

# **BiGmax Workshop 2021**

Wednesday 14 April 2021 - Thursday 15 April 2021

Virtually

## **Book of Abstracts**



# Contents

Computational material synthesis: Atomistic and molecular dynamics, and bio-inspired AI . . . . .	1
AI-driven discovery of catalytic mechanisms: application to CO <sub>2</sub> activation on semiconductor oxide surfaces . . . . .	2
Design of high-entropy Invar alloys via machine learning . . . . .	3
Data-driven search for drug-membrane permeability models . . . . .	4
Data driven approaches to design damage tolerant Dual-phase steel microstructures . . . . .	5
Model-Free Data-Driven Computing . . . . .	6
Using phase-field models to coarse-grain data sets from scanning transmission electron microscopy . . . . .	7
High-performance data-analytics and AI at MPCDF . . . . .	8
Uncovering the Relationship Between Thermal Conductivity and Anharmonicity with Symbolic Regression . . . . .	9
Resampling Base Distributions of Normalizing Flows . . . . .	10
3D image analysis . . . . .	11
Towards an automatized data analysis of large 3d volumetric data . . . . .	12
Convolutional neural network-assisted recognition of nanoscale L12 ordered structures in face-centred cubic alloys . . . . .	13
Artificial-Intelligence-Driven Characterization of Crystallographic Interfaces from Electron Microscopy . . . . .	14
Machine-learning-enhanced time-of-flight mass spectrometry analysis . . . . .	15
Robust recognition and exploratory analysis of crystal structures via Bayesian deep learning . . . . .	16
Learning Dynamics of STEM by Enforcing Physical Consistency with Phase-Field Models . . . . .	17
Consistent atom probe representation for machine learning and data mining . . . . .	18
Gigascale electron event processing for band structure mapping . . . . .	19
Knowledge-Based Approaches in Catalysis and Energy Modelling . . . . .	20

NOMAD – A FAIR Data Sharing Platform for Materials Science . . . . .	21
On Software Tools Which Assist Electron Microscopists with Sharing Metadata and Numerical Results in Accordance with the FAIR Data Stewardship Principles . . . . .	22
Approaches to FAIR TEM data . . . . .	23

**Session I / 54****Computational material synthesis: Atomistic and molecular dynamics, and bio-inspired AI**

**Author:** Markus Buehler (invited)<sup>1</sup>

<sup>1</sup> *MIT*

Nature produces a variety of materials with many functions, often out of simple and abundant materials, and at low energy. Such systems - examples of which include silk, bone, nacre or diatoms - provide broad inspiration for engineering. Here we explore the translation of biological composites to engineering applications, using a variety of tools including molecular modeling, AI and machine learning, and experimental synthesis and characterization. We review a series of studies focused on the mechanical behavior of materials, especially fracture, and how these phenomena can be modeled using a combination of molecular dynamics and machine learning. We also present various case studies of hierarchical material optimization using genetic algorithms, applied to 3D printed composites, protein design, and a translation of protein folding to music and back, to offer a broad bio-inspired AI-driven material synthesis platform. As an example we present a recent study in which we translated JS Bach's Goldberg Variations into protein form, and elucidated salient features of the resulting molecular conformations and material functions.

**Session I / 39****AI-driven discovery of catalytic mechanisms: application to CO<sub>2</sub> activation on semiconductor oxide surfaces**

**Authors:** Aliaksei Mazheika<sup>1</sup>; Sergey Levchenko<sup>2</sup>; Luca Ghiringhelli<sup>3</sup>; Matthias Scheffler<sup>4</sup>

<sup>1</sup> *TU Berlin*

<sup>2</sup> *Skolkovo Institute of Science and Technology*

<sup>3</sup> *Fritz Haber Institute of the Max Planck Society*

<sup>4</sup> *FHI*

**Corresponding Author:** ghiringhelli@fhi-berlin.mpg.de

Using subgroup discovery (SGD), an AI approach that discovers statistically exceptional subgroups in a dataset, we develop a strategy for a rational design of catalytic materials. SGD allows for the identification of distinct, possibly competing mechanisms of a catalytic activation. Here, it is applied to the problem of converting CO<sub>2</sub> into useful chemicals. We demonstrate that the bending of CO<sub>2</sub>, previously proposed as the indicator of activation, is insufficient to account for the good catalytic performance of experimentally characterized oxide surfaces. Instead, our approach identifies the asymmetric strong elongation of the molecular C-O bond as a more accurate indicator.

**Session I / 50****Design of high-entropy Invar alloys via machine learning**

**Authors:** Ziyuan Rao<sup>1</sup>; Ye Wei<sup>2</sup>; Po-Yen Tung<sup>3</sup>; Stefan Bauer<sup>4</sup>; Ruiwen Xie<sup>5</sup>; Hongbin Zhang<sup>5</sup>; Fritz Körmann<sup>3</sup>; Prithiv Sukumar<sup>3</sup>; Alisson da Silva<sup>3</sup>; Zhiming Li<sup>3</sup>; Dirk Ponge<sup>3</sup>; Baptiste Gault<sup>3</sup>; Jörg Neugebauer<sup>3</sup>; Oliver Gutfleisch<sup>5</sup>; Dierk Raabe<sup>3</sup>

<sup>1</sup> *max-planck-institut für eisenforschung gmbh*

<sup>2</sup> *Max Planck institute for iron research*

<sup>3</sup> *Max-Planck-Institut für Eisenforschung GmbH*

<sup>4</sup> *Max-Planck-Institut für Intelligente Systeme*

<sup>5</sup> *Technische Universität Darmstadt*

**Corresponding Author:** z.rao@mpie.de

Invar alloys exhibit a very low thermal expansion coefficient (TEC) below  $2 \times 10^{-6} \text{ K}^{-1}$  around room temperature. There is a strong impetus to design novel Invar alloys with better physical, mechanical and chemical properties. Here, we develop and apply an active learning strategy to accelerate the design of novel Invar alloys in a practically infinite compositional space of quaternary and quinary high-entropy alloys (HEAs). It is demonstrated that this strategy is of great potential to accelerate the discovery of novel functional materials, especially for those with large unexplored phase space such as compositions, crystal structures, and microstructures.

**Session I / 43****Data-driven search for drug–membrane permeability models****Author:** Arghya Dutta<sup>1</sup>**Co-authors:** Jilles Vreeken<sup>2</sup>; Luca M. Ghiringhelli<sup>3</sup>; Tristan Bereau<sup>4</sup><sup>1</sup> *Max Planck Institute for Polymer Research*<sup>2</sup> *CISPA Helmholtz Center for Information Security, Saarbrücken, Germany*<sup>3</sup> *NOMAD Laboratory, Fritz Haber Institute of the Max Planck Society, Berlin, Germany*<sup>4</sup> *Van 't Hoff Institute for Molecular Sciences and Informatics Institute, University of Amsterdam, Amsterdam 1098 XH, The Netherlands***Corresponding Author:** dutta@mpip-mainz.mpg.de

Passive drug–membrane permeability of a drug molecule quantifies its capacity to cross cell membranes on the way of reaching its target. In this contribution, I will present results from our work where we used sure-independence screening and sparsifying operator (SISSO) to find equations for the permeability coefficient that combine both hydrophobicity and acidity of the drugs. The predicted equations provide more accurate values of permeability, on average, than the existing models for a diverse and exhaustive class of small drug molecules. Analysis of the inhomogeneous solubility-diffusion model in several asymptotic acidity regimes further leads to a rationalization of the equations.

**Session I / 51**

## **Data driven approaches to design damage tolerant Dual-phase steel microstructures**

**Author:** Navyanth Kusampudi<sup>None</sup>**Corresponding Author:** n.kusampudi@mpie.de

Identifying representations and characterizing the microstructure features is crucial for developing the damage-tolerant dual-phase steels. These representations serve as variables to establish a structure-property relationship and to design microstructures of desired mechanical property. However, the complex nature of the DP steel microstructure, poses a challenge and the existing characterization methods are limited in encoding this information using the handcrafted features.

To tackle this challenge, I will introduce machine learning models

1. To extract features automatically from synthetic DP steel images and represent the microstructure in low dimensional latent space.
2. To conditionally generate damage-tolerant microstructure patterns with desired yield stress.

**Session II / 57**

## **Model-Free Data-Driven Computing**

**Author:** Michael Ortiz (invited)<sup>1</sup>

<sup>1</sup> *Caltech/Universität Bonn*

**Session II / 42****Using phase-field models to coarse-grain data sets from scanning transmission electron microscopy****Authors:** Ning Wang<sup>1</sup>; Christoph Freysoldt<sup>1</sup>**Co-authors:** Wenjun Lu <sup>1</sup>; Christian Liebscher <sup>2</sup><sup>1</sup> *MPI Eisenforschung*<sup>2</sup> *Max-Planck-Institut für Eisenforschung***Corresponding Author:** freysoldt@mpie.de

To extract transferable insights from scanning transmission electron microscopy (STEM), one must deal with noise arising from electron scattering and of the investigated sample. This noise hinders a quantitative analysis of the observation, notably when the features of interest lie in the gradients of the raw data. Physics-informed neural networks have been proposed as a means to incorporate compliance with physical equations that are chosen a priori. We show here that phase field models can help to efficiently coarse-grain STEM video sequences of phase transformations.

**Session II / 46****High-performance data-analytcs and AI at MPCDF****Author:** Andreas Marek<sup>1</sup><sup>1</sup> *MPCDF***Corresponding Author:** andreas.marek@mpcdf.mpg.de

Since few years, in addition to classical (simulation-based) HPC usage, we observe a steadily growing need of our users for support of high-performance data-analytics (HPDA) and AI workflows. In this presentation, we will give an overview of the HPC clusters available at MPCDF for HPDA and AI workflows, the available (HPC-optimized) software stack and we will present recently introduced services such as Jupiter Notebooks as a Service (JNaS) and containers for software deployment. Furthermore, by means of a few use cases, we will demonstrate how the user can run data-analytics or AI algorithms at MPCDF on several HPC cluster nodes.

## Session II / 48

## Uncovering the Relationship Between Thermal Conductivity and Anharmonicity with Symbolic Regression

**Authors:** Thomas Purcell<sup>1</sup>; Matthias Scheffler<sup>1</sup>; Luca Ghiringhelli<sup>1</sup>; Christian Carbogno<sup>1</sup>

<sup>1</sup> *Fritz Haber Institute*

**Corresponding Author:** purcell@fhi-berlin.mpg.de

Quantitatively understanding the link between anharmonicity and thermal conductivity,  $\kappa$ , is pivotal to the search for better thermal insulators. To help find this link we present new descriptors of  $\kappa$  based on our new measure of anharmonicity,  $\sigma^A$ . Using an updated sure-independence screening and sparsifying operator (SISSO) method, we find analytical expressions with symbolic regression and generate expressions for  $\kappa_L$  that are competitive with those previously reported in the literature using only a third of the primary features. Finally, we discuss the implications of the new models on future materials design.

## Session II / 3

**Resampling Base Distributions of Normalizing Flows****Authors:** Vincent Stimper<sup>1</sup>; Jose Miguel Hernandez-Lobato<sup>2</sup><sup>1</sup> *Max Planck Institute for Intelligent Systems*<sup>2</sup> *University of Cambridge***Corresponding Author:** vincent.stimper@tuebingen.mpg.de

Normalizing flows are a popular class of models for approximating probability distributions. However, in many tasks such as image generation benchmarks they are still outperformed by autoregressive models and generative adversarial networks. This is in part due to their invertible nature limiting their ability to model target distributions with a complex topological structure. Several approaches have been proposed to solve this problem but they sacrifice invertibility and thereby trackability of the log-likelihood as well as other desirable properties. In this work, we introduce a base distribution for normalizing flows based on learned rejection sampling, allowing them to model complex topologies without giving up on bijectivity. We applied our model to various sample problems, such as images generation and approximating Boltzmann distributions, and it outperforms the baseline qualitatively and quantitatively.

**Session III / 55****3D image analysis**

**Author:** Dagmar Kainmueller (invited)<sup>1</sup>

<sup>1</sup> *MDC Berlin*

Microscopy data is often large and 3d, and thus convolutional neural networks (CNNs) need to be applied in a tile-and-stitch manner to cope with GPU memory constraints. Concerning pixel-wise predictions obtained with UNet-style CNNs via tile-and-stitch, issues with discontinuities at output tile boundaries have been reported. However, a formal analysis of the causes has been lacking. In particular, it had not been understood how inconsistencies can arise even in case of valid padding. Our work shows that the potential for discontinuities to arise is intricately tied to the shift equivariance properties of the employed CNNs. Our theoretical analysis entails simple rules for designing CNNs that are necessary to avoid discontinuities when predictions are obtained in a tile-and-stitch manner.

**Session III / 53****Towards an automatized data analysis of large 3d volumetric data**

**Authors:** Luca Curcuraci<sup>1</sup>; Markus Kühbach<sup>2</sup>; Ronald Seidel<sup>3</sup>; Nicolas Fabas<sup>4</sup>; Andreas Marek<sup>4</sup>; Markus Rampp<sup>4</sup>; Yael Politi<sup>5</sup>; Peter Fratzl<sup>6</sup>; Richard Weinkamer<sup>6</sup>; Luca Bertinetti<sup>6</sup>

<sup>1</sup> *Max-Planck-Institut für Kolloid- und Grenzflächenforschung*

<sup>2</sup> *Fritz Haber Institute, NOMAD Laboratory*

<sup>3</sup> *B CUBE - Center for Molecular Bioengineering*

<sup>4</sup> *Max Planck Computing and Data Facility*

<sup>5</sup> *B CUBE – Center for Molecular Bioengineering*

<sup>6</sup> *Max Planck Institute of Colloids and Interfaces*

**Corresponding Authors:** luca.curcuraci@mpikg.mpg.de, kuehbach@fhi-berlin.mpg.de, luca.bertinetti@mpikg.mpg.de

Recent developments in bio-imaging technologies have allowed researchers to collect larger and larger tomographic datasets which contain an immense amount of details. To achieve a quantitative understanding, however, these datasets need to be cleaned-up and segmented. These two tasks are tedious, very time consuming, and still performed mostly manually. In our work we aim to develop a full workflow from 3D image pre-processing to DL-based 3D segmentation and analysis of large volumetric datasets. Here we present the pre-processing pipeline, including handling of metadata, the plan for the implementation of the segmentation tools and an example of large volumetric data analysis.

## Session III / 52

**Convolutional neural network-assisted recognition of nanoscale L12 ordered structures in face-centred cubic alloys****Author:** Yue Li<sup>1</sup>**Co-authors:** Xuyang Zhou <sup>2</sup>; Timoteo Colnaghi <sup>3</sup>; Ye Wei <sup>2</sup>; Andreas Marek <sup>3</sup>; Hongxiang Li <sup>4</sup>; Stefan Bauer <sup>5</sup>; Markus Rampp <sup>6</sup>; Leigh Stephenson <sup>2</sup><sup>1</sup> *Max-Planck-Institut für Eisenforschung GmbH*<sup>2</sup> *Max-Planck Institut für Eisenforschung GmbH*<sup>3</sup> *Max Planck Computing and Data Facility*<sup>4</sup> *University of Science and Technology Beijing*<sup>5</sup> *Max-Planck-Institut für Intelligente Systeme*<sup>6</sup> *Max Planck Computing and Data Facility (MPCDF)***Corresponding Author:** yue.li@mpie.de

L12-type nano-ordered structures are typically fully-coherent with FCC matrix, which is challenging to be characterized. Spatial distribution maps are used to probe local order within reconstructed APT data. However, it is almost impossible to manually analyse the complete point cloud in search for the partial crystallographic information retained within the data. Here, we proposed an intelligent L12-ordered structure recognition method based on convolutional neural networks. The approach was successfully applied to reveal the 3D distribution of L12-type nanoparticles with an average radius of 2.54nm in an Al-Li-Mg system. The minimum radius of detectable nanodomain is even down to 5 Å.

**Session III / 40**

## **Artificial-Intelligence-Driven Characterization of Crystallographic Interfaces from Electron Microscopy**

**Authors:** Byung Chul Yeo<sup>1</sup>; Christian Liebscher<sup>2</sup>; Andreas Leitherer<sup>3</sup>; Matthias Scheffler<sup>4</sup>; Luca Ghiringhelli<sup>5</sup><sup>1</sup> *Pukyong National University*<sup>2</sup> *Max-Planck-Institut für Eisenforschung*<sup>3</sup> *Fritz-Haber-Institute*<sup>4</sup> *FHI*<sup>5</sup> *Fritz Haber Institute of the Max Planck Society***Corresponding Author:** byeo@pknu.ac.kr

Characterizing crystallographic interfaces in synthetic nanomaterials is an important step for the design of novel materials. Trained materials scientists can assign interface structures of materials by looking at high-resolution imaging and diffraction data obtained by aberration-corrected scanning transmission electron microscopy (STEM). However, STEM datasets cannot be fully exploited due to the lack of automatic analysis tools. Here, we present AI-STEM, a newly developed AI tool, based on a Bayesian neural network, for accurately extracting the key features of (poly)crystalline materials from atomic-resolution STEM images. It achieves excellent predictive performance for identifying crystal structure and lattice misorientations on experimental images.

**Session III / 49**

## **Machine-learning-enhanced time-of-flight mass spectrometry analysis**

**Authors:** Ye Wei<sup>1</sup>; Baptiste Gault<sup>2</sup>; Dierke Raabe<sup>2</sup><sup>1</sup> *Max Planck institute for iron research*<sup>2</sup> *Max Planck Institute for Iron research***Corresponding Author:** y.wei@mpie.de

Mass spectrometry is a widespread approach used to work out what the constituents of a material are. Atoms and molecules are removed from the material and collected, and subsequently, a critical step is to infer their correct identities based on patterns formed in their mass-to-charge ratios and relative isotopic abundances. However, this identification step still mainly relies on individual users' expertise, making its standardization challenging, and hindering efficient data processing. Here, we introduce an approach that leverages modern machine learning technique to identify peak patterns in time-of-flight mass spectra within microseconds, outperforming human users without loss of accuracy.

**Session III / 41**

## **Robust recognition and exploratory analysis of crystal structures via Bayesian deep learning**

**Authors:** Andreas Leitherer<sup>1</sup>; Angelo Ziletti<sup>1</sup>; Matthias Scheffler<sup>1</sup>; Luca M. Ghiringhelli<sup>None</sup>

<sup>1</sup> *Fritz Haber Institute of the Max Planck Society*

**Corresponding Author:** leitherer@fhi-berlin.mpg.de

Due to their ability to recognize complex patterns, neural networks can drive a paradigm shift in the analysis of materials-science data. As a major improvement, we introduce a crystal-structure identification method based on Bayesian deep learning that is robust to structural noise and can treat more than 100 crystal structures. While being trained on ideal structures only, our method correctly characterizes strongly perturbed single- and polycrystalline systems, from both synthetic and experimental resources. Robust crystal classification, principled uncertainty estimates, and exploratory analysis of internal neural-network representations (via unsupervised learning) enable hitherto hindered investigations of noisy atomic structural data.

Poster session / 47

## Learning Dynamics of STEM by Enforcing Physical Consistency with Phase-Field Models

**Authors:** Pawan Goyal<sup>1</sup>; Christoph Freysoldt<sup>2</sup>; Christian Liebscher<sup>3</sup>; Jaber Mianroodi<sup>None</sup>; Peter Benner<sup>None</sup>

<sup>1</sup> *Max Planck Institute for Dynamics of Complex Technical Systems*

<sup>2</sup> *MPI Eisenforschung*

<sup>3</sup> *Max-Planck-Institut für Eisenforschung*

**Corresponding Author:** goyalp@mpi-magdeburg.mpg.de

In this poster, we present our research goals of a recently BiGmax funded project towards learning dynamics of scanning transmission electron microscopy (STEM) by incorporating physical consistency with phase-field models. The primary idea of this project is to develop machine learning (ML)-based modeling of an interpretable coarse-grained dynamic model utilizing in situ STEM video sequences fulfilling a suitable dynamical phase-field equation. The modeling approach aims to discover governing equations by utilizing the video sequence data and prior physics knowledge that is directly compatible with analytic theories or subsequent ML-based analysis.

Poster session / 45

## Consistent atom probe representation for machine learning and data mining

**Author:** Alaukik Saxena<sup>1</sup>

**Co-authors:** Christoph Freysoldt<sup>1</sup>; Baptiste Gault<sup>2</sup>; Benjamin Berkels<sup>3</sup>; Dierk Raabe<sup>1</sup>

<sup>1</sup> *Max-Planck-Institut für Eisenforschung GmbH*

<sup>2</sup> *Max-Planck-Institut für Eisenforschung GmbH*

<sup>3</sup> *RWTH Aachen University*

**Corresponding Author:** saxenaalaukik93@gmail.com

To correlate mechanical properties of Al alloys with chemical segregation in Atom Probe Tomography (APT), we have developed two approaches. In the first, we collect composition statistics from APT datasets for 2x2x2 nm voxels. These voxel compositions are then clustered in compositional space using Gaussian mixture models to automatically identify key phases and their corresponding statistical descriptors. In the second, we employ SOAP (Smooth Overlap of Atomic Positions) descriptors to encode local chemical and structural environment around each atom in APT dataset. Upon using a pairwise similarity criteria on SOAP vectors, atoms lying in similar atomic environments (phases) are identified.

Poster session / 44

## Gigascale electron event processing for band structure mapping

**Authors:** R. Patrick Xian<sup>1</sup>; Laurenz Rettig<sup>1</sup>; Ralph Ernstorfer<sup>1</sup>

**Co-authors:** Yves Acremann<sup>2</sup>; Steinn Y. Agustsson<sup>3</sup>; Maciej Dendzik<sup>1</sup>; Kevin Bühlmann<sup>2</sup>; Davide Curcio<sup>4</sup>; Dmytro Kutnyakhov<sup>5</sup>; Federico Pressacco<sup>5</sup>; Michael Heber<sup>5</sup>; Shuo Dong<sup>1</sup>; Tommaso Pincelli<sup>1</sup>; Jure Demsar<sup>3</sup>; Wilfried Wurth<sup>5</sup>; Philip Hofmann<sup>4</sup>; Martin Wolf<sup>1</sup>; Markus Scheidgen<sup>6</sup>

<sup>1</sup> *Fritz Haber Institute of the Max Planck Society*

<sup>2</sup> *Laboratory for Solid State Physics, ETH Zurich*

<sup>3</sup> *Institute of Physics, University of Mainz*

<sup>4</sup> *Department of Physics and Astronomy, Aarhus University*

<sup>5</sup> *DESY Photon Science*

<sup>6</sup> *Department of Physics, Humboldt University of Berlin*

**Corresponding Authors:** xian@fhi-berlin.mpg.de, rettig@fhi-berlin.mpg.de, ernstorfer@fhi-berlin.mpg.de

Mapping of the electronic band structures of materials using momentum microscopy requires processing single-electron events of a few to hundreds of gigabytes. We construct a flexible computational workflow that allows user interaction with billion-count single-electron events in these band mapping experiments. We demonstrate its compatibility with large facility and tabletop experimental setups. The workflow is open source and offers an end-to-end recipe from data source to database. Both the workflow and processed data can be archived for reuse, providing the infrastructure for documenting the data provenance for high-throughput materials characterization.

**Session IV / 58**

## **Knowledge-Based Approaches in Catalysis and Energy Modelling**

**Author:** Karsten Reuter (invited)<sup>1</sup><sup>1</sup> *FHI Berlin*

Data sciences are now also entering theoretical catalysis and energy related research with full might. Automatized workflows and the training of machine learning approaches with first-principles data generate predictive-quality insight into elementary processes and process energetics at undreamed-of pace. Computational screening and data mining allows to explore these data bases for promising materials and extract correlations like structure-property relationships. At present, these efforts are still largely based on highly reductionist models that break down the complex interdependencies of working catalysts and energy conversion devices into a tractable number of so-called descriptors, i.e. microscopic parameters that are believed to govern the macroscopic function. Generally, static predefined databases are also the norm. Future efforts will concentrate on using artificial intelligence also in the actual generation and reinforced improvement of the reductionist models, and in devising active learning approaches that generate the truly required data on demand. In this talk, I will briefly survey these developments, providing examples from our own research, in particular on data-efficient approaches to reaction kinetics and active machine learning for the design of organic semiconductors.

**Session IV / 37****NOMAD – A FAIR Data Sharing Platform for Materials Science****Author:** Markus Scheidgen<sup>1</sup><sup>1</sup> *Humboldt Universität zu Berlin / Fritz Haber Institut der Max Planck Gesellschaft***Corresponding Author:** markus.scheidgen@physik.hu-berlin.de

As an integral part of the FAIR-DI/FAIRmat initiatives, NOMAD is extending its scope. NOMAD evolves from a central repository for publishing electronic structure codes data into a federated data management network that covers all branches of materials science. Instead of just using NOMAD to publish final results, we want to show how on-site installations of NOMAD can help to manage your local daily research data, how we plan to build a distributed network of NOMAD servers, how we combine data artefacts from various material science domains, and how we plan to involve a larger community.

## Session IV / 56

## **On Software Tools Which Assist Electron Microscopists with Sharing Metadata and Numerical Results in Accordance with the FAIR Data Stewardship Principles**

Microscopy and spectroscopy experiments and the associated computational and theoretical analyses of data from such experiments are the resources of laboratory and data-processing workflows that yield numerical data and contextualization through metadata. The purpose of such experiments is ideally accurate and precise delivery of quantitative evidence in support of or against a formulated set of research hypotheses.

Faced with a large variety of hardware and software tools, and individually larger and faster acquirable data, makes a comprehensive documentation of data and metadata a challenging task. These challenges have consequences for how findable, accessible, interoperable, and how reproducible research with experiments at present is; and thus how efficiently such data can be exchanged between scientists.

To take action the German Research Foundation has made a cross-disciplinary call to form a number of national consortia to work on the building of a national research data infrastructure for experimentalists. One proposed consortium is FAIRmat, whose aim is the building of such an infrastructure for methods of the condensed-matter physics community.

In this talk, I will report an example of the methods used and the software tools which FAIRmat will develop. Specifically, the example will be on how we organize metadata of electron microscopy experiments using a common metadata schema. We will report on the role and value of parsers as tools within automated protocols for filling in the respective metadata schema. Our results support that it is possible to find a common schema to store detailed microscopy (meta)data.

**Session IV / 38****Approaches to FAIR TEM data**

**Authors:** Niels Cautaerts<sup>1</sup>; Christian Liebscher<sup>1</sup>; Christoph Koch<sup>2</sup>; Gerhard Dehm<sup>1</sup>

<sup>1</sup> *Max-Planck-Institut für Eisenforschung*

<sup>2</sup> *Humboldt university of Berlin*

**Corresponding Author:** n.cautaerts@mpie.de

Transmission electron microscopy data is rich in quantitative information about materials, information that could in theory be coupled to atomistic simulations, but extracting and harnessing that information is non-trivial. Machine learning approaches may facilitate this, but these are hampered by the limited availability and interoperability of the data. In this talk we present approaches, progress, and challenges to making TEM data more F.A.I.R. (findable, accessible, interoperable and reusable) with the ultimate aim to bridge the gap between TEM, machine learning and atomistics.