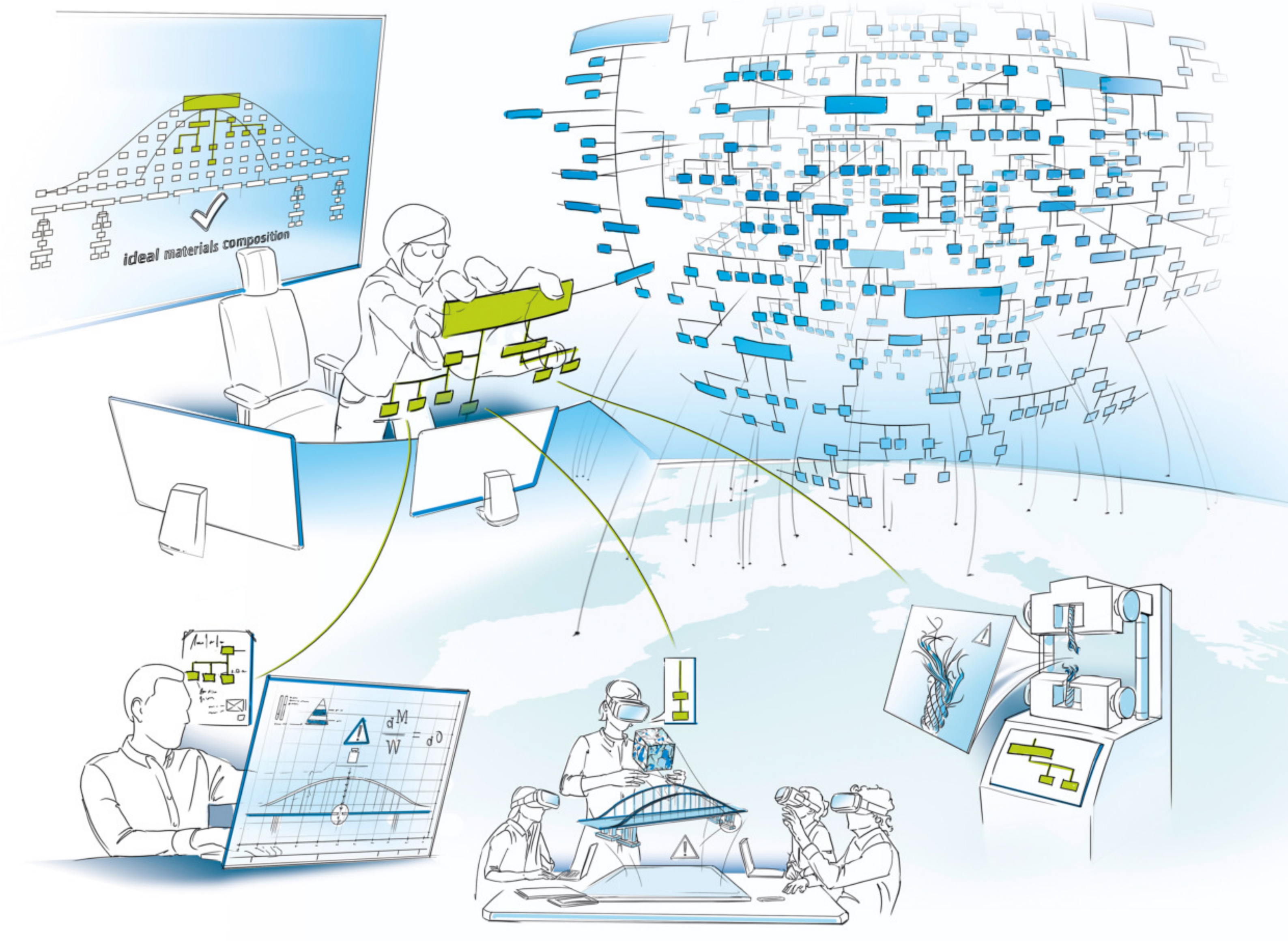
09/30/2020

NFDI Proposal NFDI-MatWerk

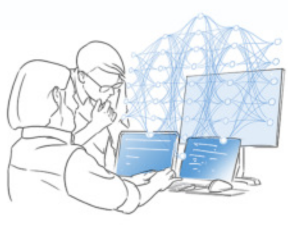
Nationale Forschungsdateninfrastruktur für Materialwissenschaft & Werkstofftechnik

National Research Data Infrastructure for Materials Science & Engineering

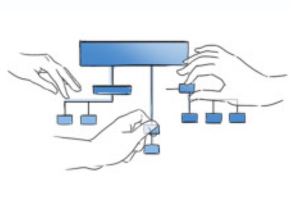


Community Interaction

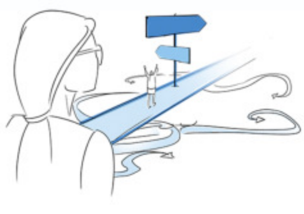
Materials Data Infrastructure



Materials Workflows and Software Development



Ontologies for Materials Science



Strategy Development

List of Abbreviations and Definitions

AAI	Authentication & Authorization Infrastructure
AI	Artificial Intelligence
ALU	Albert-Ludwigs Universität Freiburg
ASM	American Society of Materials
BAM	Bundesanstalt für Materialforschung und -prüfung
BMBF	Bundesministerium für Bildung und Forschung
BMWi	Bundesministerium für Wirtschaft und Energie
BPA	Blueprint Achitecture
CRC	Collaborative Research Center (Sonderforschungsbereich, SFB)
CRC/TRR	Collaborative Research Center/Transregio (Sonderforschungsbereich/Transregio, SFB)
DFG	Deutsche Forschungsgemeinschaft
DFKI	Deutsches Forschungszentrum für Künstliche Intelligenz
DFN	Deutsches Forschungsnetz
DGM	Deutsche Gesellschaft für Materialkunde e. V.
DiWan	Digitaler Wandel in der Werkstoffprüfung
DME	Digital Materials Environment
DVM	Deutsch Verband für Materialforschung und -prüfung e. V.
ELN	Electronic Lab Notebook
EMCC	European Materials Characterization Council
EMMC	European Materials Modelling Council
EMMO	European Materials Modelling Ontology
EOSC	European Open Science Cloud
ERC	European Research Council
EUSMAT	European School of Materials

EXC	Cluster of Excellence
FAIR	Findable, Accessible, Interoperable, Reusable
FAIR DO	FAIR Digital Object: digital representation of a research artifact carrying all information for enabling the FAIR principles
FAU	Friedrich-Alexander-Universität Erlangen-Nürnberg
FIZ	FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur GmbH
FTE	Full-Time-Equivalent (as calculated on the basis of a full year)
GAMM	Gesellschaft für Angewandte Mathematik und Mechanik e. V.
HoMMage	CRC/TRR 270 Hysteresis Design of Magnetic Materials for Efficient Energy Conversion
HZG	Helmholtz-Zentrum Geesthacht
ICAMS	Interdisciplinary Centre of Advanced Materials Simulation
ICME	Integrated Computational Materials Science
IDS Association	International Data Space Association
IUC	Infrastructure Use Case
IWM	Fraunhofer Institute for Mechanics of Materials
IWS	Fraunhofer Institute for Material and Beam Technology
KIT	Karlsruher Institut für Technologie
livMatS	EXC 2193 Living, Adaptive and Energy-autonomous Materials Systems
Mat-o-Lab	Materials-open-Laboratory
MGI	Materials Genome Initiative
MPA	Materialprüfanstalt
MPIE	Max-Planck-Institut für Eisenforschung
MSE	Materials Science and Engineering
NPO	Non-profit organization
OWL	Web Ontology Language

PID	Persistant Identifier
PP	Participant Project
PR	Public Relations
PTB	Physikalisch-Technische Bundesanstalt Braunschweig und Berlin
RDA	Research Data Alliance
RDF	Ressource-Description-Framework
RDM	Research Data Management
RUB	Ruhr-University Bochum
SaaS	Software as a service
SWZ	Simulationswissenschaftliches Zentrum Clausthal-Göttingen
TA	Task Area
TA-CI	Task Area Community Interaction
TA-MDI	Task Area Materials Data Infrastructure
TA-OMS	Task Area Ontology for Materials Science
TA-SD	Task Area Strategy Development
TA-WSD	Task Area Materials Workflows and Software Development
TUBAF	Technische Universität Bergakademie Freiberg
TUK	Technische Universität Kaiserslautern
USTUTT	University of Stuttgart
WP	Work Package

List of Contents

1 General Information

- Name of the consortium in English and German

  DE: Nationale Forschungsdateninfrastruktur für Materialwissenschaft & Werkstofftechnik

  EN: National Research Data Infrastructure for Materials Science & Engineering

English Summary:

Since the Stone Age the mastery of materials has always played a key role for societal and economic progress. Today the study of materials lies at the heart of the research field Materials Science & Engineering (MSE). This scientific discipline aims to characterize materials and study their processing and manufacturing with the ultimate goal to design materials with optimized properties and to maximize reusability at their end-of-life.

One challenge particular to MSE data is their inherent multiscale character. This is caused by the strongly heterogeneous microstructures present in virtually all materials, ranging from crystal defects at the atomistic level, through microscale secondary phases up to macroscale pores. Any process applied to a sample may change the material's microstructure and, thereby, its complete mechanical and a substantial fraction of its functional performance.

Due to the vast number of different experimental, computational and analytical methods to reveal these dependencies, essentially every lab is presently developing its own data tools and "recipes". This rapid but uncoordinated development hampers the digital transformation in MSE as well as the implementation of the FAIR principles. Therefore, the MSE-specific digital data space, as envisaged by NFDI-MatWerk, has to track the various highly complex dependencies of materials data while reducing the technological barriers within the community in order to enable synergies. To this end, NFDI-MatWerk envisages a materials ontology that is represented through a graph database infrastructure. This enables data sharing as well as highly performant, complex search queries and analysis runs over distributed and decentral data sources and presents an excellent basis for next generation AIs. The seamless integration of decentralized data and metadata, experimental and computational workflows and the materials ontology ensures maximum interoperability and reproducibility of the underlying research data processing.

The development of this infrastructure is a community-driven process. The data usage profiles of many Participant Projects from different sub-disciplines have been analyzed to identify the most relevant scientific scenarios within MSE. The resulting Infrastructure Use Cases are continuously guiding and challenging the development of our infrastructure. NFDI-MatWerk already today involves more than 80% of the MSE community, putting our stated goal of having "everyone on board" after the start of the project well within reach. As our recent survey revealed, the community regards NFDI-MatWerk as a unique chance to bundle the numerous individual materials data activities into a Digital Materials Knowledge Environment, therewith boosting the scientific productivity and satisfaction of each individual researcher within MSE.

German Summary:

Seit der Steinzeit hängt unser gesellschaftlicher und wirtschaftlicher Fortschritt von der Beherrschung von Werkstoffen ab. Heute ist die Materialwissenschaft & Werkstofftechnik (MatWerk) eine eigenständige Disziplin, deren Ziel es ist, Materialien zu charakterisieren und Herstellungsprozesse zu untersuchen, um Werkstoffe mit optimierten Eigenschaften zu entwickeln, sowie ihre Lebensdauer und Wiederverwertbarkeit zu maximieren.

Eine besondere Herausforderung stellen dabei die vielen strukturellen Skalen dar. Diese beinhalten z.B. Kristalldefekte auf der atomaren Skala, Sekundärphasen im µm-Bereich sowie Poren auf der Makroskala. Diese heterogene Mikrostruktur wird durch jedwede Bearbeitung verändert, und bestimmt die mechanischen und funktionellen Eigenschaften von Materialien.

Aufgrund der vielen verschiedenen experimentellen und numerischen Methoden zur Untersuchung dieser Mikrostruktur-Eigenschafts-Beziehungen verfügt jede Arbeitsgruppe über ihre eigenen Werkzeuge zur Datenverarbeitung. Dies ermöglicht eine schnelle, jedoch weitgehend unkoordinierte Methodenentwicklung, welche die digitale Transformation innerhalb von MatWerk sowie die Umsetzung der FAIR-Prinzipien behindert. Ein an die Bedürfnisse von MatWerk ausgerichteter digitaler Datenraum, wie in NFDI-MatWerk vorhergesehen, muss die verschiedenen hochkomplexen Zusammenhänge zwischen den unterschiedlichen Materialdaten abbilden können und, um Synergieeffekte zu entfalten, niedrige technologischen Barrieren für dessen Nutzung aufweisen. Zu diesem Zweck strebt NFDI-MatWerk eine Materialontologie an, welche über eine Graphdatenbank-Infrastruktur ein einfaches Teilen von Daten sowie hochperformante, komplexe Suchanfragen und Auswertungen über verteilte, dezentrale Datenquellen ermöglicht, und eine exzellente Basis für KI der nächsten Generation bildet. Die nahtlose Integration dezentraler Daten und Metadaten, experimenteller und numerischer Workflows und der Materialontologie erlaubt ein Maximum an Interoperabilität und Reproduzierbarkeit der Verarbeitung von Forschungsdaten.

Die Entwicklung dieser Infrastruktur ist ein Community-getriebener Prozess. Dazu wurden Datennutzungssprofile von vielen Participant Projects aus verschiedenen Unterdisziplinen analysiert, um die relevantesten wissenschaftlichen Szenarien innerhalb von MatWerk zu identifizieren. Die daraus resultierenden Infrastructure Use Cases helfen bei der kontinuierlichen Entwicklung und Überprüfung der Infrastruktur. Bereits heute deckt NFDI-MatWerk mehr als 80% der MSE Community ab. Damit ist unser Ziel, nach Beginn des Projekts “alle mit an Bord zu haben”, in greifbarer Nähe. Wie unsere jüngste Umfrage ergab, betrachtet die Community NFDI-MatWerk als eine einzigartige Chance, die zahlreichen Aktivitäten im Bereich des Forschungsdatenmanagements in einer digitalen, MatWerk-spezifischen Wissensumgebung zu bündeln, und damit die wissenschaftliche Produktivität und Zufriedenheit jedes einzelnen Forschers zu steigern.

- Applicant institution

Applicant institution	Location
Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V. **Head of the Institution:** Prof. Dr.-Ing. Reimund Neugebauer, President	Hansastraße 27 c 80686 München

- Spokesperson

Spokesperson	Institution, location
Prof. Dr. Christoph Eberl chris.eberl@iwm.fraunhofer.de Deputy Director Fraunhofer Institute for Mechanics of Materials, IWM	**Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.,** Fraunhofer IWM, Wöhlerstraße 11, 79108 Freiburg

- Co-applicant institutions

Co-applicant institutions	Location
Deutsches Forschungszentrum für Künstliche Intelligenz	Stuhlsatzenhausweg 3 66123 Saarbrücken
FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur GmbH	Hermann-von-Helmholtz-Platz 1 76344 Eggenstein-Leopoldshafen
Forschungszentrum Jülich GmbH	Wilhelm-Johnen-Straße 52428 Jülich
Friedrich-Alexander-Universität Erlangen-Nürnberg	Schloßplatz 4 91054 Erlangen
Karlsruher Institut für Technologie	Kaiserstraße 12 76131 Karlsruhe
Max-Planck-Institut für Eisenforschung GmbH, MPIE	Max-Planck-Straße 1 40237 Düsseldorf
RWTH Aachen	RWTH Aachen University Templergraben 55 52062 Aachen
Technische Universität Bergakademie Freiberg	Akademiestraße 6 09599 Freiberg
Universität des Saarlandes	Campus 66123 Saarbrücken

- Co-spokespersons

Co-spokespersons	Institution, location	Task area(s)
Prof. Dr. Erik Bitzek	**Friedrich-Alexander-Universität Erlangen-Nürnberg** Martensstr. 5 91058 Erlangen	Materials Data Infrastructure
Prof. Dr. Peter Gumbsch	**Karlsruher Institut für Technologie** Straße am Forum 7 76131 Karlsruhe	Materials Data Infrastructure
Dr. Tilmann Hickel	**Max-Planck-Institut für Eisenforschung GmbH** Max-Planck-Straße 1 40237 Düsseldorf	Materials Workflows and Software Development
Prof. Dr. Frank Mücklich	**Universität des Saarlandes** Campus D33 66123 Saarbrücken	Community Interaction
Prof. Dr. Matthias S. Müller	**RWTH Aachen** Templergraben 55 52062 Aachen	Materials Data Infrastructure
Prof. Dr. Harald Sack	**FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur GmbH**	Ontologies for Materials Science

	Hermann-von-Helmholtz-Platz 1 76344 Eggenstein-Leopoldshafen	
Prof. Dr. Stefan Sandfeld	**Technische Universität Bergakademie Freiberg** Lampadiusstraße 4 09599 Freiberg	Ontologies for Materials Science
Prof. Dr. Ruth Schwaiger	**Forschungszentrum Jülich GmbH** Wilhelm-Johnen-Straße 52428 Jülich	Materials Workflows and Software Development
Prof. Dr. Philipp Slusallek	**Deutsches Forschungszentrum für Künstliche Intelligenz, DFKI** Stuhlsatzenhausweg 3 D-66123 Saarbrücken	Materials Workflows and Software Development
Prof. Dr. Achim Streit	**Karlsruher Institut für Technologie** Hermann-von-Helmholtz-Platz 1 76344 Eggenstein-Leopoldshafen	Materials Data Infrastructure
Prof. Dr. Martina Zimmermann	**Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.** Fraunhofer IWS, Winterbergstraße 28, 01277 Dresden	Community Interaction

- Participants

Participating institutions	**Location**
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

- [redacted]

- [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

Participating individuals	Institution, location
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

Please note: The contributions of our **participating institutions** match the resulting contributions of the **participating individuals that are affiliated to these institutions**. Therefore, we refrain from listing these contributions here again. Details can be found in the participants' Letters of Commitment in the Appendix. In the following, we list the contributions of **participating individuals** that are affiliated to one of our **(co-)applicant institutions**.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

- **Names and numbers of the DFG review boards (DFG-Fachkollegien) that reflect the subject orientation of the proposed consortium**

  This Consortium applies for funds in the research areas
  405 – Werkstofftechnik and

  406 – Materialwissenschaft according to the DFG classification system.

  www.dfg.de/en/dfg_profile/statutory_bodies/review_boards/subject_areas

2 Scope and Objectives

2.1 Research Domains or Research Methods addressed by the Consortium, specific Aim(s)

2.1.1 State of Affairs and Motivation

Materials are at the core of most technological innovations of human history (and future). It is no coincidence that Human Ages were named after the materials humankind was able to master e.g. the (New) Stone Age, the Chalcolithic or Copper Age, the Bronze Age and the Iron Age. The underlying logic remains valid until today: **many technological challenges are tightly coupled to limitations materials present** – solving these problems at the same time also enables technological progress in other fields. Today the study of materials lies at the heart of the relatively young **research discipline Materials Science & Engineering (MSE)**.

Example: Th*e performance of fuel cells is limited by the membrane materials; battery charging rate, capacity and lifetime are determined by the battery material; hybrid materials made from aluminum alloys and fiber composites define the weight to strength ratio as well as the safe service time of multi copters, airplanes and cars; and, future materials need to become responsive and adaptive to the environment by processing information.*

Today's societal challenges are ***global* challenges**: climate change, resource scarcity, the transformation towards renewable energy, and the necessary introduction of a true (green) circular economy put an enormous responsibility onto Materials Science & Engineering (MSE). These challenges require not only concentrated efforts and innovative ideas; rather, the whole development and discovery cycle of new materials, of energy and resource efficient production processes, of enhancing materials' lifetime and improving the end-of-life usage or recycling possibilities need to be connected into a holistic endeavor. Consequently, isolated measurements of individual samples are not sufficient anymore to determine the performance of modern materials. Instead, MSE requires a community-driven **materials knowledge management system** to master these challenges successfully together with the greater society. With the recent developments in connecting materials data, information and knowledge through a shared materials knowledge graph, connecting relevant materials information and data, a possible materials information infrastructure solution has come within reach. While efforts already exist, the young MSE discipline has not yet initiated a central harmonization initiative. And while the collection of all available information and knowledge on materials seems to be an obvious strategy, it requires a cultural change in the MSE community. Furthermore, as an intrinsically interdisciplinary field, MSE always needs neighboring disciplines to participate in the approach. The NFDI initiative therefore provides the ideal framework to **establish a suitable Infrastructure for Materials Science & Engineering.** The necessary changes in the way we work with materials knowledge and understanding is an unprecedented task, which can only be achieved if **we, as the materials community, tackle this challenge heads on**.

2.1.2 Requirements for a Materials Science & Engineering (Inter-)national Research Data Infrastructure

Materials contain a *Microstructure*: The "inner structure" of materials is called **microstructure**. It is an agglomerate of, e.g., atomic arrangements, crystal orientations, and different crystal phases. In many cases, "defects" are a decisive part of the microstructure and are in fact - contrary to the intuitive meaning of "defect" - responsible for many useful material properties. There can be point defects (e.g., vacancies or impurities), line defects (e.g., dislocations), planar defects (e.g., boundaries between grains) or three-dimensional defects (e.g., precipitates). *The prospective knowledge management system needs to adequately reflect and support the resulting diversity of data and information on microstructures and structure-property relationships*.

Materials have a "memory": The materials' **microstructure is permanently changing along its life cycle,** while interacting with the environment**.** E.g., processing steps during manufacturing as well as mechanical, thermal or chemical interaction with the environment during service change the microstructure in an intricate and highly non-linear fashion. Therefore, the microstructural state at a given time depends on the sequence of changes beforehand. This is in stark contrast to what is needed in chemistry or condensed matter physics and is not considered in engineering. Furthermore, the **highest performance in materials** can often only be **reached by driving the materials into a thermodynamic non-equilibrium state**. Thus, making a material stronger at the same time implies that it becomes less stable and therefore needs to be tuned to their future purpose very carefully during development and manufacturing. This requires a very good knowledge of the relevant details of the microstructure for a very specific state of the material. A materials characterization in the equilibrium state, which ensures best reproducibility of the data, is often pointless for applications in materials science & engineering. *Therefore, it is of utmost importance that data in the materials knowledge management system is able to track the full materials history.*

Trying to create a specific type of microstructure that gives rise to specific material properties is a non-trivial task because the processing path to the envisioned microstructure is complicated to understand, to optimize and to implement due to the aforementioned **strong path dependency**. This means the interplay of microstructural features at various size and temporal scales are all stored within the microstructure and again affect each following step. Usually there is no way "going back" if unwanted microstructural features are developed. This requires very precise and controlled processing and manufacturing including appropriate documentation. Here, a tight integration of digital materials representations into Industry 4.0 manufacturing environments are required where information can seamlessly flow in both - manufacturing and materials - data spaces. *Therefore, the materials digital infrastructure has to allow for digital workflows representations. All information concerning machinery and process parameters have to be stored in a digital form allowing researchers to recover the path of the microstructural changes and the related changes in materials performance.*

During service, materials continue to be exposed to, e.g., elevated or high temperatures, chemical reactions and mechanical loading. Therefore, the materials microstructure continues to evolve and the changes in materials properties ultimately determines the lifetime of the application. Especially in safety relevant applications, a tight control of developing defects is used to predict the safe use of components (typical examples are turbine blades for airplanes, axels in trains, etc.). In such cases defect evolution takes priority in terms the changes analyzed and recorded. *Therefore, a digital materials infrastructure needs to contain reference materials data sets which are connected to the relevant workflows as well as their interpretation. Such data sets need to be stored in a way that all the relevant information is available and certified so that it cannot be changed or falsified. Furthermore, curation processes and as the data space becomes larger automatic plausibility checks need to be developed within the materials community and established in the infrastructure.*

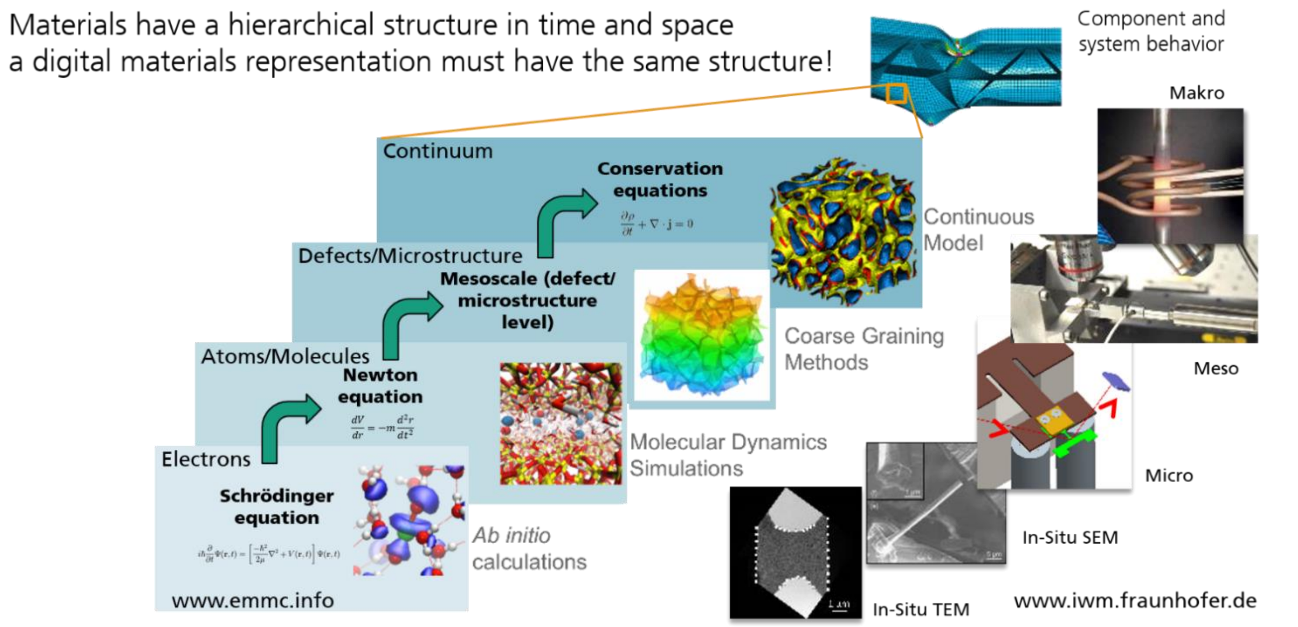


***Figure 1**. Materials are hierarchical by nature. Therefore, a digital materials representation requires the same hierarchical logic.*

Materials have an inherent *Multiscale* structure: ranging from the (sub-) atomic range up to the device level, a large number of physical, chemical and mechanical phenomena exist, operate and interact among each other in a hierarchy of different length and time scales (see the sketch below). As a consequence, materials are characterized by data at the nanoscale (10^{-10} m; 1 fs), at the microscale (10^{-6} m; 1 µs) all the way up to the macroscale (~1 cm - 1 m; 1 s – 1 yr). These scales need to be handled simultaneously to capture the materials properties and performance. This requires a combination of numerous different experimental and simulation techniques. Even more serious is, however, the strong interrelation between these scales, which can be only captured by hierarchical data and information structures. Depending on the (re-)usage scenario information on different size scales are relevant, which makes it quite impossible to define a common standard for all cases. *A digital materials infrastructure therefore needs to be flexible in terms of the data's dimensionality and its evolution.*

Materials science is highly *Multidisciplinary* with respect to the design, the fabrication and the inherent rules of materials: in simple words, chemistry is responsible for the electronic bonds of atoms and molecules, physics describes the behavior of solids, engineering is responsible for the fabrication of specimens and devices, and mathematics is the language in which all laws and relations are expressed. Equally large as the topical fields are also the numbers of methods used -- both on the experimental as well as on the simulation side. *Therefore, a knowledge management system must be able to adapt to a large and diverse spectrum of data and information sources, formats and dependencies. Furthermore, the connecting knowledge graph and its interfaces to other disciplines need to evolve continuously.*

2.1.3 How Does an MSE Research Data Infrastructure advance Prediction and Design of Complex Materials' Behavior?

The Art of Materials Science and Materials Engineering is …

… to understand and predict the ongoing microstructural evolution with all chemical and atomistic aspects on all hierarchical levels during processing and service; as well as …

… to make the link from microstructural aspects to properties of materials and thereby to the performance on the macroscopic component level.

In MSE the common format to express understanding is by formulating a hypothesis. This hypothesis contains the relevant physical or chemical mechanisms on their respective length and time scales. Based on such a hypothesis, idealized models and ontologies are formulated which e.g. describe the connection of mechanisms and data points with the microstructural feature evolution. *A digital materials research data infrastructure contains concepts (e.g., microstructural features or physical mechanisms like diffusion) and their relation (models).*

The resulting model is expressed in terms of analytical equations, which allow to describe and predict how the microstructure and related properties will develop in general. The materials models are then validated by comparison with experimental results. Owing to the high degree of complexity and the hierarchical nature of the microstructure, most analytical material models and ontologies represent the physical reality only in selected aspects. Therefore, experimental validation has to be specifically designed to each model (e.g., how materials strength depends on grain size, precipitation density and alloy content). In addition, the multiscale and 'multiphysics' character of materials requires the combination models from different length/time scales and disciplines. *Relations between concepts need to be expressed by mathematical objects within the infrastructure and data objects need to be findable within a rather specific context so they can be used for validation.*

A prerequisite to do so is to know all relevant mechanisms of the selected materials. Gaining this knowledge is challenging since it is an inherently interdisciplinary procedure that relies on a combination of vastly different experimental and simulation techniques. A collaborative approach to this is indispensable. *The infrastructure must support workflows to (automatically) evaluate materials models in different context (e.g., for different materials) and across institutions so that generalization can be tested.*

As consequence **ontologies play a much more decisive role in MSE than in many other communities**: Without a proper definition and relation of concepts defining mechanisms in materials at different scales and their relations across the scales, it is fundamentally impossible to understand the formation of microstructures and their impact on materials properties.

Detailed insights into microstructure evolution are possible through numerical materials simulations. Usually, these simulations are governed by ordinary or spatial-temporal partial differential equations which can be solved iteratively by evolving the microstructure in small time steps. Furthermore, the complex interaction between different features (e.g. dislocations with grain boundaries) can be integrated for realistic scenarios. Such models need a much tighter interaction with experiments as e.g. the microstructure at the beginning of the experiment has to be provided. In the following, the experiment (e.g. tensile test) can be conducted in the lab and the computer. Through thorough analysis, the validity of the fundamental hypotheses and mechanisms can be understood in detail. Such tight interaction between simulation and experiment is part of the so-called Integrated Computational Materials Science - ICME approach. *Workflows within the infrastructure need to allow a seamless integration of research data from lab and computer (in silico) experiments across institutions and disciplines and support data driven materials modelling (e.g. machine learning).*

The variety of numerical simulations is as broad as the field of materials which can roughly be separated into four major levels:

1. Atoms and their interplay are calculated by *ab initio* methods, i.e. by solving the Schrödinger equation and are free of any adjustable parameters (e.g. only basic information such as the number of electrons and the initial atomic positions are needed); up to 1000 atoms can be simulated for short timescales up to a few picoseconds. *On this level MSE is strongly interacting with the condensed matter community, which needs to be supported by the infrastructure.*
2. The structure and formation of individual defects can be simulated by, e.g., molecular dynamics simulations; they are based on the solution of Newton's equation of motions and therefore require the masses and positions of atoms and additionally functions for the (complicated) interatomic potentials (which can be obtained from *ab initio* methods); billions of atoms can be simulated up to the millisecond regime.
3. Interaction of microstructural features (e.g. point defects, dislocations or grain boundaries) is the focus on the next level, by using experimental and atomistic results to simulate more realistic volumes and time evolution, e.g., through a coarse-graining approach; depending on the microstructural feature, the volume can scale from cubic µm to millimeter and timescales can be hours. Sometimes such simulation methods are called mesoscale simulations.
4. On the macroscopic continuum level, Finite Element methods are often used and materials models are used which contain the knowledge and information of the smaller size scales from experiments as well as simulations; time and size scale allow to simulate full components and systems (e.g. a virtual turbine blade consisting of a superalloy under simulated service conditions). *This level is typically tightly interacting with engineering and therefore the infrastructure interface has to be seamlessly integrated in both disciplines.*

2.1.4 What are the Key Challenges in the MSE Community?

Next to the multiscale-multiphysics character inherent to all practically relevant materials, there are other key challenges in MSE: The **high dimensionality of the configuration space** that characterizes materials is not only a consequence of the huge number of possible chemical combinations, but also of the combinatorial rich microstructure. The latter is a direct consequence of the limitless variations accessible e.g. through the materials processing routes. Another challenge is that available experimental as well as simulation data **often contain substantial errors**. These errors are a direct consequence of chemical and process variations in the fabrication of materials as well as related to limitations in the measurement process, simplifications in materials models or unavoidable approximations in the simulation techniques. *Therefore, the MSE research data infrastructure needs to support curation processes, the community needs to establish metrices for data quality and develop exemplary reference materials data sets.*

To overcome some of these challenges, the availability of Artificial Intelligence methodologies enabled more **advanced data-driven materials modelling approaches**. These approaches have been implemented successfully in combination with simulated as well as experimental materials data sets. Here, typically grey box models have been used to reduce the materials related challenges in terms of dimensionality. AI based approaches will tremendously gain from larger available high-quality materials data sets and are a very useful tool towards accelerated MSE. Especially in terms of the design of new materials AI has attracted a lot of attention recently. These approaches become more and more important as the materials' complexity increases and the dimensionality of the materials parameter space multiplies. *The MSE data infrastructure needs to support and simplify the implement AI driven modelling. Furthermore, materials data needs to be connected to knowledge representations to prepare for the next generation AI approaches, where different methodologies will be combined.*

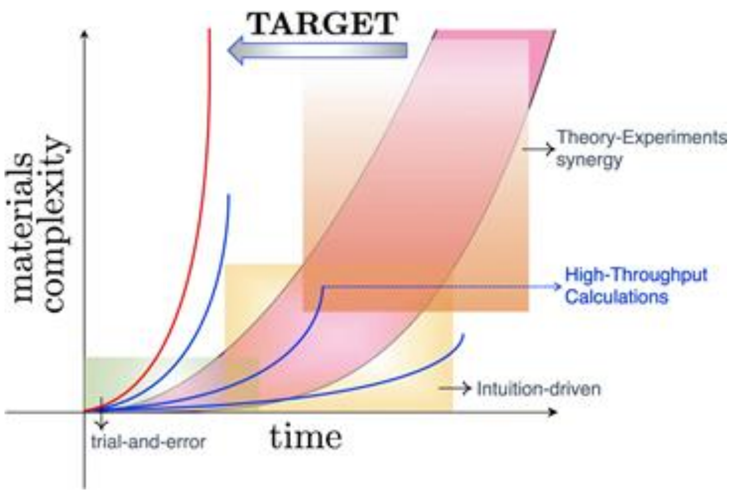


Figure 2*. Traditional, intuition-driven material development requires substantial time investment, and the materials' complexity is relatively low. Data-driven approaches promise to be able to reach much more complex materials with highly specialized microstructures (Lookman et al., 2019).*

Although artificial intelligence approaches are very promising, they are, however, still far from being a standard method for materials data applications. One of the reasons is that the materials community only has a limited number of data repositories, especially on experimental data. Important data sets related e.g. to novel materials classes or to specific material properties, and therefore **highly important materials data sets are simply not FAIR**. Today, this often prevents the application of data-driven approaches and of novel machine learning techniques since they require large databases. While some specialized high throughput simulation approaches are suitable to creating large numbers of data sets, the MSE community in general is not yet able to reliably simulate all possible situations. This is due to (i) the vast range of different physical and chemical phenomena which introduce a high complexity into simulations, and (ii) the lack of curated experimental validation data. Together, these reasons result in vast "empty regions" in the parameter space of possible materials. The situation is similar on the side of the experimental methods: again, a few specialized high-throughput methods exists, but the much larger part is rather traditional and has, by comparison, a relatively low "data-bandwidth". *The MSE infrastructure needs the ability to seamlessly integrate the existing and future materials data bases in a context-sensitive FAIR way.*

Solving these highly prominent problems of our MSE domain cannot be done in a gradual manner but requires a true paradigm shift.

2.2 Objectives and Measuring Success

The acceleration and combination of all aforementioned experimental and simulation methods is naturally far from trivial but is only one example for the current state. Hence, **any scientist within MSE is striving to answer the question: “What needs to be done in order to understand, to predict and to design better and possibly more complex materials in the future and in a shorter time?”.**

We believe that **this goal can only be reached through the following objectives related to a novel infrastructure for MSE**: we seek to drastically **increase the quality and value** of the research data that we can obtain by our well-established and highly specialized ‘traditional’ methods; we seek to enable access to and document data from as many reasonable experiments and simulations as possible, to make them findable such that future researcher can dedicate their time to filling blank spots in our high dimensional parameter space of materials by investigating new samples. Throughout, we have to consider the highly heterogeneous nature of our data in MSE, which is due to the fact that it is at the heart of different disciplines. This requires appropriate software tools as well as a formalized approach to our data and to all related processes, i.e., a materials ontology, that enables us to describe the whole materials microstructure and its history as well as to predict its future. In other words, **making our materials’ data findable, accessible, interoperable, and reusable, i.e., FAIR, will create a so-far unseen amount of added value. It will change the way how research is conducted**. This can only be done in a bottom-up approach that includes the requirements and the respective state of our community. Therefore, **our five most important objectives** are:

1. To place the individual materials scientist into the center of our efforts by **integration and inclusion of all stakeholders**, including in the decision-making processes, and the vivid exchange of opinions, ideas and experience within our community. It further comprises important education and training aspects. This will be realized amongst others through the

creation of collaborative platforms, a large number of interactive workshops and an annual conference. **Task Area Community Interaction (TA-CI)** focuses on these objectives.

2. To provide a reliable digital platform enabling the digital representation of materials data and MSE-specific metadata: the **Digital Materials Environment (DME);** and to provide services to easily store, share, search, and analyze data and metadata while ensuring data integrity, provenance, and authorship. This will be achieved by **Task Area Materials Data Infrastructure (TA-MDI).**
3. To develop a **software framework** of the DME that allows the users to **easily share workflows** in processing environments, to implement automated protocols for experimental and modelling studies with widespread MSE tools and, therewith, to unite workflows with the underlying materials data. This is the domain of our **Task Area Workflows and Software Development (TA-WSD)**.
4. To provide and integrate a **unified materials ontology**, including private materials ontologies for homegrown tools; to implement a **MSE knowledge graph** by which data characterized via their metadata becomes findable, accessible, as well as reusable. We have dedicated the **Task Area Ontologies for Materials Sciences (TA-OMS)** to this objective.
5. To allow for a **community-driven strategic development towards a digital transformation of MSE**: This requires continuously analyzing the state, drawing conclusions, developing strategies, shaping respective measures, and by controlling the effectiveness of the implementations. This will be realized in our **Task Area Strategy Development (TA-SD)**.

In general, NFDI-MatWerk will measure its success using two different methodological approaches: by **inspecting the impact of the measures in the scientific community** and the individual researchers (qualitative) and by **defining key performance indicators** (KPIs) within and across the TAs.

The former is achieved by continuously monitoring how MSE scientists' research workflows are enhanced by the services offered by the consortium, the adoptability of the services for the researchers and the rate at which the services are actually adopted within the community. The close integration of the community is implemented by 20 participant projects (PP) and related Infrastructure Use Cases (IUCs). The PPs can be derived from existing scientific projects or are related to typical scientific areas in MSE (see sect. 3.1.3). The main infrastructure usage profile is then extracted from participant projects. While the participant projects define the scientific context and provide 'real' data, workflows and access to scientists, the Infrastructure Use Cases describe the interaction with the digital infrastructure. These Infrastructure Use Cases (including the data sets from the PPs) are used to derive the requirements for the products and services as well as their interaction an interoperability. Therefore, the IUCs allow to test the implemented infrastructure (across the TAs) and form the basis for continuous delivery as automated test cases. As such, current and future PPs play an important role in this assessment as they form a cross-section of the MSE infrastructure usage profiles. NFDI-MatWerk will therefore follow a rigorous testing strategy: During design and development of the services selected

PPs that dedicate resources to the common cause of NFDI-MatWerk will serve as early beta testers to give immediate feedback. More mature services will be gradually rolled out and validated with current and future PPs as part of their IUCs and will finally be made available for entire MSE community. This service offering in so called rings from an inner circle to the wide public is a common pattern in service development. Following an agile development approach, **at each ring the services' quality will be assed in order to ensure that the services meet the requirements of the MSE scientists**. Since 20 PPs and 17 IUCs cannot fully capture the plethora of infrastructure usage profiles and therefore will be continuously developed further together with the community.

For the latter, the **TAs will monitor the success of their services with specific KPIs:**

The success of measures of **TA-CI** will be evaluated based on the absolute numbers of workshops, trainings and summer schools being held, accounting for the number of participants in these events. Furthermore, usage statistics of the interactive communication platform will be closely monitored. Professional societies (DGM, DVM and GAMM) are actively involved in the measures. They will, therefore, provide annual reports on their activities and NFDI-MatWerk-related community interaction.

TA-MDI will collect KPIs from the technical services established within the measures. On the one hand, the number of DME installations and the amount of stored data sets (number of different digital objects) will be assessed. On the other hand, KPIs are used to assess the impact within the community by considering the number of registered users, the number of search queries and the number of digital objects accessed by the users of the DME.

An ultimate goal of the **TA-WSD** is to make the interaction with the NFDI-MatWerk infrastructure user-friendly and efficient for the scientist. Therefore, the success of the TA-WSD can be measured by the number of scientists that make use of the workflow systems and software solutions developed, but also by the typical research time to generate new scientific results or to reproduce previously generated results. Another goal of the TA is to improve the exchange within the MSE community. This can be measured in terms of the number of different workflows generated by the PPs, i.e. the number of tools integrated and the number of their combinations that have been evaluated within the software infrastructure to generate scientific stories.

The coverage of the ontologies provided by **TA-OMS** can be measured in terms of (i) the covered material range (e.g. approximate number of metals covered) and (ii) the number of different materials scientific phenomena represented; this will vary for different (experimental, simulation, theoretical) methods. We will keep track of this "map" on the MFDI-MatWerk web page. The ability to match existing, ongoing and future ontology developments outside NFDI-MatWerk with the MSE ontologies of TA-OMS is another KPI that will be tracked and quantified. Through community surveys the acceptance of the ontologies will be evaluated. Further, these surveys will serve as an evaluation of the provided tools and ontology development methodologies and will provide valuable feedback for subsequent re-iterations. KPIs related to these tools include the number of downloads/installations, as well as the use within infrastructural services, e.g. in MSE related databases and in the DME of TA-MDI. Long-term KPIs include the number of scientific works referring explicitly to ontology development driven by TA-OMS and the number of queries to NFDI-MatWerk's knowledge graph developed and continuously updated by TA-OMS.

TA-SD will measure the consortium’s success through the implementation of steady community-feedback-cycles, involving both quantitative and qualitative evaluations, as it was already done for this proposal. Factors representing to which degree NFDI-MatWerk’s targets were reached may include both aspects of community satisfaction, as well as, the technological dissemination of the developed products and services. Therefore, we aim to set up viable KPIs that help to draw a roadmap about the targets for the upcoming years.

3 Consortium

Overview of consortium members’ participations in other consortia:

- Prof. Dr. Harald Sack (FIZ Karlsruhe) is also participating in NFDI4Culture, NFDI4Chem, NFDI4DataScience, MaRDI. NFDI4Memory, NFDI4Objects, NFDI4Agri, NFDI4Phys and NFDI4MobilTech
- Prof. Dr. Matthias S. Müller (RWTH Aachen) is also participating in NFDI4Ing, NFDI4Chem and NFDIxCS (under review).
- Prof. Dr. Achim Streit (KIT-SCC) is also participating in the 2nd round proposals PUNCH4NFDI and NFDI4MobilTech and is co-spokesperson in the 1st round project NFDI4Ing.

3.1 Composition of the Consortium and its Embedding in the Community of Interest

While the first proposal called NFDI4MSE of 2019 was not successful for the Materials Science and Engineering community, the support in the community so far has been overwhelmingly positive. During the past year, NFDI-MatWerk's team has been invited on many occasions to discuss the development of a shared MSE digital infrastructure. Furthermore, all materials societies have been keen on picking up the subject of the digital transformation with a strong support also from colleagues in industry. The societies’ work groups have been intensively discussing the implications on the expert level. Last but not least, the overwhelming support for the here presented effort NFDI-MatWerk can also be measured by the fact that there is no second MSE consortium, although the MSE community is rather heterogeneous as was remarked by the reviewers. Beyond that, the NFDI also has led to the implementation of strategy boards at universities and research centers which are closely related to the presidential office, respectively the top management. In several of these strategy boards materials science members are present and help connecting the different disciplines in the spirit of the NFDI.

3.1.1 Composition of the Consortium

***Figure 3**. Research institutions of various types from all throughout Germany are part of NFDI-MatWerk as either (co-) applicants or participants.*

The consortium NFDI-MatWerk consists of a broad range of universities and institutions that are active in MSE research in Germany, as illustrated in figure 3. Currently, NFDI-MatWerk consists of 12 (co-) Applicants, all of which are well-established spokespersons from both the field of MSE as well as from infrastructure providers. Another 22 participating individuals (as listed in sect.1) from German faculties and research organizations in the field of MSE have shown their commitment towards NFDI-MatWerk's objectives by expressing their willingness to collaborate and to contribute substantially. These are active in 25 participating or applying institutions (see sect. 1). As of the time of writing, the participating universities in the NFDI-MatWerk consortium cover over 80% of academic third-party funds raised in the field of MSE in 2018 (DFG, 2018). This demonstrates that Germany's most important universities and non-university research institutions in the field of MSE are going to work together in this project, which is our stated objective.

The consortium consists of stakeholders that cover essentially all materials classes, simulation methods, as well as experimental testing and characterization methods that are relevant to the MSE domain. We anticipate that more universities are going to join the consortium during the first project phase. Additionally, the consortium is closely interlinked with the most relevant expert associations of the field of MSE, all of which have expressed their support for our vision. Some of those are actively involved in NFDI-MatWerk. These links allow for a strongly augmented reach and guarantees that this reach goes far beyond the stakeholders who are already interested in "digitalization". Last but not least, we are well aware that the necessary cultural change can, in some areas, only happen, if the corresponding topics find their way into the curriculae of our universities. Our consortium consists of key persons responsible for university education and the design of curriculae on different levels who, even if this is not an explicit task or NFDI measure, will over the next years adjust their courses accordingly.

Based on the latest DFG Förderatlas (DFG, 2018) NFDI-MatWerk currently involves universities that represent more than 80% of the academic DFG funding in the subject areas materials engineering (405) and materials science (406). The consortium NFDI-MatWerk will increase their institutional reach by adding more participants after a successful funding decision.

It is fair to say that NFDI-MatWerk is thoroughly and sustainably embedded in the MSE community. Many of the authors of the highly visible white-paper "Digitale Transformation in der Materialwissenschaft und Werkstofftechnik" (Engl.: Digital transformation in Materials Science and Engineering) (Sandfeld et al., 2019) are now strongly involved in NFDI-MatWerk. Personal ties with all stakeholders have developed long before this proposal, resulting in reliable networks. These are fundamental for the success of NFDI-MatWerk, as they ensure easy and direct communication channels, especially during complex collaboration phases involving many contributors. The consortium was constituted over the course of multiple meetings of the German MSE community on research data management following a call by the DFG section for MSE. The (co-)spokespersons were identified in a community meeting on July 19th, 2019, with the aim to implement the required measures effectively and efficiently. To better include the larger community and to allow for different levels and degrees of involvement, we follow a two-pronged approach: while **Task Areas (TAs)** act as the main organizational building blocks that carry out most of the work, **Infrastructure Use Cases (IUCs)** will help inform and contribute to the development efforts of the task areas. These were derived in meetings together with the community and will be carried out by **Participant Projects (PPs)**. These IUCs will provide actual data, and ensure that NFDI-MatWerk data infrastructure will be integrated in the daily workflows at the PPs and research institutions in general. Participant projects include in particular large collaborative research initiatives such as Cluster of Excellence (CoEs), CRCs, or International Research Training Group (IRTGs), which guarantee a wide reach of NFD-MatWerk into the community.

3.1.2 The Idea behind the NFDI-MatWerk Task Areas (TAs)

Overview of NFDI-MatWerk Task Areas	
TA Name	TA Central Responsibility
TA Community Interaction (TA-CI)	This TA is responsible to stimulate the acceptance and vivid participation of the overall MSE community through communication, counselling, education / training and steady increase of the outreach.
TA Materials Data Infrastructure (TA-MDI)	This TA is responsible for developing the Digital Materials Environment (DME) and for providing services to easily store, share, search, and analyze data and metadata while ensuring data integrity, provenance, and authorship.
TA Materials Workflows and Software Development (TA-WSD)	This TA is responsible for advancing the research in MSE by providing solutions for workflow management and software foundations that will improve both productivity and quality of the data processing and simulation aspects.
TA Ontologies for Material Sciences (TA-OMS)	This TA develops Ontologies and metadata schemes in close cooperation with the stakeholders, tailored to their respective needs.
TA Strategy Development (TA-SD)	This TA is responsible for managing the development of a digital strategy for our consortium and the MSE community.

The Task Areas (TAs) are the organizational foundation of NFDI-MatWerk. Consequently, we address our challenge from an entrepreneurial viewpoint. Seeing the community as our target group and ourselves as the necessary service providers, we suggest a functional organizational form that separates the necessary tasks along their thematic areas. This allows us to anchor the ongoing and perhaps most relevant question when trying to legitimate any future measure of our consortium deeply in the setup of our cooperation: *"What serves the researcher the most?"* An overview of these Task Areas and a brief description of their contributions are given above.

The first step to resolving the challenges addressed by the TAs is to acknowledge that we as materials scientists **will not be able** to provide all the solutions fully and independently by ourselves - not even if we had the funding. Therefore, an important component for enabling our vision is that we urgently require the **best partners from outside of our community** who are not only willing to support us in terms of technical aspects, but who also are willing to consider the domain specific problems and peculiarities. Accordingly, we identified our demand and embedded their cooperation deeply into our consortium by integrating them as co-applicants. By design, each of our technical Task Areas is thus at par between materials scientists and information technology experts. We see this as one strong evidence for our honest commitment when it comes to cross-cutting capabilities of the NFDI. A more detailled overview of the resulting Task Areas and their specific work program is given in sect. 5 of this proposal. This consortium considers its direct members as among the best experts in their specific domain – both within and outside of Materials Science and Engineering – each of whom is furthermore experienced and interested in such a multidisciplinary endeavor.

3.1.3 Involvement of Community through Participant Projects

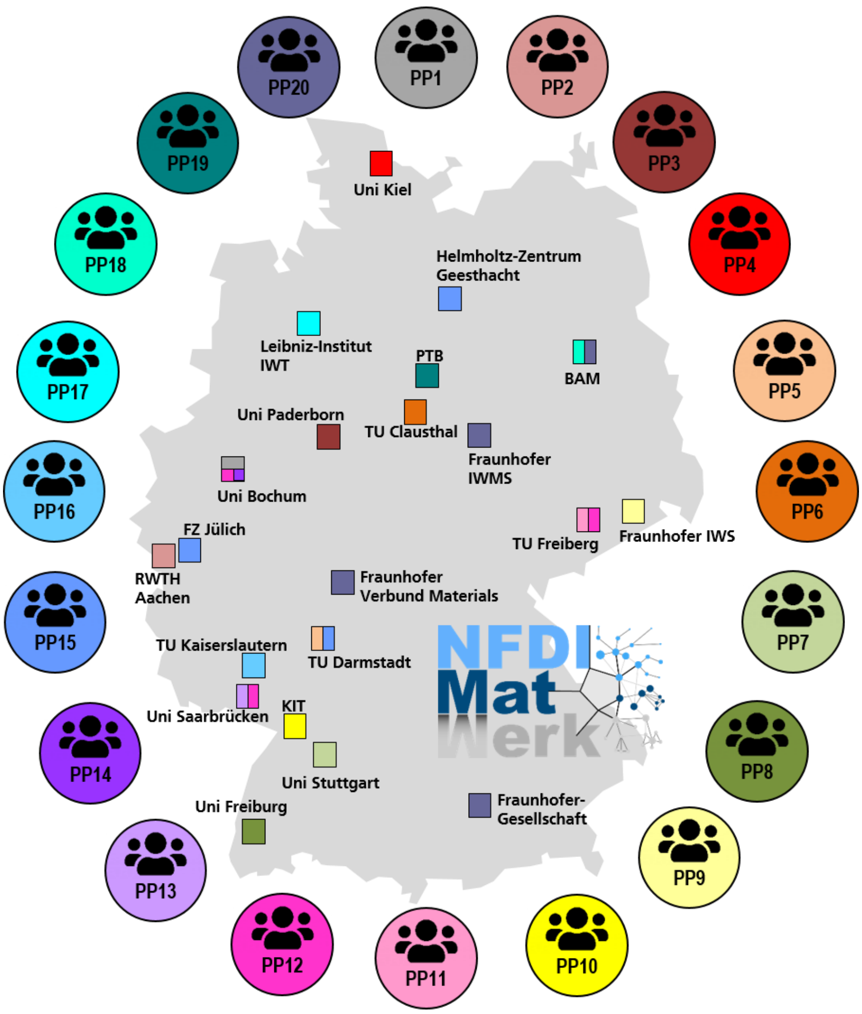


***Figure 4**. The geographic distribution of Participant Projects.*

The goal of a tight community integration into the development of a digital materials infrastructure requires processes, which allow translating 'real' scientific workflows and the related research data handling into infrastructure requirements and their implementation. The resulting infrastructure services and products need to be interoperable and easy to assemble into individual workflows, which help researchers to accelerate their work and improve the quality and reusability of their data and tools (FAIR Principles, Wilkinson et al., 2016). While many of the necessary infrastructure services and processes have been already identified (see TA-MDI and TA-WSD), their interplay in digital workflows strongly depends on the scientific usage and is specific for each sub-field of MSE. Furthermore, interoperability of available tools is still difficult and their assembly into individual digital workflows requires in-depth programming knowledge which is out-of-scope for most materials scientists. Therefore, specific **Participant Projects (PPs) have been selected**

together with the community to derive representative Infrastructure Use Cases (IUCs), which are used to guide the materials digital infrastructure development (see also sect. 4.1.2).

Role of Participant Projects (PPs): For the NFDI-MatWerk consortium, the Participant Projects exemplify typical tasks related to the management of research data in everyday scientific life (e.g. fusing data from different sources, creating digital representations of workflows) of the MSE community. The **participant projects do not claim to represent all current and future needs of the community**, which is why **they will be continuously adapted** based on feedback, input and evaluation from and by the community. This will take place according to the cycle described in sect. 3.4 of this proposal.

The jointly devised PPs set an exemplary professional context and stand for a variety of relevant scientific processes that our materials research data infrastructure must meet. Beyond that, they stand for our community's heterogeneity in context of methods, material classes, sociology, and geography, see Figure 4. Experience shows, that only the link to specific research projects (e.g., supported by DFG, BMBF, ERC) makes the practical benefit tangible for the individual scientists who, in turn, become important multipliers for our joint initiative. The PPs are confronted with recurring challenges in terms of data management and the scientists can describe the relevant, resulting demands on the central infrastructure from the user perspective.

It has to be stressed, that while the PPs through the IUCs help in their development, the products and services of NFDI-MatWerk will of course be openly and freely accessible to the entire MSE community and beyond.

Overview of the associated Participant Projects

PP01 From atoms to turbine blades (CRC/TRR 103)

Contact: *Dr. habil. Thomas Hammerschmidt, Ruhr-University, Bochum (www.sfb-transregio103.de)*	
405-01, -04 Materials Engineering; **406-03, -04** Materials Science	
Material/Methodology: Ni-based and Co-based single-crystal superalloys, metallic complex solid solutions / experiments and simulations **Related Use Cases:** • IUC02: Framework for curation and distribution of reference datasets • IUC09: Infrastructure interfaces with condensed matter physics (collaboration with FAIRmat) • IUC14: Adaptive automated characterization pipelines and meta data schemas for high throughput experiments • IUC17: Ontologies for defects in crystals	
[redacted]	**Engagement :** 24 projects in CRC/TRR103, 10 of them on 'Machine-Learning and Material Informatics'

PP02 Defect phases in structural materials (CRC 1394)

Contact: *Prof. Dr. Sandra Korte-Kerzel, RWTH Aachen; Dr. Jörg Neugebauer, MPIE Düsseldorf (www.sfb1394.rwth-aachen.de)*

405-01, -04 Materials Engineering; **406-03, -04** Materials Science
Material/Methodology: Thermodynamic and Kinetics of Defects / Simulations and Experiments **Related Use Cases:** • IUC04: Model driven data space exploration • IUC05: Digital infrastructure and workflows for labs • IUC09: Infrastructure interfaces with condensed matter physics (collaboration with FAIRmat)

[redacted]	**Engagement :** 15 project groups

PP03 Method development for mechanical joinability in versatile process chains (CRC/TRR 285)

Contact: *Prof. Dr.-Ing. Gerson Meschut; Dr.-Ing. Mathias Bobbert, Universität Paderborn (www.trr285.de)*
401-03 Production Technology; **402-01, -02, -03** Mechanics and Constructive Mechanical Engineering **405-04** Materials Science
Material/Methodology: Joining processes and loaded joints (metals, FRP) / Experiments and Simulations **Related Use Cases:** • IUC01: Web-based demonstration and teaching framework for MSE research data infrastructure • IUC05: Digital infrastructure and workflows for labs

[redacted]	**Engagement:** 16 groups

PP04 Magnetoelectric sensors: from composite materials to biomagnetic diagnostics – project A10 (CRC 1261)

Contact: *Prof. Dr.-Ing. Stephan Wulfinghoff, Thilo Paul-Stüve, Christian-Albrechts-Universität zu Kiel (www.sfb1261.de/index.php/en)*
406-01, -04 Materials Science
Material/Methodology: Magnetoelectric composite materials; coupling of magnetic, electric and mechanical phenomena; Application to magnetoelectric sensors for medically relevant questions **Related Use Cases:** • IUC06: Integrating materials data from experiments and computation into Industry 4.0 manufacturing paradigms • IUC08: Interactive visual exploration for analyzing correlations in high dimensional materials data spaces • IUC14: Adaptive automated characterization pipelines and meta data schemas for high throughput experiments

[redacted]	**Engagement :** Microstructure investigation for magnetoelectric sensors and conclusions on sensor behavior

PP05 HoMMage - Hysteresis design of magnetic materials for efficient energy conversion (CRC/TRR 270)

Contact: *Prof. Dr. Karsten Durst, Technische Universität Darmstadt (www.tu-darmstadt.de/sfb270)*
406-01, -03, -04 Materials Science
Material/Methodology: Functional magnetic and magnetocaloric materials / Additive manufacturing, artificial intelligence **Related Use Cases:** • IUC04: Model driven data space exploration • IUC06: Integrating materials data from experiments and computation into Industry 4.0 manufacturing paradigms • IUC08: Interactive visual exploration for analyzing correlations in high dimensional materials data spaces • IUC12: Alignment of application- and higher-level ontologies

[redacted]	**Engagement :** 20 groups

PP06 Working group modeling of oxygen-free production (CRC 1368)

Contact: *Jun. Prof. Dr. Nina Gunkelmann, Technische Universität Clausthal (www.sfb1368.uni-hannover.de/nocache/de/forschung/arbeitsgruppen/modellbildung)*

406-04 Materials Science

Material/Methodology: Composites between pure aluminum and iron and later alloys / MD simulations to study atomic interactions at interfaces, transfer to the continuum scale by multiscale simulations
Related Use Cases:
- IUC07: Beyond 3D: Tools for spatiotemporal microstructure studies
- IUC10: Interoperability of workflow systems (in collaboration with NFDI4Ing)

[redacted]	**Engagement :** Working group modeling consisting of 13 projects

PP07 Processing uncertain microstructural data (EXC2075-PUMD)

Contact: *Prof. Dr. Felix Fritzen, University of Stuttgart (www.simtech.uni-stuttgart.de/exc/research/pn/pn3/pn3-1)*

402-02 Mechanics and Constructive Mechanical Engineering; **405-05, -06** Materials Engineering
406-03 Materials Science; **410-05** Construction Engineering and Architecture

Material/Methodology: synthetic and experimental inclusion-matrix and porous microstructures / image-based homogenization and uncertainty prediction
Related Use Cases:
- IUC03: Storage concepts for large hierarchical datasets
- IUC06: Integrating materials data from experiments and computation into Industry 4.0 manufacturing paradigms
- IUC08: Interactive visual exploration for analyzing correlations in high dimensional materials data spaces
- IUC16: Unified ontology for matrix-inclusion microstructure and composites

[redacted]	**Engagement:** >14 EXC2075 projects 2 EXC2075 project networks Fraunhofer ITWM collaboration GAMM AG Data

PP08 - Cluster of Excellence Living, Adaptive and Energy-autonomous Materials Systems (EXC 2193-livMatS)

Contact: *Prof. Dr. Thomas Speck, Prof. Dr. Jürgen Rühe, Prof. Dr. Christoph Eberl, Albert-Ludwigs-Universität Freiburg (www.livmats.uni-freiburg.de/en)*

405 Materials Engineering; **406** Materials Science; **407-03** Systems Engineering; **108** Philosophy;
110-04 Psychology; **202-04, -05** Plant Sciences; **321** Molecular Chemistry; **327-02** Theoretical Chemistry

Material/Methodology: Polymers, functional materials, programmable materials / synthesis, additive manufacturing, simulations
Related Use Cases:
- IUC06: Integrating materials data from experiments and computation into Industry 4.0 manufacturing paradigms
- IUC12: Alignment of application- and higher-level ontologies
- IUC13: Co-creation environment for experts

[redacted]	**Engagement :** ~ 50PIs, 50PhDs, 3 Junior-Research-Groups

[redacted]	2 Fraunhofer Institutes Ecology Institute

PP09 DiWan - The Digital Transformation in Materials Testing: Necessary Actions for the Paradigm Change in Enterprises (BMBF)

Contact: *Prof. Manfred Füting, Dr. Thomas Hanke (www.fraunhofer-materials-data-space.de, www.digitales-laborbuch.de)*
405-04 Materials Engineering; **407-05** Systems Engineering; **409-06** Computer Science
Material/Methodology: various materials / digital lab book, workflow management in the lab, knowledge system **Related Infrastructure Use Cases:** • IUC01: Web-based demonstration and teaching framework for MSE research data infrastructure • IUC05: Digital infrastructure and workflows for labs • IUC06: Integrating materials data from experiments and computation into Industry 4.0 manufacturing paradigms

[redacted]	**Engagement :** 7 Partners

PP10 Integrated engineering of continuous-discontinuous long fiber reinforced polymer structures (IRTG 2078)

Contact: *Prof. Dr.-Ing. Thomas Böhlke, Karlsruhe Institut für Technologie (www.grk2078.kit.edu)*
402-01, -02, -03 Mechanics and Constructive Mechanical Engineering; **403-03** Process Engineering, Technical Chemistry; **406-03** Materials Science; **410-05** Construction Engineering and Architecture
Material/Methodology: Long fiber reinforced polymers (thermoset and thermoplastic materials) **/** Real process chain is modeled and optimized by digital twin which is based on multi-scale characterization and simulation **Related Use Cases:** • IUC01: Web-based demonstration and teaching framework for MSE research data infrastructure • IUC06: Integrating materials data from experiments and computation into Industry 4.0 manufacturing paradigms • IUC08: Interactive visual exploration for analyzing correlations in high dimensional materials data spaces • IUC15: Method- and scale-bridging workflows and data structures for tomography • IUC16: Unified ontology for matrix-inclusion microstructure and composites

[redacted]	**Engagement :** 14 doctoral projects at 8 institutes (6 KIT and 2 Fraunhofer institutes)

PP11 ERC Starting Grant MuDiLingo: Ontology Design for Dislocations

Contact: *Prof. Dr. Stefan Sandfeld, Technische Universität Freiberg (tu-freiberg.de/fakult4/imfd/mimm/erc-projekt-mudilingo)*
406-04 Materials Science
Material/Methodology: single and polycrystalline materials including superalloys, HEAs / Molekulardynamik (MD), Discrete Dislocation Dynamics (DDD), TEM **Related Use Cases:** • IUC01: Web-based demonstration and teaching framework for MSE research data infrastructure • IUC07: Beyond 3D: Tools for spatiotemporal microstructure studies • IUC12: Alignment of application- and higher-level ontologies • IUC17: Ontologies for defects in crystals

[redacted]	**Engagement :** a group of 6 PhD students and PostDocs

PP12 Data Science in MSE education, qualification and life-long learning (EUSMAT)

Contact: *Prof. Frank Mücklich, Saarland University; Prof. Stefan Sandfeld, TU Freiberg; Prof. Alexander Hartmaier, RUB Bochum (www.eusmat.net)*

406-04 Materials Science

Material/Methodology: Independent of a specific material/methodology. Supports teaching with a focus on using data science methods in MSE.

Related Infrastructure Use Cases:
- IUC01: Web-based demonstration and teaching framework for MSE research data infrastructure
- IUC13: Co-creation environment for experts

[redacted]	**Engagement :** 3 University chairs, 1 group leader at DFKI.

PP13 Tomography and Microstructure-based Modelling (Tomography)

Contact: *Prof. Frank Mücklich, Prof. Stefan Diebels, Prof. Hans-Georg Herrmann, Saarland University*

405-04 Materials Engineering; **406-03, -04** Materials Science

Material/Methodology: Mainly metallic materials, e.g. steels, aluminium alloys, Ti-6-4 / Atom Probe Tomography, nano-CT, FIB/SEM Tomography, Metallographic Serial Sectioning, Mechanical FE Simulation, Material Modeling.

Related Use Cases:
- IUC03: Storage concepts for large hierarchical datasets
- IUC05: Digital infrastructure and workflows for labs
- IUC07: Beyond 3D: Tools for spatiotemporal microstructure studies
- IUC15: Method- and scale-bridging workflows and data structures for tomography

[redacted]	**Engagement :** 3 chairs at Saarland University

PP14 Interdisciplinary Centre of Advanced Materials Simulation – Micromechancial modelling (ICAMS)

Contact: *Prof. Dr. Alexander Hartmaier, Ruhr-Universität Bochum; Dr. Franz Roters, Dr. Jaber Rezael Mianroodi, Max-Planck-Institut für Eisenforschung Düsseldorf (www.icams.de)*

405-04 Materials Engineering; **406-03, -04** Materials Science

Material/Methodology: Micromechanical and scale-bridging modeling of plastic deformation, fracture and fatigue of structural materials

Related Use Cases:
- IUC07: Beyond 3D: Tools for spatiotemporal microstructure studies
- IUC08: Interactive visual exploration for analyzing correlations in high dimensional materials data spaces
- IUC10: Interoperability of workflow systems (in collaboration with NFDI4Ing)

[redacted]	**Engagement :** 3 departments with 9 groups, Advanced Study group at MPI für Eisenforschung

PP15 Working Group “Metadata in Nanoindentation and Micromechanics” (NanoData)

Contact: *Prof. Dr. Ruth Schwaiger, FZ Jülich GmbH; Prof. Dr. Erica Lilleodden, Helmholtz Zentrum Geesthacht; Prof. Dr. Karsten Durst, TU Darmstadt (www.fz-juelich.de/iek/iek-2/DE/Forschung/Metadaten_in_der_Nanoindentation_und_Mikromechanik/Metadaten_node.html)*
405-04 Materials Engineering; **406-04** Materials Science
Material/Methodology: Nanoindentation raw data and analysis / Experiments **Related Use Cases:** • IUC05: Digital infrastructure and workflows for labs • IUC08: Interactive visual exploration for analyzing correlations in high dimensional materials data spaces

[redacted]	**Engagement :** Three scientific research groups

PP16 Profile Area Advanced Materials Engineering TU Kaiserslautern (TUK-AME)

Contact: *Prof. Dr.-Ing. Tilmann Beck; Prof. Dr. Heike Leitte, TU Kaiserslautern (www.uni-kl.de/en/ame)*
405-04 Materials Engineering; 4**09-05** Computer Science
Material/Methodology: Metallic alloys / Analysis of mechanical as well as physical properties and microstructure / Development of interactive visual data exploration and topological data analysis tools **Related Use Cases:** • IUC05: Digital infrastructure and workflows for labs • IUC08: Interactive visual exploration for analyzing correlations in high dimensional materials data spaces

[redacted]	**Engagement:** 15 Professors with their working groups involved in AME

PP17 Leibniz-Institut für Werkstofforientierte Technologien (IWT)

Contact: *Dr. Matthias Steinbacher, IWT Bremen (www.iwt-bremen.de)*
403 Process Engineering, Technical Chemistry; **405** Materials Engineering; **406** Materials Science
Material/Methodology: highly stressed metallic structural materials/ material, process and production engineering **Related Use Cases:** • IUC07: Beyond 3D: Tools for spatiotemporal microstructure studies • IUC10: Interoperability of workflow systems (in collaboration with NFDI4Ing) • IUC11: Development of coupled ontologies and workflows for thermochemical treatments • IUC17: Ontologies for defects in crystals

[redacted]	**Engagement :** 4 departments, 6 groups in department of Materials Science

PP18 Bundesanstalt für Materialforschung und -prüfung (BAM)

Contact: *Prof. Dr. Birgit Skrotzki (www.bam.de/Navigation/EN/About-us/Organisation/Organisation-Chart/President/Department-5/Division-52/division52.html)*
405-01, -04 Materials Engineering; **406-03** Materials Science

Material/Methodology: High-Temperature metal alloys / Mechanical testing, Continuum Simulations **Related Use Cases:** • IUC02: Framework for curation and distribution of reference datasets • IUC17: Ontologies for defects in crystals

[redacted]	**Engagement :** 3 Departments (about 400 people)

PP19 Physikalisch-Technische-Bundesanstalt - Persistent Identifiers (PTB)

Contact: *Dr. Joachim Meier (www.ptb.de)*
307 Condensed Matter Physics; **308** Optics, Quantum Optics and Physics of Atoms, Molecules and Plasmas; **309** Particles, Nuclei and Fields; **310** Statistical Physics, Soft Matter, Biological Physics, Nonlinear Dynamics **311** Astrophysics and Astronomy
Material/Methodology: n/a / physical standardization in metrology **Related Use Cases:** • IUC02: Framework for curation and distribution of reference datasets • IUC12: Alignment of application- and higher-level ontologies

[redacted]	**Engagement :** international network on standardization and metadata harmonization in metrology

PP20 Mat-o-Lab - Materials-open-Laboratory

Contact: *Prof. Manfred Füting, Dr. Thomas Hanke (www.fraunhofer-materials-data-space.de)*
405-01, -02, -03, -04, -05, -06 Materials Engineering; **406-04** Materials Science
Material/Methodology: Al-alloys, fiber-composites / Fraunhofer Materials Ontology Stack for Methods and Processes **Related Use Cases:** • IUC11: Development of coupled ontologies and workflows for thermochemical treatments • IUC12: Alignment of application- and higher-level ontologies • IUC13: Co-creation environment for experts

[redacted]	**Engagement:** Experts of Fraunhofer MATERIALS (18 Fraunhofer Institutes) and BAM

3.1.4 Community Support as expressed by Survey Results

To even further validate our community support beyond our direct network, NFDI-MatWerk conducted a **community survey in July 2020**, to learn more about its target group's opinions, wishes and challenges. The survey was conducted online and disseminated by the professional societies DGM, DVM and GAMM, as well as via personal channels within our consortium. This helped NFDI-MatWerk to **reach a total of 536 members of our scientific discipline** in Germany in three weeks of time and collect their feedback and crucial input on our issues and aims.

The respondents included a variety of viewpoints, covering an **almost representative landscape of our research society:** Among the repondents were 23% of professors, 45% of researchers with a doctoral degree, and 30% of a Master's degree. Age groups included a representative distribution of 10% of respondents below 29, 53% between 30 and 49 years and 37% above. Also, the occupational distributions were almost perfect with 40% of respondents from an academic field, 31% from research institutions outside of academia and an additional 27% from industry research.

Due to its effectiveness and good perception, this way of gathering feedback can be considered a representative rehearsal on **how we aim to interact with the community in the future**. This concerns both the methods and channels of conducting and receiving such surveys, as well as our internal processes and the strategy behind it. As pointed out thoroughly within this proposal, **we consider such strong ties to the community and their maintenance fundamental for the relevance, acceptance and finally the success of NFDI-MatWerk**. We can plausibly argue that we committed to the initial necessary steps to reach that aim and now have first lessons learned at our disposal.

NFDI-MatWerk's July 2020 community survey resulted in numerous relevant insights, some of which will be elaborated in the following sections, too (s. sect. 4.1.1). Besides technical questions, challenging the needs and status quo of our fellow scientists, NFDI-MatWerk also addressed questions focusing on the community's support towards this initiative.

Of all the respondents, 51% had heard of NFDI-MatWerk (formerly NFDI4MSE), which is rather impressive, considering that this initiative has not received funding and our outreach solely stems from the continuous work of the consortium's members in steadily spreading our vision. Generally, despite the MSE discipline's strong push towards digitalization measures, only 46% agreed they knew about such efforts. The reason is rather clear: Another 46% considered themselves "unexperienced" in context of digitalization. NFDI-MatWerk's strategy addresses this group of people just as well, which is one of the reasons why so much focus was put on community work, dedicating an entire TA to it.

NFDI-MatWerk is proud of its strong backing among the community respondents. After learning more and reading about NFDI-MatWerk's aims in the course of our survey and beyond, **a stunning 85% confirmed they support our initiative**, with only 7% refusals (8% neutral). Another **71% considered the initiative to have a high importance** (13% neutral, 16% refusal). For another 60% this even affected relevance to reach their own targets (17% neutral, 24% refusal). It is therefore fair to deduct that NFDI-MatWerk enjoys thorough embedding and support in our scientific community.

Naturally, with our community-driven approach and aims we will also depend on the community's own willfulness to contribute and interact with this initiative. It is one aspect to ensure the channels for doing so, but, as pointed out before, in the end the community itself must be mature enough to participate actively, themselves. This notion also received strong support, as 50% of respondents confirmed they would wish to participate in NFDI-MatWerk (21% neutral, 29% refusal), which can be considered a strong backing, given that such participation does not involve funds and our initiative is still at the very beginning. It also furthermore confirms that NFDI-MatWerk hits the right spot and its addressed issues are so important and pressuring that large parts of the affected scientific community are ready to contribute voluntarily.

Generally, we understand the presented survey results as

- **further prove that** NFDI-MatWerk enjoys thorough community support
- **further demonstrate that** NFDI-MatWerk addresses the community's true pain-points
- **another perspective highlighting NFDI-MatWerk's** multiple ways of interacting with and involving our community
- **first learnings and insights confirming our** appropriateness to take on this responsibility and manage this community **through sophisticated methods and channels**

3.2 The Consortium within the NFDI

Materials Science and Engineering is a relatively large subject within the science community and a highly interdisciplinary field of research. MSE integrates mathematics, physics, chemistry, mechanical and electrical engineering to develop novel materials, characterize and optimize them, develop processes to manufacture materials, evaluate and predict their lifetime and generate strategies to maximize the recyclability at the end-of-life. Therefore, NFDI-MatWerk represents the node connecting fundamental research to engineering. The consortium NFDI-MatWerk has already identified and initiated starting points for ensuring long-term interoperability with other consortia. As discussed in many preparation meetings and DFG organized conferences all consortia agreed to develop together tomorrow's interdisciplinary research data environment, see "Berlin Declaration on NFDI Cross-Cutting Topics" (Glöckner, 2019) and "Leipzig-Berlin Declaration" (Bierwirth, 2020). Modern, efficient and adaptive data infrastructures are the cornerstones of this future that will reshape research throughout Germany on a fundamental level. This effort will synchronized with our colleagues on an international level. The consortium NFDI-MatWerk shares this mission wholeheartedly and will contribute to that cause.

We consider our consortium's funding needs appropriate and worthy of such a strong commitment at this time. We argue that the NFDI's support is crucial in the context of our discipline's current status to take the next substantial steps of the digital transformation. We equally acknowledge the initiatives of several other disciplines as noteworthy and even potentially supportive for our own aims. Thus, NFDI-MatWerk actively sought for professional exchange and the identification of more concrete points of cooperation than what the "Berlin Declaration" currently manages to establish. This work's results are listed below. However, these potential fields of collaboration do not indicate dependencies on others when it comes to accomplishing this consortium's tasks. We rather intend to point out that there are overlapping fields of interest between related research domains.

3.2.1 Our Connection to Engineering

Regarding engineering sciences, **NFDI4Ing** received a funding admission to implement common interfaces between the different disciplines of engineering sciences. Generally, NFDI4Ing is intending to help a broad spectrum of disciplines, including all different domains of engineering sciences. One focus of NFDI4Ing will be therefore on providing a metadata and terminology service, which **NFDI4Ing** will develop, deploy and sustainably operate. This includes the technical platform and the generic tools, while the subject-specific requirements will remain parts of each specialized NFDI consortium. To this end, NFDI-MatWerk agrees with **NFDI4Ing** to collaborate on this semantical harmonization. Thus, domain consortia for the larger disciplines must be set up that address and implement their specific needs and directly connect to their communities. Therefore, NFDI-MatWerk aims to develop a digital infrastructure that allows for the representation of materials microstructure and its path dependence. **The two initiatives NFDI4Ing and NFDI-MatWerk are therefore highly complementary.** Accordingly, **NFDI4Ing** and NFDI-MatWerk jointly initiated MSE-specific working groups within a workshop in July 2020 and identified joint interests, namely the development of software infrastructure, an ontological interface between materials information, as well as Industry 4.0 standards describing the related machinery and processes.

Within their archetype ELLEN of **NFDI4Ing** the exchange of information between heterogeneous disciplines is addressed, with a focus on diverse data sources. We have therefore initiated the joint infrastructure use case IUC15, to extend this concept to the exchange of workflows. More specifically, an integrated development environment for the specific needs of NFDI-MatWerk is developed within measure WSD-4. At the same time **NFDI4Ing** uses the workflow system KaDI4Mat within the community cluster “Materials Science and Engineering” for simulations as well as for tracking samples (in archetype CADEN). Ensuring the interoperability between both systems, optimally connects NFDI-MatWerk to the whole framework of engineering science.

In the closely related **scientific discipline** of **computer and data sciences**, NFDI-MatWerk has already started discussions with **NFDIxCS** and **NFDI4DataScience**. Since NFDI-MatWerk is relying on large-scale data and computational resources, the collaboration with dedicated expert consortia will be highly valuable in the ontology field as well as reproducible use of Artificial Intelligence and the related training data.

3.2.2 Our Connection to Chemistry

NFDI4Chem is developing and maintaining the infrastructure and software for electronic lab notebooks ELN as open source. NFDI-MatWerk will tailor the ELN to the requirements of the MSE community and collaborate with NFDI4Chem to ensure interoperability. Furthermore, NFDI-MatWerk will collaborate with **NFDI4Chem** to establish interoperable workflows and provide tools for domain ontology development. **NFDI4Chem** and NFDI-MatWerk will collaborate on developing semantic interfaces between the chemistry and materials ontologies. Here, processes and tools for ontology matching will be implemented.

3.2.3 Our Connection to Biology

NFDI-MatWerk and **DataPLANT**, which received funding in the first NFDI round 2020, share personal ties and the same innovative spirit. NFDI-MatWerk acknowledges the advanced starting point that **DataPLANT** can base its efforts upon, already having a steadily growing platform at its disposal through the so-called Galaxy Project. NFDI-MatWerk looks forward to user their technical infrastructure as a learning example and will transfer experiences gained into the shape of the MSE digital materials environment (DME). Here, especially Galaxy's modular tool box for data processing and computing, allowing for a flexible integration of newly added tools promises an interesting starting point for the development of comprehensive interfaces, while at the same time ensuring the necessary adaptability. At the same time, NFDI-MatWerk has advanced concepts at its disposal for a comprehensive approach to infrastructure development, some of which **DataPLANT** could profit from. This includes, e.g., the thematic fields of persistent identifiers, decentral raw data access points, authentication infrastructures, but also the more organizationally relevant approaches to education or business and incentive models.

3.2.4 Our Connection to Physics

In the field of natural sciences, NFDI-MatWerk has started collaborations with different consortia from Physics (FAIRmat), and Chemistry (NFDI4Chem). In Physics, we aim to collaborate closely with **FAIRmat**, **NFDI4Phys** as well as **DAPHNE4NFDI**.

Here, the interface to **FAIRmat** is most relevant for NFDI-MatWerk, since condensed-matter physics is the basic science perspective on solid-state phenomena. **FAIRmat** focuses their effort on fundamental chemo-physical effects. In the case of atom probe tomography (APT), for example, the correlation of details of the electric field and the electronic density in the sample needs to be analyzed and exploited in order to predict the emission of atoms from steps at the surface. Such predictions are the key to arrive at more accurate and precise predictions of the ion launch and trajectories; and thereby are a route to improve current methods for backing out atomic positions. In contrast, NFDI-MatWerk addresses the relevance of the microstructure for engineering materials and their applications. Therefore, the focus of NFDI-MatWerk is on the correlation of atomic positions and defect states as a function of the chemical potential and the sample history (processing, heat treatment). For both perspectives, several gigabyte of data needs to be handled and needs to be evaluated with machine learning approaches. Therefore, the infrastructure use case IUC09 is used to design in coordination with **FAIRmat** such that the exchange of data, metadata, ontologies, and concepts for data curation between two NFDI consortia is ensured without a loss of information.

NFDI4Phys aims to harmonize the semantic description of physical units and quantities, which will be closely related and relevant to the ontologies of NFDI-MatWerk. NFDI-MatWerk and **NFDI4Phys** have initiated talks towards a joint perspective on the semantical standardization of all aspects that characterize materials as the foundation of their digital representation. From Physics point of view, MSE represents a strongly application-oriented research domain, which aims to collect specific materials data by different experimental methods (and also by simulations). These (determined) factual data about a specific physical material, sample or probe are also vital to improve the theoretical modelling of the respective material characteristics and its behavior. **NFDI4Phys** confirms the shared need for a coherent representation (description) of materials through common unified ontologies, to which persistent identifiers are assigned. Furthermore, the two consortia aim for common unified metadata schema-modules and common uniform documentation rules on how schema-modules and ontologies are to be applied to build the digital representation of the researched material. Since **NFDI4Phys** is looking forward to submitting its proposal within the 2021 call, a possible integration of first elements of the described approach into NFDI-MatWerk has already been discussed. This led to the concrete cooperation of the two consortia in the field of ontology development and beyond, supported by and under cooperation with the PTB (Physikalisch-Technische Bundesanstalt), which thus became a Participant Project in NFDI-MatWerk.

DAPHNE4NFDI intends to provide interfaces to large experimental facilities, both ontologically and in regards of software, for which a collaboration with NFDI-MatWerk will also be mutually beneficial. NFDI-MatWerk aims to set up a digital platform for the materials community investigating the microstructure of materials and its relation to mechanical and functional properties. Its community constitutes an important user sub-community of **DAPHNE4NFDI**. In consequence, MSE-related metadata and catalogue specifications will be reconciled. While DAPHNE4NFDI focuses on managing provided raw / reduced data, NFDI-MatWerk will develop ontologies for materials definition along the complex details of a sample life path, which could be inherited in the DAPHNE4NFDI (MSE-related) infrastructures and effect the developments for other sub-communities.

3.2.5 Our Connection to Mathematics

NFDI-MatWerk and **MaRDI** share their interest in the mathematical description and integration of naturally hierarchical data concepts. In this context, the two consortia are looking forward to discuss their scientific approaches with each other to learn mutually from the perspective of a related field of science. Practically, this is supposed to be done by interacting through participant projects and Infrastructure Use Cases that serve both of the two consortia. For instance, IUC16 “Unified ontology for matrix-inclusion microstructure and composites” linked to measures OMS-2 and OMS-3 of TA-OMS is operating on the intersection of both consortia and could benefit from the cross-community interaction. IUC16 is tailored to serve material-specific demands in the abstract methodologies opted for within MaRDI (measure 2.4; TA4 Cooperation with other disciplines). It could prove as a good starting point to establish an ongoing interaction of the two consortia: While NFDI-MatWerk use cases will serve MaRDI as relevant applications, **MaRDI**’s unifying representation of data and modelling will contribute to NFDI-MatWerk's goal of establishing a knowledge graph for MSE data.

3.3 International Networking

Through the connections of the individual researchers, the NFDI-MatWerk consortium is well embedded in international MSE community (e.g. USA, Japan, China, Korea) on the one hand, and in research data initiatives on the other hand. Hence, it is ensured that the NFDI-MatWerk activities will not be insular but tightly integrated with above-mentioned and future research data management activities in MSE:

The Research Data Alliance (RDA) (RDA, 2020) is a worldwide organization with the vision that researchers and innovators openly share data across technologies, disciplines, and countries. Scientists of NFDI-MatWerk are strongly involved in RDA in the governance bodies (2013 - today: Rainer Stotzka (KIT), elected member of the technical advisory board) and as co-chairs of working and interst groups, e.g. WG Research Data Repository Interoperability, IG Data Fabric, to harmonize research data management technologies and policies. The IG Research Data Management in Engineering (RDM4Eng, 2020) aims to connect scientific and industrial stakeholders from all relevant sectors to discuss their legal and technological challenges around Research Data Management (RDM) practices, and the provision of a forum for exchanging knowledge, options, and experiences on a national and international level. Researchers from RWTH and KIT are among the initial members of the IG and will thus will represent the NFDI-MatWerk consortium within the international activities in the RDA.

Cooperation with EOSC (EOSC Portal, 2020) is established via KIT, who is a partner in EOSC-pilot, EOSC-hub, EOSC-secretariat, EOSC-Pillar and EOSC-synergy.

The European Materials Modelling Council (EMMC, 2020) is an initiative to bring materials modelling closer to the demands of industry. Due to the connection of Adham Hashibon from Fraunhofer IWM to the operational management board of EMMC, the NFDI-MatWerk consortium will benefit from the international activities and networks. Furthermore, Adham Hashibon is part of the European Materials Modelling Ontology (EMMO) developer team.

The European Materials Characterization Council (EMCC) (EMMC, 2019) is the equivalent initiative on the experimental materials characterization side. Christoph Eberl is part of the Organisational Management Board as well as head of the Working Group 3 ‘Characterisation Data and Information Management’.

The International Data Space Association (IDSA, 2020) network defined a reference architecture for creating and operating virtual data spaces. The architecture is based on common models facilitating secure exchange and easy linkage of data within business ecosystems. This can be accessed and collaborations initiated through the Fraunhofer society by Christoph Eberl.

The German Materials Society (DGM) is currently establishing a worldwide collaboration between the American Society of Materials (ASM). This effort is managed by Frank Mücklich, currently president of the German Materials Society (DGM). He is also fellow of the American Society of Materials (ASM). The NFDI-MatWerk will profit from this relation on a societal level by sharing experience on change processes and communication.

The European project AARC2 (AARC, 2020) offers a Blueprint Architecture (BPA) for access management solutions for international research collaborations. This can be used as a starting point for implementing the Authentication & Authorization Infrastructure (AAI) for federated data access in the NFDI-MatWerk.

NFFA EUROPE (NFFA, 2020) in Horizon 2020 sets out a platform to carry out comprehensive projects for multidisciplinary research at the nanoscale extending from synthesis to nano-characterization to theory and numerical simulation. It includes an integrated Information and Data Management Repository Platform (IDRP) covering the full research cycle by the users. The results and gained experiences building up the platform-wide data model metadata management services will be integrated in the NFDI-MatWerk, whenever suitable.

EUDAT CDI (EUDAT, 2020) offers services, support and best practices for data management and training for various research communities. Service instances for data hosting, registration, management, and sharing can be reused and adapted for the requirements of NFDI-MatWerk. Currently, Achim Streit (KIT) is the chair of the EUDAT CDI Executive Board.

3.4 Organisational Structure and Viability

The consortium NFDI-MatWerk has proceeded to lay out its organizational structure. Generally, the consortium is represented by the General Assembly, which consists of the consortium’s (co-)applicant spokespersons and participants.

The General Assembly decides on the overall strategy of the NFDI-MatWerk. Furthermore, it prioritizes measures of the TAs based on Infrastructure Use-Cases (IUCs) and other strategic aims. As representatives for their institutions, members of the General Assembly are required to act and decide for the MSE-community. This appears fair to argue, since at the moment, NFDI-MatWerk's (co-)spokespersons and participants jointly represent institutions with more than 80% of DFG funding in Materials Science and Engineering 2018 (DFG, 2018).

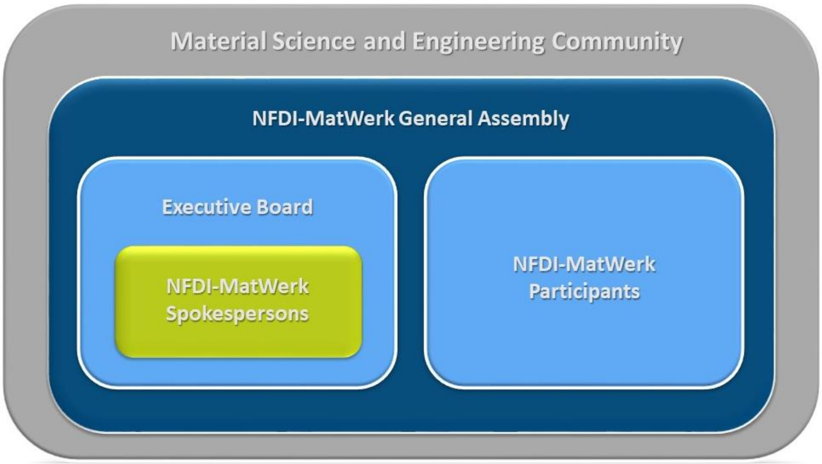


Figure 5. *Schematic overview of NFDI-MatWerk's decision-making entities.*

The General Assembly's meetings are tied to the three main meetings of NFDI-MatWerk, which take place three times a year (see figure 6 on p. 45). If required the Executive Board can summon in between.

The **Executive Board consists of the (Co-)Applicants** with all NFDI-MatWerk spokespersons. NFDI-MatWerk spokesperson is responsible for organizing and moderating meetings and the decision processes. The co-spokespersons are responsible to organize and moderate meetings and decision processes within their specific TAs. Decisions within TAs affecting other TAs need to be handled by the Executive Board. The TA (co-)spokespersons are responsible to develop the TA-specific strategy, matching the needs of the community and the overall strategy of the NFDI-MatWerk. The Executive Board meets monthly.

The NFDI-MatWerk Office will consist of a Permanent Office managed by TA-Strategy Development and additional TA Offices.

The Permanent Office is managed by the Applicant Institution and is the single point of contact for the MSE community as well as for the greater NFDI and e.g. international partners. The TA Offices will manage TA relevant processes and will be run by TA Coordinators which aid the TA Speakers in their tasks of e.g. managing the specific TA, developing the TAs strategy, controlling the strategic measures and work packages, overseeing Infrastructure Use-Cases and community interactions (together with the TA-CI).

Decision Making Processes and Community Interaction

Our decision-making process is based on a tight interaction with the NFDI-MatWerk community. To ensure a due process for enabling this interaction, NFDI-MatWerk will follow a clear schedule beyond our numerous parallel activities.

NFDI-MatWerk's governance process starts with a period of gathering community input and requirements (e.g. PPs, as mentioned in sect. 3.1.3, Annual Conference, further workshops, events, hackathons...). The resulting input is structured and then translated into our Infrastructure Use Cases (IUCs) which are sected at the strategy meeting by the General Assembly. This decision will be based on various considerations, including (a) the TAs evaluation, (b) the community's priorities and (c) the available resources. The TAs then tune their measures and work packages to the selected IUCs. For this proposal NFDI-MatWerk has conducted this process up to this point for a flying start. Next, at the All-Hands-On meeting, all working groups present their Measures and Work Packages to initiate operative implementation. Based on the resulting work program specific goals, user stories, backlogs, timelines and DoDs (definitions of done) are developed till the end of the year. Accordingly, NFDI-MatWerk's operations are based on time-boxed agile management methods with typically four-month-cycles. The strategy process parallelly starts anew.

Throughout the project runtime, NFDI-MatWerk aims at a growing number of consortium members to ensure a broad representation of university and non-university institutions. The admission of new members (representing their respective institutions) is decided by simple majority at the annual strategy meeting. All members become important multipliers.

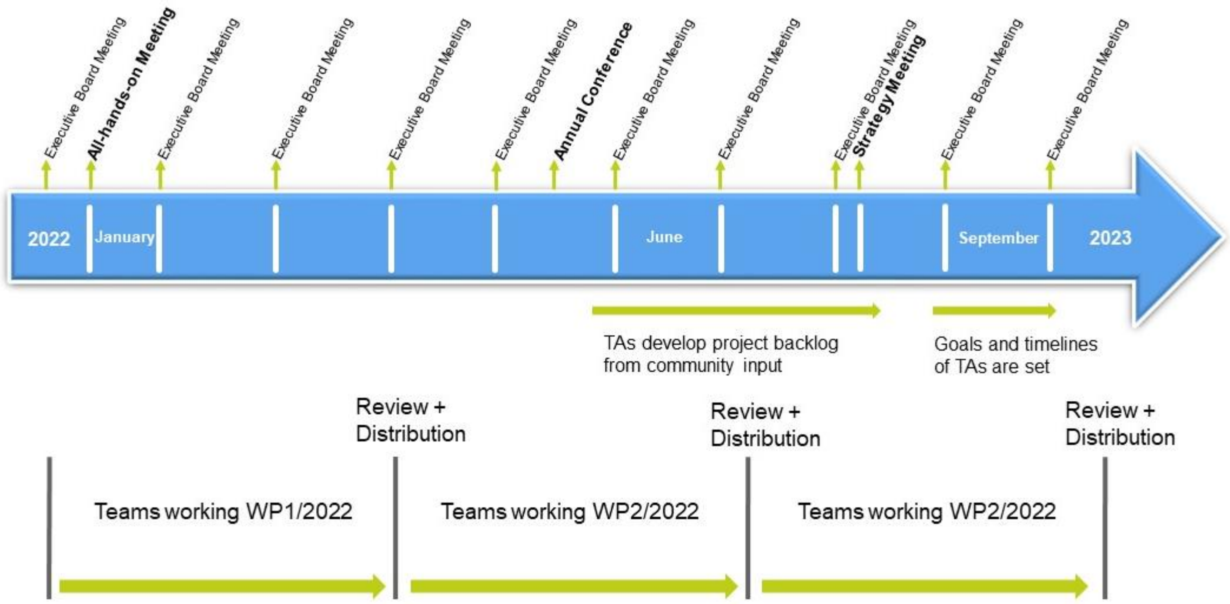


Figure 6*. The timelines for strategy development, decision making and community interaction follow a yearly cycle, while technological development is, e.g., on 4-month cycles.*

3.4.1 Legal Setup of the Consortium

This consortium **refrained consciously from constituting itself into a legal entity**. Accordingly, NFDI-MatWerk is currently a mere personal coalition among its contributors. This open approach has two principal reasons. Firstly, **it was the declared outcome of the consultation** within the seminar that the DFG conducted on governance on August 30th, 2019 in Bonn, that the NFDI should rather eventually become its own legal body, preferably an association, with both the federal and state governments as its decision-makers (Schauhoff, S. & Kirchhain, C., 2019). According to the then present experts in the field of tax law, responsibles from the side of the DFG and to our own experience, a **legal constitution would require an additional massive amount of resources** stemming from the complexity of coordination and red tape. It should be added, that to our best knowledge most consortia that received funding in the first round of proposals also refrained from a legal institution and instead formally organised their collaboration via cooperation agreements. Accordingly, we see no benefit in a legal setup at the current state of proceedings, which leads us to the second reason to decide against a formal setup.

3.4.2 Legal Integration of Key Stakeholders

The Fraunhofer IWM applies for a grant for all (co-)applicants and some of the participants within the NFDI-MatWerk's proposal. The applicant is responsible for **requesting the grant and for immediately forwarding the provided funds/grant** to the other partners in accordance with the respective partial expenditure plans specified in the grant notification and based on the reviewed proposal. According to this consortium's information, **only one grant decision will be issued**. Details of the call for funds will be given in the notification of the award. Thus, regulations on the contractual modalities **must be contained in a contract** between applicant and co-applicant, which regulates the cooperation within the consortium NFDI-MatWerk fundamentally. Due to the legal uncertainty in context with the NFDI's overall process and setup we add the possibility for every (co-)applicant to opt out of this process if further legal consolidation should lead to unfavorable outcomes.

The collaboration contract (based on the template by the NFDI directorate) will specify the legal interaction between the applicant. co-applicants and participants.

3.4.3 Community Empowerment

The consortium NFDI-MatWerk **aims to maximize the empowerment of our community** while at the same time ensuring high-quality results and stringent proceedings. This claim is complex to meet. Obviously, the formal project responsibility stays with the (Co-)Applicants and further contributors receiving direct funding. At the same time, however, **previous approaches lacked the necessary community commitment** in our view, a pitfall that we want to avoid.

It is an ongoing discussion of how to avoid the potential source of ambiguity when intending to **separate formal from factual decision-making**, or accountability from power. We proceed by ensuring a strategic base funding for any of the identified Task Areas that provides them with the chance to focus both personnel and scientific capacities on NFDI-MatWerk. This answers the question about the working nucleus, the interconnecting units that stay at the center of the project and enables continuous coherence in our efforts. At the same time, we set up a Use-Case oriented approach for giving our community the chance to not only contribute their projects but also decide which ones to work on in the first place. Theoretically, this construct will allow our Task Areas to become the necessary expertise providers for NFDI-MatWerk that the community can subsequently rely on for general oversight while Use Cases naturally remain more specific. However, the decision, which Use Cases to select for NFDI-MatWerk, independently of it being for temporal, occupational or financial commitment, must be answered by a community-wide mechanism through maximum possible bottom up democracy. Also, the RfII has suggested that the consortia of the NFDI need to stay agile and accompany and support their own community throughout the digital transformation (Gehring, 2019).

Generally, who controls the use cases will control the focus of our infrastructure in the mid-term time frame. Thus, for us, the use cases play the central role when we ask for community feedback. We suggest that the integration of new Use Cases starts with their representatives' address to our community within the Annual NFDI-MatWerk Conference. In the following months, the TAs have to analyze the different proposals or suggestions they received, in the light of their feasibility, reasonableness, contribution to the infrastructure and demand of resources. With the readily analyzed Use Cases at hand, the TAs then proceed to share their views with the General Assembly during the Strategy Meeting. Per TA, a poll among our community representatives takes place subsequently that decides about the choice of Use Cases admitted into each TAs working program. The final number of selected Use Cases then results from the amount of deferred project funds for Use Case integration per TA, for as long as funds are available.

Having the Use Cases both harmonized and selected, requests for funding via rededication (if needed) is initiated with the DFG. Here, this consortium's exact proceedings rely on a process for a potential rededication of funding, that the DFG is yet to suggest and develop at the time of this writing.

Insight: Suggestions towards a Community-rooted Governance for NFDI-MatWerk
We will implement a clear governance mechanism that considers these different aspects and allows to provide the necessary agility. (1) New participants can join the Consortium by application. The General Assembly will decide on the basis of a simple majority. (2) A governance process will allow to change either the work packages in use-cases or exchange use-cases if the community enforces a different prioritization. A reallocation of funding within the same use-case can be done without a direct interaction with the DFG. In case of the change of use-case and the necessity to reallocate funding to different participants, the responsible (Co-) Applicant will apply for a reallocation of funds through the DFG. While such a mechanism carries definite eligibility for another potential NFDI-wide crosscutting topic, with other consortia having to tackle the same questions, **we proceed to put this mechanism into practice before we can rely on a central solution**. However, we **stay open** and even **ask for general approaches** to community empowerment.

3.4.4 Personnel Recruitment Strategy

As pointed out during last year's review procedure, the ability to attract a work force with the specific skill set required by NFID-MatWerk is crucial for the success of this initiative. NFDI-MatWerk is in a unique position to have multiple direct recruiting channels. The spokespersons lead or are directly involved in several master's programs that educate computational engineers in materials science (CE, 2020) or teach computational skills, software engineering and programming to materials scientists (TUBAF, 2020; ICAMS, 2020; FAU, 2020). Furthermore, the applicants are also involved in post-graduate and continuing education regarding Information and Data Science, e.g. through the Helmholtz Information and Data Science Academy (HIDA, 2020), GridKA (SCC, 2019) or data.RWTH (RWTH-DL, 2020). Germany's famed dual education system also includes apprenticeships for programmers or system administrators that take place at the computer centres involved in NFDI-MatWerk. E.g., RWTH Aachen trains DataScience-MATSE as a specialization of the "Mathematical-technical Software Developer" (ITC, 2020) to educate young professionals in programming and mathematical skills and especially focusses on data science, data management and curation.

3.5 Operating Model

The operating model behind the NFDI-MatWerk is derived from community driven non-profit organizations (NPO) **even though the consortium itself is no legal entity**. While this might be an ambiguous claim in terms of legal aspects, it helps to **describe the intention behind our motivation,** which is the main driver for the (co-)applicants and participants.

Priority of this project are both the **user experience** and **usefulness**, in which the users are namely our fellow colleagues. NPO Project management aims to maximize the impact by continuously reviewing and improving performance through the **tight interaction with the community**. Here our goal is to meet the community demands with the available capacity within the NFDI-MatWerk and the related use-cases. Our **consortium** and **use-cases** are **expected to change continuously over time**, so that necessary competences can be added as needed or requested by the community. Furthermore, the NFDI-MatWerk, and especially the TA-SD, is responsible to track and adapt the project to the change process within the Digital Transformation of MSE. Therefore, the processes on how the goals of the NFDI-MatWerk are set and decisions made will be continuously developed, together with the materials community as described in section 3.4.

It is in the sense of an NPO operating model that its contribution is supposed to add value to something more fundamental, being it services in the area of e.g. ecological, humanitarian or animal well-being. We claim this to be the case to our interpretation of the NFDI-MatWerk operating model, equally: The change process that our consortium's infrastructure wants to trigger and support is way more fundamental than its contribution can be, especially in initial project phases. We accept this notion as the overall vision, but see the start for any such necessary change in creating the first working environment as its initial condition, while no one can say where the way might go subsequently and in the long term.

Furthermore, just as it is the case in an NPO sense, our operations rely on the community's involvement and its readiness to contribute. Our operating model thus must be centrally evolving around fellow MSE researcher's adaptiveness and inclusion into our framework. It is of maximum relevance to us to create according processes that ensure such smooth operating integration, as it can be seen in shapter 4 by the manifold capacities that the TAs are ready to commit to this challenge.

4 Research Data Management Strategy

The consortium NFDI-MatWerk will follow the FAIR data principles. FAIR data meets standards of findability, accessibility, interoperability, and reusability according to the FAIR Principles (Wilkinson et al., 2016) within our approach (cf. sect. 4.3). We intend to closely follow international metadata standards and will interact with the international governing bodies (e.g. RDA, developments in the EOSC, ASM initiatives on metadata, EMMC and EMMO). The international relations in that respect are described in section 3.3. NFDI-MatWerk is envisioning to implement control mechanisms, monitoring features, data protection plans and policies, which will be embedded into our delocalized data infrastructure. To this end, the NFDI consortium relies on the long experience of the FIZ as well as the two computing centers from KIT and RWTH. Nevertheless, it is our firm believe that it will be worthwhile to go well beyond metadata standards and to develop a materials' ontology. This view has been shared across many NFDI consortia.

Materials ontology as unifying solution

The **heterogeneity of data** is a central characteristic in MSE that **makes data analytics rather difficult**. Given the additional hierarchical structure of materials data, information and knowledge as well as their strong inter-dependencies, the envisioned data space fabric needs to be based on an adaptive ontology/strong semantics approach to connect the various data types, formats and database sites that are commonly used in MSE (e.g. data from experiment, simulations, various length and time scales). The metadata within this connecting structure will not only contain materials-specific metadata, but also information on the used data types and formats.

Typical data types are (tabulated) time series from a vast variety of sensors and machines as well as simulation codes; furthermore images, image stacks (3D tomography), image time series, metadata on objects, processes, environmental information, infrastructure information, text in various formats.

The amount of data produced by a single measurement device or theoretical simulation can be enormous, but often only a small fraction of the total data might be required for the joint and comprehensive representation for a material at hand. Secondly, data sources that have to be maintained are **distributed** over many labs in various locations. Due to these technical challenges, it will be more economical to bring the computational workflows to the data than operating vice versa on a huge shared data server.

The materials data processing is usually individualized, due to the heterogeneous materials data and depending on the scientific question driving the research. Typical workflows in MSE are based on a large variety of tools ranging from standard software tools (e.g. Excel, Origin) to languages also used for scripting (e.g. Python, Matlab) and computationally highly efficient codes (e.g. Fortran, C, C++) running on high-performance supercomputer centers.

Therefore, at the very core of our approach we envisage a **Materials Ontology** with a dedicated Task Area (TA-OMS, sect. 5.4) devoted to this challenge. The ontology-based concept **graph database infrastructures** (TA-MDI, sect. 5.2) can be used as a representation of data collections providing high performant, complex search patterns over distributed data sources. The APIs (application programming interfaces) available for most of these databases allow scientists to easily automate the access to such data structures. An example would be that data search is implemented into experimental software, so that scientists can test their samples and have them as on overlay of real time accessed external data. More generally, a close interrelation of **experimental and computational workflows** with the materials ontology is foreseen for data analysis as well as data generation (TA-WSD, sect. 5.3). With regards to **materials science**, this ontology will not only be able to characterize the microstructural, chemical or atomic state of a material, it will furthermore be able to capture the time-dependent behavior of evolving microstructures along with the respective properties.

4.1 State of the Art and Needs Analysis

There are multiple methods for conducting an analysis of both our target group's state of the art and of requirements in context of the digital transformation of MSE. We acknowledge the inherent biases of each approach and evaluate various principle sources of information to determine the status of our community to the best of our knowledge:

1. quantitative surveys (as shown in sect. 4.1.1)
2. qualitative expert interviews (e.g. resulting in the PPs described in sect. 3.1.3 and translated into Infrastructure Use Cases in sect. 4.1.2),
3. input from conferences, workshops, hackathons, personal networks and more (as planned in NFDI-MatWerk's work program).

We apply the resulting community needs to form our comprehensive vision of tomorrow's Digitalized Workflow as our overall perspective (in sect. 4.1.3.). It is the sum of both our analyses and their combination with state-of-the-art technological possibilities. In the following two sections we show examples for a survey conducted in July 2020 (sect. 4.1.1), as well as expert interviews resulting in 17 Infrastructure Use Cases (IUCs, sect. 4.1.2).

4.1.1 MSE-Needs Analysis (Quantitative)

The MSE community has expressed its research data infrastructure needs that form the basis of the NFDI-MatWerk Task Areas. In July 2020, 536 participants replied to a Germany-wide survey about their requirements in the field of digitization. The survey went on to query the most pressuring needs more precisely, as expressed by specific sub-topics of digitalization (s. Table 4.1.1.1). As expected, the respondents show a broad variety of knowledge of research data management that also supports the NFDI-MatWerks mission in training and education.

Most urgent needs - Replies to NFDI-MatWerks July 2020 community survey						
"Please indicate on a scale from 1 (lowest) to 7 (highest) in what area of Digitalization of the MSE you see the most urgent needs for unified solutions"						
Infrastructure		**Sum of Replies**	**Average Approval (1-7)**	**Refusal**	**Neutral**	**Approval**
#1	Base Infrastructure	220	5.2	18.2%	11.8%	70.0%
#2	Hardware Interfaces	223	4.7	22.9%	15.2%	61.9%
#3	Datacenter Access	213	5.0	18.8%	11.7%	69.5%
Data Creation		**Sum of Replies**	**Average Approval (1-7)**	**Refusal**	**Neutral**	**Approval**
#4	Measuring Device Interfaces	244	5.7	9.8%	7.8%	82.4%
#5	Documenting Context Information	244	6.0	6.1%	4.9%	88.9%
#6	SI-Units	255	6.2	8.2%	3.9%	87.8%
#7	Digital Twin per Specimen	183	3.0	70.5%	12.0%	17.5%
#8	Harmonized Raw Data	242	5.6	9.9%	5.8%	84.3%
#9	Data Bases	218	4.8	24.8%	11.0%	64.2%
Data Usage		**Sum of Replies**	**Average Approval (1-7)**	**Refusal**	**Neutral**	**Approval**
#10	Barrier-free Reading Access	217	4.3	35.9%	11.5%	52.5%
#11	IT-secure Access Protocols	207	5.6	8.2%	10.1%	81.6%
#12	Multi-Scale Simulation	195	5.3	12.3%	11.8%	75.9%
#13	Workflow Management Software	235	5.6	11.9%	8.5%	79.6%
#14	Workflow Software Transparency	217	4.5	35.0%	13.4%	51.6%
#15	Workflow Management Software in Python	168	5.2	16.1%	16.1%	67.9%
#16	GUI-Compatibility	229	5.5	14.4%	7.9%	77.7%
#17	Interoperable Data Evaluation	239	5.9	9.2%	3.8%	87.0%
#18	Machine Learning	229	5.6	12.2%	5.7%	82.1%

Table 4.1.1.1*. Results of the community survey*

In summary, the participants generally agreed that (1) implementing the FAIR-principles is a central objective (78%), (2) unified context data formats to document the history of specimens for both experiment and simulation are needed (89%), (3) harmonization of units towards SI-standards (88%) and raw data (84%) are required, and (4) the setup of interoperable evaluation scripts for applicability on data from different sources is important (87%).

For NFDI-MatWerk, it is an ongoing transfer effort to translate this comprehensive and multilayered feedback in formulating its products and services. The reported needs were also the starting point when consortium members went on to moderate the discussion to define Infrastructure Use Cases together with the community in the scope of the process defined below in sect. 4.1.2. Within the final list of our IUCs the impact of this feedback can therefore still be identified. Here, first, the common agreement that we need to define context information (#5) is centrally represented in the idea of the

MSE knowledge-graph within the DME, itself. NFDI-MatWerk thoroughly commits to this aim with an entire TA dedicated to ontology development. This TA will also include the harmonization of units (#6) and therefore the foundation to create interoperable evaluation scripts (#17) as well as unified workflow management (#13). This is the condition for a barrier-free reading access (#10).

Aspects of the survey were incorporated into the IUCs for building a “framework for curation and distribution of reference datasets” (IUC02), “interoperability of workflow systems” (IUC10), “development of coupled ontologies and workflows for thermochemical treatment” (IUC11), and the “alignment of application- and higher-level ontologies” (IUC12). Furthermore, the need for measuring-device interfaces (#4) as well as the harmonization of raw data (#8) can be discovered in “digital infrastructure and workflows for labs” (IUC05). Further implemented needs involve GUI compatibility (#16) and Workflow Software Transparency (#14) as represented in IUC08 “Interactive visual exploration for analyzing correlations in high dimensional materials data spaces”, harmonized databases (#9) as represented in IUC03 “Storage concepts for large hierarchical datasets”, and Multi-Scale-Simulation (#12) as represented in IUC07 “Beyond 3D: Tools for spatiotemporal microstructure studies” and in IUC15 “Method- and scale-bridging workflows and data structures for tomography”.

4.1.2 Community-driven Infrastructure Use Case Selection and Results

Infrastructure requirements resulting from various research scenarios (e.g., PPs) are collected and specified in the form of Infrastructure Use Cases (IUCs). For this purpose, PPs and other research scenarios are analyzed as community inputs regarding their individual data management challenges. The resulting specification of the technological requirements are compared and related to the relevant measures in all task areas. The consortium will continuously broaden the scope of the participant projects (or PPs) and therefore has adopted a democratic and user-oriented selection process:

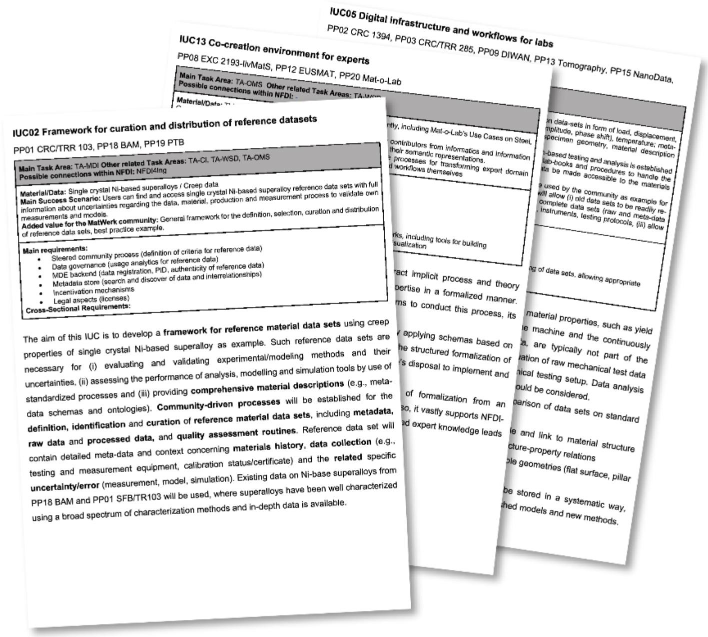


***Figure 7**. Infrastructure Use Cases (IUCs) as prepared by NFDI-MatWerk*

1. The selected Participant Projects (PPs) and related Infrastructure Use Cases (IUCs) are presented online and can be commented directly by the community.
2. At the Annual Conference of the consortium in June, the TAs present the status of the Infrastructure in a generally accessible way. At the meeting, the "Open Space" method will help to create an open environment for everyone to readily exchange ideas and bring in own

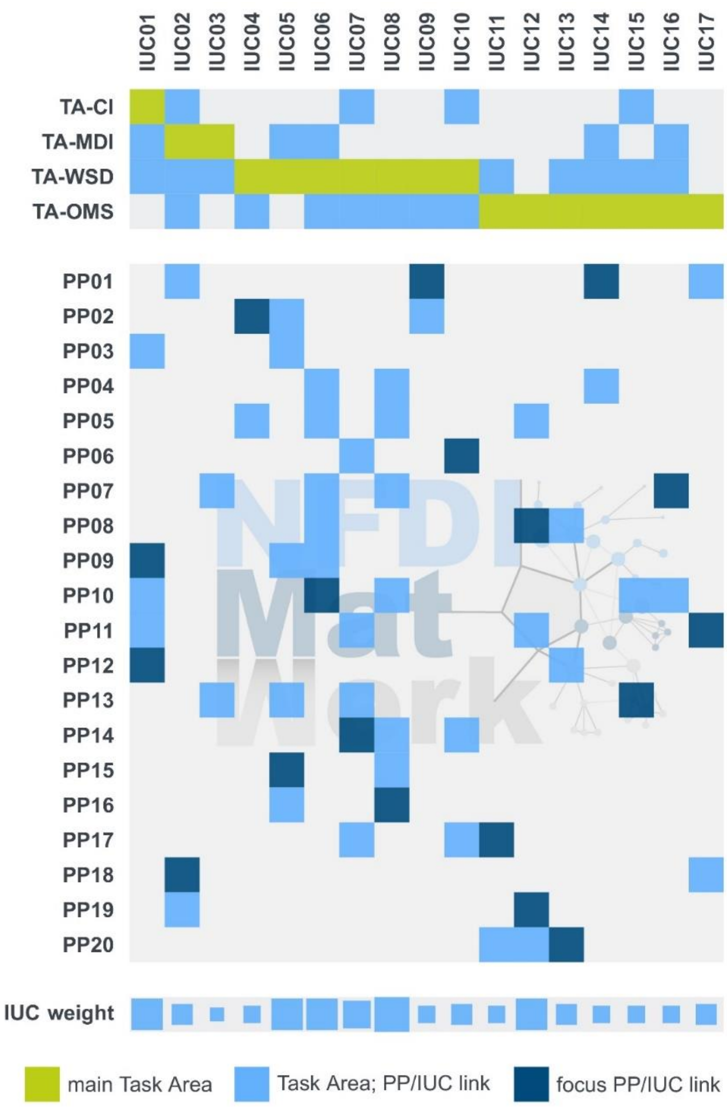


***Figure 8**. Matrix plot of Participant Projects (PPs) and Infrastructure Use Cases (IUCs) emphasizing their dependency. The visualization represents the result of the conducted expert interview process.*

Participant Projects (PPs). Internet-based technologies allow a broad participation for jointly building a consensus on the community's requirements and their priority.

3. In preparation for the strategy meeting, and based on the prioritization of the Infrastructure Use Cases (IUCs) by the community, the TAs will estimate the required resources for their implementation. Furthermore, technical dependencies are analyzed and coordinated among the TAs (s. fig. 7).
4. At the strategy meeting, the TAs present the collected IUCs together with the necessary resources and technical dependencies. The General Assembly decides on the prioritization of the IUCs based on (a) the TAs evaluation, (b) the community's priorities and (c) the available resources.
5. Based on this decision, the TAs work out the resources dedicated to the measures and time schedule as well as concrete milestones.
6. In preparation of the following All-Hands-On meeting at the beginning of the year, the consortium presents the developed working plans internally to manage their joint work for the upcoming period. This includes methods from agile management.

Infrastructure Use Cases (IUCs) as valid tests in the form of "queries" to the NFDI-MatWerk infrastructure: The different IUCs provide research-relevant usage profiles ("queries") with real data and workflows that allows realistic testing during and at the end of development periods. On this basis, NFDI-MatWerk implements (a) Quality assurance for implementation and deployment, (b) a constant alignment of all activities with overarching goals and (c) a clear future perspective for the next development steps of the research data infrastructure.

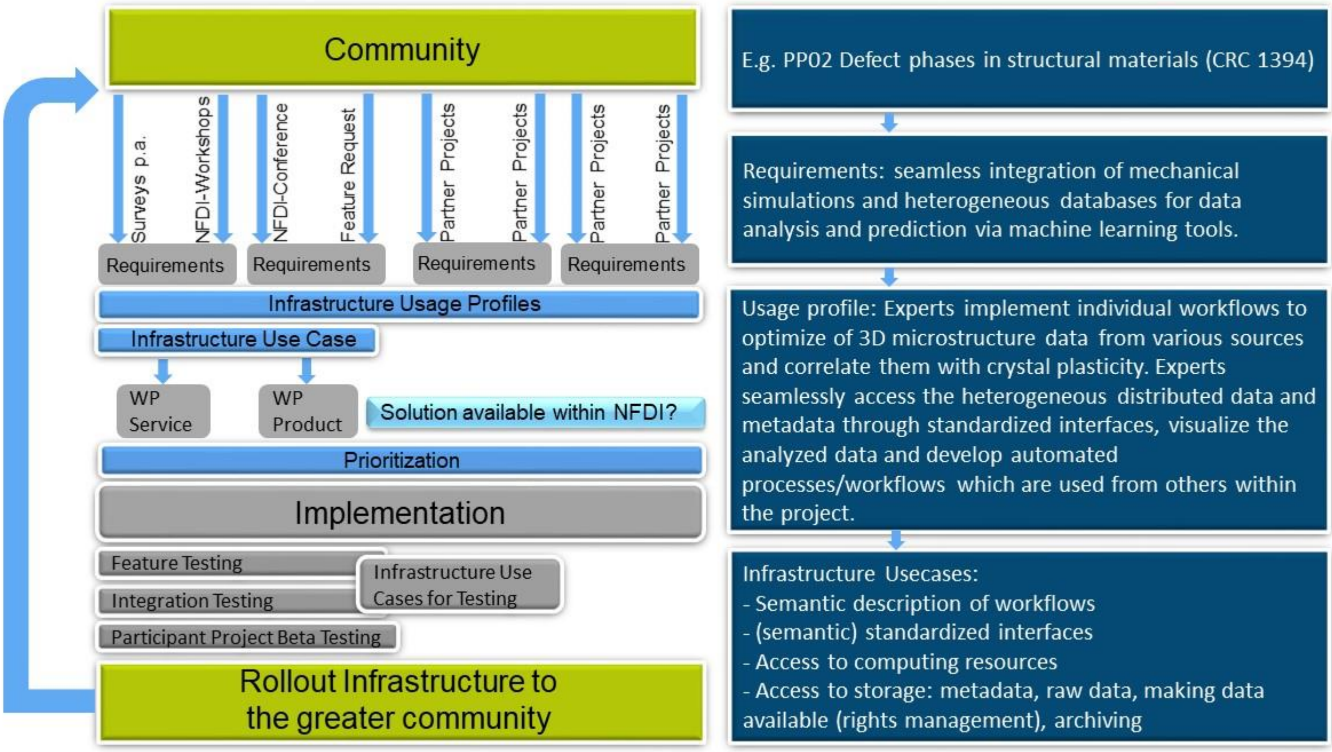


***Figure 9**. The flow chart shows the process how Infrastructure Use Cases (IUCs) are derived from community input. Furthermore, IUCs allow validating infrastructure feature implementations as well as their deployment.*

Overview of NFDI-MatWerk's resulting Infrastructure Use Cases

IUC01 Web-based demonstration and teaching framework for MSE research data infrastructure

Participant Projects: PP03 CRC/TRR 285, PP09 DIWAN, PP10 IRTG2078,PP11 ERC MuDiLingo,PP12 EUSMAT **Main Task Area:** TA-CI **Other related Task Areas:** TA-MDI, TA-WSD **Possible connections within NFDI:** NFDI4Ing, FAIRmat
Material/Data: Teaching **Main Success Scenario:** Students can experience the the full functionality of the DME and learn to use workflows with pre-existing datasets and ontologies. **Added value for the MatWerk community:** teaching personnel can tailor the demo framework as well as datasets and workflows to match their study courses.
Main requirements: Support for the development of academic teaching materials (TA-CI), DME (TA-MDI), Validated and well-documented workflows (TA-CI)

IUC02 Framework for curation and distribution of reference datasets

Participant Projects: PP01 CRC/TRR 103, PP18 BAM, PP19 PTB **Main Task Area:** TA-MDI **Other related Task Areas:** TA-CI, TA-WSD, TA-OMS **Possible connections within NFDI:** NFDI4Ing

Material/Data: Single crystal Ni-based superalloys / Creep data

Main Success Scenario: Users can find and access single crystal Ni-based superalloy reference data sets with full information about uncertainties regarding the data, material, production and measurement process to validate own measurements and models.

Added value for the MatWerk community: General framework for the definition, selection, curation and distribution of reference data sets, best practice example.

Main requirements: Steered community process (definition of criteria for reference data), Data governance (usage analytics for reference data), MDE backend (data registration, PID, authenticity of reference data), Metadata store (search and discover of data and interrelationships), Incentivation mechanisms, Legal aspects (licenses)

IUC03 Storage concepts for large hierarchical datasets

Participant Projects: PP07 EXC2075-PUMD, PP13 Tomography

Main Task Area: TA-MDI **Other related Task Areas:** TA-WSD, TA-Strategy
Possible connections within NFDI: NFDI4Ing, FAIRmat

Material/Data: (mainly) metallic materials/tomographic data and related simulations with focus on large datasets

Main Success Scenario: Large amounts data from various, scale-bridging tomography methods of a given sample as well as simulations can be fused, specific volume elements can be accessed at different scales.

Added value for the MatWerk community: An exemplary implementation of a storage concept for large fused and linked datasets which is accessible through the DME and a web-based interface.

Main requirements: Digital Materials Environment (DME) Backend (MDI), Materials Data Infrastructure Support (MDI), Connection of workflows and ontologies (WSD)

IUC04 Model driven data space exploration

Participant Projects: PP02 CRC 1394, PP05 CRC/TRR 270

Mainly related to Task Areas: TA-WSD **Other related Task Areas:** TA-OMS
Possible connections within NFDI: NFDI4Ing, FAIRmat

Material/Data: Mg-Al-Ca alloys / Thermodynamic and structural data of defects

Main Success Scenario: User can automatically generate defect phase diagrams that can be used to predict the performance of materials

Added value for the MatWerk community: General framework of model-driven thermodynamic databases that combine computation and experiment

Main requirements: Workflows combining theoretical and experimental structural data of defects, Adaptive databases for high dimensional data structures containing sparse data, Visualization (of thermodynamic dataspace from micro-structure, chemical or mechanical perspective), Multiscale simulation (of multiphysics data), Electronic lab book (for defect data)

IUC05 Digital infrastructure and workflows for labs

Participant Projects: PP02 CRC 1394, PP03 CRC/TRR 285, PP09 DIWAN, PP13 Tomography, PP15 NanoData, PP16 TUK-AME

Main Task Area: TA-WSD **Other related Task Areas:** TA-MDI
Possible connections within NFDI: NFDI4Chem

Material/Data: Mechanial characterization workflows - nanoindentation data-sets in form of load, displacement, time, possibly extended to dynamic response (load and displacement amplitude, phase shift), temperature; meta-data associated with instrument specifications and calibration files, specimen geometry, material description (possibly based on material/microstructure ontology)

Main Success Scenario: (i) a data base and workflow for nanoindentation-based testing and analysis is established for appropriate comparison and/or re-evaluation of data, (ii) electronic lab-books and procedures to handle the synthesis of materials and history of samples implemented, (iii) the data be made accessible to the materials community.

Added value for the MatWerk community: The implementation can be used by the community as example for many characterization lab workflows and the integration of equipment. It will allow (i) old data sets to be readily re-analyzed with updated algorithms from new knowledge, with access to complete data sets (raw and meta-data previously not of interest), (ii) comparison of data across different groups, instruments, testing protocols, (iii) allow robust statistical analyses, (iv) contribute to updatable material data base

Main requirements: Exemplary machine interface implementations, Storage of data, meta data, based on a materials ontology, Automated analysis workflows, Flexible data labeling system (can be updated, changed) for sorting of data sets, allowing appropriate comparisons of data

Cross-Sectional Requirements: Electronic Lab Books with NFDI4Chem

IUC06 Integrating materials data from experiments and computation (ICME) into Industry 4.0 manufacturing paradigms

Participant Projects: PP04 CRC 1261, PP05 CRC/TRR 270, PP07 EXC 2075-UMD, PP08 EXC 2193-livMatS, PP09 DIWAN, PP10 IRTG 2078

Main Task Area: TA-WSD **Other related Task Areas:** TA-MDI, TA-OMS
Possible connections within NFDI: NFDI4Ing

Material/Data: Polymeric matrix material reinforced with glass and carbon fibers / Computed tomography data, thermo-physical properties on different scales (e.g. stiffness, heat capacity), process data (e.g. molding pressure or temperature), product design data (e.g. part, bead, ribs geometry)

Main Success Scenario: Users can integrate materials data into digital representations of process and manufacturing chains and make the data accessible for the community.

Added value for the MatWerk community: An infrastructural framework coupling and materials data and industry-relevant manufacturing processes with their model representations (digital twins).

Main requirements: Data storage infrastructure and DME platform for data from different levels of the process chain, Workflows combining physical and synthetic data on different levels of the process chain

Cross-Sectional Requirements: Industry 4.0 Standards and Interfaces (NFDI4Ing), Cross-discipline ontology-matching (NFDI4Ing)

IUC07 Beyond 3D: Tools for tracking spatiotemporal microstructure evolution

Participant Projects: PP06 CRC 1368, PP11 ERC MuDiLingo, PP13 Tomography, PP14 ICAMS, PP17 IWT

Main Task Area:TA-WSD **Other related Task Areas:** TA-OMS, TA-CI
Possible connections within NFDI: NFDI4Ing, FAIRmat

Material/Data: Spatiotemporal point, triangle cloud/mesh, and tensorial field data from computer simulations of microstructure evolution at the microscopic and atomic scale

Main Success Scenario: The user can characterize the geometry (and topology) of the crystal defects and describe their motion and interaction in time by flexibly exchanging spatiotemporal data from representative-volume-element (RVE) models, dislocation dynamics codes and force-field/MD codes.

Added value for the MatWerk community: A computational geometry description for crystal defects (point, line, and surface patches) is connected with a corresponding ontology. This enables users to identify crystal defects agnostic of a specific implementation in a simulation code, enables a seamless encoding between the atomic and the microscopic/macroscopic scales and provides access to the study of the time evolution of defects.

Main requirements: , orkflows to store/extract e.g. atomic positions, atom types, material point data (continuum, voxel data) including metadata and contextualization, Workflows to execute existing modeling tools for microstructure evolution (plasticity, recrystallization, grain growth, precipitation, dislocation dynamics, and force-field codes)

Workflows how to use existent MatWerk post-processing tools including metadata and contextualization
Development of ontology for defects, Tools for 3D visualization, Integration in institutional education

Cross-Sectional Requirements: Workflow description

IUC08 Interactive visual exploration for analyzing correlations in high dimensional materials data spaces

Participant Projects: PP04 CRC 1261, PP05 CRC/TRR 270, PP07 EXC 2075-PUMD, PP10 IRTG 2078, PP14 ICAMS, PP15 NanoData, PP16 TUK - AME

Main Task Area: TA-WSD **Other related Task Areas:** TA-OMS
Possible connections within NFDI: NFDI4Ing, NFDI4Chem and DataPlant

Material/Data: Metallic alloys / microstructural features and mechanical as well as physical properties

Main Success Scenario: Users can identify and validate correlations for highly complex, sparse materials data sets and can predict materials properties based on microstructural features.

Added value for the MatWerk community: Development interactive data and topology analysis tools based on a materials knowledge management system. This will be used to support and validate the materials ontology developed within the NFDI4-MatWerk.

Main requirements: Uniform (Meta)data formats to enable a utilization of heterogenous data from different test devices, Metadata scheme to ensure comparable information for different test methods, Standardized units to facilitate data transformation, Electronic lab book for uniform data documentation and management, Integration of measurement devices to simplify utilization of different test devices, Visualization tools for interactive correlation analysis

Cross-Sectional Requirements: Curation software, Collaboration on Interoperability with DataPlant

IUC09 - Infrastructure interfaces with condensed matter physics (collaboration with FAIRmat)

Participant Projects: PP01 CRC/TRR 103, PP02 CRC 1394

Main Task Area: TA-WSD **Other related Task Areas:** TA-OMS
Possible connections within NFDI: FAIRmat

Material/Data: Atom Probe Tomography (APT) for metallic alloys

Main Success Scenario: Through the physical modeling of the measurement process, users get more precise information and a deeper understanding of their experimental results, in particular to distinguish defect states in the microstructure.

Added value for the MatWerk community: Workflows are designed in coordination with FAIRmat such that the exchange of data, metadata, ontologies, and concepts for data curation between two NFDI consortia is ensured without a loss of information. In this way, both consortia can mutually benefit from ongoing improvements in the infrastructure for condensed-matter physics (FAIRmat) and the interpretation of microstructure data (NFDI-MatWerk).

Main requirements: Workflows how to process measured data (e.g. detector hit positions, time-of-flight per ion) into atomic position and atom types (chemical species) including all metadata with contextualization (e.g. sample preparation, electric field) (FAIRmat), Close interaction with device software via open-source parsers and converters, Machine learning/statistical analysis of APT atomic data to detect and interpret defect structures, Interface between formats and queries of (meta)data in FAIRmat and NFDI-MatWerk, Ontology matching between both communities, Workflow combining experimental data, image processing and theoretical models / properties of defects in microstructured materials (NFDI-MatWerk)

Cross-Sectional Requirements: Data format and ontology standards

IUC10 Interoperability of workflow systems (collaboration with NFDI4Ing)

Participant Projects: PP06 CRC 1368, PP14 ICAMS, PP17 IWT

Main Task Area:TA-WSD **Other related Task Areas:** TA-OMS, TA-CI
Possible connections within NFDI: NFDI4Ing

Material/Data: Multiscale simulation of microstructure formation

Main Success Scenario: The user will be able to connect the tools and concepts employed by different communities to complete simulation protocols. Therewith, a multi-scale simulation of microstructures in materials becomes accessible, which connects, e.g. atomistic and phase field simulations.

Added value for the MatWerk community: The purpose of the IUC is to demonstrate the transfer of simulation protocols from one workflow system to another one. This gives the MatWerk community the chance of accessing tools designed for the needs of the whole engineering community captured in NFDI4Ing.

Main requirements: Integrated development environment that allows flexible adaptations, Special workflow formats for multiscale materials simulations, Ontology development at the interface of different communities, Data formats for materials data infrastructure

Cross-Sectional Requirements: Workflow-Ontology interaction

IUC11 Development of coupled ontologies and workflows for thermochemical treatments

Participant Projects: PP17 IWT, PP20 Mat-o-Lab

Main Task Area: TA-OMS **Other related Task Areas:** TA-WSD
Possible connections within NFDI: NFDI4Ing

Material/Data: high strength structural materials / time-temperature-transformation relations to microstructures

Main Success Scenario: Users can utilize a knowledge graph populated by process and material data to predict microstructures and their alteration by thermochemical heat treatments. Using a graph based experimental data structure enables a FAIR datastorage to allow for a new level of reproducibility and comparability of experimental results from different experimental setups and groups.

Added value for the MatWerk community: Concepts for the combined ontologies will be developed, combining thermo-chemical processing and materials transformation. General framework for heat treatment representation and resulting erroneous and intended experimental impact on materials microstructure.

Main requirements: Ontology for representation of heat treatment and transformation in structural materials, Workflow management systems can be adapted during use (e.g., primary data acquisition, assertion augmentation, heterogeneous post-processing), Tools for automated data assembly, Data augmentation and condensation for advanced data handling

Cross-Sectional Requirements: Data standards with NFDI4Ing

IUC12 Alignment of application- and higher-level ontologies

Participant Projects: PP05 CRC/TRR 270, PP08 EXC 2193-livMatS, PP11 ERC MuDiLingo, PP19 PTB, PP20 Mat-o-lab

Main Task Area:TA-OMS **Other related Task Areas**: TA-SD
Possible connections within NFDI: all consortia, in particular NFDI4Ing, FAIRmat, NFDI4Chem

Material/Data: Other ontologies, e.g. EMMO, ontologies from NIST

Main Success Scenario: Users are able to easily connect application and higher level ontologies with general ontologies.

Added value for the MatWerk community: Consistently interface with all neighboring communities and benefit from their expertise.

Main requirements: Existing application domain or general ontologies, Basis ontology advisory, Ontology development support, Software Interfaces for ontology alignment

Cross-Sectional Requirements: Generally necessary ontology frameworks, formalized units, interfaces, ontology visualization

IUC13 Co-creation environment for experts

Participant Projects: PP08 EXC 2193-livMatS, PP12 EUSMAT, PP20 Mat-o-Lab

Main Task Area: TA-OMS **Other related Task Areas:** TA-WSD
Possible connections within NFDI: -

Material/Data: This depends on the selected expert scenarios, currently, including Mat-o-Lab's Use Cases on Steel, Copper and fiber-composites

Main Success Scenario: MSE-domain experts and interdisciplinary contributors from informatics and information science can jointly work on digitizing scientific workflows and develop their semantic representations.

Added value for the MatWerk community: Known and teachable processes for transforming expert domain knowledge into formalized structures as well as the resulting formalized workflows themselves

Main requirements: Basis Ontology advisory, Ontology development support,Multi-scale bridging ontologies, Software Interfaces for accessing workflow data semantically, Implementation of DME for applying the representation

Cross-Sectional Requirements: Generally necessary Ontology frameworks, including tools for building knowledge graphs for non-experts, formalized units, interfaces, ontology visualization

IUC14 Adaptive automated characterization pipelines and meta data schemas for high throughput experiments

Participant Project: PP01 CRC/TR 103

Main Task Area: TA-OMS **Other related Task Areas:** TA-MDI, TA-WSD

Possible connections within NFDI: FAIRmat, NFDI4Chem

Material/Data: Combinatorial thin films synthesis and automatized X-ray diffraction data analysis

Main Success Scenario: A toolkit for the automated handling of all meta data during sample characterization is set up. This allows for a high-throughput analysis of structural data for materials design and materials discovery

Added value for the MatWerk community: Data management for automated experimental characterization techniques; Concepts and software solutions for a continuous adaptation of data management workflows in conjunction with a developing meta data schema

Main requirements: Standardized formats for data and metadata, Flexible storage of large amount of data obtained in high-throughput measurements, Workflow management systems that can be adapted during use, Combination of automated experimental characterization and automated data analysis, Device software, ELN

IUC15 Method- and scale-bridging workflows and data structures for tomography

Participant Projects: PP10 IRTG 2078, PP13 Tomography

Main Task Area: TA-OMS **Other related Task Areas:** TA-CI, TA-WSD
Possible connections within NFDI: NFDI4Ing, FAIRmat

Material/Data: Tomographic data and related mechanical simulations of mainly metallic materials

Main Success Scenario: A description of scale- and method-bridging ontology and workflows has been developed which allows to connect tomographic (meta-)data and related mechanical simulations from a given sample**.**

Added value for the MatWerk community: A general framework for workflows and metadata of materials tomographies and related microstructure features is supplied.

Main requirements: Ontology development support (TA-OMS), Multiscale-bridging ontologies (TA-OMS), Support from professional societies to identify relevant data by involvement of expert committees (TA-CI), Metadata Store (TA-MDI), Graphical programming language (TA-WSD) **Cross-Sectional Requirements:** Common data standards and formats (FAIRMat, NFDI4Ing)

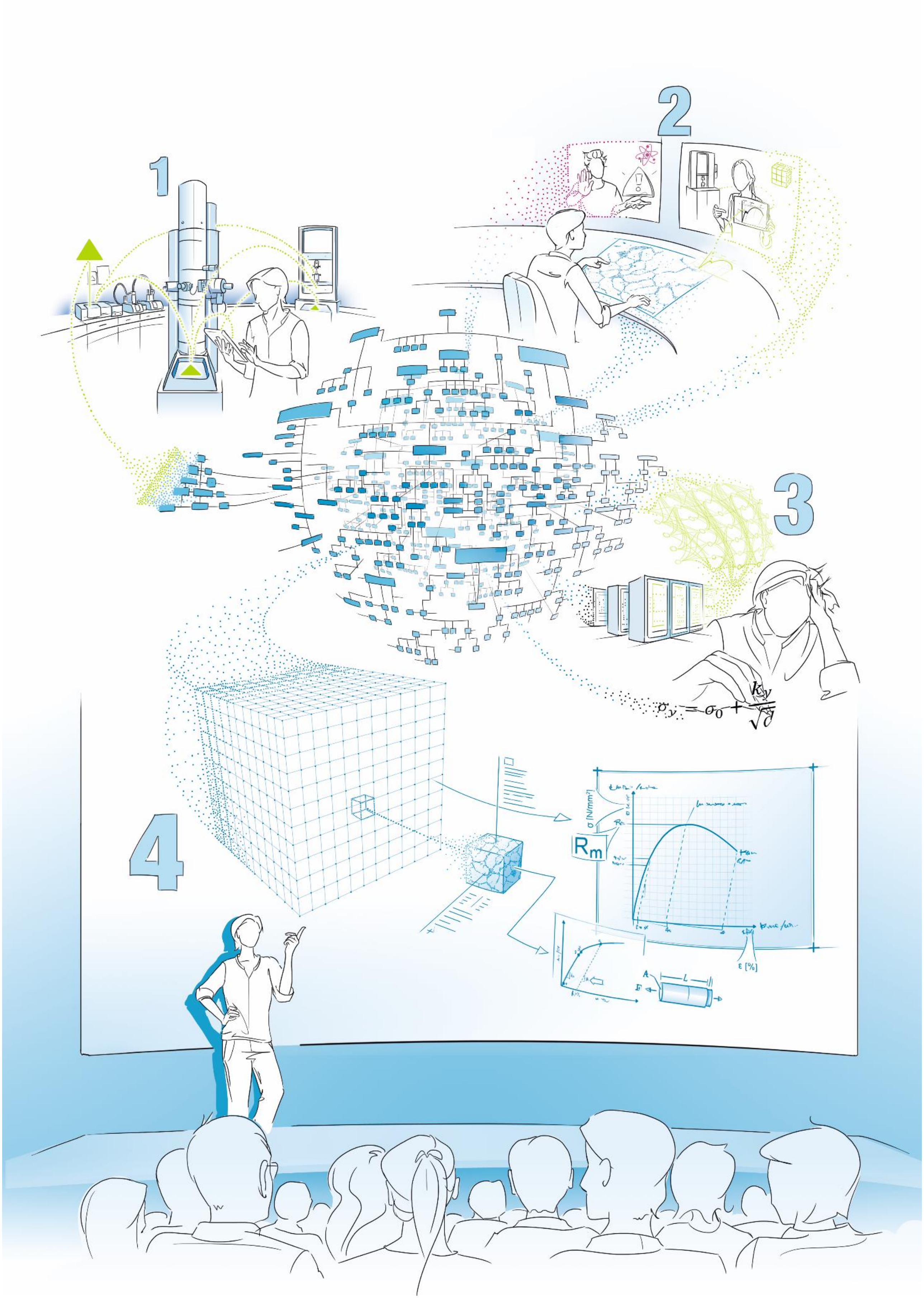
IUC16 Unified ontology for matrix-inclusion microstructure and composites

Participant Projects: PP07 EXC 2075-PUMD, PP10 IRTG 2078 **Main Task Area:** TA-OMS **Other related Task Areas:** TA-MDI, TA-WSD **Possible connections within NFDI:** MaRDI, NFDI4Ing
Material/Data: various materials with matrix-inclusion morphology (containing, e.g., particles, fibers, whiskers) / microstructure data from experiments, simulations and abstract models **Main Success Scenario:** users from different disciplines can query matrix-inclusion morphology information from various sources of data; the main application is the exploitation of the microstructural data in simulations, surrogate modeling and in connecting experimental evidence with models **Added value for the MatWerk community:** The ontology can be used in many applications, both for existing data and newly captured data sets and can be generalized to other materials.
Main requirements: ontology metadata schema, ontology development support, multiscale-bridging ontologies

IUC17 Ontologies for defects in crystals

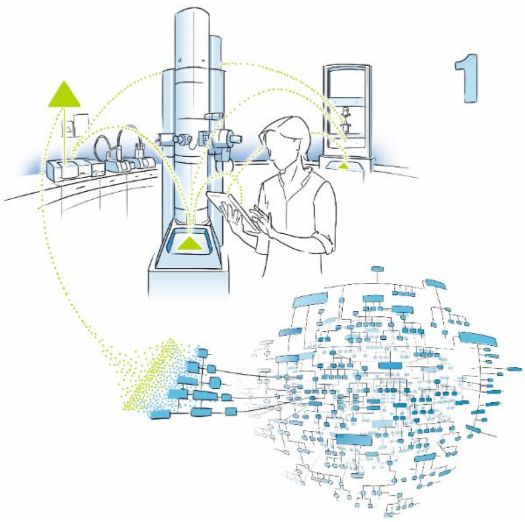
Participant Projects: PP01 CRC/TRR103, PP11 ERC MuDiLingo, PP18 BAM **Main Task Area:** TA-OMS **Other related Task Areas:** TA-WSD, TA-CI **Possible connections within NFDI:** NFDI4Ing
Material/Data: Single- and polycrystalline materials/metals including superalloys and HEAs / Simulation data as well as postprocessed microscopy data **Main Success Scenario:** Users can describe the defect structure with a defined vocabulary and quantitative mathematical representation that is comparable between experiments and transferable to simulations **Added value for the MatWerk community:** A consistent and standardized description of crystalline structures and defects that can easily be extended and adjusted
Main requirements: Metadata store, Software tools for creating and visualizing ontologies, formalized units **Cross-Sectional Requirements:** Common data standards, generally necessary Ontology frameworks

4.1.3 NFDI-MatWerk's Infrastructure Vision: The Digitalized Workflow

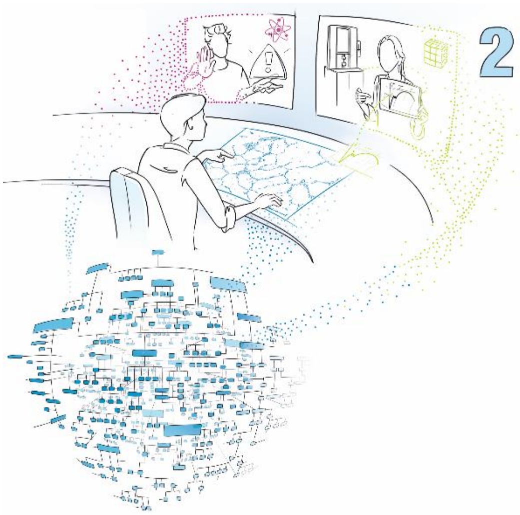


NFDI-MatWerk addresses the community-needs from a perspective that focuses on the way our fellow scientists work and interact today. We assume that our infrastructure will only receive usage and support from them if it offers clear and unique benefits within this given research environment. Therefore, we explain our infrastructure's advantage in the same context the scientist already works in and proceeds today: The workflow. However, we suggest it to change into a comprehensively Digitalized Workflow, indicating that today we obviously work with digitized or computational methods already, but do not apply them consequently to the overarching workflow perspective.

At the heart of NFDI-MatWerk's vision of the Digitalized Workflow lies the MSE-knowledge graph as represented by the Digital Materials Environment, or DME. The knowledge graph is a variable and universal ontological representation that standardizes information storage, access, and retrieval for our discipline and offers ties to other disciplines beyond it. We separate our model workflow into the four subsequently described steps.

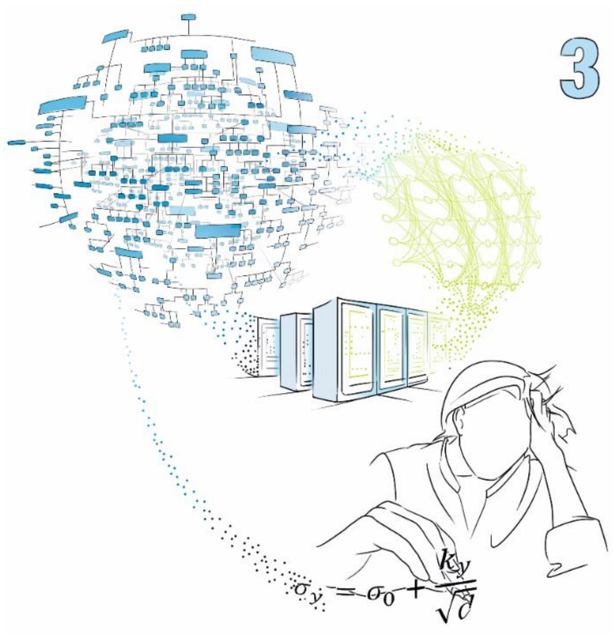


1 - Experiment: As it is the case today, the future researcher will need to conduct experiments, gathering real-world data as inputs to any subsequent model-setup or machine-based processing. However, in contrast to today, the highly interconnected laboratory will enable the testing devices and experimental setups to communicate with each other through standardized interfaces and data structures. This will enable them to aggregate and edge-compute their joint results prior to infrastructure exposure. With the relevant testing results, excluding many kinds of irrelevant raw data, single laboratories will then be able to offer their experimental information in the DME, which works via an input processing of the locally diverse information towards the agreed standards of the knowledge-graph. For any interaction with the knowledge-graph, such interfaces will need to be developed. This does not affect the data itself, which remains decentral. The DME rather represents where what information lies.

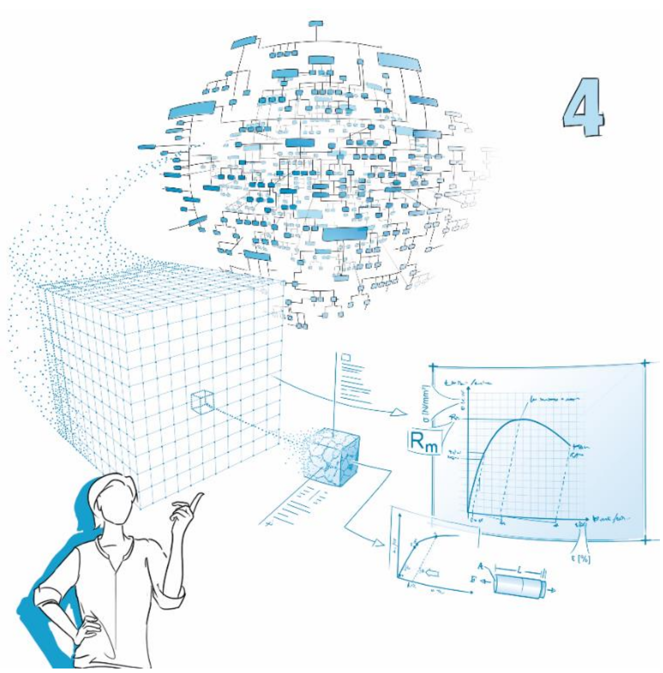


2 - Results: The researcher then continues to review the testing results. This is common in today's proceeding and the resulting exchange with colleagues is a part of the scientific discourse. The Digitalized Workflow enables the scientist to access her experimental outputs and results from anywhere at any time, as well as to provide this information to others through the same standardized channels of the DME knowledge graph. This makes it possible to share any necessary insight easily and both distribute one's own measurements as well as access the contribution of others in real time. Naturally, this results in a highly accelerated exchange that does not end within a single scientific discipline, but is expandable to others if they rely on the same standardized description methods.

3 - Evaluation: The real science takes place when the researcher analyses the results systematically and reviews and combines them to develop new models, theories, and predictions. In interaction with an MSE knowledge-graph, this will be possible under consideration of a vastly more comprehensive data basis, of all the properly stored information from prior research. The ontological nature of a knowledge graph will furthermore allow for the incorporation and representation of given pre-defined human knowledge (like equations) in a machine-readable shape. Future software will be able to use this knowledge and combine it with state-of-the-art statistical techniques, like machine-learning, to uncover fundamental new insight.



4 - Application: The knowledge-graph of the MSE will always stay holey and incomplete, which owes to the excessive rotational variability of our discipline. However, by relying on a maximum of representable knowledge and data, the DME will be able to gather enough information to create fundamental new insight from correlations, interpolations, or transitivity. It is unpredictable what variety of applications will therefore become possible through such a structured knowledge-representation. Like the researcher uses according findings today for presenting them in publications and in front of scholars, it will be the case once the research is increasingly based on the knowledge-graph itself. However, there is one significant difference: The research and the addition of delta-knowledge that a scientific contribution created will itself become part of the knowledge representation. In the end, the Digitalized Workflow that the scientist set up for and applied for storing all her necessary information along the way is one important new research output, itself.



4.2 Metadata Standards

In order to facilitate internal and external coordination and collaborations, metadata can be distinct into following categories:

1. Bibliographic metadata, e.g., creator of data, time stamps and version history, used by/in, published in, or cited in. For the definition of this kind of metadata, we will follow the standards established, for example in the DataCite (DataCite, 2019) or DublinCore (DCMI, 2003) XML or JSON schema for storing metadata (Metadata Schemas), and work together with the local libraries at the participant institutions.
2. Administrative metadata required for the administration of the research data in a broader sense, e.g., location of data (filename, url, down to internal HDF5 level), file format, location of physical specimen, specimen ID, provenance or access restrictions, but also including provenance. Here, we will use and extend existing Schemas like PROV-O (Lebo, 2013), DCAT (W3C, 2020) and the metadata element sets recommended by the Research Data Alliance.
3. MSE specific metadata, which will directly derive from the knowledge graphs established in TA Ontology. Examples include material chemical composition, material process history, type of data (e.g., stress-strain data or data obtained from APT), units, instrument data and Instrument ID, error, used data processing pipeline, etc. Examples of existing metadata models that are used for materials scientific tasks are the Core Scientific Metadata Model (CSMD, 2013) as well as upcoming recommendations from the materials groups of the Research Data Alliance.
4. Other domain specific metadata models exist in abundance in particular for computer science (e.g., semantic interfaces to databases, metadata for machine learning), but also for disciplines from the natural sciences and engineering. EngMeta (Universität Stuttgart, 2020) is one example for the latter one. We plan to closely collaborate with regards to metadata and ontologies with a number of related consortia such as NFDI4Ing, FAIRmat, MaRDI and NFID4Phys.

4.3 Implementation of the FAIR Principles and Data Quality Assurance

Findable: The planned data infrastructure will be strongly linked to the underlying materials science and arranged such that the digital data objects will be represented as FAIR Digital Objects that are uniquely and persistently identifiable. The overarching data structure will be developed by the materials experts together with ontology and structured data scientists and in close interaction with international efforts (e.g. European Materials Ontology EMO, European Materials Modelling Ontology EMMO (EMMO 2020) as well as derivatives from non-European countries). The strategy of NFDI-MatWerk to support ontologies that are based on physical principles will minimize the effort for the development of automatic mapping strategies. This includes the handling of consecutive version changes of the official ontologies as long as the governing bodies grant enough resources to initiatives like the NFDI.

Accessible: The goal of NFDI-MatWerk is to make materials data accessible in a human and machine-readable format by providing the infrastructure, tools and metadata. While the metadata (uploaded by the individual scientist into the overarching data structure) is centrally accessible, the raw data and optionally also the processed data, as well as workflows are stored locally (e.g. similar to “edge computing”). Access to this data will be granted within the client by the operating institution and the responsible scientist (e.g. author or scientific group leader). The (meta)data will be accessible through open protocols. Data plans for sustainable data storage will be developed which can be implemented by the responsible institutions.

Interoperable: TA Materials Data Infrastructure will harmonize all services with other NFDI consortia according to the Research Data Commons (Bierwirth, 2020) and the EOSC Interoperability Framework (EOSCFAIR, 2020). The **technical interoperability** will include (meta)data formats, service specifications and interfaces, supporting materials as well as the introduction of the principles of FAIR Digital Objects.

Semantic interoperability will be ensured through dedicated materials science ontologies. Ontologies ensure that (different) terminologies can be used in an unambiguous manner. This is achieved by providing explicit, consistent and systematic specifications of the respective semantics.

The **organizational interoperability** will be ensured by implementing agile methods for project management, working with flat hierarchies and flexible team compositions and a tight collaboration between NFDI-MatWerk and all other NFDI-stakeholders, including the DFG, the NFDI-directorate, as well as other consortia.

Finally, **legal interoperability** will be established by closely monitoring and implementing developments of the NFDI’s common contractual foundation, maintaining, joint communication channels and fostering networks.

Organizational and legal interoperability both already started their implementation through our efforts to harmonize our ideas and approach with other consortia and the NFDI as a whole, where the members of NFDI-MatWerk actively participated in discussions about the common challenges and according solutions.

Reusable: The goal of NFDI-MatWerk is to use the ontology-based metadata structure to achieve well described and rich enough metadata including license information that can be used for linking and integration into novel workflows, simulations and advanced materials data analytics. Tools and metadata standards for recording the data provenance will be provided by (TA MDI, TA WSD).

4.4 Services Provided by the Consortium

MSE scientists are already situated in a service landscape that consists of local, university-wide and national infrastructure services. The core problem is that these services are often isolated or only loosely connected. NFDI-MatWerk aims to incorporate these existing infrastructures with MSE specific semantics (TA-OMS) applications (TA-WSD) and interfaces (TA-MDI) to allow for an exploitation of the MSE research data treasure trove.

NFDI-MatWerk therefore brings together scientists from the MSE community and existing service providers. The local and national service providers within the consortium already provide a set of research data management services as part of their institutional mission. These services are, however, of general nature and are targeting the broadest possible scientific community. **Based on the consortiums' objectives, these services need to be enhanced and recombined in a way that the resulting service landscape fits to the scientific workflows of MSE scientists.** Only this will ensure that scientists will adapt research data management best practices within their daily duties.

In TA-CI the professional societies DGM, DVM and GAMM are engaged in achieving the goals of NFDI-MatWerk. They represent important parts of the MSE community and give NFDI-MatWerk access to their established networks. Among their inherent aims are promoting interaction within the community, organizing conferences, seminars, workshops and professional trainings and contribution to education in MSE. This makes them valuable partners for NFDI-MatWerk with a solid expertise and established structures that will be used in specific measures of TA-CI to disseminate knowledge and ideas of NFDI and FAIR research, gather feedback from the community and stimulate exchange within the community on NFDI-MatWerk topics.

TA-MDI will provide technical services for data and metadata management, user interfaces, service administration as well as the technical users' support. From an MSE scientists' perspective the TA Materials Data Infrastructure (TA-MDI) provides a reliable digital platform enabling the digital representation of materials data and MSE-specific metadata: the Digital Materials Environment (DME). From a technical perspective, this combines: a common web frontend, creation of persistent identifiers (PID), metadata and storage services, authentication and authorization infrastructures (AAI), rights management, proofs of authorship and usage analytics for compensation and incentivisation. The web frontend furthermore offers the possibility to share tools and workflows and will act as a community hub, offering amongst others educational resources.

TA-WSD will provide software solutions to handle complex experimental and simulation protocols with the goal to connect the aggregation and processing of materials data with the corresponding ontology. Software solutions handle the central storage of raw data, metadata, and materials data. In addition, the data processing environments will be virtualized and version-controlled to make data processing workflows repeatable and shareable. In order to handle the various multiscale and multidisciplinary approaches in MSE, TA-WSD will provide connectors between different incompatible codes to ensure data exchange between methods.

TA-OMS will provide ontologies that are able to represent all relevant aspects typically encountered in the materials science and materials engineering domain. This ranges from properties of materials including microstructure descriptions on possibly multiple scales and goes up to representing time dependent behaviour of materials and possibly of typical processes as well. Furthermore, TA-OMS will also enable the MSE community to interconnect their ontologies with ontologies of adjacent scientific domains. Last but not least, in close collaboration with TA-WSD visualization tools for ontologies and tools for working with/creating ontologies will be provided.

5 Work Programme

Table 5.0: Overview of task areas

Task Area	Measures	Responsible Co-spokesperson(s)
Community Interaction (CI)	CI1: Qualification CI2: Institutional Education CI3: Interactive Communication Platform CI4: Psychology of Change Management CI5: Community Counseling CI6: Stimulate and Enable Community Interaction CI7: Definition of Data Quality Standards CI8: Learning Examples for Data Curation CI9: Current State of Data Handling CI10: Task Area Management	Prof. Dr. Frank Mücklich Prof. Dr. Martina Zimmermann
Materials Data Infrastructure (MDI)	MDI1: Digital Materials Environment (DME) Backend MDI2: Web Frontend MDI3: Metadata Services MDI4: Data Governance Services MDI5: Materials Data Infrastructure Support	Prof. Dr. Erik Bitzek Prof. Dr. Peter Gumbsch Prof. Dr. Matthias M. Müller Prof. Dr. Achim Streit
Materials Workflows and Software Development (WSD)	WSD1: Coordination and Service WSD2: Graphical Programming Language WSD3: Version Control, Centralized Environments, and Access Control WSD4: Integrated development environments WSD5: Application workflows in experiment and simulation WSD6: Data curation, numerical precision and statistical/systematic errors WSD7: Machine Learning Platform	Dr. Tilmann Hickel Prof. Dr. Ruth Schwaiger Prof. Dr. Philipp Slusallek
Ontologies for Materials Science (OMS)	OMS1: MSE Ontology Working Groups OMS2: MSE Ontology Design OMS3: MSE Knowledge Graph and FAIR Linked (Open) Data OMS4: MSE Ontology Dissemination and Education	Prof. Dr. Harald Sack Prof. Dr. Stefan Sandfeld

Strategy Development (SD)	SD1: Continuous Strategy Development and Optimization SD2: Culture Change and Target Group Feedback Analytics SD3: Incentive mechanisms strengthening Good Scientific Practice SD4: Overall Project Management SD5: Legal compliance	Prof. Dr. Christoph Eberl

5.1 Task Area Community Interaction (TA-CI)

5.1.1 General Objectives

The main goal of this TA is to stimulate a broad acceptance from and to ensure a vivid participation of the overall MSE community with its highly interdisciplinary field of research areas and diverse working cultures. In order to achieve this goal, the following objectives must be addressed:

- stimulate and moderate active participation and vivid exchange of opinions and ideas,
- consult and guide individual members of the community as well as institutional groups interested in participating,
- establish the proper basis to train and educate all target group levels,
- elaborate guidelines for data curation based on a broad community consent and to prepare support for the necessary change process regarding data property and data handling mind-sets.

In addition, the TA will ensure that a steady growth of the circle of participants and users will be stimulated and channelled to all other TAs. In order to achieve these goals, the TA wants to address as many MSE stakeholders as possible, be it professional associations, review board members, research facilities, libraries or even trade associations. Hence, manifold measures must be initiated to stimulate the interaction with the NFDI-MatWerk consortium covering a vast spectrum of communication and interaction instruments. This can be achieved by offering virtual and actual platforms and events for information dissemination and exchange of views and to pursue a change in mind-set and working culture regarding the curation and exchange of materials data and its ontologies. Based on existing structures such as expert committees, material related conferences and fairs, national and international interactions between professional associations as well as by establishing new communication channels, maximum transparency for all community members and beyond shall be created. Within this aim the TA will stimulate integration of collaborative research efforts and work out new training programs considering the particular needs of target groups, from technician to advanced researcher. The goals are establishing an interactive and sustainable platform for the interaction with the MSE community. The benefit is that the goals, actions and outcomes of the NFDI-MatWerk project are constantly scrutinized, fed with input and are broadly and directly implemented in the target community with a highly multiplying factor.

5.1.2 Relevance for the Consortium's Objectives

Without achieving a broad acknowledgement of the overall MSE community, digital transformation driven by the NFDI-MatWerk initiative will not be able to manifest a long-term viability or sustainability. Hence, only a professional, comprehensive and structured approach to integrate the overall MSE community into the activities will ensure that the infrastructure created within the consortium will meet the actual requirements for future application scenarios. A constant information flow and discourse between the consortium and the overall community has to be guaranteed and quality standards for the curation of data have to be defined which meet the consensus of the community. At the same time, not only information from the consortium must be transferred to but also participant projects from the community are vital for NFDI-MatWerk. As precursor of a participant project this TA will organize events, info flyers and will establish counselling centers for those interested to become active but not yet ready to set up new participant projects. Since different community members may have to be addressed throughout the overall NFDI-MatWerk project phases, a persistent link and directory for the interchange of information is necessary in order to obtain a neutral, continuous, flexible while sustainable workflow with low boundaries for all NFDI-MatWerk partners. While the other task areas focus on the actual realization of the data infrastructure and the boundary conditions necessary for the realization, the task area “Community Interaction” will focus on the human/technology interface, amongst other identifying the roles and needs of actual and potential participants, e.g. early adopter, decision maker, student researcher, speakers of collaborative research initiatives, reviewer panels, technicians, expert committee leaders and many more. According to the roles of the different participants specific communication tools and channels have to be created, from Newsletter to NFDI-MatWerk-apps to Hackathons to specific sessions and information booths at conferences.

5.1.3 Status of this Task Area's Field in MSE

Professional societies like German Materials Society (DGM), International Association of Applied Mathematics and Mechanics (GAMM), German Society for Materials Research and Testing (DVM) and Federal Association Materials Science and Engineering (BVMatWerk) are the base to build upon. In these societies MSE and MSE-related professionals are well organized and have structural and personal connections to the MSE community on national and international level. They have the ability to build a trustworthy approach for extensive dissemination activities. Especially the institutional frameworks such as expert groups have a longstanding experience in working on specific tasks and disseminating project outcomes within the community, which will act as highly valuable contribution in this TA.

Besides the professional societies, BAM will be a cornerstone contributing impulses to the initiative with its background in standardized and accredited characterization and testing procedures. One of the tasks of BAM is to provide reference materials and reference data. Through its interdisciplinary, international network, the BAM can mediate a matching on a European and international level (e.g. NPL, EMPA, NIST, NIMS etc.) and initiate and moderate processes for the interdisciplinary compatibility of data sets and formats. A special focus of the work in different domains is the development of data formats and specific ontologies that meet the different expectations of academic, industrial and political stakeholders. In addition, BAM is a member in the Versailles Project on Advanced Materials and Standards (VAMAS), which promotes world trade by innovation and adoption of advanced materials through international collaborations, providing the technical basis for harmonization of measurement methods, leading to best practices and standards. In addition, the BAM will also support the curation and validation of external data sets through appropriate process knowledge and concepts.

With Martina Zimmermann, member of the presidency of the DGM, being one of the two Co-Applicants a close interaction with the review panels of the German Research Foundation will be guaranteed with her being vice-spokesperson of one of the two review panels of the MSE-forum. The other Co-Applicant Frank Mücklich will contribute with his network as former president of the DGM as well as member of Acatech (National Academy of Science and Engineering) to broaden the interaction beyond MSE community.

5.1.4 Measures and Work Packages

To address the objectives of this task area, a number of measures will be realized.

Measure CI-1 Qualification (*DGM, GAMM, DVM, Saarland University):*

Researchers on different levels of professionality need to be qualified in order to apply the ideas, concepts and tools of NFDI-MatWerk in their daily work.

WP CI-1.1 Organizing: Workshops (1 day), Summer Schools (3 days) and Professional Trainings are organized by the professional societies. Topics are focused on the ideas and concepts of NFDI-MatWerk and not limited to the scope of this Task Area but can also cover other topics or on request other Task Areas e.g. software development or ontologies. This work package also organizes workshops for specific participant projects.

WP CI-1.2 Training courses for international students: With the integration of NFDI-MatWerk concepts into education (see Measure 2), MSE students in Germany will “grow up” with the ideas behind NFDI. However, a significant number of (graduate) students come from other (non-)European countries, each having their own cultural background and education from their home country. Therefore, dedicated training courses for international students will be offered annually at four different universities across Germany to make them familiar with the concepts and tools developed by

NFDI-MatWerk. Special attention will be paid to intercultural aspects. Co-Applicant Frank Mücklich is founder and chairman of EUSMAT (European School of Materials), part of the department of Materials Science and Engineering at Saarland University. Its long-standing experience in intercultural trainings and summer schools for international MSE students is a significant benefit for this measure.

[redacted].

Measure CI-2 Institutional Education (*Fraunhofer IWS, DGM):*

Making the digital transformation of MSE part of the institutional education is key for a sustainable deployment of NFDI concepts. This measure covers both academic and technical education.

WP CI-2.1 Academic teaching material: Develop teaching material for undergraduate students integrating the ideas, concepts and methods of NFDI-MatWerk in curricula. This includes guidelines, slide decks, scripts and apps. A key role will be played by Studientag Materialwissenschaft und Werkstofftechnik (StMW e. V.) a national network of MSE universities which will help us to standardize teaching approaches and disseminate teaching material.

WP CI-2.2 Technical teaching material: Develop teaching material for education of lab technicians and organize workshops to train lab technicians in the use of NFDI-MatWerk infrastructure and develop a common understanding of how to define material ontologies and to establish an awareness of the significance of the documentation of metadata. Lab technicians rely in particular on knowledge based experimental strategies established on continuously refined workflows. On the basis of this educational measure this empirical knowledge has to be scrutinized according to NFDI-MatWerk quality standards while at the same time harnessed for the overall NFDI-MatWerk activities.

[redacted]

Measure CI-3 Interactive Communication Platform (*DGM):*

TA Materials Data Infrastructure (see Work Package MDI-2.2) will set up an interactive website. TA Community Interaction is responsible for following content: Interactive communication platform (e.g. moderated forum), offer teaching material for download and dissemination of information (important dates, upcoming events, etc.). A quarterly newsletter will be deployed.

WP CI-3.1 Content management: Information must be gathered from the consortium which requires action from the person responsible for this WP. Communication skills on MSE expert level are necessary along with technical skills and experience in running a website.

[redacted]

Measure CI-4 Psychology of Change Management *(Fraunhofer IWS):*

Handling the necessary change management in the MSE community is crucial for the success of NFDI-MatWerk. This measure will be closely coordinated with TA Strategy Development.

WP CI-4.1 Stimulating awareness and critical discourse: Within this WP change management as a consequence of the digital transformation will be addressed with support from work organization experts. Requirements from the organizational background and the management level will be dealt with as will be the creation of an awareness why and how to share data and knowledge. Lessons learned from DIWAN and other Participant Projects will be transferred into NFDI-MatWerk by utilizing the discussion platforms established within the TA. This measure will implement concepts developed by TA Strategy Development.

[redacted]

Measure CI-5 Community Counseling (*DGM, BAM, GAMM, DVM):*

Professional societies serve as contact for people in MSE who are interested in participating in NFDI-MatWerk, thus increasing the outreach of NFDI.

WP CI-5.1 Contact person: Provide a contact person to which questions and concerns from members of the MSE community can be addressed. For each mentioned professional society one person will be the general contact for individual NFDI-MatWerk community members and their issues, e.g. if workshops or teaching material are requested or if newsletter articles have to be sent as an update for the community.

WP CI-5.2 Templates: Provide templates for raw and metadata. Templates for various test methods (e.g. hardness test, tensile test), measurement methods (e.g. density measurement, dilatometry) and imaging methods (e.g. microscopy, tomography) are created. This includes minimum basic information which should be given in the metadata, e.g. investigated material including processing history, how (used method, test standard, test parameters, specimen geometry) and by whom was the data generated, date/time of generation, location, link to raw data, file size, data quality. Templates for data of different quality are generated, which depend on the amount and completeness of the metadata. Templates for raw data include the file format (reduced formats may be necessary), which measurement data must be captured, and which can be captured.

[redacted]

Measure CI-6 Stimulate and Enable Community Interaction (*DGM, GAMM, DVM):*

Interaction between NFDI-MatWerk and the MSE community as well as interaction among community members on NFDI topics helps to identify the needs of researchers and creates awareness for the digital transformation.

WP CI-6.1 Conference and Meetings: As described in Section 3.4, regular meetings and workshops are being held on a regular schedule. TA-CI will organize the annual NFDI-MatWerk conference (national level in the first year, international thereafter) including invited speakers. Furthermore, the annual all-hands-meeting of the broader consortium incl. participants and related community members will be organized by this TA. The all-hands-meeting is dedicated to discussing work on and current status of IUCs and PPs.

WP CI-6.2 Events: Organizing and moderating events and pollings to gather input from the community on relevant topics (e.g. hackathons, knowledge cafes). Stimulate discussion of "data ownership" and anonymization of data. The aim is to stimulate active participation of the community in shaping the future development of NFDI-MatWerk and to identify “hot topics”. While this WP is partly similar to TA Strategy Development’s Measure SD-2, it focuses on the rather operative implementation of strategic surveys, while the former focuses on their concepts.

WP CI-6.3 Expert Committees: Get expert committees of professional societies involved e.g. for defining relevant data in participant projects.

[redacted]

Measure CI-7 Definition of Data Quality Standards (*BAM, DGM, GAMM, DVM):*

Interdisciplinary guidelines for the collection of high-quality, well-documented data in standardized workflows and quality criteria for data are created.

WP CI-7.1 Elaboration of guidelines and standards: High-quality data contain complete information about the measuring/testing object (i.e. sample history including processing, machining parameters, microstructure, sample geometry) as well as complete documentation of the measuring/testing. This includes the naming of all measuring equipment with associated calibration certificates to determine the uncertainty of measurement and compliance with measuring/testing standards. Manual entries in the documentation process should be reduced to a minimum in order to avoid transcription errors. Raw data from the testing/measuring instrument, including instrument settings, should be transferred directly. A similar approach should be developed for modeling and simulation. Expert committees are involved via professional societies.

[redacted]

Measure CI-8 Learning Examples for Data Curation *(Fraunhofer IWS, BAM):*

Best practice examples for data curation are developed to help the community maintain high data quality standards.

WP CI-8.1 Exemplary Guidelines: Elaborate “How to” for participant projects from other Task Areas: Based on its long-standing experience in setting standards for the harmonization of measurement methods, leading to best practices and standards the BAM will work out guidelines how to define, document and monitor workflows that serve as basis for proper data curation. The learning examples will be identified by Fraunhofer IWS on the basis of participant projects and will be derived from experimental technologies with maximum potential of interfaces within the community (e.g. classical experimental setups such as tensile tests, hardness measurements etc.). Documentation strategies will include the prerequisites to the use of electronic lab books. This task is not to be confused with TA Strategy Development's Measure SD-3, that will provide the necessary theoretical foundations to data curation mechanisms on an infrastructural rather than processual level.

[redacted]

Measure CI-9 Current State of Data Handling *(DGM):*

Constant monitoring of data handling in MSE and related disciplines is required due to the strong momentum of digital transformation.

WP CI-9.1 Current state: Find out about current developments in the community and related disciplines, subdivided according to expert committees. This information will serve the TA Strategy Development in creating its regular reviews about the Community Development as described in Measure SD-1.

WP CI-9.2 Journal interaction: Raising the awareness of editors and boards of professional journals regarding data quality.

[redacted]

Measure CI-10 Task Area Management (*Saarland University and Fraunhofer IWS):*

WP CI-10.1 Management: Coordinate and supervise Task Area activities and ensure communication among the involved institutions. Support Task Area Representatives on a national and international level. Coordinate cooperation with other task areas.

[redacted]

5.1.5 Interfaces and Interdependencies with other Task Areas

In addition to regular meetings with the participants of the other TAs to discuss and coordinate overlapping implementations, an online communication platform, which will be established for this Task Area is supposed to foster information accessibility and to inform about the ongoing adjustments. The communication platform will be part of the website supplied by TA MDI as Single Point of Entry. In detail – results from discussion panels, hackathons, opinion polls and other measures to monitor the opinion making within the community will be communicated to and discussed with the TA "Strategy Development". The identification of additional participant projects throughout the initiative will be actively pursued in close collaboration between these two TAs as well. Regarding the TA MDI key topics for interfaces and interdependencies of each TA will be data safety and data curation. In addition, Measure CI-9 related to continuous screening of the data situation (raw and metadata) from the literature and getting into touch with editorial boards of scientific journals shall serve as support for TA MDI in order to stimulate an exchange of ID management and tracking data use questions with the persons in charge. The digital platform developed in the framework of this TA will be communicated to the community via TA “Community Interaction” and in a later stage the organizational basis for training sessions and workshops will be provided. With regard to TA WSD the initial activities in TA CI are supposed to encourage both experimental as well as simulation experts to contribute to NFDI-MatWerk with their current tools and software solutions, codes, work flows and raw and metadata management. This will be the major goal of Measure CI-3. Community interaction with regard to standards for the curation of raw and metadata will also be a mutual topic of the two TAs. Besides the setup of an interactive digital platform these topics will be addressed in side events within established workshops and conferences organized by the associations involved. These topics are also good starting points for hackathons. In addition to the project-related cooperation with the project partners, there is a regular and intensive exchange with the members of the professional associations and the universities of the Studientag Materialwissenschaft und Werkstofftechnik. In addition, maintaining close contact with the Materials Science and Materials Engineering thematic network of the German Academy of Science and Engineering (acatech).

5.1.6 Contribution to Cross-Cutting Topics of the NFDI

Community interaction and the dissemination of provided services and frameworks are a central topic for all NFDI consortia. However, the concrete culture regarding sharing and handling of data strongly varies from field to field. Therefore, a close monitoring and exchange with other consortia is required. This will be managed together with TA Strategy Development. Results of this exchange will be communicated inside NFDI-MatWerk and to the community.

This exchange has at least two major advantages. On the one hand, we expect mutual learning and thus work could be made more efficient. On the other hand, strengthening of personal contacts beyond their own community borders leads to a consolidation and acceptance of the NFDI project results.

In coordination with the governance of NFDI-MatWerk and the other consortia or the project governance as a whole, cross-consortia events can be organized together with TA Strategy Development. For this task a large event per year is appropriate. Additional online measures such as newsletters, discussion forums or the like are useful additions.

5.1.7 Appropriateness of the Task Area’s Representatives

The Co-Applicant, Prof. Martina Zimmermann, with her double-position at the Technical University of Dresden and the Fraunhofer Institute of Material and Beam Technology will contribute to the TA by building bridges between the research community and industrial needs, while at the same time acting as a multiplier in her function as a member and vice spokesperson of the expert forum of the German Research Foundation. With the ongoing BMBF-project “DIWAN” she will bring into the TA the particular needs of the target group of material testers and technicians. With the scientific insight into user interfaces and the acceptability of digital expert systems such as an electronic lab-book addressed in DIWAN also from a work psychology perspective, a substantial added value will be entered to NFDI-MatWerk.

The Co-Applicant, Prof. Frank Mücklich, has many years of experience in leading scientific community activities and also as initiator of national and international activities of such communities. As head of the advisory board of the German Materials Society (DGM), he developed with this board the idea, the structure and the design of the largest event in materials research in Germany, the international MSE Conference Series, and was its conference chair both in Nuremberg in 2008 and in Darmstadt in 2018. In addition, he chaired the largest European materials research conference for the Federation of European Materials Societies (FEMS), EUROMAT2016 in Seville, Spain. For the term 2016-2019, Frank Mücklich has been also Director in the Board of Directors of IMS in the American Society for Materials (ASM) and is now with ASM and FEMS active in particular for the establishment of an international network of the largest professional societies in materials science and engineering worldwide. From 2019 to 2020, Prof. Frank Mücklich served as President of the DGM and actively promotes the digital transformation in all subject areas and sub-communities of this society. To achieve this, Frank Mücklich co-initiated and co-authored the strategy paper "Digital Transformation in Materials Science and Engineering" (DGM, 2018). The implementation of this development also requires timely integration into national and international teaching. Already in 2008, Frank Mücklich founded the European School for Materials (EUSMAT) in Saarbrücken on the basis of the EU's Erasmus Mundus program, with partnerships in the four most important European languages with Barcelona (Spanish), Nancy (French) and the Swedish Lulea (English). EUSMAT is now well established internationally, making it the ideal partner for modern teaching to implement digital transformation in all levels of academic education with international appeal.

5.1.8 Connection to Infrastructure Use Cases

Infrastructure Use Cases (IUCs) serve on the demands of Participant Projects. IUCs are realized by contributions from one or more task areas and serve as a query for the infrastructure developed by NFDI-MatWerk. In case of TA Community Interaction, the measures described above can be divided into two categories: IUC-related and general measures. IUC-related measures contribute directly to the realization of certain IUCs. General measures are not directly related to an IUC but serve the development of underlying, community-oriented infrastructure or assist other TAs.

Examples for IUC-related measure are:

- Workshops, professional trainings and Summer Schools as being organized for education and qualification by Measure CI-1 and CI-2. This will be extensively used by IUC01 “Web-based demonstration and teaching network for the creation of educational material” and for contributions to Participant Projects such as e.g. “IRTG 2078 Integrated engineering of continuous-discontinuous long fiber reinforced polymer structures” where the educational background of this measure comes into play.
- Interaction with expert committees via professional societies (Measure CI-6) is important to achieve broad acceptance within the MSE community when it comes to defining standards and developing ontologies. As an example, this is highly relevant in IUC15 “Method and scale-bridging workflows and data structures for tomography” but it also applies to all IUCs and PPs where input from a broader range of experts plays a role.

Examples for general measures are:

- Measure CI-3: Interactive communication platform. This measure is of high relevance for the outreach of NFDI-MatWerk. It elevates the internet presence of the consortium from a one-way distribution and information platform to a two-way interactive platform that will exchange between the consortium, participants, other members of the MSE community and even researchers from other consortia.
- Measure CI-5: Community counseling. Professional societies are important partner for the dissemination of information and outreach to more specific sub-groups of the MSE community. Therefore, DGM, DVM and GAMM are equipped with staff to facilitate topic-specific actions and act as bridge builders between NFDI-MatWerk and expert committees within the community.

5.1.9 Funding Request for TA Community Interaction

Table 5.1.9.1: Funding Request for Task Area Community Interaction by Institution

	2021 (Oct-Dec)	2022	2023	2024	2025	2026 (Jan-Sep)	Total in €
Institution		Totals in €					
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

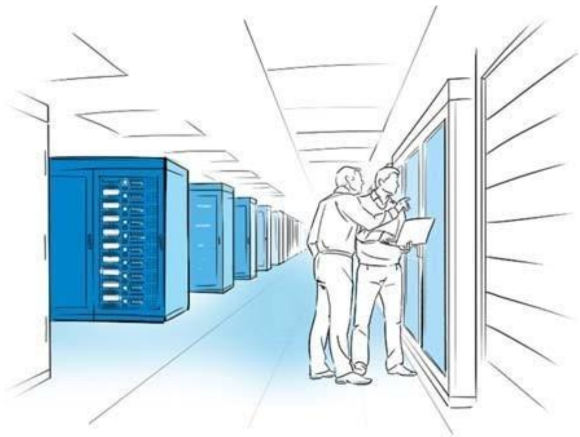
Table 5.1.9.2: Funding Request for Task Area Community Interaction by Funding Category

Funding category	**2021 (Oct-Dec)**	**2022**	**2023**	**2024**	**2025**	**2026 (Jan-Sep)**	**Total**
Staff by category	**Number of persons (full-time equivalents)**						
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Totals in €						
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

(*) this corresponds to the DFG staff category "Postdoctoral researchers and comparable"

(**) this corresponds to the DFG staff category "Doctoral researchers and comparable"

5.2 Task Area Materials Data Infrastructure (TA-MDI)



5.2.1 General Objectives

From the viewpoint of an MSE scientist the TA Materials Data Infrastructure (TA-MDI) provides a reliable digital platform enabling the digital representation of materials data and MSE-specific metadata: The Digital Materials Environment (DME) in which researchers, engineers and technicians can

- search, access, and analyse data without worrying about the physical location or file format of the data
- effortlessly, ideally automatically, store data together with the corresponding metadata
- easily access related information, e.g., the entire sample processing history, different measurements on the same material, etc.
- securely share data with collaborators by granting selective access to data
- know if, how and how often data was used and be assured of their authorship.

From the viewpoint of the infrastructure, TA-MDI provides reliable data and metadata services that interoperate with many other research data infrastructures worldwide.

The main objective of the TA-MDI is to set up the DME such that it fulfills the vision outlined above. To that end, the TA-MDI will establish

1. a **reliable digital platform (DME)** enabling the digital representation of materials data and MSE-specific metadata fulfilling the FAIR principles. This will be realized through **a data space that feeds from a linked data knowledge graph**, and (meta-)data formats, protocols and interfaces meeting the requirements of the community, thereby reducing the barrier for data sharing and allowing analytics on shared data sets, including information on reuse;
2. **comprehensive metadata services,** where metadata ontologies are implemented, allowing the full processing history of specimens to be recorded, data to be easily searched and collected for analysis, metadata to be edited and added in accordance with individual researchers' needs;
3. **a user identity management** via an Authentication & Authorization Infrastructure (AAI), which enables cryptographic proofs of authorship and provides the historical record of the data (Data Provenance), as well as the access to data and computational resources across organizations;
4. for all the above: **long-term technical support** for users and local installations and a **sustainable, cost efficient operation and compensation model** and platform that integrates individual storage providers, central institutional providers and "cloud" providers.

The DME will serve as a general framework for the tools developed in TA-WSD, with which TA-MDI is closely collaborating, e.g., regarding the definition of interfaces and testing. As the envisaged materials data framework DME is based on a linked knowledge graph, its structure is directly related to materials ontologies, and all developments within the TA-MDI will be closely coordinated with the TA-OMS.

The currently available data and computing service landscape for the MSE community is highly decentralized and mostly operated by researchers themselves or by local infrastructure providers. Moreover, these infrastructures are generally only available to researchers of the local institution. To reach the vision of a common, distributed data space each participant can contribute research data as well as parts of their storage & compute infrastructures. The use of, e.g., Docker containers (Docker, 2020) allows each participant to operate and integrate their own data infrastructure into the DME, thereby enabling the scaling of the DME to national as well as international level. **The TA-MDI will thus seamlessly integrate the data on multiple servers/computers, and provide researchers access to local and centralized data and computing services.**

Although the enrichment of materials data by metadata itself cannot be completely automated, **tools and templates will be provided to help the researchers, engineers and technicians to add metadata (Metadata Editor)** and to profit from metadata management services that work in the background. It will allow users to create and benefit from the advantages of FAIR Digital Objects (RDA DFIG, 2020) without perceiving the underlying complexity.

A **central web frontend** will be installed that will serve as an entrance to the DME and as community hub. It will offer a comprehensive metadata service for **data search, registration** of research data sets, provide compute resources and enable **easy sharing and accessing of data** through intuitive user interfaces. It will furthermore offer an "**App-Store**" (run by TA-WSD) where tools and workflows can be shared and rated. **Reducing the technical barriers to data sharing, search, access and analysis is key for the acceptance by the users and is at the center of the TD-MDI.**

Together with the TA-CI the TA-MDI will furthermore **provide best practices** and implementations for researchers and local infrastructure providers to connect existing scientific workflows with the DME and offer support for users and use cases, including **on-site support and training**. Additional features or functionalities will be considered according to the guidelines developed in TA-WSD.

The building blocks of TA-MDI are summarized in Fig.10, highlighting the overlap with NFDI's cross-cutting topics.

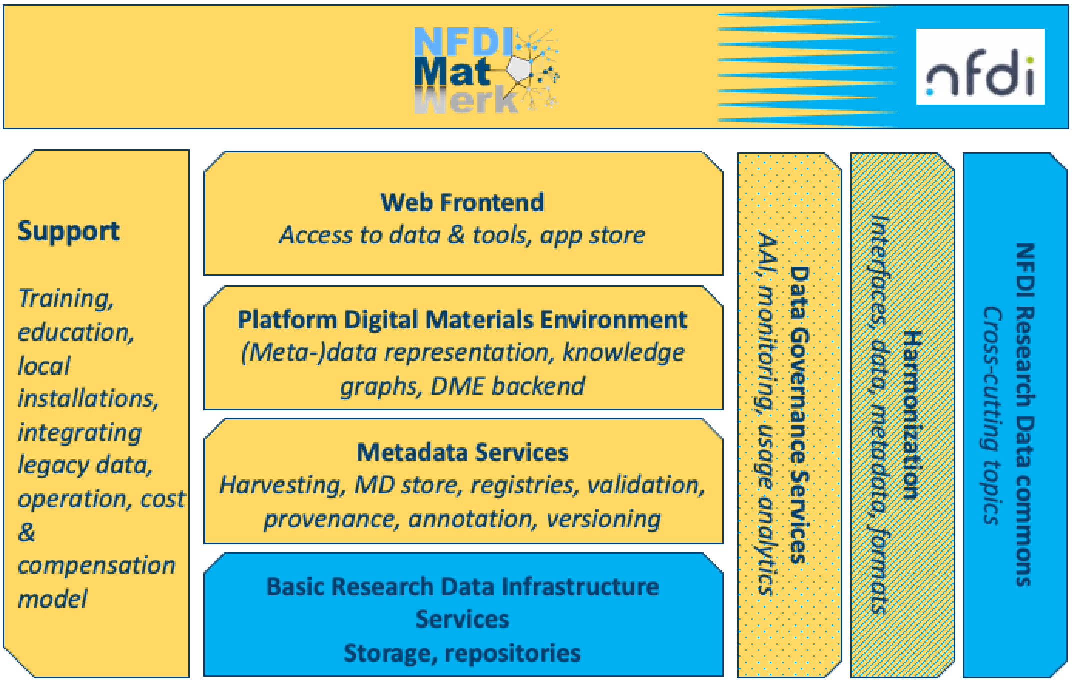


Figure 10. *Building blocks of the TA Materials Data Infrastructure. The yellow blocks and services will be provided by NFDI-MatWerk, the blueish ones interact accordingly with the crosscutting topics of NFDI.*

5.2.2 Relevance for the Consortium's Objectives

Materials Science and Engineering is by its nature highly multidisciplinary, requiring close cooperation between materials scientists and engineers, mechanical engineers, process and manufacturing engineers, physicists, and chemists, usually across multiple groups, institutes and institutions. These collaborations generate a wealth of data of vastly different nature (e.g., information on processing, microstructure, properties…), origin (experiments, numerical simulations) and formats (tables, pictures, 3D datasets, time series...). **Making this treasure trove of materials data from the technical perspective *easily* findable, accessible, interoperable and reusable (FAIR) is at the heart of the TA-MDI. It will provide the basic, unified (software) architecture to store, share, search, access, and cite data as well as to perform analysis and computations on data all while shielding the user from technical details.**

In particular, the DME will allow to share and access materials' raw data and metadata across institutional boundaries while securing data integrity and authorship. Through it, researchers will be able to quickly and easily set up local data spaces with integrated data sovereignty. It will furthermore provide the data infrastructure for digital workflows and tools, e.g., for visualization, analysis or machine learning developed by the TA-WSD. The organization of data and metadata through knowledge graphs allows for direct implementation of the materials ontologies derived in the corresponding TA-OMS.

The relevance of the TA-MDI for the consortium's objectives and the MSE community in general is highlighted in the Infrastructure Use Cases. While all IUCs and PPs will benefit from the DME and DME-related tools, large collaborative projects like CRCs that span a multitude of institutions and combine data from a vast range of processing, characterization and testing techniques as well as simulation methods will particularly profit from the common data infrastructure.

5.2.3 Status of this Task Area's Field in MSE

Data on materials properties, of their performance as well as on the processing leading to these properties and performances as well as on their chemical composition and microstructure **are of fundamental importance for the development of new materials and the design of components.** Historically, data on materials was available to researchers and engineers in academia and industry through: (1) own collections of data, e.g., in individual research institutes or companies; (2) open publications, including scientific and engineering articles as well as patents; (3) commercially available database, e.g. ASM Handbooks (ASM Digital Library, 2020) and Springer Materials (Springer, 2020).

In the last years, several repositories have emerged where materials data can be contributed by individual researchers and shared with the community. Examples for general MSE-specific repositories are the National Institute of Standards and Technology (NIST) Materials Data Repository (NIST, 2020), the Materials Data Facility (Blaiszik, 2016), or MatDB (European Commission, 2020). All of them have, however, currently way less than 1000 datasets each. There exists also domain- or subject-specific repositories, such as NOMAD (NOMAD, 2020) and Materials Cloud (Talirz, 2020) for simulation data by quantum mechanics, and the thermodynamic database at NIST and National Institute of Standards and Technology (NIMS) for the CALPHAD community, which contain, however, only a very limited subset of material data. Due to the highly interdisciplinary nature of MSE, materials data can also be found on repositories of related disciplines such as physics, chemistry, nanotechnology (e.g. NFFA Europe (NFFA, 2020) or of large experimental facilities. For a full overview of materials-related repositories, see re3data.org. Furthermore, materials data can also be found on generic data repositories like RADAR (RADAR, 2020). This multitude of repositories reflects the interdisciplinary nature of materials science as well as the broad variety of materials, materials data type, and characterization, processing and simulation methods. **We value this diversity and foster the utilization and linking these diverse data.**

5.2.4 Measures and Work Packages

Measure MDI-1 Digital Materials Environment (DME) Backend *(RWTH, KIT, FAU)*

The DME backend will provide the framework to register, address, search and access research data and corresponding metadata on local as well as distributed data & compute resources. It will be able to run locally on individual computers, group or institute servers as well as on cloud servers and on large academic computing centers, independently of the underlying OS. Multiple DME instances can be linked together using peer-to-peer interfaces, and, if desired, linked to the central front end, thus becoming part of the public DME ecosystem. The DME will thus form the bases for implementing the FAIR principles and access to data & compute resources through machine operable interfaces. We will mostly rely on existing technologies comparable and compatible with existing infrastructures like the International Data Space (IDSA, 2020) Galaxy (Galaxy, 2020) or NFFA Europe (NFFA 2020). To support the unique requirements of the MSE community, like accessing extremely different types of interrelated data sets with sizes up to multiple TBs, we will extend and adapt these solutions. The central challenge in this measure will be to develop a set of interfaces that allows distributed instances running on the researchers computing environments, like Pyiron (Pyiron, 2020) interact with another. Therefore, a set of defined interfaces will be designed that allow researchers to execute validated analysis workflows to a server with direct access to the stored data storage and then return the results back to the user. Technologies like docker (Docker, 2020) and Anaconda (Anaconda, 2020), already allow for the packaging of complete computing environments and their dependencies and thus can serve as a technological basis. To access, register or manipulate data sets in the DME, researchers will use a peer-to-peer client that is installed on local environments, or can use the central web frontend. Defined interfaces will allow interaction and communication of different decentralized clients.

Sustainable data & compute resources are the basis of the DME. We will therefore provide blueprints for consistent and high standards for IT operations in local environments. This will enable decentralized IT service providers and individual researchers to set up the hard- and software environment necessary to run the DME and - if desired - share selected parts of their resources in a controlled and secure way. Based on the current state in the research community we focus on services that provide support for large binary files, like HDF5 (HDF Group, 2020), and that are accessible from scripting programming environments like Python or Jupyter (Jupyter, 2020).

WP MDI-1.1 Administrative Graph Database: Implementation of an efficient, machine-operable representation of the Resource Description Framework (RDF) based knowledge graph for the information necessary to access the contents of files of different format and location and to perform computations.

WP MDI-1.2 Data Registration and Ingestion Interfaces: Provide interfaces for the minting of persistent identifiers (PIDs), cryptographic proofs of authorship and different access levels that allow MSE researchers and technicians to register data sets & compute workflows in the DME.

WP MDI-1.3 (Meta-) Data Formats: Leverage community agreement on (meta-)data schemas, formats, protocols and interfaces for exchange of data sets & compute workflows, like RDF, HDF5 or docker containers or conda packages.

WP MDI-1.4 Compute & Data Sharing Interfaces: Extend existing solutions for sharing local data & compute resources with peer researchers of other organizations. This will require close coordination with the measure MDI-5 “Materials Data Infrastructure Support” and the TA-WSD, using common frameworks and technologies.

WP MDI-1.5 Operating Model: Development of a best practice operating model and software stack for decentralized compute & storage resources for sharing with peer researchers based on standardized interfaces that also regulate range & rate accessible, if necessary, also for High Performance Computing (HPC) access.

WP MDI-1.6 Peer-to-peer Client: Together with the TA-WSD development of ready-to-use implementations that allow integration of existing infrastructures into the DME using python and docker technologies.

WP MDI-1.7 Local Version Control: Implement solutions for local version control and long-term archiving according to decisions made in a data management plan to assure data that is referenced in the DME is not altered and compute workflows are reproducible while data sets remain in the local environments. Local versioning will be done in accordance to WP MDI-3.4

[redacted]

Measure MDI-2 Web Frontend *(RWTH)*

The TA-MDI will set up and host a common website for the NFDI-MatWerk and all TAs, who will provide the contents for the web presence. In particular, the website will act as a low-threshold user interface and gateway to the DME. While the core of the DME is a decentralized infrastructure that uses local data & compute resources, the web frontend will serve as a lightweight central user interface for initial access. It will feature different functionalities: an app store like interface for shared data sets & compute workflows; access to statistics and analytics capabilities and meta-search interface across different connected repositories and the connected decentral environments. The web frontend will support the decentralized registration of research data resources according to FAIR principles and serve as a community hub for sharing and accessing data. It will make use of existing software products or cloud offers that are actively used in the MSE community like GitLab (GitLab, 2020) or Anaconda (Anaconda, 2020) and metadata services established by measure 3. To provide a low-threshold initial access to the DME and provide a consistent user experience, this measure will be closely coordinated with TA-CI, in particular also to provide web-based teaching and demonstrations.

WP MDI-2.1 Website: Setup, host and operate the main website with content provided by TA-WSD and TA-CI.

WP MDI-2.2 Community-Hub: Set up forums, download areas, etc. available for the general public as well as only for registered participants, as specified by TA-CI. Will be used for polling, for dissemination of information about NFDI-MatWerk and related events (e.g., conferences, workshops etc.).

WP MDI-2.3 App Store Infrastructure: Setup, operate and maintain app store infrastructure using existing products and SaaS like GitLab, anaconda, docker registry that allows sharing data sets and compute workflows for, for example from the TA-WSD but also from researchers of the MSE community. Include feedback and feature requests in coordination with TA-CI.

WP MDI-2.4 Meta-Search Interface: Build a comprehensive web interface that allows users from the MSE community to search or browse existing data & compute workflows within the DME and from existing external repositories if relevant for the community (see also WP MDI-3.4).

[redacted]

Measure MDI-3 Metadata Services *(KIT, RWTH)*

Metadata are one of the key elements to implement a human readable- as well as machine actionable representations of materials-related information. Whilst the TA-OMS will derive the necessary vocabularies and metadata schemes specific to MSE, the measure Metadata Services will implement their technical representations in FAIR Digital Objects (RDA DFIG, 2020) and together with TA-WSD provide easy-to-use tools for the integration into the DME as well as user-friendly tools for metadata enrichment. It will provide common interfaces to ensure technical interoperability and the infrastructure and tools enabling data search, provenance, annotation and analytics. In particular, it is important to foster interoperability with the Terminology Services that are planned to be established in NFDI4Ing, NFDI4Culture and NFDI4Chem. NFDI-MatWerk will adopt services and tools that are recommended and developed by other projects and initiatives, e.g. Research Data Alliance (RDA, 2020), Helmholtz Metadata Collaboration (HMC) Platform (HMC, 2020), and the DFG projects “Applying Interoperable Metadata Standards” (AIMS, 2020) and “Metadata for Applied Sciences” (MASi, 2015).

WP MDI-3.1 Infrastructure Services: Provide services for storing and accessing metadata with a focus on MSE-specific and bibliographic metadata as well as for registering research data, using the RDF based interfaces from WP MDI-1.2. These general metadata infrastructure services will be integrated in close collaboration with other NFDI consortia according to the Research Data Commons described in “Berlin Declaration on NFDI Cross-Cutting Topics” (Glöckner, 2019) and the EOSC Interoperability Framework (EOSCFAIR, 2020) and adapted to NFDI-MatWerk. It will include, e.g., registries for metadata schemas, vocabularies and data types, storage infrastructure for metadata (metadata stores), etc., and provide interfaces to external repositories and search engines.

WP MDI-3.2 Integration of Ontologies: Implementation and integration of materials metadata schemas and vocabularies developed in TA-OMS. This will include, e.g., the establishment of a metadata schema registry and a formal validation of metadata according to the metadata schemas.

WP MDI-3.3 Search: This includes metadata harvesting, indexing, search, and discovery. It will provide tools and services to collect metadata from various sources, including established MSE-specific and general repositories and store them centrally for searches through the web frontend. Metadata harvester will compile all metadata from existing MSE-specific repositories supporting standardized interface protocols, e.g. OAI-PMH (Open Archives, 2015). That way, users can perform searches not only over data registered in the DME, but also use other data sources that will enable powerful discovery tools.

WP MDI-3.4 High-level Services: Integration of additional functionalities like data versioning, data provenance and data annotation. Data versioning will be reflected in the administrative metadata sets and demands special treatment in accessing as well as in modifying the metadata and data entries. Data provenance will be recorded in the workflows and stored in an W3C standard for provenance interchange PROV (W3C, 2013) using the metadata infrastructure services. Data annotations will be treated according to the Web Annotation Data Model (W3C, 2017).

[redacted]

Measure MDI-4 Data Governance Services *(KIT, RWTH)*

In a distributed environment like the DME, access, restrictions and monitoring are key factors to allow the enforcement of data governance policies. The reliable identification of users and implementation of their permissions to access data and resources forms the basis of the DME. That way, research groups can easily grant fine-grained access to specific data, storage and computing resources administered by them or their institutions to selected individuals or groups. Thereby, research data can remain within the organization where it was created, while still being accessible. Explicit access rights will be granted on a “per user” basis. Thus, users need to be identified within the DME and connected software applications; as for example developed within the TA-WSD or originating from the MSE community. We will therefore provide comprehensive ways to integrate AAIs into individual software products. Reference architectures, like AARC (AARC, 2020) allow joining existing federated infrastructures to make user profiles available in other services. The existing federations like DFN-AAI (DFN, 2020) or EDUGAIN (GEANT, 2020) form a well-established starting point for an AAI for users of member institutions, however infrastructures like ORCID allow better long-term identification of users as they change institutions during their career. It is therefore necessary to combine and extend these existing approaches. As this is a current challenge throughout all scientific fields, this measure will be carried out in close collaboration with the other NFDI consortia and the NFDI roof association.

WP MDI-4.1 Identity Space: Establishment of a base scheme for the common identity space using existing identity providers, e.g., DFN-AAI / EDUGAIN and / or ORCID tailored to the circumstances and workflows within the community.

WP MDI-4.2 Role and Group Management: Provisioning of a service to manage distributed access rights and make local and centralized data & compute resources accessible by peer researchers across organizations.

WP MDI-4.3 Monitoring & Data Usage Analytics: Collect usage analytics for data sets which are accessed through the DME to compute metrics like data reuse count or data cites and in coordination with TA-CI define and collect metrics that are especially important for MSE researchers.

WP MDI-4.4 Reference Applications: Development of reference applications and libraries that implement the authentication & authorization reference architecture with use cases from the MSE community and the TA-WSD using common programming languages like Python.

Measure MDI-5 Materials Data Infrastructure Support *(FAU, RWTH, KIT)*

This measure subsumes all tasks related to the internal coordination between the task areas, applicants and participants as well as the external coordination regarding the NFDI consortia and other national, European and international initiatives concerning data infrastructures, see sect. 3.3. Domain experts will serve as links between the MSE community and the IT service providers. The task force will in particular provide support to users of the DME and works together with TA-CI towards the dissemination of the DME infrastructure and related knowhow. A central task is the coordination with and the support of the Infrastructure Use Cases (IUCs), see sect. 4.2, 4.1.2 and sect. 5.2.8. A further long-term goal together with TA-WSD and TA-CI is the establishment of cost and compensation models that enable the sustainable provision of decentralized storage and computing resources related to the DME.

WP MDI-5.1 Coordination: Coordination of the activities in this task area and with the other TAs as well as other initiatives related to research data infrastructures.

WP MDI-5.2 Requirements Analysis & Feedback: The thorough and detailed analysis of the technical, structural and user requirements for the DME is key to its successful implementation and wide acceptance within the MSE community. The technical requirements are related to the existing local IT infrastructure as well as to the local data generating devices and workflows. The requirements regarding the structure of the DME are closely related to the underlying ontology (TA-OMS) and need to address process-property-performance data and relationships as well as data relating to composition, microstructure and its evolution. Further requirements stem from the user, e.g., with respect to bibliographical metadata and usability aspects. These requirements will be analysed in detail at the beginning of the NFDI, in particular together with the PPs in the framework of the IUCs, and together with TA-CI we will set up an ongoing feedback process to adjust the requirements, including to newly developed technologies.

WP MDI-5.3 Knowledge Base & Dissemination: Besides the documentation for the implemented services, interfaces, software and installation procedures we will provide and FAQ, tutorials, training materials and best practice guidelines. These will be designed together with TA-CI, who will also be in charge of organizing the workshops, outreach and dissemination events, where members of the TA-MDI will serve, e.g., as trainers. A special focus of this WP will be together with TA-WSD to provide recommendations and examples for integrating existing data collections into the DME.

WP MDI-5.4 Support: We will offer 1st and 2nd level support for the DME including a ticket system and a telephone hotline. We will also offer on-site installation support to selected early adopters who will serve as multipliers and directly involve research groups of the use-cases.

WP MDI-5.5 Infrastructure Use Cases: While most IUCs require the implementation and basic adaptations of the DME that will be managed in WP MDI-5.4, selected PPs and IUCs, see sect. 5.2.8, will be an integral part of the TA-MDI. They will be directly involved in the development and testing of the DME as well as in all other measures. Special attention will be paid to the handling and integration of legacy data to allow exploitation of existing data collections. All corresponding activities will be coordinated through this WP. This includes providing funding for personnel in PP02 and PP18, which is matched through funds from the PPs.

WP MDI-5.6 Operation Model, Cost & Compensation Model: Development of a cost and compensation model for peer researchers accessing data & compute resources across organizations and together with TA-SD and TA-CI establishment of a consortium and community agreement to support the model. Such models are required to keep a record of services provided, their respective costs and legal terms under which they are available. The kind of compensations can be twofold: On the one hand monetary to directly compensate the usage, however in a “mutual-consideration” scenario third party usage may be scheduled in idle times and therefore does not necessarily inflict direct costs. More importantly the scientific compensation and acknowledgement need to be incorporated into the sharing workflows.

[redacted]

5.2.5 Interfaces and Interdependencies with other Task Areas

As the TA Materials Data Infrastructure is in charge of the technical implementation of a digital representation of materials which included information on their (micro-)structure, properties and all related processing steps as well as information related to characterization, testing and simulation methods, it is most closely connected to TA-OMS and TA-WSD.

Specifically, the implementation of the DME relies on the knowledge graph developed in TA-OMS. To ensure early operationality of the DME, a simplified/basic knowledge graph developed by TA-OMS will serve as basis for the DME, see WP OMS-3.1. Information obtained during the testing and roll-out of the DME in the use cases will directly feed into the ontology development. The implementation, e.g., through a graph database will allow for the flexibility to adapt and extend the underlying the knowledge graph according to the developments of TA-OMS. The use of the DME for tools and workflows developed by the TA-WSD or the greater community will be enabled through the definition of interfaces in TA-MDI. The web fronted implemented and hosted by TA-MDI will furthermore feature and 'App-Store' in which software and workflows developed TA-WSD and the community can be shared. It will furthermore be the basis for the community hub run by TA-CI. The community will furthermore be included in the development of cost models for shared computational and storage resources.

The DME and all services will be elaborated in close collaboration with TA-WSD and TA-CI. With both we will ensure constant alignment of the design of the DME and community requirements. Strong links between the broader MSE community and the infrastructure providers are furthermore assured by the co-applicants forming the TA-MDI which combines material scientists and heads of computer and data centers.

5.2.6 Contribution to Cross-Cutting Topics of the NFDI

The objectives and measures in the task area materials data infrastructure form the basis for the standardization and easy sharing of research data from the MSE community, which allows for data searches and reuse, e.g. through performing local computations, all while securing data integrity, authorship and correct attribution by performing analytics on data usage. While we take the unique requirements from this community very serious in the requirement engineering process, other scientific communities will without doubt benefit from the enacted measures. Apart from reusing the resources and best practices created in the different measures, we most prominently see the following contributions to the cross-cutting topics of the NFDI initiative:

- Authentication and authorization are a central challenge in federated infrastructures, while we focus on a highly decentralized setting, the NFDI can profit from the base scheme for authentication.
- Sharing of data and compute resources will focus on user-interfaces specific to the MSE community. The interfaces defined for packaging and exchanging data and compute jobs or packages, however, can form a common infrastructure used by infrastructure providers or researchers of all scientific disciplines.
- Metadata services in principle allow storage and retrieval for arbitrary disciplines. Sharing metadata in standardized formats like RDF, JSON and XML using standardized interfaces and protocols will

therefore become a central building block for the overall NFDI initiative according to the Research Data Commons (Glöckner, 2019). The seamless integration of the FAIR Digital Object concept (RDA DFIG, 2020) is an essential component of the EOSC Interoperability Framework (EOSCFAI, 2020) that will influence the NFDI as a whole.

- Storage of research data is central to all scientific and engineering disciplines. We will establish the basis for a highly scalable storage infrastructure that allows for the sovereign participation of infrastructure providers and research groups.
- Tracking of data use and data citations (analytics) is similarly a topic, which is central to the NFDI.

Consequently, we will develop our approach in close collaboration with other initiatives.

5.2.7 Appropriateness of the Task Area's Representatives

Co-Applicant **Prof. Matthias S. Müller** is the head of the Chair for High Performance Computing, director of the IT Center of RWTH Aachen University and on the Board of Directors of Jülich Aachen Research Alliance Center for Simulation and Data Science (JARA-CSD). His research interests include programming methodologies, software development tools and computational science on high performance computers. The University of Excellence, RWTH, is home to two Clusters of Excellence (CoEs) with engineering focus. RWTH's RDM team formed by IT Center, University Library and researchers collaborate on joint projects, consulting and governance since 2014. The resulting RDM strategy includes data stewards in three CoEs and three CRCs and central coordination by the IT Center to improve digital support for research (IdM.nrw, HPC.nrw, AcademicGroupware.nrw) and cooperate with RDM experts from TU9, CESAER, DH-NRW, DINI/nestor, and RDA. Within JARA-CSD, NFDI4Ing and NFDI4Chem, compute- and data-infrastructures are made available to a broad scientific user spectrum across Germany. As part of NFDI-MatWerk Prof. Müller will connect researchers and operators of existing research infrastructures on national and international level.

The Institute for Applied Materials (IAM) at the Karlsruhe Institute of Technology (KIT) with its links into the chemical, electrical and mechanical engineering faculties and its tight connections with physics and chemistry acts as liaison between materials research and the application of materials information in engineering. IAM is fundamental to the two CoEs at KIT (EXC 2082 '3D Matter Made to Order' and EXC 2154 'POLiS - Post Lithium Storage') and hosts several Research Training Centers. **Prof. Peter Gumbsch** from IAM has his scientific background in connecting materials modelling and simulation with experiment and application. He is also director of the Fraunhofer Institute for mechanics of materials IWM and chair of the Fraunhofer Group MATERIALS. In his dual function in academia and application-oriented research he provides the important links between the academic materials science community, generating materials data by experiment and simulation and the users of materials data in the academic engineering communities and in industry. He is spokesperson of the BMBF platform MaterialDigital (2019) and one co-author of the Fraunhofer Materials Data Space Initiative (2016).

The Steinbuch Centre for Computing (SCC) is the computing and data center of KIT with cutting-edge R&D in data science and research data management. With the Grid Computing Centre Karlsruhe (GridKa) SCC is responsible for the storage and analysis of a significant part of data from the LHC (Large Hadron Collider) experiments. **Prof. Dr. Achim Streit** is director at SCC and professor for “Distributed and Parallel High-Performance Systems”. His expertise and research focus lies in data management, data-intensive computing, distributed & federated computing, data federations, parallel systems and distributed resource management. He is Topic Speaker “Data-Intensive Science and Federated Computing” in Helmholtz Programme “Supercomputing & Big Data”, coordinator of the Helmholtz Data Federation, member of the Executive Board of EUDAT and of the Helmholtz Information & Data Science Academy (HIDA) as well as of the Helmholtz Federated IT Services (HIFIS). Through continuous contributions in RDA (RDA, 2020), EUDAT-CDI (EUDAT, 2020), AARC2 (AARC2, 2017), NFFA Europe (NFFA, 2020), re3data (re3data, 2020), RADAR (RADAR, 2020), DARIAH-DE (DARIAH, 2020), many EOSC-projects, e.g. (EOSC Portal, 2020)(EOSC Hub, 2020)(EOSC Secretariat, 2020), and NFDI consortia, e.g. NFDI4Ing and NFDI4Chem, SCC has been at the core of data infrastructure services and policy developments in Germany and in Europe. Thus, the expertise stemming from existing and internationally approved solutions for data infrastructure will be used to build the Materials Data Infrastructure.

The Co-Applicant **Prof. Erik Bitzek** at the FAU Erlangen-Nürnberg will act as liaison between materials scientists and engineers as users and providers of data and the computing centers as providers of materials data services. He will furthermore closely coordinate the activities of the TA-MDI and the TA-WSD. His expertise lies in the integration of experimental and simulation data. He is part of the research training network GRK1869 “in situ microscopy” and co-speaker of the GRK 2423 FRASCAL - Fracture across the scales. As PI in the PPs CRC/TRR 103 (PP01) and CRC 1394 (PP02) he will furthermore directly interact with the corresponding IUCs. He is part of the steering committee of the Working Group “Digital Research Data and Research Information” (AGFD, 2020) of the FAU, which is responsible for developing FAU’s RDM strategy, including the establishment of new research data/information services. Currently, AGFD employs two FTEs, who will support the NFDI-MatWerk coordinator and computational engineer at FAU.

5.2.8 Connection to Infrastructure Use Cases

All IUCs will in one way or the other benefit from the DME developed in this TA.

The DME will feature prominently in **IUC02 - Framework for curation and distribution of reference datasets,** which will mostly focus on implementing, testing and highlighting the concepts of the DME relating to data integrity, provenance and analytics. In particular, in IUC02 a strategy and a framework for the trusted identification and distribution of reference data as well as procedures to assess data quality will be developed. A researcher partly financed through TA-MDI (50% TA-MDI, 50% BAM) will be in charge of the technical implementation of the metadata schemes for full documentation of material history and how the data were collected. Together with the TA-MDI she or he will furthermore work on developing methods to guarantee the integrity of the data and its authorship through cryptography and methods for tracking of data use, data views and data cites. The TA-MDI will provide defined interfaces and protocols to register local stored datasets (and repositories) and recommendations and support for data management and workflows for sustainable maintenance (to harvest changing local datasets). [redacted]

Quality assurance through IUCs: In addition to the usual unit and integration tests during the development of the DME, all individual components of the DME, as well as their interplay, will be acceptance-tested with selected PPs according to the strategy described in Section 2.2. These initial tests will be closely coordinated with the quality assurance measures in TA-WSD to validate the interfaces between the DME and developed software and workflows and to assess their interplay and performance. To demonstrate that the DME can be integrated throughout multiple institutions with different IT and research environments, large collaborative PPs like the CRC/TRR103 (PP01) and CRC 1394 (PP02) that contribute with their personnel will be the first beta-testers. With their multitude of experimental as well as computational methods and types of data, they will be able to validate the universal applicability of DME to materials data. Furthermore, the different material classes and applications covered by the PPs allow to examine challenges posed by different ontologies. The different stages of the PPs allow to assess the appropriateness of the DME for legacy and “living” data in recently started projects, and the large number of involved researchers and labs in the PPs allow to test performance and scalability of the DME. The PPs will contribute to the further development of the DME, provide best-practice examples and act as multiplicators. [redacted]

5.2.9 Funding Request for TA Materials Data Infrastructure

Table 5.2.9.1: Funding Request for Task Area Materials Data Infrastructure by Institution

	2021 (Oct-Dec)	2022	2023	2024	2025	2026 (Jan-Sep)	Total in €
Institution		**Totals in €**					
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 5.2.9.2: Funding Request for Task Area Materials Data Infrastructure by Funding Category

Funding category	2021 (Oct-Dec)	2022	2023	2024	2025	2026 (Jan-Sep)	Total
Staff by category	Number of persons (full-time equivalents)						
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Totals in €						
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

(*) this corresponds to the DFG staff category “Postdoctoral researchers and comparable”

(**) this corresponds to the DFG staff category “Doctoral researchers and comparable”

5.3 Task Area Materials Workflows and Software Development (TA-WSD)

5.3.1 General Objectives

The vision of the NFDI-MatWerk initiative is to establish a national Digital Materials Environment (DME), in which research data, corresponding metadata and workflows in MSE are shared between institutions, made long-term available after projects, and are semantically and quantitatively described in a reliable and machine usable form. In this vision, computers can answer the question about a property of a material X. The successful realization ultimately depends on the availability of software that supports this vision. Today, a wide variety of software tools exists in the community, while exactly this abundance makes interoperability and the exchange of comparable and compatible data problematic. TA-WSD will provide software interfaces for a common ontology to facilitate links between different tools, methods, and data. The software foundations provided will boost both productivity and quality of the data processing and simulation aspects of MSE research. We will address the following issues:

(1) **Reproducibility** is a major concern **in experimental MSE**. Frequently, in publications, the data processing steps of experiments are described in a rather qualitative way, which in practice often render results “unreproducible”. Even if sufficient detail is provided, the data processing environment

is typically not readily available to everybody. For example, review of data processing is, thus, typically limited to a plausibility check. To solve this issue, the data should be processed in a **virtualized processing platform,** i.e. a digital notebook. Such virtual, standardized environments can be executed on servers (personal, per team or institution), which allows controlled-access sharing. Data processing tools can then be accessed by simply following a link to the scripts. Comparable workflows have already been implemented in other fields using, e.g., Jupyter notebooks. Since Python programming is not sufficiently widespread in MSE, we will provide a graphical programming language as an interface between Jupyter and widely used software tools.

(2) The **simultaneous application of various codes to combine materials simulations** on the atomic, meso and continuum scales represents another challenge in MSE. For a seamless connection and interoperability, generic (i.e. not code-specific) user interfaces are required. While software-specific interfaces as well as parsers ensure the communication with the code, unifying standards of all input and output data are needed. Ideally, they should be automatically determined by the materials ontologies. The **software environment** should also allow for an automated submission of individual calculations as well as complex simulation protocols and an automated analysis of the output data based on user-defined materials models.

(3) A challenge commonly encountered in both modelling and experimental studies are errors due to **numerical precision and statistical/systematic errors**. Other than software bugs, such errors are an inherent part of research in MSE. In simulations, implementations of most software packages do not allow for the determination of such errors. In experiments, novel high-throughput approaches produce large amounts of data, while tools to efficiently curate them, consistently identify outliers or inconsistencies do not yet exist. Furthermore, the propagation of errors in complex simulation and experimental workflows can nowadays not easily be considered. We want to contribute a solution to this problem by implementing simulation protocols and developing the required software components for this to provide efficient and user-friendly tools that address all aspects of error estimation and propagation.

(4) With the increased importance of **machine learning techniques** in MSE, quality control of research relying on these techniques becomes an issue and a data infrastructure for machine learning is required. While statistical data requirements are well understood and most scientists are aware that, for example, computing a reliable p-value implies minimal requirements on the sample population size, comparable demands are frequently ignored when using Deep Learning. We will contribute a solution to this emerging issue by providing an **online platform for research data management in Deep Learning** that automatically performs plausibility checks and enforces certain minimum standards on data handling. For example, the platform will ensure that the available data is split in training, validation and holdout sets and that the split is properly randomized. It will enforce that the network is not changed

after accessing the holdout data, give warnings if the total data size is insufficient or if the data contains duplicates.

Our objectives will be pursued in close interaction with TA-OMS, since the ontologies define and standardize the physical parameters and the relations in-between. A central theme of the TA is the implementation of ontologies in a data-driven way, i.e. by reading from files rather than hard-coding dependencies into the workflow. Ontology information could then be exchanged on the fly. Furthermore, TA-WSD will use the knowledge graph of metadata defined in TA-MDI.

TA-WSD is supported by the Max Planck Institute für Eisenforschung in Düsseldorf, the German Research Center for Artificial Intelligence in Saarbrücken, and the Forschungszentrum Jülich. A number of different Infrastructure Use Cases will be considered to align the tasks with the demands of the community. On one hand, techniques and software solutions that are of general interest for the whole community will be developed. On the other hand, the new tools will be applied, tested, and improved in exchange with specific materials science problems.

5.3.2 Relevance for the Consortium's Objectives

The concept for digital platforms as envisaged by the NFDI-MatWerk consortium, has to take the hierarchical structure of materials and the history of samples and data processing into account. The semantic description of data in terms of ontologies provides the conceptual framework. This intrinsically requires software solutions that can handle the underlying diversity. Different tools are currently used by the participants, in order to conduct research on the different length and time-scales, e.g., codes for density functional theory, computational thermodynamics, phase field simulations, crystal plasticity or finite element simulations. Furthermore, software solutions exist to, e.g., analyse chemical gradients in atom probe tomography, create microstructural maps in electron backscatter diffraction, or visualise stress-strain curves determined from different mechanical tests.

The wide variety of tools used makes the reproducibility of results and the transparency of data problematic underscoring the necessity to handle the entire input/output data of the codes in the digital platform. In contrast to many other communities, the inherent multiscale and multidisciplinary character of applied materials requires a framework that can combine tools that are *per se* not interoperable. The few currently available software solutions are by far not sufficient for the manifold needs of the community. To meet these challenges, the community sees a demand for improved software solutions in areas related to experimental data processing and simulation. The following contributions have been identified:

(1) The software solutions should combine experimental results with metadata of the measurements (e.g., in the form of a digital twin) and to automate complex workflows for data processing in a way that is compatible with the community standards, i.e. without writing source code for every task.

(2) Integrated development environments for complex simulation protocols are needed that connect the aggregation and processing of materials data with the corresponding ontology.

(3) A seamless connection between the different inherently incompatible codes ensuring a straightforward data exchange between them is needed, to handle the variety of multiscale and multidisciplinary approaches in MSE.

(4) Data processing environments should be virtualized and version-controlled to make data processing workflows repeatable and shareable. This represents an important step towards limiting the impact of software bugs on research results.

(5) Software solutions are required for the central (per group or institution) storage of metadata (including data processing) in a self-explanatory, understandable format to avoid a loss of methods and sample histories as researchers change labs.

(6) Standards and a software foundation to connect numerical simulations with quantitative error bars need to be established.

Most points listed above address software solutions and data management simultaneously. It is a central idea of NFDI-MatWerk to unify both aspects, requiring a close link with TA-MDI.

5.3.3 Status of this Task Area's Field in MSE

Addressing both experiment and simulation, this TA aims at the integration of the wide variety of available software tools used in MSE into a common environment. Due to the inherently digital nature of computational simulation data, to date many more frameworks for numerical data analysis pipelines exist than for experimental ones.

To foster the development of simulation protocols, their dissemination and use, a framework called pyiron (Janssen et al., 2019) has been developed at MPIE that allows for the automation of routine tasks, stores all input/output data of large numbers of individual jobs in a generic format together with the simulation protocols and the hierarchy of the tasks. While originally targeting thermodynamic simulations, the framework is highly versatile and can be applied by a broad community.

Similar abstraction layers have previously been built, starting with the atomic simulation environment (ASE) (Larsen, 2017) and high frameworks such as AFLOW (Curtarolo, 2012), the materials project infrastructure with Fireworks (Anubhav, 2015), pymatgen (Ong, 2013) and ATOMATE (Mathew, 2017) or AiiDA (Pizzi, 2016) and commercial frameworks like Materials Studio (BIOVA, 2020), or MedeA (MedeA, 2020). Based on one or several of these frameworks, more specialized tools have been developed. Examples are Materials Simulation Toolkit (MAST) (Mayeshiba, 2017), MPInterfaces (Mathew, 2016) and Jarvis (Choudhary, 2017), which are all based on pymatgen and compatible with ASE. Additionally, stand-alone tools that have not been derived from any of these platforms exist. Examples are pylada to setup and analyse defect calculations or openkim and atomman to provide user-friendly interfaces for empirical potentials. All of these tools aim at automating calculations and implementing best practice workflows of complex, error prone, and labour-intensive steps.

Calphad codes (i.e., Calculation of Phase Diagrams, e.g., ThermoCalc), which evaluate large amounts of thermodynamic data obtained mainly from experiments, represent some of the most successful software solutions. However, automation in data processing is limited, the data format (TDB files) very strict, the flexibility for implementing new physical models very limited, and there is hardly any handling of uncertainties. The same shortcomings are typical of most software solutions in MSE, in particular with respect to the integration of experimental data.

5.3.4 Measures and Work Packages

Measure WSD-1 Coordination and service (*MPIE)*

To support the general philosophy of NFDI-MatWerk, several software solutions need to be available for the community. They will be implemented, maintained and further developed by TA-WSD, in close collaboration with TA-MDI.

WP WSD-1.1 Coordination: Coordination of activities in this TA and with other TAs, with other initiatives related to workflows and software development, and with participant projects.

WP WSD-1.2 Version Control: The TA will ensure that a distributed version control system (e.g., Git) for software solutions is available for the community.

WP WSD-1.3 Automated Installation: An automated installation of tools together with their framework will reduce the barrier for applications in the community. The TA will set up an **Anaconda** environment for this purpose.

WP WSD-1.4 Execution Framework: The development of a minimalistic framework to setup and benchmark workflows in connection with ontology solutions is required. Hence, the operation of **Docker containers** at local or centralized servers is supported.

WP WDS-1.5 Support: Continuous support of participants, e.g., in terms of Infrastructure Use Cases, within the NFDI-MatWerk to make use of the following measures.

[redacted]

Measure WSD-2 Graphical Programming Language (*DFKI)*

To support objective (1), we will develop a graphical programming language that allows researchers in MSE to perform image processing and data analysis tasks related to the microstructure characterization in an automated, repeatable and transparent way based on Jupyter notebooks and the SciPy stack, both of which have become increasingly popular within the MSE community. Since normal operations of Jupyter notebooks require writing Python code for every task, we will instead provide a graphical programming language as an interface to the functionality. This language will use a graph-based metaphor to programming, an approach that has shown very good user acceptance, for example in the well-known LabView environment and in shader networks in computer graphics. Additionally, we will integrate functionality provided by the widely-used image processing software ImageJ as nodes in the system.

WP WSD-2.1 Basic System: We will setup the development environment and perform a technical integration (vertical slice) of the software. This includes the setup of code repositories, build server, and test automation infrastructure.

WP WSD-2.2 Graphical Language: We will implement the basic features and user interface of the graphical programming language, including a basic version of the type system and data transfer system, i.e. implementing methods to transfer data from one compute node to the next.

WP WSD-2.3 Jupyter and SciPy Integration: We will integrate Jupyter and SciPy and make a selection of SciPy compute nodes available to the programming language.

WP WSD-2.4 ImageJ (Fiji) Integration: We will integrate the ImageJ (Fiji) software to the graphical programming language.

[redacted]

Measure WSD-3 Version Control, Centralized Environments, and Access Control (*DFKI)*

We will develop a platform that provides version controlled configurable access to a data processing environment. Both version- and access control relate to the data processing platform described in measure 1 providing controlled access to the *software,* in contrast to the Digital Material Environment (DME) developed in TA-MDI, which provides controlled access to *data*. The platform will be based on virtualization and provide operating system instances with pre-installed software versions on a central server (installed per research group).

WP WSD-3.1 Basic System: We will implement a system that automatically performs a transparent version control on any data processing script that the user creates in the context of the systems developed for measure 1.

WP WSD-3.2 Access Control: We will extend the system to functions for sharing and managing access to its data processing to external users (for example reviewers) in a simple and usable way.

WP WSD-3.3 Citeability: We will implement procedures to make the graphical representations of the data processing workflows citeable. This part of the measure will be implemented in cooperation with Saarland University library. A shared data processing sheet can then obtain a permanent identifier like a DOI and be cited directly.

[redacted]

Measure WSD-4 Integrated development environments *(MPIE)*

To support objective (2) of the task area we will develop frameworks for the handling of simulation workflows. The software architecture should follow the philosophy of a software bus with well-defined connections and communications of software tools. A Python solution via Jupyter notebooks is foreseen as the best solution. To capture the multiphysical character of materials, a centralized framework for all materials science applications is planned. On top of this, the hierarchical nature of materials will be reflected by hierarchical software solutions accounting for the specific workflows.

WP WSD-4.1 Job management: For the execution of individual calculation tasks (=jobs) on high performance computer (HPC) clusters, a generalized connection to queuing systems, both for serial as well as parallel jobs will be provided.

WP WSD-4.2 Data storage: In collaboration with TA-MDI a generalized interface for communication with the digital materials environment (DME) will be developed, which takes care of the hierarchical and often unstructured character of data in MSE. A flexible and efficient frontend will be developed that ensure a seamless and user-friendly navigation through various data storage formats such as file systems, databases, HDF5 etc. For the analysis of aggregated data, a generalized data management in terms of Python Pandas tables will be developed.

WP WSD-4.3 Interface to Ontology: Software solutions will be developed that allow a direct interpretation of ontology files (e.g., OWL) to categorize and store the data into the framework. Categorizing data directly by ontology files will provide enough flexibility and user-friendliness to cover the full range of research tools within MSE.

WP WSD-4.4 Parser: Parsers will be written to translate the code specific in- and output into a generic format according to the ontologies specified by the TA-OMS.

WP WSD-4.5 IDE: A command-line driven integrated development environment (IDE) will be provided that incorporates the aspects given in WPs 4.1 – 4.4. It will allow the combination and flexible exchange of tools, the automated submission of jobs in high-throughput simulations, the collection and the analysis of data. The workflow will be made consistent with the graph-based RDF (Resource Description Framework) provided by TA-MDI.

[redacted]

Measure WSD-5 Application workflows in experiment and simulation (*FZJ)*

The establishment of robust workflows represents a very fundamental need of the MSE community, which will improve the transparency, comparability, and reproducibility of both experiment and simulation thereby accelerating the development of new materials. The core research tasks consist of synthesis, processing, characterization and performance evaluation, all of them related to very specific techniques, and have to be combined with data management, data interpretation, and quality control. The field of modelling and simulation provides new insights into fundamental processes as well as quantitative predictions often in combination with experimental analysis and microstructural characterization. The need for well-defined workflows in MSE is reflected in a number of IUCs, through which we will test and continuously adapt the tools and methods developed.

WP WSD-5.1 Mechanical testing and characterization workflows: Experimental workflows for selected, widely-used experimental devices, e.g., tensile testing equipment, nano/microindenters, and SEMs/FIBs, will be designed and implemented. In coordination with TA-CI and collaboration with TA-MDI, we will create and develop the tools facilitating generalization and standardization of analysis methods accounting for the different data formats and automatically extracting all metadata.

WP WSD-5.2 Workflows for high-throughput and data-intensive experiments: The basic experimental workflows will be extended to cover high-throughput methods, experiments conducted at large-scale infrastructures, and data-intensive characterization methods, such as tomography.

WP WSD-5.3 Electronic lab books (ELNs): In cooperation with NFDI4Chem and building upon the open source ELN developed within NFDI4Chem, an ELN for MSE will be developed tailored to the requirements of MSE. The ELN will be designed to track the planning, execution and results of experiments including the raw data, analysis results, and metadata. Additional modules, e.g., for the creation of reports and the provision of research data will be provided.

WP WSD-5.4 Workflows for multiscale simulations: A generic workflow for multi-scale simulations will be developed. The simulation methods included are diverse covering different length and time scales, including phase field methods to model the mesoscale behaviours, molecular dynamics methods for the fundamental atomic-scale processes, crystal plasticity and finite element simulations describing the macroscale. Focus will be put upon identification and automatic extraction of relevant data and metadata.

[redacted]

Measure WSD-6 Data curation, numerical precision and statistical/systematic errors *(MPIE)*

To support objective (3) of the task area, various concepts for the analysis of errors will be implemented.

WP WSD-6.1 Error bars: A software infrastructure will be provided that supports or enforces the joint handling of data points with their corresponding error bars.

WP WSD-6.2 Parser for experiment: A highly flexible framework will be developed to semi-automatically write adapters/parsers that convert commonly highly unstructured data arising from multiple experiment-specific data sources into a common database system (NoSQL, MongoDB)

WP WSD-6.3 High-throughput: Software tools together with an easy-to-use graphical user interface will be developed that allows also non-experts to quickly assess and curate the quality of huge experimental data sets that result e.g. from high-throughput materials screening experiments.

WP WSD-6.4 Convergence: Specifically, for materials simulation tools a framework will be provided to analyse the sensitivity of results with respect to convergence parameters (e.g., mesh sizes) intrinsic to the employed tool, but controllable by input files.

WP WSD-6.5 Error propagation: The integrated development environment in Measure 3 will be extended such that an automized evaluation of error propagation throughout the workflow is possible.

WP WSD-6.6 Data consistency: Available concepts will be established for the evaluation of data veracity, exploring data inconsistencies, data incompleteness, ambiguities, the latency up to the deception of data.

[redacted]

Measure WSD-7 Machine Learning Platform *(DFKI)*

To support objective (4) of the application area, we will develop an online platform for the management of machine learning data and tasks. The platform will allow a user to upload data, manage data, and execute deep learning tasks like training and evaluation while enforcing some scientific standards.

WP WSD-7.1 Basic Platform: We will implement a first version of the platform that allows the upload of image data. The system will support to automatically split data in training, validation and holdout data and enforce proper randomization.

WP WSD-7.2 Integration with DME: We will integrate the basic platform in the Digital Materials Environment (DME, see description of TA-MDI). From the point of view of the machine learning platform developed in measure 6, the DME will take the role of an additional storage backend.

WP WSD-7.3 Data Splitting Support: We will extend the system to allow the specification of the launch of training and evaluation tasks. The system will ensure that the model remains fixed after the holdout data has been accessed.

WP WSD-7.4 Validation Engine: We will implement a rule-based engine comparable to a static code analysis (or warning system) in conventional programming. When launching a training task, the system will check the task against a rule engine that provides feedback of the amount of training data is suitable for a given network architecture, suitability of the architecture and meta-parameters.

[redacted]

5.3.5 Interfaces and Interdependencies with other Task Areas

The Task Area Materials Workflows and Software Development (TA-WSD) is connected to all other task areas of the consortium. A strong link to the Task Area **Materials Data Infrastructure** (TA-MDI) is required to ensure consistent handling of data. TA-MDI creates the (ontology-specific) framework for the handling of metadata, which is subsequently filled by the software solutions of TA-WSD. To this end, the graph-based RDF format used in TA-MDI needs to be closely linked to seamless and user-friendly navigation through various data storage formats envisaged in WP4.2 of TA-WSD. Primarily, the metadata required for data infrastructure (e.g., location of files of physical samples) as well as the transdisciplinary/generic metadata (e.g., creator of data, timestamp, publication) will be handled by TA-MDI. The MSE-specific metadata (e.g., chemical composition, sample history, units, errors) will be accounted for by the workflows developed in TA-WSD.

The consideration of MSE-specific metadata requires close interaction with the Task Area **Ontologies for Materials Sciences** (TA-OMS). Ontologies, i.e., all definitions of data types, physical units, interdependencies between data, etc. in the software solutions will follow the suggestions of TA-OMS. The hierarchical nature of materials data and corresponding software solutions, render the reliable, transparent and efficient treatment of ontologies indispensable. While a number of separate tools for ontologies and data- and workflow management systems exist, it is the goal of NFDI-MatWerk to establish a unified framework facilitating the seamless connection between them. The hardcoded ontologies in software solutions will be replaced by information transfer from ontology representations, e.g. Web Ontology Language (OWL) files. Software solutions required for the development and operation of ontologies (e.g., VoCol) will be supervised and further developed by TA-WSD.

To establish the developments of TA-WSD in the MSE community, intensive training programs and information of the community is necessary. Therefore, strong interaction with the Task Area **Community Interaction** (TA-CI) will be ensured. The technical content of the workshops will be provided by the TA-WSD, while the organizational issues are handled TA-CI.

Within TA-WSD, the TA coordinator will be mainly responsible for the interactions with the other TAs. He/she will be in close contact with the Task Area **Strategy Development** (TA-SD) to ensure concerted actions within the NFDI-MatWerk initiative.

5.3.6 Contribution to Cross-Cutting Topics of the NFDI

The objectives and measures described within the TA-WSD form the basis for interaction of users with materials data infrastructure. The software solutions provide an interface to individual experimental setups as well as simulation codes. Within the software frameworks, workflows for the processing of data are described, connecting the raw data with the published results rendering them transparent and reproducible facilitating, e.g., the knowledge transfer between different generations of PhD students. The software solutions ensure that aggregated data and metadata are automatically incorporated into the digital materials environment (DME) and therewith can be shared between research units. Although these developments are designed for the MSE community, they address cross-cutting topics that are of general interest to the NFDI. In addition to these points, the following objectives will have an impact on other communities:

- Software solutions will be developed that allow a direct interpretation of ontology files (e.g., OWL) to categorize and store the data in the framework. This provides flexibility and user-friendliness to cover a full range of research tools ensuring at the same time the consistency of data.
- The software development in NFDI-MatWerk will focus on the consideration of data curation, numerical precision and statistical/systematic errors. This includes concepts of version control of software solutions.
- The quality control of research relying on machine learning techniques represents an important issue in data science in general. The MSE community faces special challenges, such as small sample population sizes, but the planned platform for the management of machine learning data and tasks will be of general use.

A close interaction with other consortia, e.g. FAIRmat, NFDI4Ing, DataPLANT, and MaRDI (see sect. 2.3), on these topics has been agreed. Joint workshops are planned to ensure an intensive exchange of knowledge and consistency of the implementations.

5.3.7 Appropriateness of the Task Area’s Representatives

The **Max Planck Institut für Eisenforschung GmbH** (MPIE) conducts basic research on high-performance materials, in particular metallic alloys and related materials. MPIE develops and applies novel experimental and simulation to explore the correlation between the highly complex micro- and nanostructure of advanced materials and their properties under realistic environmental conditions. **Dr. Tilmann Hickel** is head of the group “Computational Phase Studies” in the department “Computational Materials Science”. His expertise is the development of complex automatic workflows for the simulation of thermodynamic properties of structural and functional materials and the combination of *ab initio* methods with thermodynamic databases, such as Calphad. A core development in the department, headed by **Prof. Jörg Neugebauer,** is an integrated development environment (IDE) for the creation, editing and conservation of complex simulation protocols. This Python-based framework, called pyiron, is specifically designed for interactively developing and testing complex physical concepts and extending them to high-throughput calculations in combination with a hierarchical data management. It has led to a standardization of the handling of data and software tools within MPIE and in cooperation with various partners. Pyiron was made available to the scientific community in 2018 through an open-source license (www.pyiron.org).

The **German Research Center for Artificial Intelligence (DFKI)** is a leading organization for innovative software technology in Germany. Focused on artificial intelligence in scientific and industrial applications, the institute provides cutting edge research in artificial intelligence topics with industrial grade software engineering processes. In this TA, DFKI is represented **by Prof. Dr.-Ing. Philipp Slusallek,** who is Scientific Director of the research area *Agents and Simulated Reality* and Site Manager. His research focuses on the application of high-performance computing and real time raytracing to the synthetic generation of training data for machine learning applications. The operational responsibility for the TA will lie with the team *Computational 3D Imaging* (headed by **Dr. Tim Dahmen,** whose research focuses on machine learning applications and image reconstruction in sparse- and adaptive scanning microscopy) and the team *Distributed & Web-based Systems* (headed **René Schubotz,** who is an expert in scalable full-stack platform solutions for compute- and data-intensive applications). Tim Dahmen and René Schubotz co-authored the DGM strategy paper “Digitale Transformation in der Materialwissenschaft und Werkstofftechnik”.

The **Institute of Energy and Climate Research – Microstructure and Properties of materials (IEK-2)** of the Forschungszentrum Juelich GmbH, develops and characterizes high-performance materials and composites for energy conversion and storage systems. By combining experimental and simulation-and-modelling approaches, IEK-2 investigates microstructure-properties relations of novel materials on length scales ranging from the nano to the macroscale. **Prof. Dr. Ruth Schwaiger** is director of the IEK-2 and Professor for Energy Engineering Materials at RWTH Aachen University. Her research interests include deformation and failure mechanisms in materials, the mechanics of small-scale structures, and mechanical metamaterials, and aims to develop a mechanism-based understanding of deformation and failure by establishing a strong link between experiments and simulations. The operational tasks for the TA will be headed by **Dr. Steffen Brinckmann**, who is head of the group *Micromechanics and Microtribology*. He is an expert for the development of complex tailored experiments and data analysis methods, and aims at establishing microstructure-property correlations together with a digital, mechanism-based description of the materials investigated.

5.3.8 Connection to Infrastructure Use Cases

The Infrastructure Use Cases (IUC) associated with TA-WSD serve two main purposes: (i) provide techniques and software solutions that are of general interest to the whole MSE community complementing the measures established by the TA, (ii) apply the software solutions and workflows developed to specific materials science problems and give feedback to the TA. In this way, the combination of data infrastructure and workflows will be tested and established in the community. In the following the relevance of selected IUCs for TA-WSD is described:

IUC04: This IUC requires the development of completely new workflows to create defect phase diagrams, since such an approach has not been established earlier. It is an ideal case for combining novel ontologies with novel software solutions for workflows that generate data according to the requirements of underlying models. For the management of experimental and simulation data and the entire research workflow a Python based tool and Jupyter Notebooks will be used (see WSD-4). By reusing metadata, the huge number of required phase diagrams can be determined simultaneously. [redacted]

IUC05: This IUC will test and evaluate a full experimental workflow applied to nanoindentation testing, and support its optimization. The goal is to provide a tool for storing and evaluation of raw mechanical test data while maintaining the typically "hidden" metadata, based on an appropriate ontology. Data analysis frameworks according to standards as well as new, emerging methods will be considered. An electronic lab book gathering all information regarding material synthesis, sample histories, methods and results will be adapted to the needs of the MSE community (in collaboration with NFDI4Chem). This IUC will serve as an example for general workflows appropriate for different characterization labs.

IUC06: This IUC provides a framework for the integration of materials data from experiments and computation into manufacturing data infrastructures for two polymer systems. The materials data will be integrated in a material database infrastructure allowing for the continuous evaluation of the compatibility of simulation models, which are part of a model hierarchy, and materials data. A framework coupling materials data and manufacturing processes with their digital twins will be provided.

IUC07: In this IUC the spatial and temporal (i.e. 4D) distribution of crystal defects and their embedding into three-dimensional field quantities like temperature, mechanical stress and strain, or chemical concentration fields will be considered. The aim is to develop an integrated post-processing software package that interfaces with existing developing software packages (e.g., Damask, OpenPhase) for full-field simulation studies. This will therefore support Measure WSD-4, but with a focus on unstructured (material) point cloud data for microstructures. Since a large number of time-dependent snapshots can make the volume and acquisition velocity quickly impractical, machine-learning techniques (WSD-7) will become necessary. A code-agnostic representation of the simulated defects can be achieved by a connection to the defect ontology that is developed in IUC17. [redacted]

IUC08: This IUC will develop visually based data analysis strategies and appropriate tools to identify, quantify and evaluate complex correlations between mechanical behavior and evolution of physical materials properties as well as microstructural features. The major challenge for data analysis apart from the large volume of data is the comparative analysis and integration of various target variables and analysis procedures. While subsets of data can be analyzed with existing mathematical techniques, the holistic analysis requires a combination of automated data analysis and interactive data exploration, visual communication, and knowledge-management systems. [redacted]

IUC09: In this IUC, workflows are designed in coordination with FAIRmat such that the exchange of data, metadata, ontologies, and concepts for data curation (measure WSD-6) between two NFDI consortia is ensured without a loss of information. Both consortia benefit from ongoing improvements in the infrastructure for condensed-matter physics (FAIRmat) and the interpretation of microstructure data (NFDI-MatWerk).

IUC10: Multiscale simulations represent a challenge for workflow systems in materials science (see WSD-5). They require the connection of various distinct data types as well as computer codes of different communities. For example, the cross-scale model for the processes of joining technology, which is developed in PP06 (Nina Gunkelmann), will be used to benchmark the workflow system in TA-WSD. The coverage of available tools is substantially improved, if a connection to workflow systems established in other NFDI communities is possible. This is here demonstrated for NFDI4Ing.

IUC11: In this IUC, a tool for condensing large datasets derived from dilatometry will be developed -- based on mathematical representations of the transformations occurring during heat treatment. An ontological representation encompassing both heat treatment courses and material transformation is designed to provide a common and unified description. Condensing data via parametrization will provide significantly reduced data sets and a direct representation of the measured data to be used in common FE-solver based mesoscopic materials simulation. [redacted]

IUC15: This IUC demonstrates how scale-bridging tomographic data of a sample can be combined and made accessible as a "digital twin". This includes modelling and simulation data that is derived from experimental data obtained with various techniques. The intended approach makes sure that results from one method can later be complemented by those from other methods. Data must be taken both from metadata supplied by the instrument or software used and from manual user input. A user-friendly interface which allows easy integration of new data into the digital twin, making use of the developed ontology, is the aim here (measure WSD-2). [redacted]

Quality assurance through IUCs: In addition to the usual unit and integration tests during the software development, all individual components of TA-WSD, as well as their interplay, will be integration tested with selected PPs. These tests will be closely coordinated with the quality assurance measures in TA-MDI to validate the interfaces between the DME and developed software and workflows and to assess their interplay and performance. Large collaborative PPs like the CRC/TRR 103 (PP01) and CRC 1394 (PP02) that contribute with their IT personnel will be the first testers, give immediate feedback and allow the personnel to adopt the software solutions. The different material classes and applications covered by the PPs allow to examine challenges posed by different ontologies. The PPs will contribute to the further development of the software solutions provide, best-practice examples and act as multiplicators.

5.3.9 Funding Request for TA Materials Workflows and Software Development

Table 5.3.9.1: Funding Request for Task Area Materials Workflows and Software Development by Institution

	2021 (Oct-Dec)	2022	2023	2024	2025	2026 (Jan-Sep)	Total in €
Institution		**Totals in €**					
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 5.3.9.2: Funding Request for Task Area Materials Workflows and Software Development by Funding Category

Funding category	2021 (Oct-Dec)	2022	2023	2024	2025	2026 (Jan-Sep)	Total
Staff by category	**Number of persons (full-time equivalents)**						
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Totals in €						
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

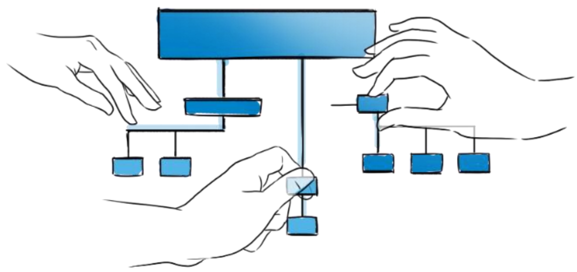
(*) this corresponds to the DFG staff category "Postdoctoral researchers and comparable"

(**) this corresponds to the DFG staff category "Doctoral researchers and comparable"

(***) This includes student research assistants at the height of 20% FTE

5.4 Task Area Ontologies for Materials Science (TA-OMS)

5.4.1 General Objectives



The microstructure of materials consists of information that can "live" on very different length and time scales. Furthermore, an important characteristic is that the microstructure permanently evolves in a complex manner, introducing a history dependence of the material properties. Therefore, characterizing the microstructural state is not sufficient for a complete description of a material or sample: the representation of the processes influencing the evolution is equally important. In this task area the following main objectives are identified:

- First, a **minimal description** of material properties, samples, experiments and simulations is developed. This enables our stakeholders **to develop, implement, standardize and use metadata in a systematic manner**. This minimal description will be accessible to material scientists as well as theoreticians and simulators at a low technical barrier. This first step is a prerequisite for the subsequent development of ontologies.
- Building on the minimal description, the major aim of TA - OMS is the development and integration of a global **materials ontology**. This ontology will enable interoperability across heterogeneous materials data, analytics tools and material models. The different **material specific length and time scales** have to be incorporated into the ontology in a flexible and seamless manner. A key objective of the base ontology is to ease the **implementation of** *private, more specific,* **materials ontologies**. This includes building homegrown tools as well as digital and experimental workflows. All of these benefit from the structure provided by the ontology in terms of interoperability and the ability of sharing results within the MSE community, particularly within NFDI-MatWerk. This comprises the **distribution and extension of private ontology branches** with partners and the greater community including a stringent version control.

Access to research data will be enabled via the implementation of a **material sciences knowledge graph**. It renders data characterized via their metadata **findable, accessible, interoperable on the basis of commonly used ontologies as well as reusable and, thereby, sustainable**.

Through these main objectives, the TA OMS will **create a unified description of materials and related experiments and simulations.** To reach this goal, it is additionally important to create a certain **cultural change**: An awareness for the value of data and for the necessity for "documenting" data and the concomitant processes (simulations, experiments) has to be created. This cultural shift will only take place if the stakeholders experience benefits by adding metadata or populating an ontology that outweigh the additional effort. By providing outstanding examples, tools and training (for all levels from the PhD student to the group leader and Professor) we will foster this change. Our vision is to build important and essential foundations for implementing the **FAIR principles** in the whole MSE community by fully exploiting existing research data to enable scientific progress because the FAIR principles are at the heart of this project. The proposed MSE Ontology and its implementation within the MSE knowledge graph will enable web-based access, findability, re-use, as well as efficient data integration.

5.4.2 Relevance for the Consortium's Objectives

Today, most experiments or simulations are – if at all – only partially described, e.g., mentioning only the physical phenomena (e.g. microkinetic), the length and/or time scale (e.g. atomic, mesoscale), the used machine or software, and some numerical details of the used simulation method or, e.g., details of the experimental sample preparation. Additionally, even in the materials science and engineering community, each sub-domain and sometimes even each institute uses its own terminology. This fundamentally impedes or even prohibits data integration, reuse, as well as replication on any significant scale. Therefore, a **generic and standardized language** with common terminology (i.e. definition of concepts and vocabulary) and which **all stakeholders** in the information chain are able to understand and to agree upon has a high relevance for the whole consortium. Clearly, **standardized metadata schemes** need to be machine readable and ultimately machine understandable, i.e. machines must be able to interpret data correctly, so that data access can be implemented in an automatic fashion without the need of additional context information or user interference. Furthermore, **ontology-based schemes that can be extended and adapted over time,** without the need of expensive change or maintenance of the applications dealing with it, will guarantee that they always will be in line with the overall requirements of the consortium. The objective of TA-OMS to develop machine readable data with built-in semantics is a key to making material related data FAIR. The structuring of metadata information and the development of an extensible base ontology are prerequisites for the description of material, process and simulation data. The reusability of data will be largely boosted and the technical burden will be lowered. The guided structure will lead to more complete datasets. Thereby, the additional efforts for the curation and capturing of the data will pay off for individual community members as well as for all MSE stakeholders. This message will be communicated through TA-CI alongside OMS related training. By linking the developments within TA-OMS to dedicated workflows and software (e.g. within TA-WSD) services and products will be established that help in the curation and capturing of the data.

5.4.3 Status of this Task Area's Field in MSE

Many MSE working groups have realized that tracking metadata is useful, and that the effort to implement an according system is indispensable in view of the FAIR principles. This is still not the case to a satisfying extent, despite some progress in the U.S., e.g. by the Materials Genome Initiative (MGI).

A prototype of a "MGI Ontology" was (co-) developed by the NIST (National Institute of Standards and Technology, https://www.nist.gov/) for specific applications but did not receive enough application to become common.

Today, the importance of designing a MSE ontology has been recognized by all major international groups and organizations, while its development still happens to be overdue. This is mostly owing to the complexity of such an effort and community ontology development in general. Among the ongoing efforts is the activity of the European Materials Modelling Council (EMMC, 2020) with the aim to develop a European Materials Modelling Ontology (EMMO, 2020), which was recently published in its first iteration. The EMMO uses the established W3C Web Ontology Language (OWL). EMMO is designed in a top-down manner from an upper-level ontology. Hence, within this Task Area we will **re-use, adapt and extend the EMMO for use within the MSE community**. Thereby, one of our measures includes a detailed requirement analysis carried out by domain experts from MSE in close collaboration with knowledge engineering experts to learn about the necessary process for doing so. Apart from EMMC/EMMO and MGI there are a number of smaller efforts, among others the work by S. Zhao and Q. Qian (2017) or Cheung (2009) which, however, are quite specialized and are only used within a small niche of MSE. Nevertheless, all existing efforts to model the domain of material sciences will be taken into account. Generally, selecting starting points from the plethora of available approaches is a major challenge to this TA.

5.4.4 Measures and Work Packages

Measure OMS-1 Establish and maintain MSE Ontology Working Groups *(FIZ, TUBAF)*

To coordinate the design and development of the MSE Ontology (a) within the consortium with other task areas, (b) for the NFDI-MatWerk consortium as a whole, as well as (c) with other NFDI consortia and (d) the international MSE community, an initial materials ontology working group will be established to bundle domain knowledge with ontological engineering knowledge.

WP OMS-1.1 Support foundation and manage interaction of ontology working groups: Awarely, that ontology development must take place decentralized, guided by knowledge and domain expert groups, this WP aims to foster the formation of such nuclei throughout our community. Therefore, it provides guidance with the necessary professional tools and teaching and coordinate inter-group communication. This helps streamlining the manifold current (inter-)national efforts for MSE ontologies.

WP OMS-1.2 Link and coordinate ontology working groups with existing community efforts: At the moment, there are several ongoing efforts creating ontologies for material sciences, such as e.g. BMBF Innovationsplattform MaterialDigital, the RDA, or the EMMC. This WP aims to link these efforts professional to benefit of each other. This also provides early visibility to this TA and NFDI consortium. Furthermore, external experts and stakeholders shall be invited to NFDI-MatWerk working groups to contribute their views.

[redacted]

Measure OMS-2 MSE Ontology Design *(FIZ, TUBAF, ALU, USTUTT)*

This measure comprises the design, implementation, and evaluation of the MSE Ontology as coordinated by and with the MSE Ontology Working Group (OMS-1). Based on already existing ontologies for MSE, domain experts together with knowledge engineers will analyze and determine the requirements any ontology has to fulfill for the field of MSE.

WP OMS-2.1 Requirement Analysis: This analysis comprises the collection of already existing MSE related ontologies, as well as an analysis of weak points and missing aspects. Then we will be able to formulate the requirements for our ontology design.

WP OMS-2.2 A Collaborative Ontology Design Environment and Tools fulfilling the requirements for the ontology design devised in OMS-2.1 shall be created. For this we start with the evaluation of existing tools. While reuse would be the preferred strategy, we anticipate that a specialized software implementation might be necessary. In this case, such a tool will be designed together with the TA-WSD and/or TA-MDI.

WP OMS-2.3 Metadata Schemes: Metadata schemes tailored to the needs of the MSE community shall be developed. As a first step, we will evaluate in how far existing Metadata Schemes can be reused or have to be extended. Based on this analysis, appropriate metadata schemes will be designed. Initially, these will be specialized schemes applicable only for a specific experiment or use case but a generalization as well as a standardization of metadata schemes in coordination with the ontology working group as well as the international material science community is envisaged as well.

WP OMS-2.4 Ontology Design: One of the core tasks of this TA is the design of ontologies for material sciences together with the community. In analogy to the previous WP we aim to reuse already existing ontologies as far as possible and will extend them where necessary. In all other cases, new ontologies will be designed by domain experts and knowledge engineers in close coordination with the ontology working group.

WP OMS-2.5 Demonstrators for MSE Ontology: Demonstrators are an important aspect of NFDI-MatWerk in order to promote the use of the MSE ontology throughout our community. While parts of this MSE community will be involved through their PPs already in the ontology design (WP OMS-2.4), in this WP our goal is to create, design and implement a representative set of demonstrators comprising private, material-specific ontologies beyond the base ontology. Seeing, how ontology development leads to uncertainty throughout unexperienced target groups, we suggest that tangible demonstrators serve as valuable learning examples to foster their understanding and support.

WP OMS-2.6 Ontology Evaluation: encompasses the (i) Definition of appropriate Evaluation Criteria together with an Evaluation Plan, (ii) Execution of qualitative and quantitative Evaluation, and (iii) Reiteration of the Ontology Design Cycle as often as necessary.

[redacted]

Measure OMS-3 MSE Knowledge Graph and FAIR Linked (Open) Data *(TUBAF, FIZ, ALU, USTUTT, MPIE)*

Based on the MSE ontologies (OMS-2), existing MSE research data together with background knowledge will be provided in terms of assertional data to compose a MSE knowledge graph. This measure (in close coordination with TA-MDI) comprises the design and implementation of an MSE knowledge graph under consideration of W3C standardized technologies to enable an 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.

WP OMS-3.1 A “Simplified” Knowledge Graph: will be developed already at an early point of time in the project. It will enable TA-MDI as well as TA-WSD to begin their implementations without having to wait for a fully-fledged and more sophisticated knowledge graph (WP 3.3). Wikibase (https://wikiba.se/) will be used as a platform for deployment and curation of the MSE knowledge graph. It enables low barrier access for knowledge graph population.

WP OMS-3.2 Distributed Infrastructure and Index: Within this WP the evaluation of centralized and distributed solutions for the implementation of the MSE knowledge graph in coordination with all other TAs shall be done. Furthermore, we will define requirements for an implementation strategy (e.g., distributed vs. centralized index). Such implementations will be realized by TA-MDI and TA-WSD.

WP OMS-3.3 MSE Knowledge Graph: This WP considers the development of NFDI-MatWerk's knowledge graph. This is a continuous effort, which results in a steadily growing and evolving representation of MSE. The link to external (e.g., NIST-, NFDI-) knowledge graphs will be created by linking to authority files (Integrated Authority File, GND) and hubs (Wikidata). Finally, we plan the integration of MSE ontologies and our MSE knowledge graph in existing web-based platforms conforming to international standards.

WP OMS-3.4 NFDI-MatWerk FAIR Linked (Open) Data: Based on existing MSE Linked (Open) Data resources, available NFDI-MatWerk data and metadata, e.g. from participant projects, will be transferred into Linked (Open) Data. In order to guarantee data sovereignty, access restrictions have to be defined alongside intellectual property and authorship management (with TA-MDI). TA-

OMS ensures their ontological implementation. Hence, publication of Linked (Open) Data in accordance with the authoring institutions is enabled. This is closely coordinated with TA-MDI.

[redacted]

Measure OMS-4 MSE Ontology Dissemination and Education *(TUBAF, ALU, USTUTT)*

To get the scientists of the MSE discipline involved in the creation process of the MSE ontologies (OMS-2) and MSE knowledge graph (OMS-3) and to ensure their continuous development, workshops will be organized and established. These bring together domain experts and knowledge engineers. Furthermore, applications of the MSE ontologies and the MSE knowledge graph will be implemented in PPs demonstrating the usefulness of the approach and fostering its acceptance within the research community.

WP OMS-4.1 Workshops for Ontology Working Group interaction: We will organize and design workshops tailored for the ontology workings groups and the MSE community for the matching of application ontologies and the continuous development of NFDI-MatWerk's knowledge graph in coordination with TA-CI. The workshops provide a forum for the exchange of MSE ontology experts.

WP OMS-4.2: MSE Ontology applications: We enable the linked PPs to make use of the currently available MSE ontologies, knowledge graphs, and of MSE Linked (Open) Data in coordination with the other TAs. Demands and newly gained knowledge reported to the TA by the PPs through the IUCs will be considered in the ontology design (OMS-2). This corresponds to a bottom-up process for our community-inspired ontology development.

WP OMS-4.3 MSE Ontology education and training: This WP ensures the creation of (online) teaching material for MSE Ontologies, knowledge graphs, and their application in MSE practice, as well as the organization and implementation of workshops and seminars on the MSE Ontology and knowledge graph in close coordination with TA-CI. Training events will enable the use of the most recent tools. The ontology working groups assist in defining the content for the continuous training program.

[redacted]

5.4.5 Interfaces and Interdependencies with other Task Areas

Ontologies and knowledge graphs are the state-of-the-art technology to enable data integration. Therefore, TA-OMS has strong links to the other task areas: Data integration is one of the most important aspects for the implementation of a national research data infrastructure. By using standardized W3C technologies and interfaces, unified access to research data beyond the boundaries of a specific consortium can be enabled. To ensure the necessary information exchange, regular meetings are in place to discuss and coordinate overlapping implementations via the Ontology Design Working Groups (OMS-1). Moreover, the actions of the Ontology Design Working Groups will actively be coordinated with other NFDI consortia as well as with the scientific community and standardization organizations.

This TA will develop tailored tools to enable the seamless connection between ontologies, data- and workflow management in close collaboration with TA-WSD and TA-MDI. Thereby, TA-OMS contributes to one of the main goals of NFDI-MatWerk. In particular, a close collaboration is required with TA-WSD: while the TA-OMS is mainly focusing on the development of the ontologies itself, as well as on the development of software tools for generating the corresponding, e.g., OWL files, TA-WSD will develop tools for the visualization and analysis of ontologies. The tight coupling of the two TAs not only requires regular video conferences but also meetings in person where the responsible staff members shall interact and collaborate by, e.g., jointly developing the required software interfaces and by testing them out on small benchmark problems and in the related IUCs and PPs.

NFDI-MatWerk's ontologies and the NFDI-MatWerk MSE knowledge graph will be the core elements for many future MSE related software and tools (as, e.g., developed in MDI-3 "Metadata Services" in TA-MDI). They enable the integration of heterogeneous data and, subsequently, its analysis. Furthermore, the knowledge graph will greatly improve visibility of and accessibility to research data over the world wide web based on automated methods. Workshops will foster sharing of achieved results, discussions and will assist in the solution of related problems in data representation, structuring and analysis based on ontologies.

5.4.6 Contribution to Cross-Cutting Topics of the NFDI

Metadata and vocabularies, but even more so ontologies, play a central role for **all** NFDI consortia. They are the foundation for the realization of the FAIR data principles. Therefore, they represent a truly cross-cutting topic within the NFDI. The objectives and measures of TA-OMS are designed for the MSE community while, at the same time, they relate to all neighboring NFDI consortia. While we will use metadata as a “stepping stone” for parts of the MSE community, the work in this consortium will mainly focus on the development of ontologies. These ontologies will showcase the benefits and inherent potentials of ontologies to other NFDI consortia. The related tools and methodologies can be used or adapted by these consortia. Other consortia as, e.g., NFDI4ING or NFDI4Chem, also focus efforts on ontology development for their domain. Different domain ontologies and especially knowledge graphs building on these ontologies can easily be interconnected via common concepts, processes, or entities, thereby enabling cross domain data accessibility, findability, data integration as well as reusability as well as semantic interoperability. To further enhance efficient ontology design, the design and implementation of the MSE Ontology has to be coordinated with further consortia and with the entire MSE research community (addressed also via our “MSE Ontology Working Groups”). This interaction will be improved by the organization of common workshops. The importance of the link to all other neighboring consortia is cemented in our IUC12: Alignment of application- and higher-level ontologies: It will create the connection between domain and application ontologies – a topic that is of particular interest at the interface between related NFDI consortia.

5.4.7 Appropriateness of the Task Area’s Representatives

The responsible persons for this Task Area are Prof. Dr. Stefan Sandfeld, TU Freiberg and Prof. Dr. Harald Sack, FIZ Karlsruhe. The expertise of the spokespersons of the TA cover the whole range of required knowledge and experience from materials science to computer science.

Prof. S. Sandfeld is a professor for Micromechanical Materials Modelling at the Technical University for Mining and Technology in Freiberg and is an expert for computational materials science and materials informatics, with an emphasis on bridging between simulations and experiments. He is teaching topics ranging from fundamental topics of materials science and simulations methods to advanced python programming, data analysis and topics of software engineering. Together with the DFKI, he offers training seminars in Deep Learning for Materials Scientists and therefore is well acquainted with the status quo in the materials community in terms of computational aspects. Furthermore, since 2019 he is chairman of the Expert Committee “Materials Modelling, Simulation and Data” of the DGM (German Materials Society).

Prof. H. Sack is Professor of Information Service Engineering at FIZ Karlsruhe, Leibniz Institute for Information Infrastructure and Karlsruhe Institute of Technology (KIT). His area of research spans from ontological engineering to knowledge mining including natural language processing, semantic technologies, and deep learning. His experience in ontology design including knowledge graph creation is shown by his involvement in publicly funded large research projects, as e.g. THESEUS research program - New technologies for the Internet of Services (BMWi research program for the development of new search technologies including ontologies and knowledge graphs), dwerft (BMBF regional growth core for designing and implementing ontologies for the media value chain), QPTDat (BMBF, Data Quality Assurance in the plasma technology including ontology design and implementation) as well as Platform Material Digital. He has published more than 150 scientific papers in international journals and conferences, including several standard textbooks. Since 2014 he has successfully complemented his university teaching with Massive Open Online Courses (MOOCs) on Semantic Web, Linked Data Technology and Information Service Engineering via the OpenHPI platform.

Due to the central role of ontologies in this consortium, all participants will at one point or another collaborate with this task area. Actively involved participants are currently: Prof. Dr. Felix Fritzen (University of Stuttgart), Prof. Dr. Alexander Hartmaier (RUB), Dr. Fabian Spreng (MPA Stuttgart), Dr. Joachim Meier (PTB).

5.4.8 Connection to Infrastructure Use Cases

These examples show how our ontology development is connected to Infrastructure Use Cases:

IUC12 Alignment of application- and higher-level ontologies: This IUC is mainly dedicated to all aspects touching upon the question how to create the connection between domain and application ontologies with general ontologies. Such “ontology matching” is crucial for enabling the interoperability between different (sub-)domains and therefore important for linking different sub-domains within NFDI-MatWerk. Furthermore, it is one of the prerequisites for creating a well-connected network of ontologies across different NFDI consortia.

IUC16 Unified ontology for matrix-inclusion microstructure and composites: In this IUC an ontology for matrix-inclusion materials is developed. This includes particulate, fibrous and porous, natural as well as artificial materials. Building on the base ontology and by collaborating with the MSE Ontology Working Groups, this ontology will interconnect heterogeneous data from experiments, synthetic microstructures and theoretical models. Further, this IUC is tailored to serve material-specific demands in the abstract methodologies opted for within MaRDI (measure 2.4; TA4 Cooperation with other disciplines).

IUC17 Ontologies for defects in crystals: This IUC is dedicated to developing metadata and ontologies describing crystalline structures and crystalline defects (such as point defects, line defects, area and volume defects) as well as their temporal evolution. This will be important for different types of microstructure simulations, experiments and microscopy data. For example, this has high relevance for defect structures for superalloys developed in the CRC/TRR103 (PP01) which is also closely linked with this IUC. Interconnections to existing ontologies will be created in based on IUC12.

Further, the following IUC are also linked to TA-OMS as main task area:
IUC11 Development of coupled ontologies and workflows for thermochemical treatments
IUC13 Co-creation environment for experts
IUC14 Adaptive automated characterization pipelines and meta data schemas for high throughput experiments
IUC15 Method- and scale-bridging workflows and data structures for tomography.

5.4.9 Funding Request for TA Ontologies for Materials Science

Table 5.4.9.1: Funding Request for Task Area Ontologies for Materials Science by Institution

	2021 (Oct-Dec)	2022	2023	2024	2025	2026 (Jan-Sep)	Total in €
Institution		**Totals in €**					
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 5.4.9.2: Funding Request for Task Area Ontologies for Materials Science by Funding Category

Funding category	2021 (Oct-Dec)	2022	2023	2024	2025	2026 (Jan-Sep)	Total
Staff by category	**Number of persons (full-time equivalents)**						
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Totals in €						
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

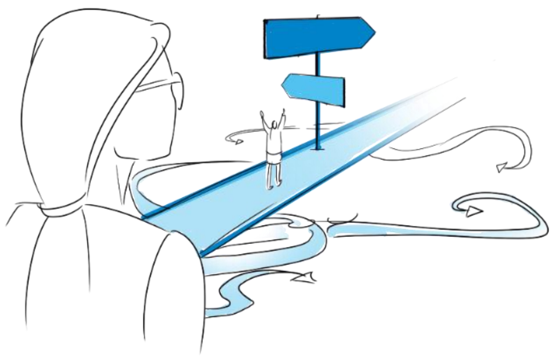
(*) this corresponds to the DFG staff category “Postdoctoral researchers and comparable”

(**) this corresponds to the DFG staff category “Doctoral researchers and comparable”

(***) This includes student research assistants at the height of 20% FTE

5.5 Task Area Strategy Development (TA-SD)

5.5.1 General Objectives



The NFDI initiative paves the way for a vastly more comprehensive shift in science than the mere establishment of an infrastructure. We are convinced that tomorrow’s scientific methods will **diverge in many ways from today’s reality**. This paradigm shift brought by the digital transformation in science will need to go much deeper than just data harmonization or common metadata schemas. For

NFDI-MatWerk, the NFDI cannot simply be reduced to its technical elaboration – a notion frequently agreed upon when discussed within the community in the course of this proposal's development. In fact, our community of MSE already experienced **the rise and abandonment of other mere technocratic initiatives** for data storage, like repositories. Learning from these experiences, it becomes clear that a **unified strategy** for the **scientific community is needed**, and we commit ourselves **with a dedicated Task Area**.

As in any well-managed organization, an elaborate strategy should address the overall challenges from different viewpoints and ensure a harmonized target implementation. Therefore; with the creation of a comprehensive infrastructure for MSE, our strategy is supposed to involve activities to tackle different possible risks that could prevent a successful implementation and usage our digital MSE infrastructure. The general objectives are:

- Establish **transparent processes and governance** to enable **community-driven decision-making.**
- Ensure **relevance** by closely involving the scientific community into the infrastructure's development process.
- Develop **monitoring** and **feedback procedures** for the community.
- Develop and transfer today's mechanisms for **incentives and good scientific practice** into the digital realm for a successful change process.
- Ensure **information sovereignty** as well as the legal framework for the intended infrastructure.
- Connect to international efforts

Monitoring these challenges greatly improves the chance for a successful infrastructure implementation, independently of its technological advancement: If the MSE community's needs are taken into account and met, then general acceptance will be high and the overall goal of the NFDI will be achieved.

Naturally, the right approach is only possible in an environment, in which the community is ready for such a close interaction and contributes the necessary time investments. For the MSE community, this is the case according to our community wide survey, described in sect. 3.1.4.

5.5.2 Relevance for the Consortium's Objectives

This task area serves as the central unit, responsible to set up the processes and structures within NFDI-MatWerk to ensure the **acquisition and distribution of information relevant for *all* task areas**. Furthermore, the task area ensures a **user focussed infrastructure development** and **integrates the community into decisions** concerning the resource distribution within the NFDI-MatWerk. The task area is also responsible to setup processes within the greater MSE community to **establish a community-driven strategy process** concerning the shared effort to tackle the digital transformation in the discipline of Materials Science and Engineering. While strategic management is not able to predict the future, the shared vision within the MSE community will greatly help providing aims for the NFDI-MatWerk. The outcome of this process will be the fundament for all TA's goal setting.

The IUCs clearly show that the usage profiles of the infrastructure typically depend on concerted implementations from different TA's. Therefore, quality assurance and requirement fulfilment need to be controlled throughout the project. The need for a comprehensive project monitoring and strategy formulation is fundamental for NFDI-MatWerk's success and reach. The relevant project overview will be gathered by TA SD and also shared with the community.

Developing an infrastructure without regard for the social and organizational dimension of the challenge is a common pitfall. The development of a coherent, reliable and community approved approach to our community's change process relies on both *professional research* of what is needed on a practical scientist level and *professional development*. This task area serves the consortium in terms of the strategic unit, which keeps watch over the different dimensions and interdependencies between technology, incentive models for Good Scientific Practice, cultural change and, among other things, an international foundation.

5.5.3 Status of this Task Area's Field in MSE

In science, change is a continuous process, that is rarely proactively managed. Change in scientific methods on the other hand is a natural part of science but typically affects a rather limited group. As a result, in the context of materials data management, researchers often react pragmatically by developing their own hands-on solutions for challenges that sometimes result in an advantage within the competitive field of science. In the past, this caused decentralised and inconsistent solutions towards research data management within parts of our community. Obviously, this is far from what is necessary to fulfil FAIR principles.

The digital transformation itself is too comprehensive of a challenge for anyone to manage individually and currently leads to both disappointments and wasted capacity. Lately, a fundamental readiness to commit to novel unified ways can be witnessed in our community, which is also confirmed by the s feedback and support demonstrated by the results ofour survey (sect. 3.1.4). The NFD-MatWerk consortium nurtures this development as it will only be successful when a strong community commitment is available. **It also means that the implementation of the NFDI in the materials community is a pressing matter, since community support can be lost when central platforms and projects are stalled.**

5.5.4 Measures and Work Packages

Measure SD-1 Continuous Strategy Development and Optimization (*Fraunhofer IWM)*

This measure ensures the development and continuous adaption of NFDI-MatWerk's internal strategy processes, in close coordination with the community. As a result, the adoption of this measure will enable guidance to the consortium and the community trough differing viewpoints, knowledge bases, and activities.

WP SD-1.1 Develop Processes for Translating Community Needs into Infrastructure: With the development and adoption of the Infrastructure Use Cases in sect. 4.1.2 NFDI-MatWerk followed a methodical path for transparently processing needs expressed by the MSE community and transferring them into workable modules that this consortium can address. Fostering and standardizing this steady communication process will be key to this consortium in ensuring its own activities are always based on its customer: the community.

WP SD-1.2 Develop and Implement Consortium Governance: Based on the existing experience in developing strategies for projects with a variety of contributing organisations, NFDI-MatWerk aims to implement common governmental mechanisms that will help us in decision-making (see sect. 3.4). Furthermore, a governance helps to disseminate responsibility and empower individuals with outstanding commitment. NFDI-MatWerk's governance will be jointly developed with all stakeholders and regularly reviewed and improved along the way, according to newly encountered challenges. TA-SD will work with TA-CI to implement these measures through dedicated events.

WP SD-1.3 IUC-Management: After developing and implementing IUCs act as bridge to between the community and NFDI-MatWerk's activities, our consortium will have to steadily manage their progress and adaptions throughout the infrastructure development. This will take place in close collaboration with the scientists which the IUCs aim to serve. This activity will enable us to implement FAIR standards that define the maturity of an IUC, its fulfilment, overall management embedding and finally its conclusion.

WP SD-1.4 Community Participation in Strategy Development: For TA-SD, community participation takes place in close collaboration with TA-CI, whereas TA-SD develops workshop contents with strategic relevance, analyses and documents the output and feedback created, e.g., by expert committees. Therefore, this activity will require close overlap with all other TAs. NFDI-MatWerk's yearly Strategy Meeting, organized by TA-SD in this WP, will be key in empowering the community as described in sect. 3.4.

[redacted]

Measure SD-2: Culture Change and Target Group Feedback Analytics *(ALU)*

This measure relies on NFDI-MatWerk's steadily generated community feedback acquired by TA-CI and other channels (e.g. strategy process of Measure SD-1, workshops hackathons, annual conference…) and analysis it for defining culture change approaches and appropriate learnings to tackle possible inhibitors of the infrastructure's advance.

WP SD-2.1 Community Needs, Surveys and Evaluation: Like the survey presented in sect. 3.1.4, TA-SD will rely on steady user and community feedback in learning more about the infrastructure's relevance and its adoption. Therefore, we apply both state-of-the art social surveying methods and their statistically extracted messages, as well as qualitatively created feedback methods through community expert interviews. In fact, while creating this proposal both of these general methods found their first applications in gathering community feedback, which is exactly what this WP aims to operationalize.

WP SD-2.2 Stakeholder Management: General stakeholder management, both internally, externally and third-party, relies on the establishment of steady, precise and professional communication channels. Ensuring this will be this WP's aim.It will be achieved via flexible and transparent guiding methods based on short channels, flat hierarchies and a service attitude. Therefore, we assume a variety of differing stakeholders that we all need to address, like fellow scientists with and without IT-knowledge, other scientific disciplines, the general public or the political arena. Developing precise and appropriately tailored communication methods will be a core activity of this TA in close collaboration with TA-CI for its subsequent implementation.

WP SD-2.3 Development of Consortium Branding: As known from the commercial context, a unified branding and identity focused on the users is the key in uniting people and achieving long-term community goals. Therefore, this TA assumes the challenge of defining and maintaining a consortium branding for NFDI-MatWerk, e.g., the way we explain and position ourselves, our corporate identity or the development of visions and applicable user scenarios and user experiences.

WP SD-2.4 Public Affairs: NFDI-MatWerk's activities never happen in an empty space but will always be embedded into the legal, political and social context that we operate in. Therefore, the TA-Strategy Development acknowledges the need to develop our own stances towards pressing societal questions in the context of digitalization, which interact with our aim of developing a community-wide infrastructure and beyond. We assume our opinions, angles and insights to be relevant to the public and the political realm and therefore implement this WP's to represent our consortium's aims there accordingly.

[redacted]

Measure SD-3 Incentive mechanisms strengthening Good Scientific Practice *(ALU)*

A highly interconnected infrastructure fundamentally changes the way we conduct science today, which is not only a consequence but a declared aim of NFDI-MatWerk. The resulting new methods will rely on fundamentally new mechanisms, on how and why a researcher on the one hand applies them, and on the other hand personally contributes to them. . This measure aims to anticipate resulting issues and propose solutions for tackling them while conserving the Good Scientific Practice we obey today.

WP SD-3.1 Sharing and Curating Incentives: Today, sharing data or software does not offer inherent scientific value, which is why many researchers tend to disregard this activity and instead focus on publications. This WP therefore aims to develop new incentive mechanisms for making it valuable to produce and share one's digital scientific output in a FAIR way with other fellow scientists. Here, TA-Strategy Development intends to not only find causes and methods of tackling this problem, but it also ensures together with TA-MDI their technological implementation within the developed infrastructure.

WP SD-3.2 Research Game-Theory Approaches for Interactions: Social group dilemmas like loathing, free-riding or leeching result from situations in which there are no mechanisms in place to prevent them. Economic and social theory, however, offer a variety of approaches of addressing such issues, which should be considered from day one when developing an open data infrastructure. This WP aims to research and develop solutions based on game theory and nudging to prevent or diminish social misalignment and contribution problems within our digital infrastructure.

WP SD-3.3 Business models: Although NFDI-MatWerk itself represents an open scientific infrastructure, many stakeholders that it interacts with will require either scientific or commercial compensation models for ensuring their contribution. This Work Package therefore aims to develop stakeholder-harmonized solutions that lead to business models for commercial service and context providers, whose involvement NFDI-MatWerk profits from.

[redacted]

Measure SD-4: Overall Project Management *(ALU)*

As in any consortium project of this size, NFDI-MatWerk will need a strongly operationalized project management for guiding all its different multi-disciplinary entities. This must take place beyond the internal management that the TAs implement themselves and ensures the overall harmonization of approaches. It is this measure's aim to develop and supervise such comprehensive project management.

WP SD-4.1 Consortium Coordination: General consortium coordination relies on implementing a single point of information, bundling and disseminating information internally as needed.

WP SD-4.2 Quality Assurance: NFDI-MatWerk's overall project management must ensure its fulfilment of community-expressed needs, as represented by our IUCs. This WP will address the need to develop according Quality Assurance techniques and routines for querying and challenging developed products and services of NFDI-MatWerks infrastructure under the constraint of predefined functionalities.

WP SD-4.3 Project Management and Reporting Toolbox: Successful project management relies on standardized formalizations among all members of an organization, being it standard processes and procedures, reporting frameworks or coordinating tools. The same applies to a large scientific consortium. This WP assumes responsibility for setting up a toolbox of internally shared technologies and standards for conducting unified process management among all partners.

WP SD-4.4 Fostering International Relations: The development of NFDI-MatWerk's infrastructure takes place at a moment when numerous similar initiatives are germinating around the world. To provide the best services to our community, this WP acknowledges the need to stay on track with parallel international developments, which are relevant to our own proceedings due to their technological, organizational or cultural approach. This WP enables such coordination through the systematic development of international networks and contributions to international forums and formats.

WP SD-4.5 Infrastructure Issue Tracking: No matter the variety of quality assurance and user-centred mechanisms in place, development of a novel infrastructure will inevitably come with technological flaws, bugs, and issues. This WP acknowledges our rocky path ahead and takes responsibility for making things the right way, even if it only works on a second try. Therefore, we develop an issue tracking service and processes to the community that enables it to give us low-threshold feedback about any technological shortcomings and steer such information into the channels, where the issues will be solved.

[redacted]

Measure SD-5: Legal compliance (*Fraunhofer IWM)*

NFDI-MatWerk's infrastructure development takes place under times of vast legal uncertainty when it comes to digital environments. This affects dimensions such as privacy, sovereignty as well as ownership issues, up to national security, to only name a few. Naturally, the regulators and authorities define the legal scope, in which this infrastructure will exist. The hence resulting fast-paced nature and complexity of the legal context can cause uncertainty among stakeholders. This measure aims to provide them guidance in managing such legal issues.

WP SD-5.1 Ensure compliance with NFDI: At the time of this writing, important aspects of the NFDI-context are still under development and therefore uncertain. This affects, e.g. the oversight by the DFG, operational guidance by the NFDI directorate, cross-cutting topics and organizational embedding of the consortia. This Work Package therefore manages this interface to the common NFDI-guidance and ensure this consortium's compliance through close cooperation with both consortium members and the NFDI supervisors.

WP SD-5.2 Data Protection Guidelines: Protection of data, being it personal, industrial or scientific, relies on the precise and transparent definition of rights and permissions but also obligations and responsibility. This is not only relevant for infrastructure operations but also for clearly defining responsibility for cases of misuse or data theft. This WP therefore aims to develop shared understandings and guidelines for implementing according standards for NFDI-MatWerk and explain them to all affected stakeholders.

WP SD-5.3 Ownership and Sovereignty: Every stakeholder contributing and relying on information in a digital infrastructure relies on personal legal constraints and interests when interacting with the provided data. Research data infrastructures lead to important questions when it comes to ownership and sovereignty, stemming e.g. from uncertainty such as if and to what extent the producer of a machine, the conductor of experiments or the provider of a specimen may own and use the eventually generated data. This WP aims to develop guidance on uncertain ownership and sovereignty issues, both legally and technologically and ensure the implementation of developed solutions.

[redacted]

5.5.5 Interfaces and Interdependencies with other Task Areas

Beyond the obvious interface through TA-Strategy Development's central strategic efforts for the entire consortium, there is a particular focus on all TAs' strategic coherence and the implementation of our Infrastructure Use Cases.

The strategic coherence will affect the way we disseminate and acknowledge our core values and aims internally. This interface represents a two-sided interdependency, since strategy building depends on analyzing the present situation in the TAs and the community themselves and thereafter plausibly deducting a focus for resulting fields of engagement. The measures listed above regard this interface thoroughly, since they strongly emphasize the desired exchange. The overall process of ensuring and guiding the interdependency between both the community's needs and the work of the TAs is guaranteed by this TA-SD's process for defining and implementing IUCs.

Another particularly strong interface will exist between this TA and the TA-CI. This comes from the latter's efforts in channelling information from NFDI-MatWerk to the community and vice versa. The TA-SD will rely on TA-CI's expertise for both feedback and dissemination of its developed strategic concepts, similar to the way Strategy- and PR-departments interact within a company.

Finally, this TA will set up a central interface to the other TAs when it comes to legal and financial integration. This will play a foremost role when the consortium aims to involve further participants and subcontractors according to the programs developed and agreed upon beforehand. The planned informational interface must enable an explicit description of participants' services and subsequently formalize them into contractual agreements, while ensuring appropriate funding and communication channels in direct coordination with the other according TA(s) that requested the integration. Providing the processual and legal frameworks will be the one of this TA's responsibilities.

5.5.6 Contribution to Cross-Cutting Topics of the NFDI

The aims addressed in TA-Strategy Development are specific in their configuration but comprehensive in their nature. As expressed above, the need for establishing and maintaining close community ties, careful process design and due legal foundation is paramount to all NFDIs. However, under vastly technology-driven approaches, it is often disregarded. This stems from the incorrect notion that technology, not people, will decide about the organization, implementation, and finally the success of the NFDI-initiative. We are therefore ready to contribute and advocate our viewpoint and resulting approach and thus create value for all other consortia and thereby the NFDI as a whole.

In fact, most approaches we see coming up in context of the NFDI have a largely community-driven foundation, which confirms the NFDI initiative's relevance and urgency. Hence, what our TA-Strategy Development will elaborate and contribute in a structured shape, no matter if it is process development, mechanism design or legal evaluations, will be relevant to other community-driven consortia. Since we addressed these aspects comprehensively from day one, we will also be proficient to create tangible results, applicable in slightly different shapes to other scientific communities. We look forward to doing this and commit ourselves to finding unified solutions to pressuring strategic challenges of the NFDI.

Finally, the consortium NFDI-MatWerk is committed to participate in the measures expressed by the joint "Berlin Declaration" (Glöckner, 2019) and "Leipzig-Berlin Declaration" (Bierwirth, 2020), together with numerous other NFDI consortia, which now already serves as a first step towards self-regulation and a collaboration basis. The declaration also addresses fundamental strategic and cultural issues, and therefore confirms their decisive relevance.

5.5.7 Appropriateness of the Task Area's Representatives

The Fraunhofer IWM's deputy director Prof. Dr. Christoph Eberl, is the responsible spokesperson for representing this TA. Prof. Eberl and his staff already played a key role in developing the collaboration infrastructure and channels that led to the consortium's foundation, which resulted in him becoming the overall spokesperson of this consortium's proposal. The status of this consortium and its proposal hence might be the best proof for the appropriateness of this TA's leadership in facing our strategic challenges. The fact that we are able to present NFDI-MatWerk today with such a unified vision stems from an ongoing community discourse about the domain-specific advance, which was already guided by this TA and will have to be expanded in the course of the project.

The Fraunhofer IWM considers itself pioneering in the digital representation of materials and the according efforts. The status of this institute's projects and insights on the digital depiction of material relationships is more advanced than most of what is found in research today, and definitely of what is found among industry standards. Accordingly, the Fraunhofer IWM is in close contact with a number of partners and further stakeholders that support it in this endeavour and thus has an exceptionally broad strategic viewpoint at its disposal.

When it comes to the strategy behind past and ongoing joint digitalization projects, the Fraunhofer IWM can rely on its long-term experience. With the “Innovation Platform MaterialDigital” and the “Landesprojekt MaterialDigital” the Fraunhofer IWM was already set in charge of directing two other endeavours that intend to create digital environments for exchanging materials data. While generally smaller and more specific in scope, these efforts and this consortium’s plans have in common that they include a number of stakeholders from different scientific backgrounds that all expect guidance throughout the projects’ progress. Just as the challenge comes with this TA, the Fraunhofer IWM already demonstrated its skills in culture change and dissemination, stakeholder integration and the development of administrative and legal compliance and guidelines within the other projects. Finally, the Fraunhofer IWM can rely on a generally strong community-network, which results from its management’s strong anchorage and years of engagement in the field of MSE. This network is unique, ranging from scholars to industry stakeholders, educational institutions and non-university research organizations. It is at our disposal when implementing any measures in the context of this TA.

Throughout these and further challenges in context of materials digitalization, Prof. Dr. Chris Eberl has been the leading voice inside the institute and towards its partners. Since his appointment as deputy head of the Fraunhofer IWM in 2014, Mr. Eberl has put the digital transformation of MSE prominently on his agenda in an exceptional way. Mr. Eberl has created a comprehensive and unique vision for this aim and shares it on a daily basis with his numerous colleagues and counterparts, gradually but sustainably shaping and influencing a cultural change within a traditional scientific discipline. After all, the existence of these above-mentioned efforts stems primarily from his ongoing work inside the community. Overall, it is thus a logical consequence of this engagement to rely on Mr. Eberl as the responsible force for driving this consortium’s strategic efforts.

5.5.8 Connection to Infrastructure Use Cases

There are no Infrastructure Use Cases (IUCs) connected to TA-SD, since it is a comprehensive TA that does not directly implement directly implement products and services to the community

5.5.9 Funding Request for TA Strategy Development

Table 5.5.9.1: Funding Request for Task Area Strategy Development by Institution

	2021 (Oct-Dec)	2022	2023	2024	2025	2026 (Jan-Sep)	Total in €
Institution		**Totals in €**					
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 5.5.9.2: Funding Request for Task Area Strategy Development by Funding Category

Funding category	2021 (Oct-Dec)	2022	2023	2024	2025	2026 (Jan-Sep)	Total
Staff by category	**Number of persons (full-time equivalents)**						
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Totals in €						
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

(*) this corresponds to the DFG staff category “Postdoctoral researchers and comparable”

(**) this corresponds to the DFG staff category “Doctoral researchers and comparable”

6 Overall Funding Request

Table 6.1: Overall Funding Request by Task Area

	2021 (Oct-Dec)	2022	2023	2024	2025	2026 (Jan-Sep)	Total in €
	Totals in €						
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 6.2: Overall Funding Request by Institution

	2021 (Oct-Dec)	2022	2023	2024	2025	2026 (Jan-Sep)	Total in €
Institution	Totals in €						
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 6.3: Overall Funding Request by Funding Category

Funding category	2021 (Oct-Dec)	2022	2023	2024	2025	2026 (Jan-Sep)	Total
Staff by category	Number of persons (full-time equivalents)						

[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Totals in €						
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

(*) this corresponds to the DFG staff category “Postdoctoral researchers and comparable”

(**) this corresponds to the DFG staff category “Doctoral researchers and comparable”

(***) This includes student research assistants at the height of 20% FTE

7 (Co-) Applicant Contributions

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

8 Appendix

Bibliography and List of References

Please list all sources and data repositories, information infrastructures and software that you have used to describe the status quo in your proposal. Indicate which sources were written or developed by members of the consortium (for example, by highlighting the source in bold).

AARC (2020). Authentication and Authorisation for Research Collaborations, https://aarc-project.eu/

AGFD. (2020). AG Digitale Forschungsdaten und Forschungsinformationen. https://www.agfd.fau.de

AIMS. (2020). Applying Interoperable Metadata Standards. https://gepris.dfg.de/gepris/projekt/432233186?context=projekt&task=showDetail&id=432233186&

Anaconda. (2020). Anaconda Data Science technology. https://www.anaconda.com/

Anubhav, J. et al. (2015) . Fireworks: a dynamic workflow system designed for high throughput applications, Concurr. Comput.: Practice Exp. 27(17), 5037–5059.

ASM Digital Library. (2020). ASM Handbooks Online. https://dl.asminternational.org/handbooks

Bierwirth, M., Glöckner, F. O., Grimm, C., Schimmler, S., Boehm, F., Busse, C., Degkwitz, A., Koepler, O., & Neuroth, H. (2020). Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung. https://doi.org/10.5281/ZENODO.3895209

BIOVA, 2020. http://accelrys.com/products/collaborative-science/bioviamaterials-studio/

Blaiszik, B., Chard, K., Pruyne, J., Ananthakrishnan, R., Tuecke, S., & Foster, I. (2016). The Materials Data Facility: Data Services to Advance Materials Science Research. JOM, 68(8), 2045–2052. https://doi.org/10.1007/s11837-016-2001-3

CE (2020) FAU-Erlangen-Nürnberg. Computational Material Science – Master Program. https://www.ce.studium.fau.eu/prospective-students/technical-application-fields-taf/computational-material-science/

Choudhary, K. & Kalish, I., & Beams, R., & Tavazza, F. (2017) High-throughput identification andcharacterization of two-dimensional materials using density functional theory, Sci. Rep. 7 (1), 5179.

CSMD. (2013). Core Scientific Metadata Model (CSMD). http://icatproject-contrib.github.io/CSMD/

Curtarolo, S. et al. (2012). AFLOW: An automatic framework for high-throughput materials discovery, Comput. Mater. Sci. 58, 218–226..

DARIAH. (2020). The Pan-European Infrastructure for Arts & Humanities Scholars. https://www.dariah.eu

DataCite. (2019). DataCite Metadata Schema. https://schema.datacite.org/

DCMI (2020). Dublin Core Metadata Initiative, https://www.dublincore.org/schemas

DFG, Deutsche Forschungsgemeinschaft (2018): DFG Förderatlas 2018. WILEY. ISBN 978-3-527-34520-5. https://www.dfg.de/sites/foerderatlas2018/download/dfg_foerderatlas_2018.pdf

DFN. (2020). Deutsches Forschungsnetz, DFN-AAI. https://www.dfn.de/dienstleistungen/dfnaai/

Docker. (2020). Docker.com. https://www.docker.com

DublinCore. (2003). Dublin Core Metadata Initiative. https://www.dublincore.org/schemas/

EMCC (2020). European Materials Characterisation Council, http://characterisation.eu/)

EMMC (2020). European Materials Modelling Council, Link: https://emmc.info/

EMMC (2020). The European Materials Modelling Council, https://emmc.info/

EMMO (2020). The European Materials Modelling Ontology, https://github.com/emmo-repo/EMMO

EOSC Hub. (2020). European Open Science Cloud Hub. https://www.eosc-hub.eu

EOSC Portal. (2020). European Open Science Cloud Portal. https://www.eosc-portal.eu

EOSC Secretariat. (2020). European Open Science Cloud Secretariat. https://www.eoscsecretariat.eu

EOSCFAIR Executive Board Working Group. (2020). EOSC Interoperability Framework (v1.0). https://www.eoscsecretariat.eu/sites/default/files/eosc-interoperability-framework-v1.0.pdf

EUDAT. (2020). EUDAT Collaborative Data Infrastructure. https://www.eudat.eu

European Commission. (2020). MatDB. https://odin.jrc.ec.europa.eu/alcor/Main.jsp

FAU. (2020). FAU Erlangen-Nürnberg. https://www.map.tf.fau.de/about-map/structure/#collapse_8

Galaxy. (2020). Galaxy Community Hub. https://www.galaxyproject.org

GEANT. (2020). EduGAIN. https://edugain.org/

Gehring, P. (2019). Putting the NFDI into Practice. Impulses of the German Council for Scientific Information Infrastructures. Slides at the NFDI Conference of the DFG, as presented on May 13 and 14, 2019 in Bonn.

GitLab. (2020). https://about.gitlab.com

Glöckner, F., Diepenbroek, M., Felden, J., Overmann, J., Bonn, A., Gemeinholzer, B., Güntsch, A., König-Ries, B., Seeger, B., Pollex-Krüger, A., Fluck, J., Pigeot, I., Kirsten T., Mühlhaus, T., Wolf, C., Heinrich, U., Steinbeck, C., Koepler, O., Stegle, O., Weimann, J., Schörner-Sadenius, T., Gutt, C., Stahl, F., Wagemann, K., Schrade, T., Schmitt, R., Eberl, C., Gauterin, F., SchultzM., Bernard, L. (2019): Berlin Declaration on NFDI Cross-Cutting Topics, https://doi.org/10.5281/zenodo.3457213

HDF Group. (2020). The HDF5 Library & File Format. https://www.hdfgroup.org/solutions/hdf5

HMC. (2020). Helmholtz Metadata Collaboration Platform. https://www.helmholtz.de/en/research/information-data-science/helmholtz-metadata-collaboration-plattform-hmc/

ICAMS. (2020). Interdisciplinary center for advanced materials simulation. Masters Course Materials Science and Simulation. http:/www.icams.de/content/master-course-mss/

IDSA. (2020). International Data Spaces Association. https://www.internationaldataspaces.org

ITC. (2020) RWTH IT-Center. https://www.itc.rwth-aachen.de/go/id/obqb

Janssen, J. & Surendralal, S. & Lysogorskiy, Y. & Todorova, M. & Hickel, T. & Drautz, R. & Neugebauer, J. (2019). pyiron: An integrated development environment for computational materials science. Computational Materials Science 163, 24-36. [Accessed 25 09 2019].

Jupyter. (2020). Jupyter. https://jupyter.org/

Larsen, A.H. et al. (2017). The atomic simulation environment-a python library for working with atoms, J. Phys.: Condens. Matter 29(27), 273002.

Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., & Zhao, J. (2013). Prov-o: The Prov Ontology. W3C Recommendation, 30.

Lookman T., et al. (2019): Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design, npj Computational Materials 5(1)

MASi. (2015). Metadata Management for Applied Sciences. https://gepris.dfg.de/gepris/projekt/262962976

Mathew, K. et al. (2016) Mpinterfaces: A materials project based python tool for high-throughputcomputational screening of interfacial systems, Comput. Mater. Sci. 122, 183–190.

Mathew, K. et al. (2017) A high-level interface to generate, execute, and analyze computational materials science workflows, Comput. Mater. Sci. 139 (2017), 140–152.

Mayeshiba, T., & Wu, H., & Angsten, T., & Kaczmarowski, A., & Song, Z., & Jenness, G., Xie, W., & Morgan, D. (2017) The MAterials Simulation Toolkit (MAST) for atomistic modeling of defects and diffusion, Comput. Mater. Sci. 126, 90–102.

MedeA, 2020. Link: http://www.materialsdesign.com/medea

NFFA. (2020). Nanoscience Foundies and Fine Analysis. https://www.nffa.eu

NIST. (2020). NIST Materials Data Repository. https://materialsdata.nist.gov/

NOMAD. (2020). NOMAD Repository & Archive. http://nomad-repository.eu/

Ong, S.P. , & Richards, W.D., & Jain, A., & Hautier, G., & Kocher, M., & Cholia, S., & Gunter, D., & Chevrier, V.L., & Persson, K.A., & Ceder, G., (2013) Python materials genomics (pymatgen): A robust, open-source python library for materials analysis, Comput. Mater. Sci. 68, 314–319.

Open Archives. (2015). The Open Archives Initiative Protocol for Metadata Harvesting. http://www.openarchives.org/OAI/openarchivesprotocol.html

Pizzi, G., & Cepellotti, A., & Sabatini, R., & Marzari, N., & Kozinsky, B. (2016) AiiDA: automated interactive infrastructure and database for computational science, Comput. Mater. Sci. 111, 218–230.

Pyiron. (2020). Pyiron, The Materials Science IDE. https://pyiron.org

RADAR. (2020). A Data Repository for the Research Community. https://www.radar-service.eu/en

RDA DFIG. (2020). RDA IG Data Fabric: FAIR Digital Objects.https:/www.rd-alliance.org/group/data-fabric-ig/wiki/rda-ig-data-fabric-fair-digital-objects

RDA. (2020). Research Data Alliance. https://www.rd-alliance.org

RDM4Eng. (2020). RDA IG Research Data Management in Engineering. https://www.rd-alliance.org/groups/research-data-management-engineering-igEOSC Portal, 2020

re3data. (2020). Registry of Research Data Repositories. https://www.re3data.org

RWTH-DL. (2020). RWTH Center für Lehr- und Lernservice. Link: https://cls.rwth-aachen.de/data-literacy/

Sandfeld, S., et al. (2018): Strategiepapier: "Digitale Transformation in der Materialwissenschaft und Werkstofftechnik", https://www.dgm.de/medien/print-medien/strategiepapier-digitale-transformation/

SCC. (2019). KIT-SCC. GridKa School 2019. https://indico.scc.kit.edu/event/460/

Schauhoff, S. & Kirchhain, C. (2019). Die Struktur der NFDI aus steuerrechtlicher Sicht. FlickGockeSchaumburg. Slides at the Governance Workshop of the DFG, as presented on August 30, 2019 in Bonn.

Springer. (2020). Springer Materials. https://materials.springer.com

Talirz, L., Kumbhar, S., Passaro, E., Yakutovich, A. V., Granata, V., Gargiulo, F., Borelli, M., Uhrin, M., Huber, S. P., Zoupanos, S., Adorf, C. S., Andersen, C. W., Schütt, O., Pignedoli, C. A., Passerone, D., VandeVondele, J., Schulthess, T. C., Smit, B., Pizzi, G., & Marzari, N. (2020). Materials Cloud, a platform for open computational science. Scientific Data, 7(1), 299. https://doi.org/10.1038/s41597-020-00637-5

TUBAF. (2020). International M. Sc Program "Computational Materials Science. https://tu-freiberg.de/international-masters-program-in-computational-materials-science

Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data.(2018). Publications Office. https://data.europa.eu/doi/10.2777/1524oogl

Universität Stuttgart. (2020). engMeta—Beschreibung von Forschungsdaten. https://www.izus.uni-stuttgart.de/fokus/engmeta

W3C. (2013). An Overview of the PROV Family of Documents. https://www.w3.org/TR/prov-overview/

W3C. (2017). Web Annotation Data Model. https://www.w3.org/TR/annotation-model/

W3C. (2020). Data Catalog Vocabulary (DCAT)—Version 2. https://www.w3.org/TR/vocab-dcat-2/

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18