Semantic data integration using the dataspace management system (DSMS)

Tobias Huschle, Yoav Nahshon (Fraunhofer IWM, Freiburg)

Semantic data integration using the DSMS

Why semantic data?

Current situation

- Wide range of **subdomains** with **various requirements**
- **Large volume of data**
- Located in multiple **“Silos”**
- **Heterogeneous** data

Challenges

- **Limited access** to existing data
- High number of **tailored solutions** (most of the time incompatible)
- Need for **resources**

Which materials are already tested?

What is the correlation between carbon weight fraction and tensile strength?

Goal

- **Interoperability –** break up silos
- **Data-Driven** decision making

Solution

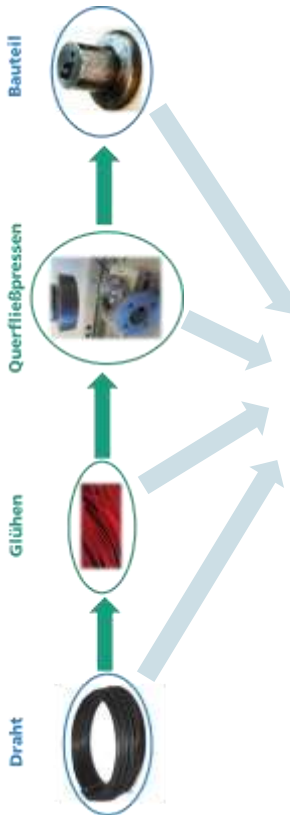
- **Semantic data integration**
- **Ontology-driven** tools (*e.g.*: Triple-store)

Semantic data integration using the DSMS

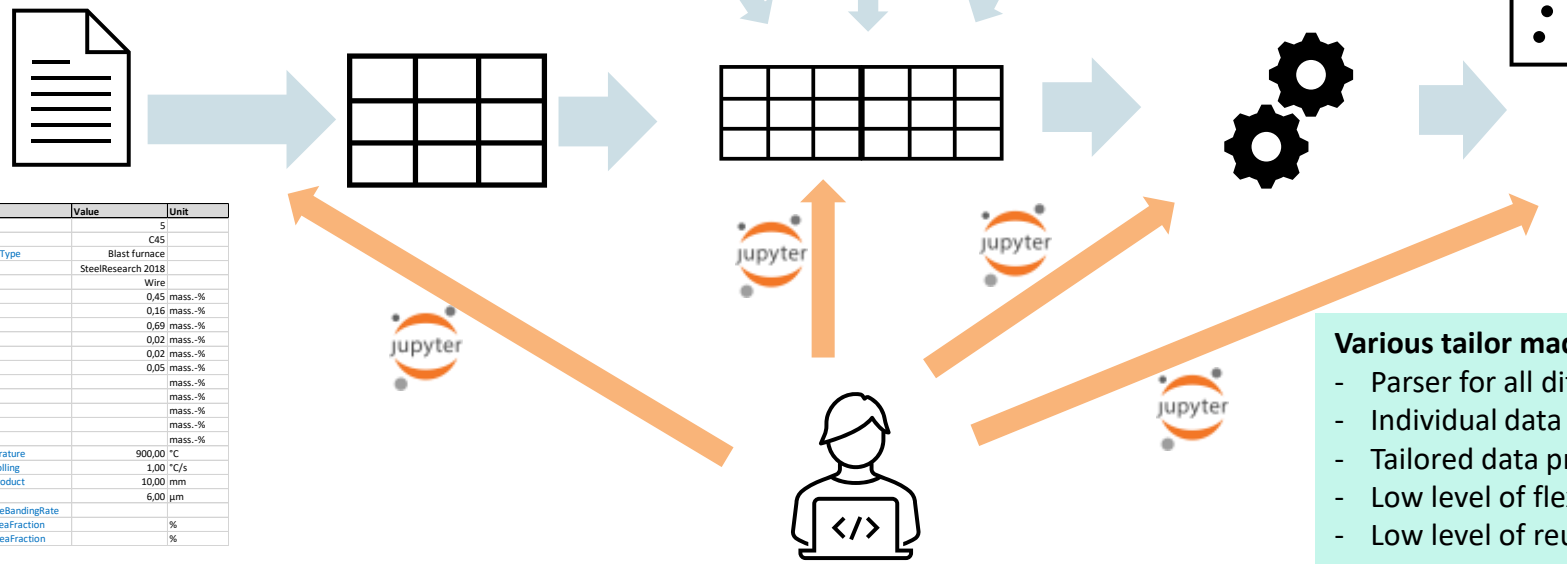
Classic workflow of scientific data processing

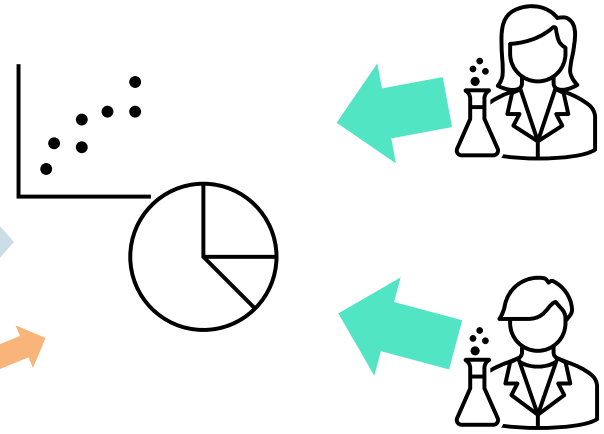
> "Most data scientists spend **only 20% of their time on actual data analysis** and **80% percent of their time finding, cleaning, and reorganizing huge amounts of data,** which is an **inefficient data strategy**"
>
> Armand Ruiz, Lead Product Manager IBM Data Science and Watson @ IBM

Process chain	Data sources	Read in data	Structure data	Process data	Evaluate data	Scientist



Term	Value	Unit
Part-ID	5	
MaterialName	C45	
MaterialProductionType	Blast furnace	
MaterialBatchID	SteelResearch 2018	
ProductType	Wire	
CWeightFraction	0,45	mass.-%
SiWeightFraction	0,16	mass.-%
MnWeightFraction	0,69	mass.-%
PWeightFraction	0,02	mass.-%
SWeightFraction	0,02	mass.-%
CrWeightFraction	0,05	mass.-%
MoWeightFraction		mass.-%
NiWeightFraction		mass.-%
NWeightFraction		mass.-%
AlWeightFraction		mass.-%
CuWeightFraction		mass.-%
FinishRollingTemperature	900,00	°C
CoolingRateAfterRolling	1,00	°C/s
InitialDiameterOfProduct	10,00	mm
FerriteGrainSize	6,00	µm
InitialMicrostructureBandingRate		
NmITotalSurfaceAreaFraction		%
MnSTotalSurfaceAreaFraction		%





Various tailor made IT-solutions required

- Parser for all different data files
- Individual data schema for each project
- Tailored data processing
- Low level of flexibility
- Low level of reuse

Semantic data integration using the DSMS

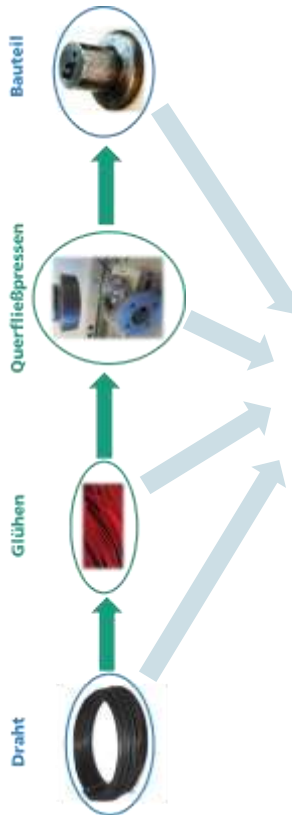
Architecture using the DSMS

Process chain

Data sources

Data Space Management System

Scientist



Term	Value	Unit
Part-ID	5	
MaterialName	C45	
MaterialProductionType	Blast furnace	
MaterialBatchID	SteelResearch 2018	
ProductType	Wire	
CWeightFraction	0,45	mass.-%
SiWeightFraction	0,16	mass.-%
MnWeightFraction	0,69	mass.-%
PWeightFraction	0,02	mass.-%
SWeightFraction	0,02	mass.-%
CrWeightFraction	0,05	mass.-%
MoWeightFraction		mass.-%
NiWeightFraction		mass.-%
NWeightFraction		mass.-%
AlWeightFraction		mass.-%
CuWeightFraction		mass.-%
FinishRollingTemperature	900,00	°C
CoolingRateAfterRolling	1,00	°C/s
InitialDiameterOfProduct	10,00	mm
FerriteGrainSize	6,00	µm
InitialMicrostructureBandingRate		
NmlTotalSurfaceAreaFraction		%
MnSTotalSurfaceAreaFraction		%

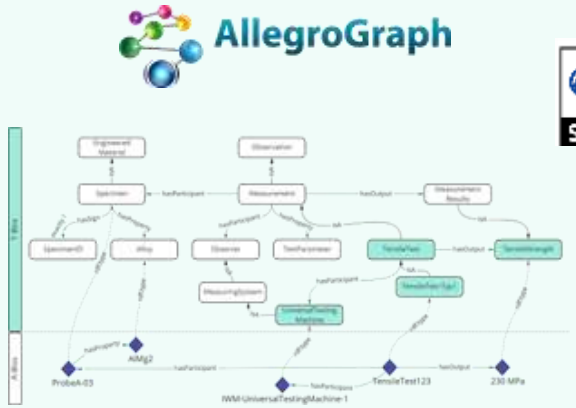
Andere Versuchsdaten

Input

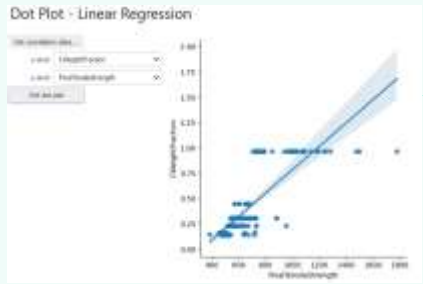
Transform data

Store data

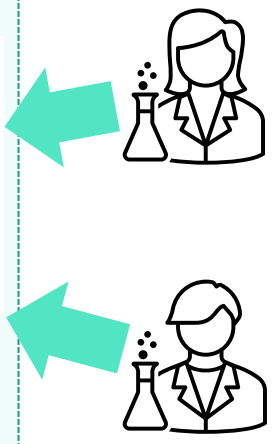
Process data



Structured data enriched by semantic meanings
→ Easy data access



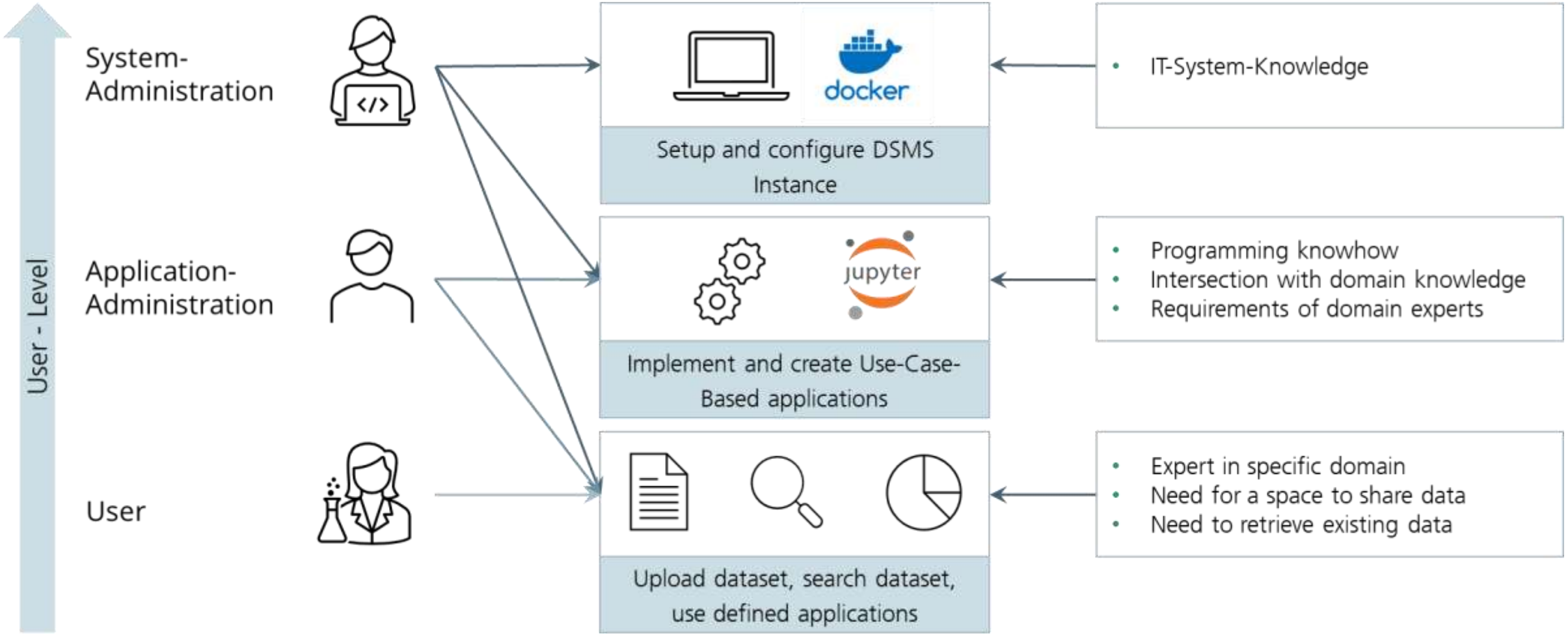
Standardized interface for data retrieving



Fraunhofer IWM

Semantic data integration using the DSMS

Roles in DSMS



Semantic data integration using the DSMS

BAM Brinell-Hardness data

BAM

ID	Test Piece Identifier	Test Piece Composition	Test Piece Producer	Indentation Repetition	Indentation Horizontal Diameter	Indentation Vertical Diameter	Indentation Average Diameter	Brinell Hardness	Total Average Diameter	Average Brinell Hardness	Standard Deviation of Brinell Hardness	CRM Average Brinell Hardness	CRM Standard Deviation Brinell Hardness	CRM Uncertainty (UCRM)	Testing Machine Uncertainty (Uh)	Measurement Resolution Uncertainty (Ums)	Permissible Uncertainty (Umpe)	Brinell Hardness Uncertainty
1	A	CuZn38As	Copperalli	1	0,86316	0,83682	0,84999	106,8933	0,833716	111,4762	6,710004098	141,390453	0,549079652	0,995	0,625950803	0,123289022	4,26	5,45755974
2	A	CuZn38As	Copperalli	2	0,84611	0,82287	0,83449	111,0279	0,833716	111,4762	6,710004098	141,390453	0,549079652	0,995	0,625950803	0,123289022	4,26	5,45755974
3	A	CuZn38As	Copperalli	3	0,86161	0,86781	0,86471	103,1706	0,833716	111,4762	6,710004098	141,390453	0,549079652	0,995	0,625950803	0,123289022	4,26	5,45755974
4	A	CuZn38As	Copperalli	4	0,81512	0,79962	0,80737	118,8424	0,833716	111,4762	6,710004098	141,390453	0,549079652	0,995	0,625950803	0,123289022	4,26	5,45755974
5	A	CuZn38As	Copperalli	5	0,80892	0,81512	0,81202	117,4468	0,833716	111,4762	6,710004098	141,390453	0,549079652	0,995	0,625950803	0,123289022	4,26	5,45755974
6	B	CuZn215Si3P	Copperalli	1	0,64311	0,65551	0,64931	185,5625	0,64828	186,4511	9,33984791	141,390453	0,549079652	0,995	0,625950803	0,261937263	4,26	5,47709797
7	B	CuZn215Si3P	Copperalli	2	0,64776	0,62351	0,635635	193,776	0,64828									
8	B	CuZn215Si3P	Copperalli	3	0,68185	0,671	0,676425	170,7244	0,64828									
9	B	CuZn215Si3P	Copperalli	4	0,62606	0,64931	0,637685	192,511	0,64828									
10	B	CuZn215Si3P	Copperalli	5	0,647776	0,63691	0,642343	189,6815	0,64828									
11	C	CuNiSi	fem	1	1,11266	1,09561	1,104135	61,93647	1,071297									
12	C	CuNiSi	fem	2	1,04602	1,05377	1,049895	68,87533	1,071297									
13	C	CuNiSi	fem	3	1,01193	1,02433	1,01813	73,46174	1,071297									
14	C	CuNiSi	fem	4	1,08321	1,08476	1,083985	64,39339	1,071297									
15	C	CuNiSi	fem	5	1,111266	1,08941	1,100338	62,38918	1,071297									
16	D	CuSn6	fem	1	0,94374	0,92825	0,935995	87,55401	0,967232									
17	D	CuSn6	fem	2	0,93444	0,94529	0,939865	86,80624	0,967232									

Metadata	
Test Standard	DIN EN ISO 6506-1
Test Date	21.04.2022
Test Data File	?
Test Piece Thickness	8
Test Piece Processing	casting and rolling
Test Piece Preparation	all samples prepared through steps of smoothing, polishing, and cleaning
Testing Machine	Emco Test M4C 025 G3
Optical Measurement	Emco Test M4C 025 G3
Indenter Identifier	3688
Indenter Composition	Tungsten Carbide Composite
Indenter Shape	Ball
Indenter Diameter	2,5
Test Temperature	22,7
Test Force	612,9
Force-Diameter Index	10
Test Points Distance	in the standard range
Test Point Edge Distance	in the standard range
Loading Time	14
Indentation Shape	circle
Constant Load Unit Conversion	0,102
Constant Pi	3,141592654
Hardness Symbol	HBW 2.5/62.5
Certified Reference Material (CRM)	15808010607
CRM Certified Brinell Hardness	142
CRM Indentation Reputation	5
Constant Sigma ms	0,00155
Constant t for CRM	1,14
Permissible Error	-0,43
Constant Erel/per	0,03
CRM Uncertainty (Ucrm)	1,99

Meta data:

- General information to reproduce the experiment

Measurement data:

- Measurement specific information
- Resulting data of the measurement process

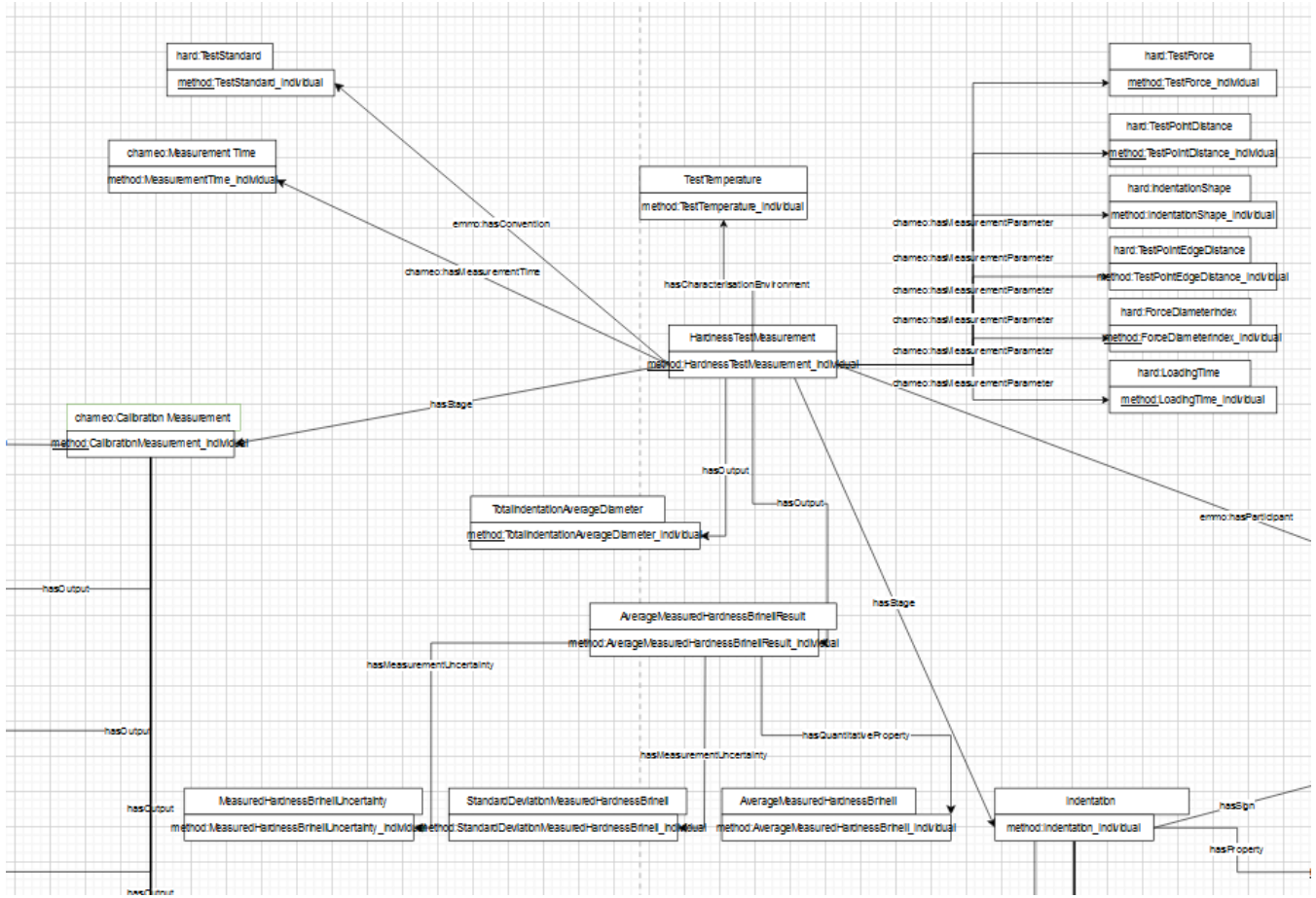
Semantic data integration using the DSMS

Ontology

Ontology development

Ontology to provide data schema used in SPARQL queries

- Domains:
  - Hardness measurement
  - Tensile test measurement
- Top-Level/ Domain-Ontology
  - EMMO - Elementary Multiperspective Material Ontology
  - CHAMEO - Characterisation Methodology Domain Ontology
  - Mechanical Testing Ontology basierend auf EMMO



Semantic data integration using the DSMS

Live Demo

https://kupferdigital.materials-data.space/

Conclusion

- DSMS provides a framework for semantic data integration and data processing
- Benefits of DSMS:
  - Different roles for specific needs
  - User management and access control
  - Flexible installation of e.g. pipelines for RDF-Generation
  - Integration of application e.g. data analysis
  - Low hurdle for material scientist for scientific data research due to UI
  - Integration of technical innovation e.g. AI-based text processing

Thank you for your attention!