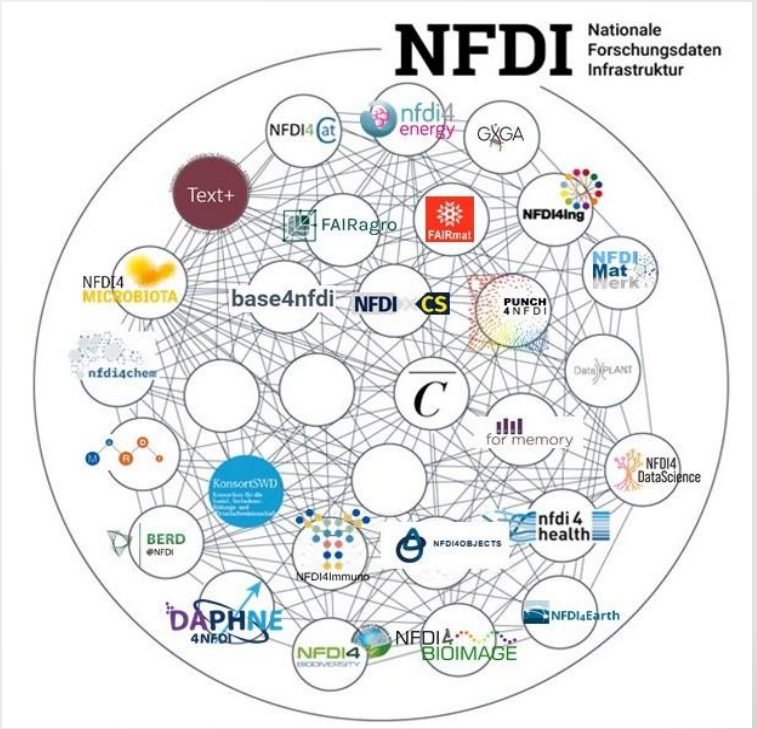
Ebrahim Norouzi, Jörg Waitelonis, Heike Fliegl, Abril Azócar Guzmán, Said Fathalla, Ahmad Zainul Ihsan, Volker Hofmann, Stefan Sandfeld, Felix Fritzen, Amir Laadhar, Harald Sack

Patents4Science Workshop, 05th Oct 2023

NFDI

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



Starting Point: Information Needs

(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**

[3]

TA-OMS

Team:

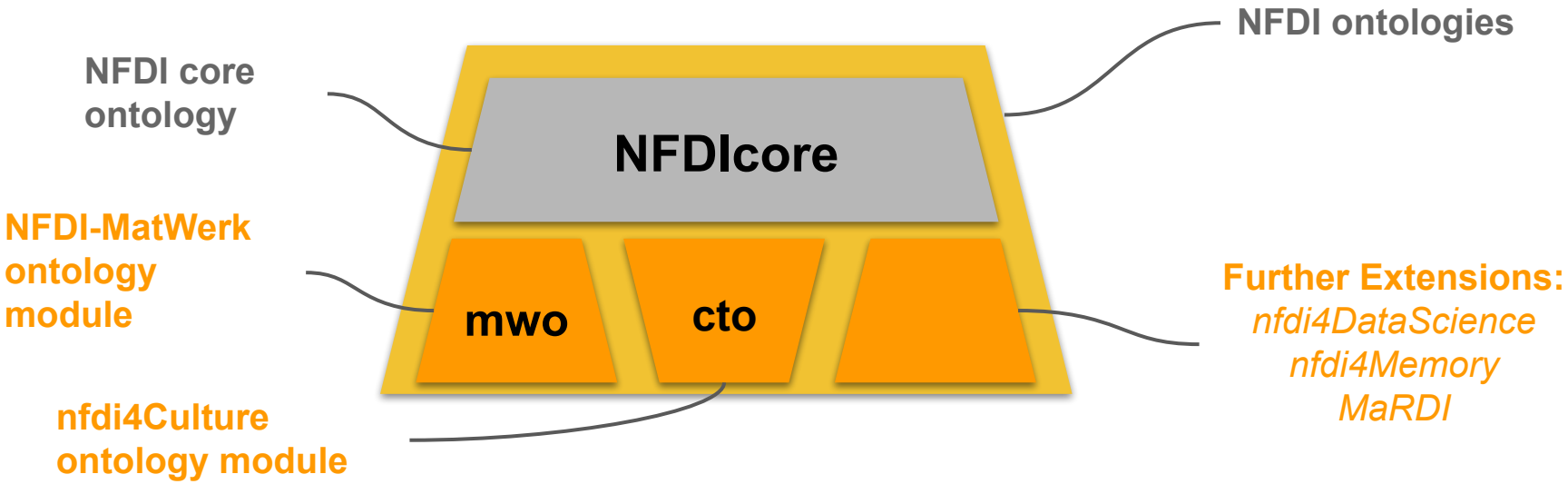
- Prof. Harald Sack
- Prof. Stefan Sandfeld
- Prof. Felix Fritzen
- Dr. Volker Hofmann
- Dr. Abril Azocar-Guzman
- Dr. Said Fathalla
- Dr. Jörg Waitelonis
- Dr. Heike Fliegl
- Dr. Amir Laadhar
- Dr. Angelika Gedsun
- Ebrahim Norouzi
- Ahmad Ihsan
- Mirza-Mohtashim Alam

Agenda @ Patents4Science Workshop

- **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 Architecture

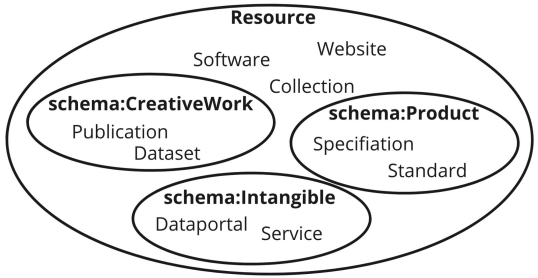
- **NFDI ontologies** follow **modular approach**
- **NFDIcore v1.1.0** represents **NFDI consortia in general**
- **NFDI modules** represent **specific consortia**, as e.g. NFDI-MatWerk
- Specific requirements of consortia lead to **ontology extension**

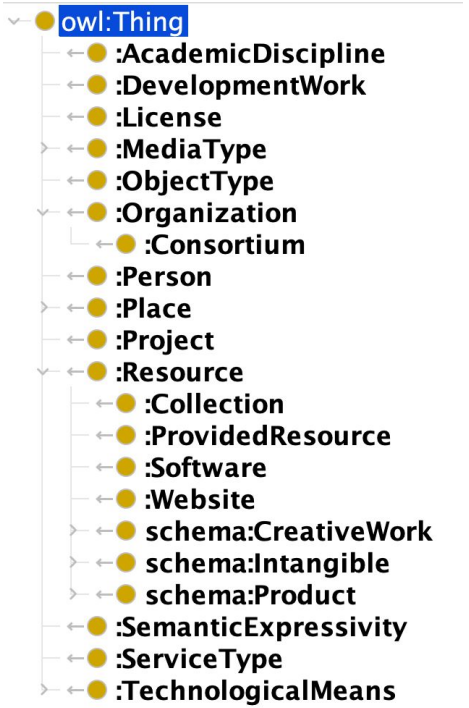


NFDI Core Ontology (NFDI Core)

- 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

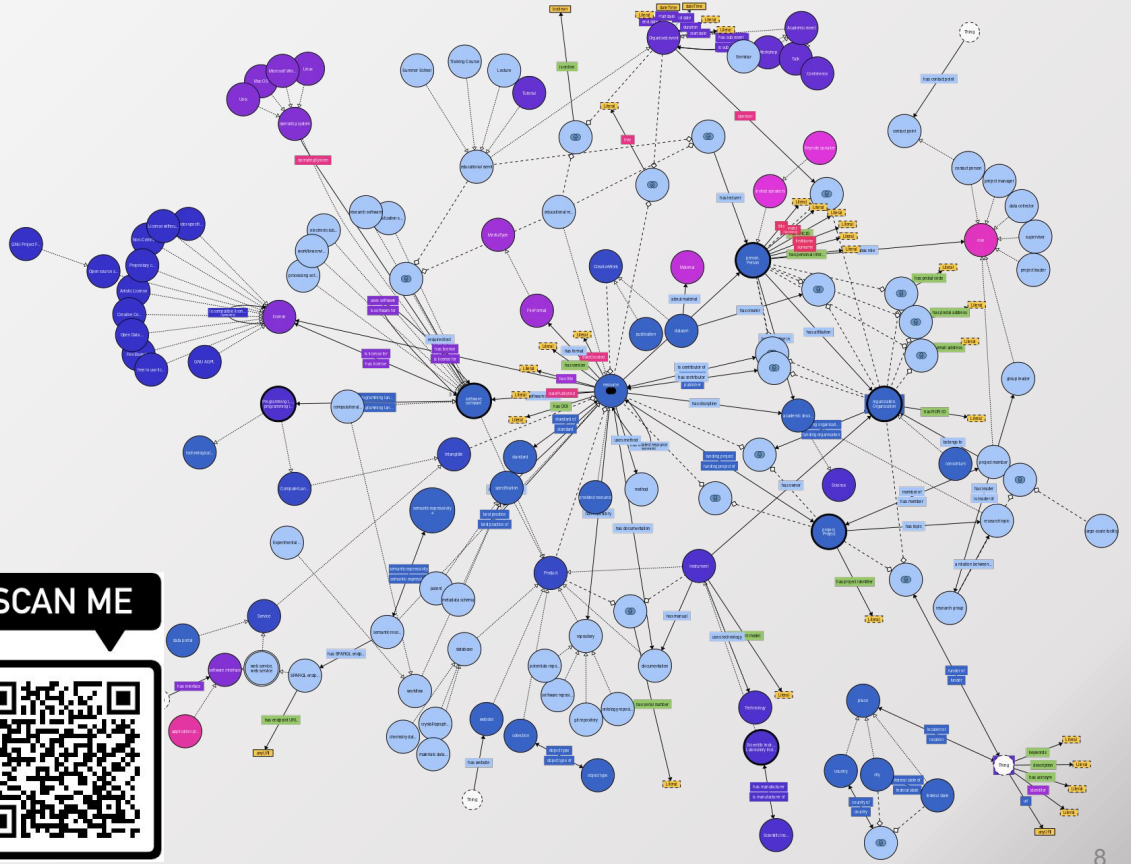
Temporary namespace: https://nfdi.fiz-karlsruhe.de/ontology





MatWerk Ontology (MWO)

- **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.



SCAN ME

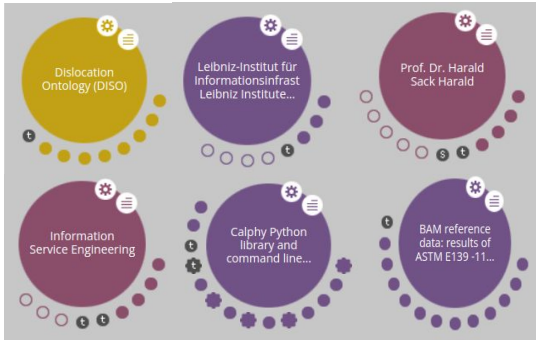
Agenda @ Patents4Science Workshop

- 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

The MSE KG v1.0

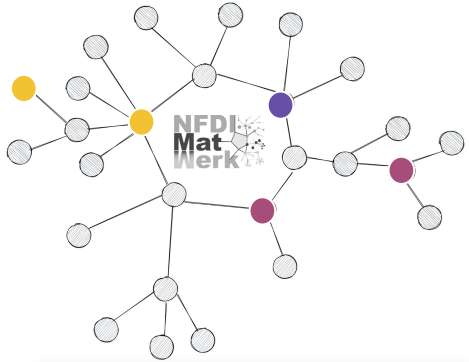
CONTENT:

- 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:

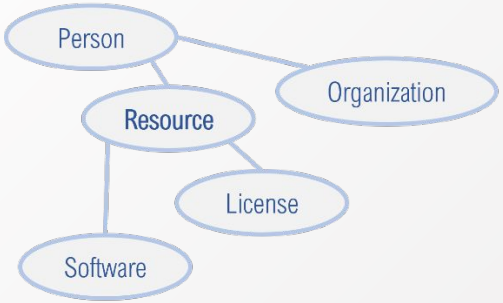
- Backend (Meta-)Data Resource for the **MatWerk** Portal



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

The MSE KG v1.0 - Workflow

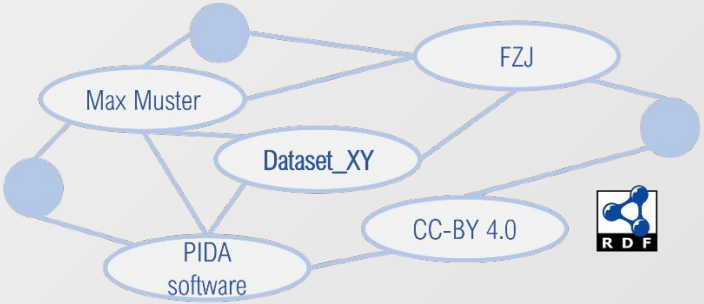
MatWerk Ontology



Collected Data

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

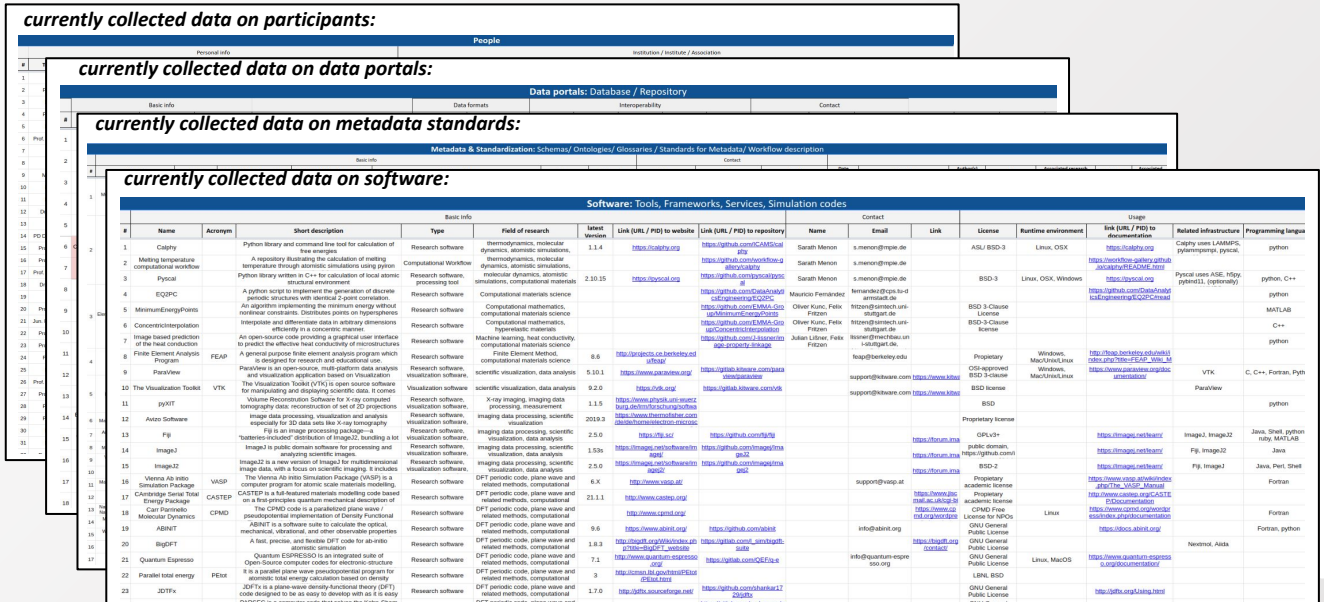
MSE KG



https://git.rwth-aachen.de/nfdi-matwerk/ta-oms/kg_demonstrator

The MSE KG v1.0 - Data Collection

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 381
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

The MSE KG v1.0

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

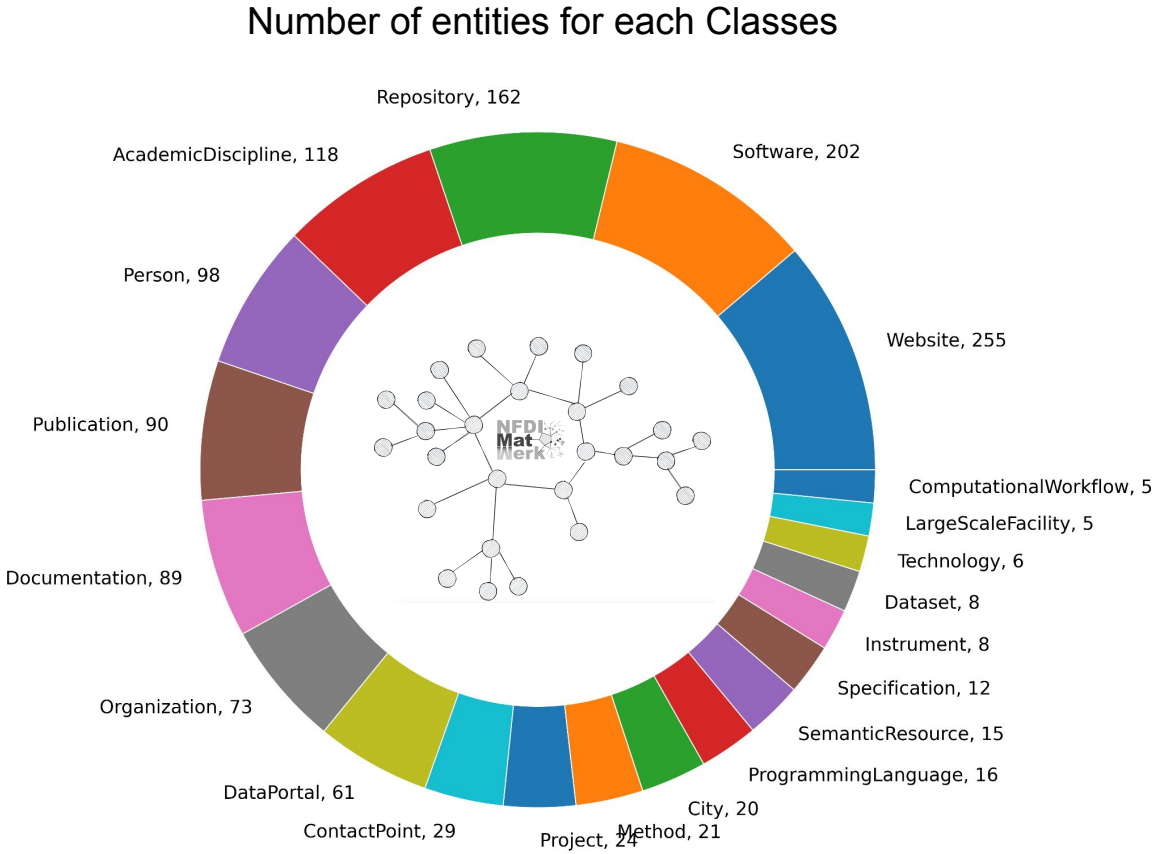
Purpose

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

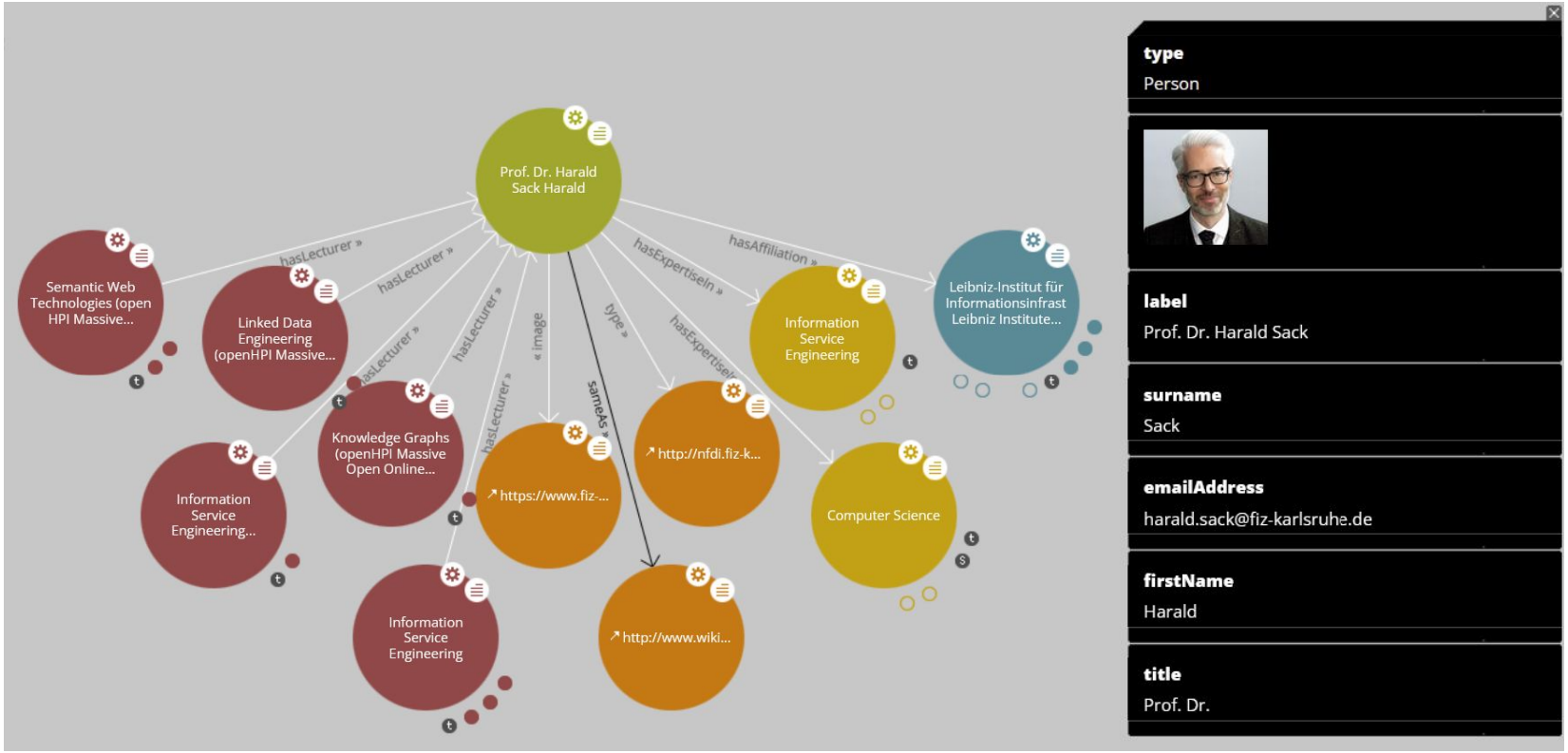
The MSE KG v1.0 - Statistics

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



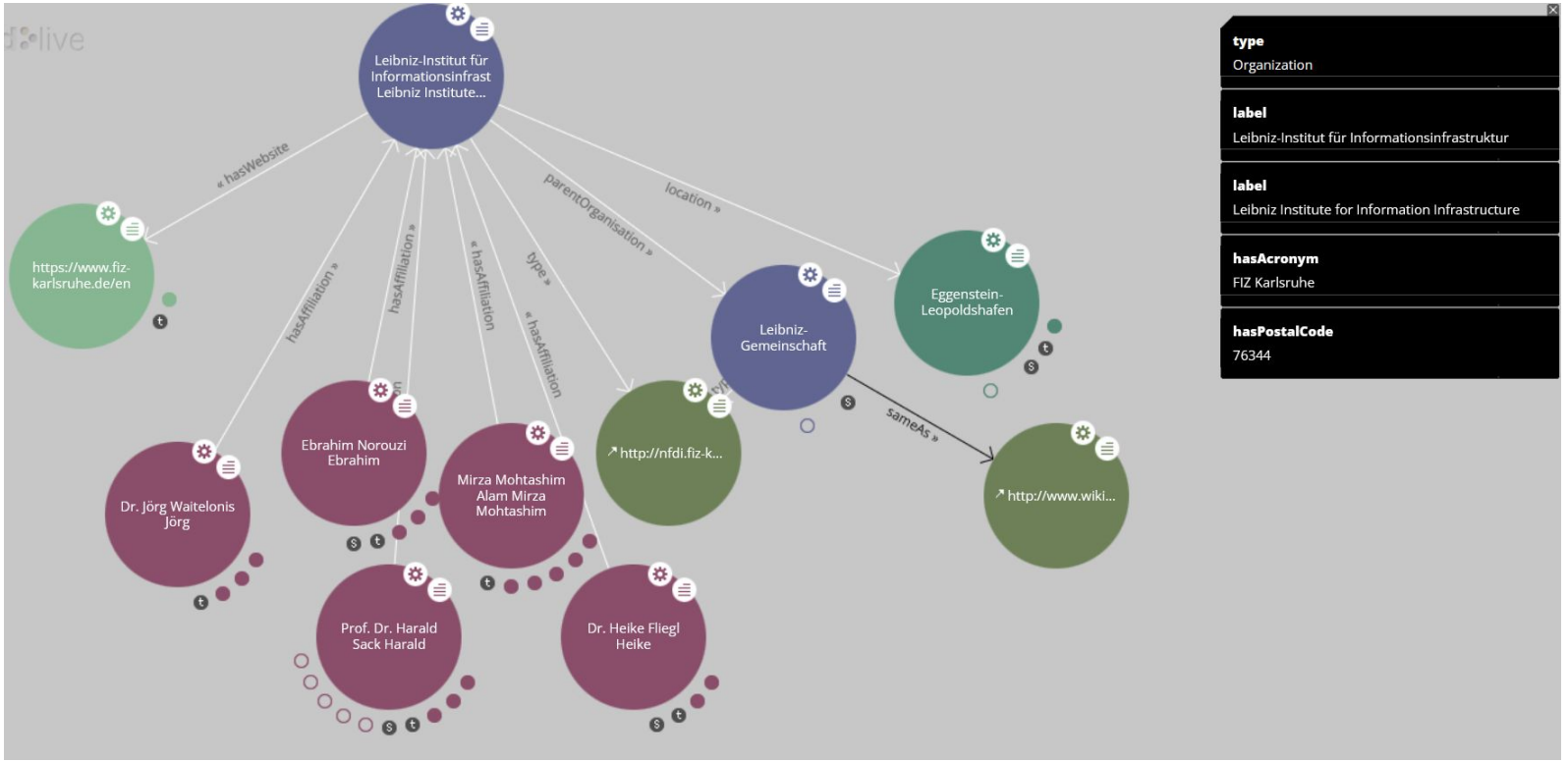
https://demo.fiz-karlsruhe.de/matwerk

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



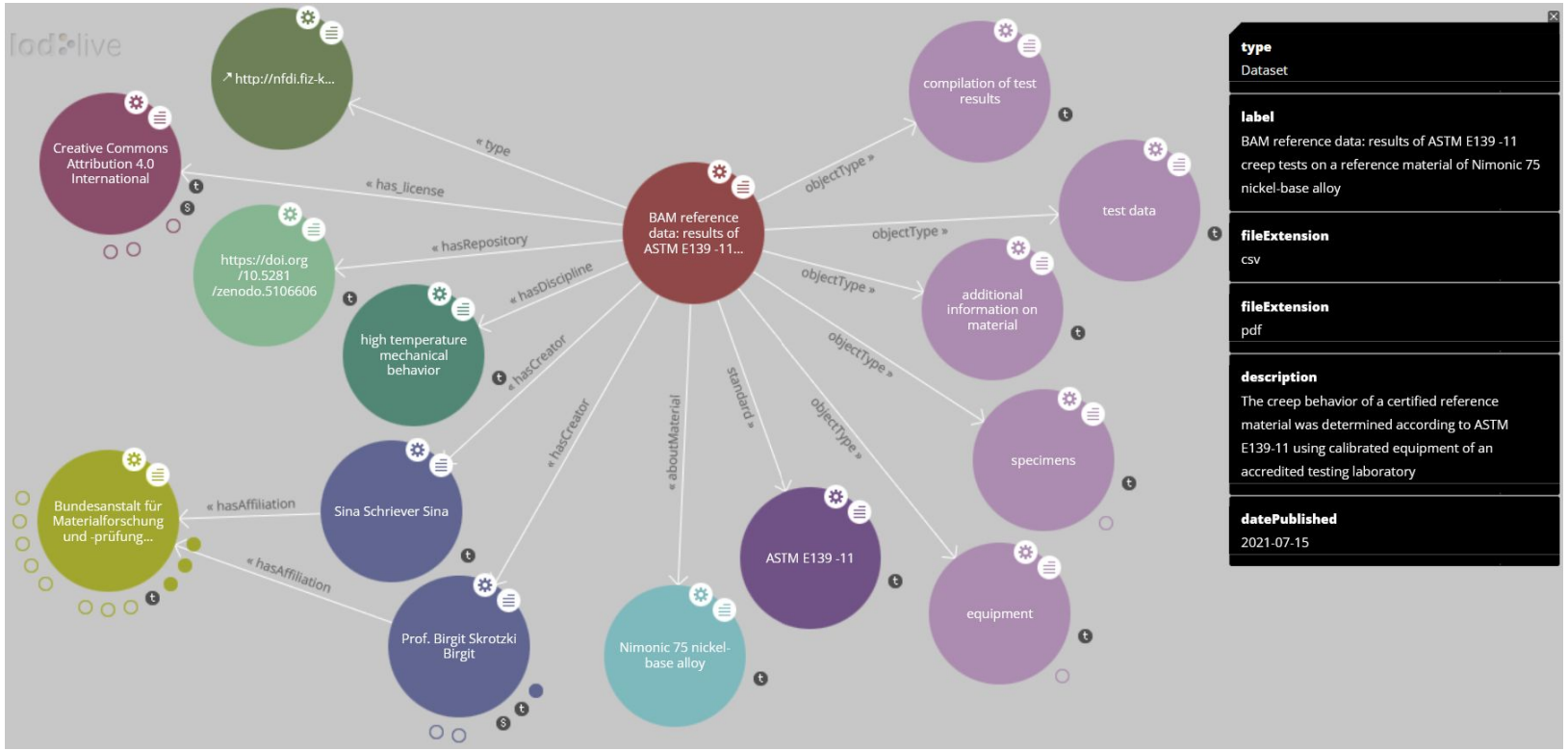
https://demo.fiz-karlsruhe.de/matwerk/E14862

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



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

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



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

The MSE KG v1.0

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

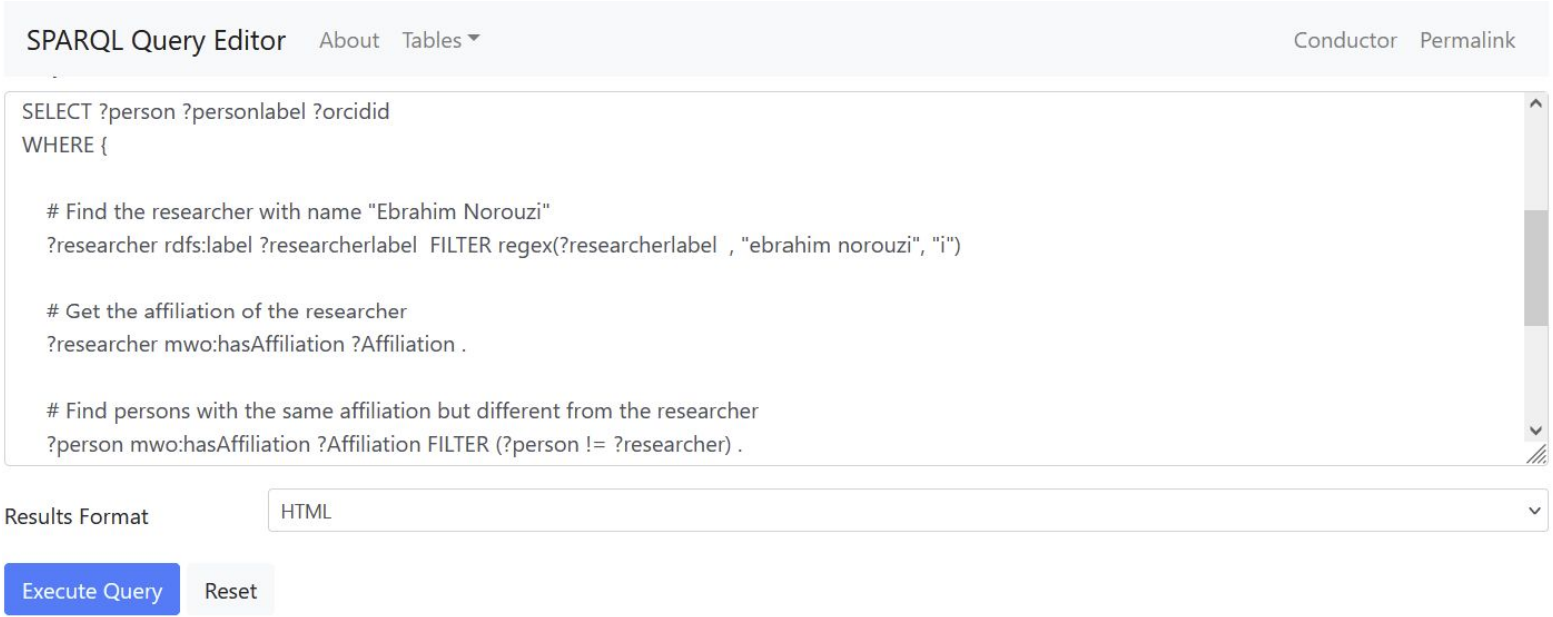
Sample resources

- **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

The MSE KG v1.0 - SPARQL query examples

Who is working with Researcher "Ebrahim Norouzi" in the same group? Return the ORCID IDs.



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

The MSE KG v1.0 - SPARQL query examples

Who is working with Researcher "Ebrahim Norouzi" in the same group? Return the ORCID IDs.

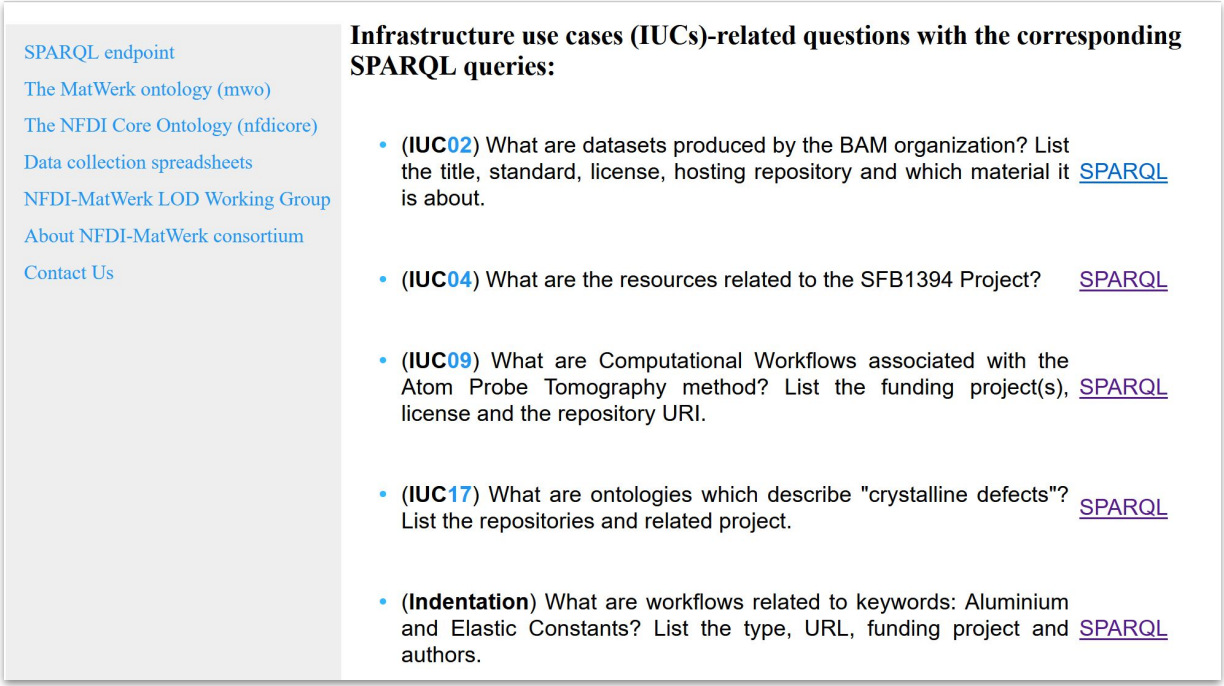
SPARQL | HTML5 table

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

https://demo.fiz-karlsruhe.de/spargl

The MSE KG v1.0

LodView, in conjunction with a **SPARQL endpoint**



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

Agenda @ Patents4Science Workshop

- 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

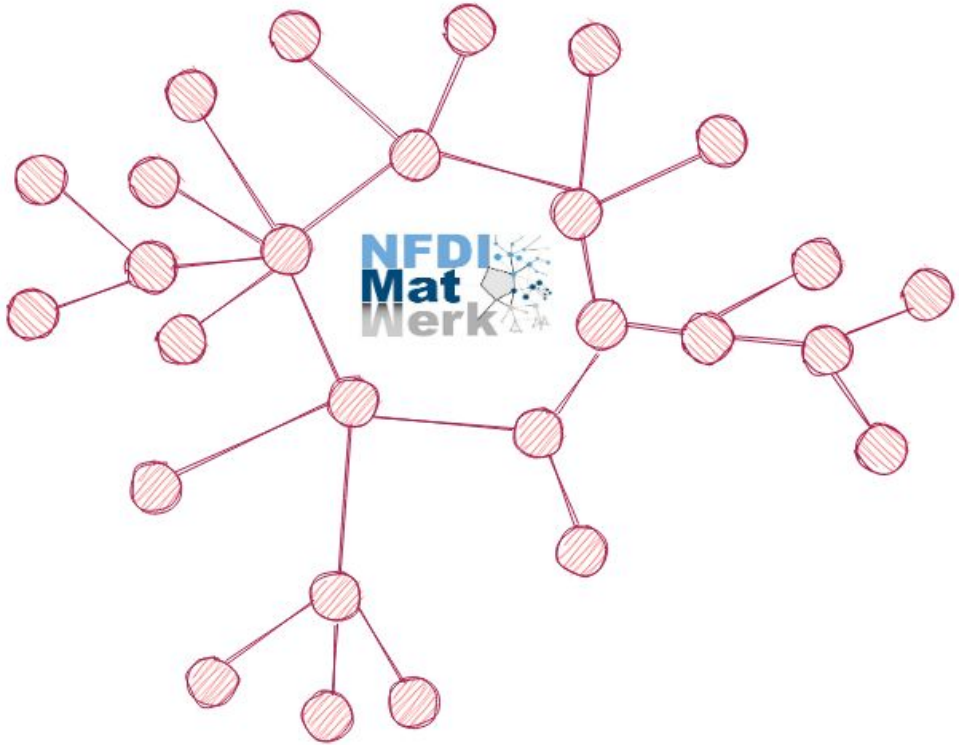
NFDI MatWerk KG v2.0

CONTENT:

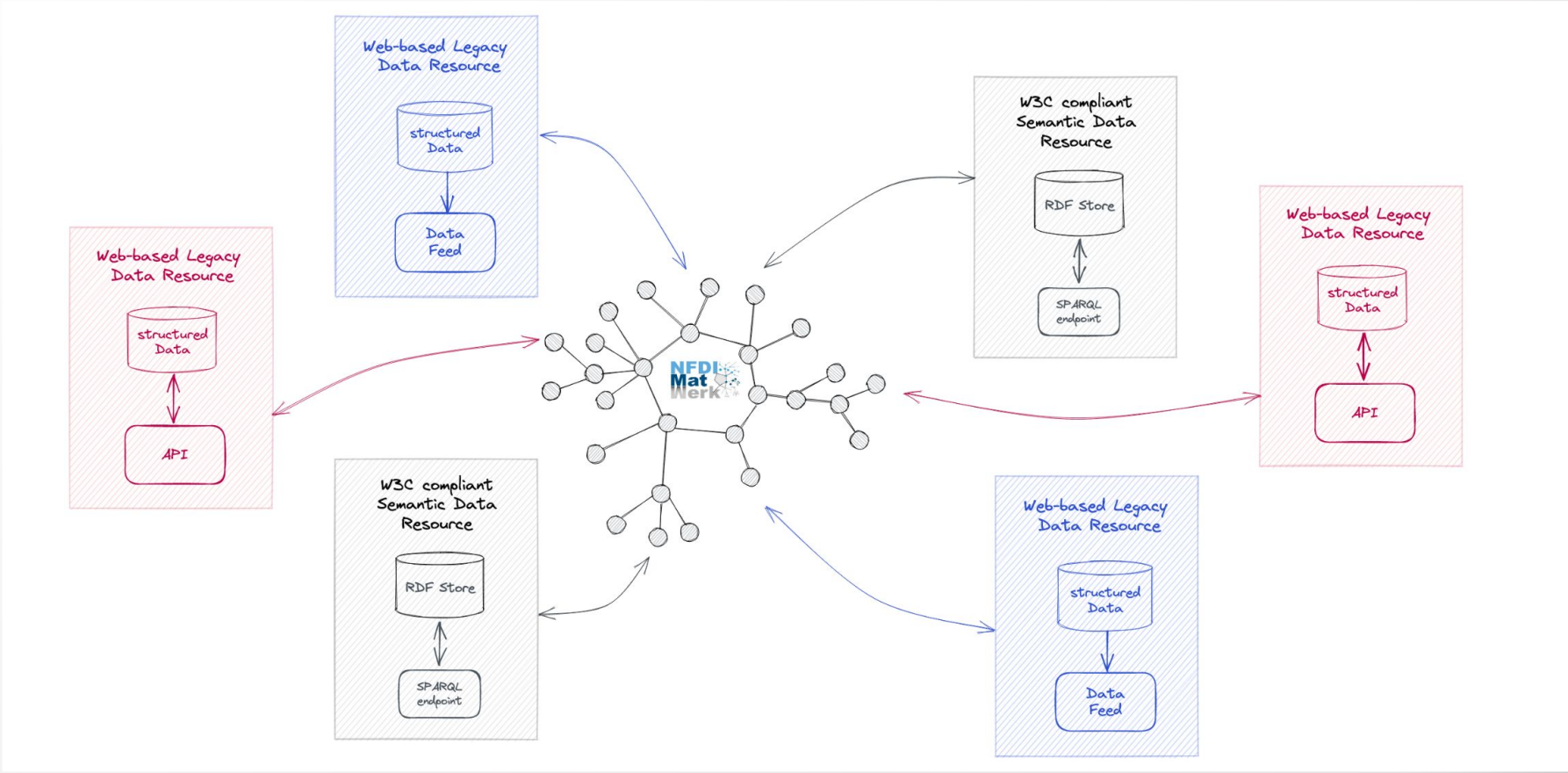
- Index and Metadata for **MatWerk** Project Resources (v1.0)
- plus Metadata for **MatWerk** content as e.g., from
  (1) Fully fledged Semantic Data resources
  (2) Structured Data resources (Legacy Data resources)
  (3) Wikibase-based resources

PURPOSE:

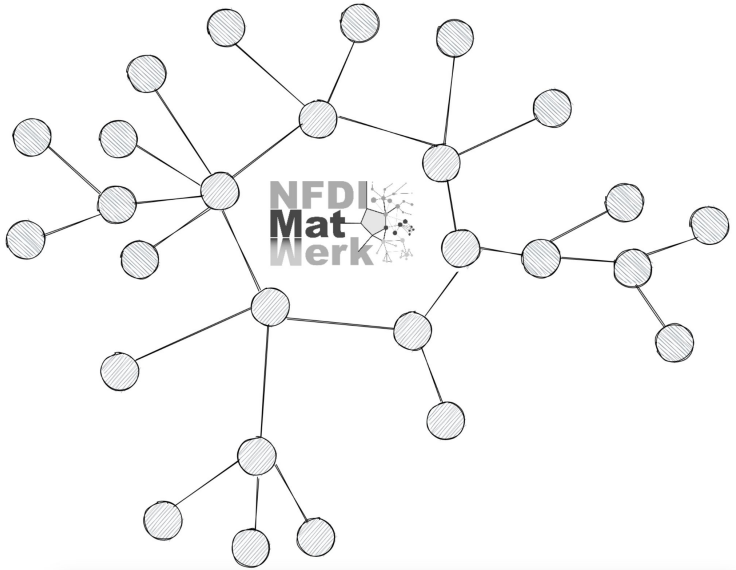
- Metadata for retrieval, exploration, and analysis of **MatWerk** research data resources

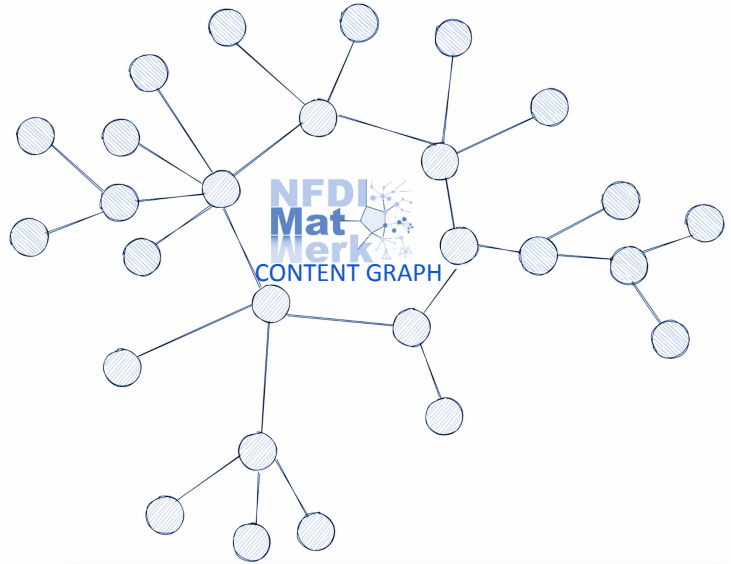


MSE KG and MatWerk Data Resources



NFDI MatWerk KG: Content Integration



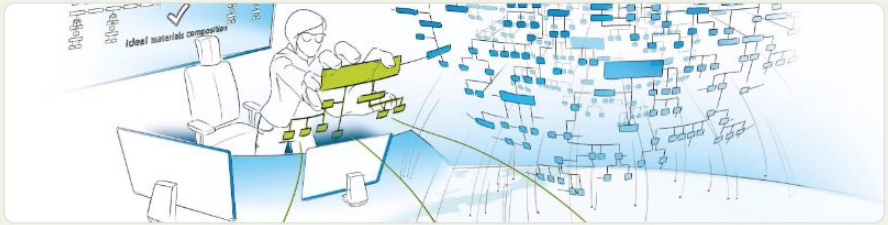


MSE KG v1.0 + Content Graph = MSE KG v2.0

Agenda @ Patents4Science Workshop

- 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



NFDI-MatWerk LOD Working Group

Linekd Open Data defines a vision of globaly 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)

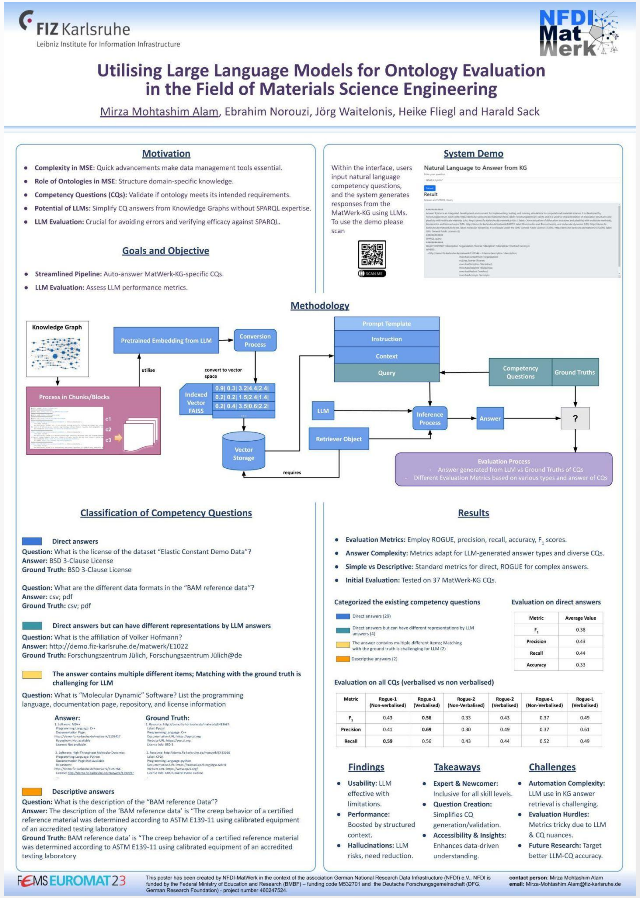
NFDI MatWerk LOD Working Group

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?

Should domain experts know SPARQL?



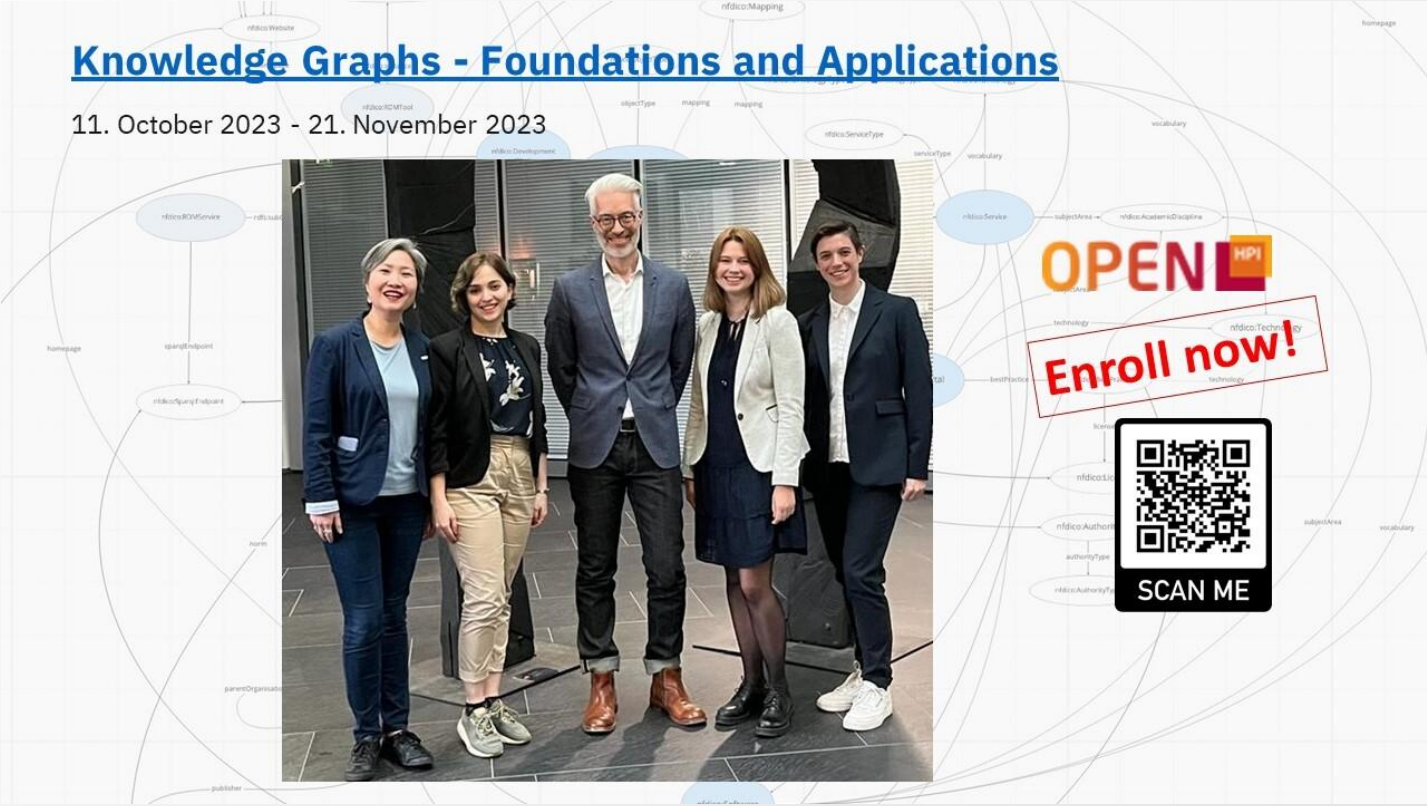
Utilising Large Language Models for Ontology Evaluation in the Field of Materials Science Engineering

Summary

- 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.**

Free Open MOOC: Knowledge Graphs - Foundations and Applications

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



Thank you very much for your attention!