FAIRmat users meeting

7 June 2023

Contents

Schedule of talks

Time	Talk	Speaker
13:00	Introduction to FAIRmat	Christoph T. Koch
13:30	Molecular electronics and spintronics as a challenge for first-principles methods	Carmen Herrmann
14:00	Data management plans: The foundation for proper handling of research data	Ahmed Mansour
14:20	The Electronic Lab Notebook functionality in NOMAD applied to an epitaxial synthesis use case	Andrea Albino
14:40	On Software Tools for Reproducible Atom Probe Research	Markus Kühbach
15:00	Poster session	

List of posters

1.	DAPHNE4NFDI	DAPHNE4NFDI: DAta from PHoton and Neutron Experiments for NFDI
2.	FAIRmat	FAIRmat: FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids
3.	NFDI4Cat	NFDI for catalysis related sciences: Shaping the digital future of catalysis
4.	PUNCH4NFDI	Particles, Universe, NuClei and Hadrons for NFDI: Update on current technical implementations and work in progress
5.	Adrianna Wojas	FAIR Data Research with NOMAD and NOMAD Oasis
6.	Fadwa El Mellouhi	Accelerating Materials discovery: Best practices for Research Data Management
7.	Florian Dobener	Bringing Scientific Communities and Technology Partners Together
8.	Gabriel Preuß	How are we making Matter data FAIR?

9.	Gerrit Günther	Towards Intelligible Metadata
10.	Hampus Näsström	An Ontology for Sample Synthesis by Physical Vapor Deposition
11.	Henriette Wiehl & Hoa Nguyen	FID Materials Science - Your Open Pathway to the World of Materials
12.	Holger von Wenckstern	A2 Use case: Acquisition of data and metadata during pulsed laser deposition
13.	Johannes Lehmeyer	A Configurable Application for Measurements, Experiments and Laboratory Systems
14.	Jonas Lähnemann	LumiSpy - Luminescence Spectroscopy Data Analysis in Python Using HyperSpy
15.	Jonathan Noky	Integration of the Labfolder and elabFTW ELNs in NOMAD
16.	Julia Schumann	Development of a FAIR Data Repository Infrastructure for Catalysis Data
17.	Luca Ghiringhelli	FAIRifying Computational Materials-Science Data: Workflows, Ontologies, and Data

18.	Lucia Rotheray	Writing a research data management plan
19.	Mariano Forti	Steps Taken Towards Reproducible and FAIR Atom Probe Research
20.	Mohammad Khatamirad	Piloting a Local Research Data Infrastructure for Heterogeneous Catalysis
21.	Mojeeb Rahman Sedeqi	HMC Dashboard on Open and FAIR Data in Helmholtz
22.	Rubel Mozumder	Development and Implementation of Data Model for Scanning Tunneling Microscopy and Scanning Tunneling Spectroscopy
23.	Rukeia El-Athman	BAM Data Store: An Institutional Approach to Research Data Management for MSE
24.	Sebastian Brückner	Data governance framework for a sustainable change of research data management
25.	Sebastian Brückner	Towards Standard Data Structures and Tools for FAIR Synthesis Data

Abstracts

Talks

Molecular electronics and spintronics as a challenge for first-principles methods

Carmen Herrmann
Institute for Inorganic and Applied Chemistry, University of Hamburg, Germany

Molecular conductance is measured in different experimental setups such as scanning tunneling microscopes (STMs), molecular break junctions, and nanoparticle arrays [1,2,3]. The motivation behind these experiments is not only studying potential reproducible nanoscale building blocks for electronics or spintronics but also learning about molecules under unusual conditions. Using the spin degree of freedom in such settings offers fascinating options for nanoscale functionality, and also provides new experimental data for improving our insight into fundamental aspects of nonequilibrium physics at that scale [3]. Compromises between accuracy and computational feasibility imply that for molecular electronics and spintronics, a quantitative first-principles description may be elusive. We illustrate the resulting challenges as well as successes for examples such as chiral induced spin selectivity [4,5,6,7,8], length-dependent charge delocalization in molecular wires [9], conductance histograms [10,11], or single-molecule magnetoresistance [12,13], and the Kondo effect in molecular adsorbates [14,15].

[1] N. Xin, J. Guan, C. Zhou, X. Chen, C. Gu, Y. Li, M. Ratner, A. Nitzan, J. Stoddart, X. Guo, Nat. Rev. Phys., 1, 211-230 (2019) [2] Y. Liu, X. Qiu, S. Soni, R. Chiechi, Chem. Phys. Rev., 2, 021303 (2021) [3] P. Gehring, J. Thijssen, H. van der Zant, Nat. Rev. Phys., 1, 381-396 (2019) [4]R. Naaman, Y. Paltiel, D. Waldeck, Nat. Rev. Chem., 3, 250-260 (2019) [5] C. Aiello, J. Abendroth, M. Abbas, A. Afanasev, S. Agarwal, A. Banerjee, D. Beratan, J. Belling, B. Berche, A. Botana, J. Caram, G. Celardo, G. Cuniberti, A. Garcia-Etxarri, A. Dianat, I. Diez-Perez, Y. Guo, R. Gutierrez, C. Herrmann, J. Hihath, S. Kale, P. Kurian, Y. Lai, T. Liu, A. Lopez, E. Medina, V. Mujica, R. Naaman, M. Noormandipour, J. Palma, Y. Paltiel, W. Petuskey, J. Ribeiro-Silva, J. Saenz, E. Santos, M. Solyanik-Gorgone, V. Sorger, D. Stemer, J. Ugalde, A. Valdes-Curiel, S. Varela, D. Waldeck, M. Wasielewski, P. Weiss, H. Zacharias, Q. Wang, ACS Nano, 16, 4989-5035 (2022) [6] M. Zöllner, S. Varela, E. Medina, V. Mujica, C. Herrmann, J. Chem. Theory Comput., 16, 2914-2929 (2020) [7] M. Zöllner, A. Saghatchi, V. Mujica, C. Herrmann, J. Chem. Theory Comput., 16, 7357-7371 (2020) [8] S. Naskar, V. Mujica, C. Herrmann, J. Phys. Chem. Lett. 14, 694-701 (2023) [9] S. Kröncke, C. Herrmann, J. Chem. Theory Comput., 16, 6267-6279 (2020) [10] Z. Li, L Mejia, J. Marrs, H. Jeong, J. Hihath, J., I. Franco, J. Phys. Chem. C 125, 3406-3414 (2021) [11] M. Deffner, M. P. Weise, H. Zhang, M. Mücke, J. Proppe, I. Franco, C. Herrmann, J. Chem. Theory Comput. 19, 992-1002 (2023) [12] G. Mitra, J. Z. Low, S. Wei, K. R. Francisco, M. Deffner, C. Herrmann, L. M. Campos, E. Scheer, Nano Lett. 22, 5773-5779 (2022) [13] H. Zhang, C. Herrmann, ChemrXiv 2022, DOI 10.26434/chemrxiv-2022-gdqvc [14] P. Wahl, L. Diekhöner, G. Wittich, L. Vitali, M. Schneider, K.

Kern, Phys. Rev. Lett., 95, 166601 (2005) [15] M. Bahlke, M. Schneeberger, C. Herrmann, J. Chem. Phys., 154, 144108 (2021)

Data management plans: The foundation for proper handling of research data

Ahmed Mansour
Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin, Germany

As scientific laboratories continue to generate vast volumes of data, the need for proper handling and management of research data has become increasingly critical. The heterogeneous nature of research data further emphasizes the importance of adopting a comprehensive data management process across various career levels. This process begins with planning before project initiation, followed by ongoing monitoring and control throughout the project's duration, and concludes with defining the long-term fate of the data after project completion. Data management plans (DMPs) are the foundation of this approach. For this reason, funding agencies now commonly mandate the inclusion of a DMP in research proposals. In this talk, I will describe the fundamental elements of an effective DMP, while providing practical insights and tips specifically tailored to the fields of condensed-matter physics and materials science. Moreover, we will highlight how to prepare DMPs that align with the rigorous requirements set by the German Research Foundation (DFG) and ensures proper handling of research data.

The Electronic Lab Notebook functionality in NOMAD applied to an epitaxial synthesis use case

Andrea Albino[1], Hampus Näsström[1], Jose Marquez Prieto[1], Sebastian Brückner[2], Lauri Himanen[1], Mohammad Nakhaee[1], Amir Golparvar[1], Markus Scheidgen[1], Luca Seravalli[3], Piero Mazzolini[3], Matteo Bosi[3]

[1]Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin, Germany

[2]Leibniz-Institut für Kristallzüchtung (IKZ), Berlin, Germany

[3]IMEM-CNR Institute, Parma, Italy

Approaching the era of big data-driven materials science, one crucial step to collecting, describing, and sharing experimental data is the adoption of ELNs. The project FAIRmat (fairmat-nfdi.eu) is offering such tools by developing and operating the open-source software NOMAD. The NOMAD ELN aims at offering a secure environment to protect the integrity of both data and metadata, whilst also affording the flexibility to adopt new synthetic processes or changes to existing ones without recourse to further software development. We find that to promote an early adoption, it is important to adapt to a single user's needs and workflows. An inductive approach, going from a particular set of experiments to a general description of the similarities recurring in each of them, led us to adopt a common data structure as a standard. The state-of-the-art ELN features for a synthetic process will be shown in the talk, highlighting the development of both data modeling and specific implementation solutions.

In this talk, we will show how the NOMAD platform can be applied on the whole RDM life cycle in epitaxy, meeting FAIR data standards. On the example of data from MOVPE synthesis process from a research group in IMEM-CNR institute in Parma we will demonstrate the state-of-the-art ELN features for a synthesis process in NOMAD, focusing both on the underlying data model concepts and on the implementation of a use case. We will show how structured data facilitates and automatizes data analysis and visualization, and allows to routinely access machine learning tools for process optimization in synthesis.

On Software Tools for Reproducible Atom Probe Research

Markus Kühbach[1], Sandor Brockhauser[1], Alexander Reichmann[2], Lorenz Romaner[2], Sarath Menon[3], Alaukik Saxena[3], Mariano Forti[4], Baptiste Gault[3,5], Tilmann Hickel[3,6], Thomas Hammerschmidt[4]

[1] Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin, Germany

[2]Montanuniversität Leoben, Lehrstuhl für Metallkunde, Franz Josef-Straße, 18 A-8700 Leoben, Austria

[3]Max-Planck-Institut für Eisenforschung GmbH, Max-Planck-Straße 1, Düsseldorf, Germany

[4]ICAMS Ruhr-Universität Bochum, Department of Atomistic Modelling and Simulation, Universitätsstraße 150, 44801 Bochum, Germany

[5]Royal School of Mines, Imperial College London, South Kensington Campus

[6]Materials Informatics, Bundesanstalt für Materialforschung und -prüfung (BAM), Unter den Eichen 87, 12205 Berlin, Germany

Atom probe tomography and related field-ion microscopy became mature nanoscale structure and composition characterization methods[1, 2]. Commercial as well as open-source software is available which can profit from improvements to support research data management according to the FAIR[4] and FAIR4RS[5] principles. We are convinced this brings benefits such as a stronger contextualization and automated provenance tracking for atom probe experiments and analyses, exactly reproducible and

more automatable workflows, and possibilities for sharing datasets between laboratories[3] using a common description. In this talk we deliver an update of steps and software tools developed to support the atom probe community. This will include a summary of domain-specific data schemas for describing the data collection for all relevant instruments and include a proposal of about a dozen schemas for describing computational-geometry-focused data analysis steps whereby reconstructed atom positions are transformed into engineering-relevant descriptors of materials' micro- and nanostructure. Software development of domain-specific parsers and an electronic lab notebook template which can be used as a standalone or service in NOMAD, as well as containerization of applications for atom probe practitioners will be covered.

References:
[1] B. Gault et al., Nat. Rev. Methods Primers 1 (2021) 52, https://doi.org/10.1038/s43586-021-00054-x
[2] S. Katnagallu et al., New J. Phys. 21 (2019) 123020, https://doi.org/10.1088/1367-2630/ab5cc4
[3] M. Kühbach et al., Microsc. Microanal, 28(4) (2021), p.1038, https://doi.org/10.1017/S1431927621012241
[4] M. D. Wilkinson et al., Scientific Data 3 (2016) 160018, https://doi.org/10.1038/sdata.2016.18
[5] M. Barker et al., Scientific Data 9 (2022) 622, https://doi.org/10.1038/s41597-022-01710-x

Posters

NFDI consortia

A. DAPHNE4NFDI: DAta from PHoton and Neutron Experiments for NFDI
B. FAIRmat: FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids
C. NFDI for catalysis related sciences: Shaping the digital future of catalysis
D. Particles, Universe, NuClei and Hadrons for NFDI: Update on current technical implementations and work in progress

FAIR Data Research with NOMAD and NOMAD Oasis

Adrianna Wojas[1], Markus Scheidgen[1]
[1] Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin, Germany

The NOMAD poster provides a comprehensive overview of two powerful tools, NOMAD and NOMAD Oasis, designed to revolutionize the landscape of scientific data research. Developed by the FAIRmat consortium, NOMAD and NOMAD Oasis promote the principles of Findable, Accessible, Interoperable, and Reusable (FAIR) data, enhancing transparency and collaboration in materials science. NOMAD is a centralized material science data repository addressing the needs of researchers from various domains and fields. It seamlessly manages diverse data formats, including raw files, facilitating efficient data analysis and exploration. The poster highlights the key features and uses cases of NOMAD and NOMAD Oasis while thoroughly exploring their capabilities.

The project outlines the data workflow within the NOMAD ecosystem, illustrating how uploaded data is transformed into processed and modeled data. Subsequently, it offers three options for utilizing the parsed data: publish, analyze, and explore. Researchers can leverage NOMAD's data publishing capabilities to share their findings with the scientific community, obtain a DOI, and support open collaboration. By promoting FAIR data research, the NOMAD poster underscores the importance of data integrity, accessibility, and interoperability. It is a valuable resource for scientists, data centers, and scientific organizations, showcasing the potential of NOMAD and NOMAD Oasis in everyday workflow.

Accelerating Materials discovery: Best practices for Research Data Management

Johanne Medina[a], Abdul Wahab Ziaullah[b], Heesoo Park[c], Ivano E. Castelli[d], Arif Shoan[e], Halima Bensmail[f] and Fedwa El-Mellouhi[b]

[a]College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

[b]Qatar Environment and Energy Research Institute, Hamad Bin Khalifa University, Doha, Qatar

[c]Centre for Material Science and Nanotechnology, Department of Chemistry, University of Oslo, Oslo, Norway

[d]Department of Energy Conversion and Storage, Technical University of Denmark, Kgs. Lyngby, Denmark

[e]Qatar National Library,Qatar Foundation, Doha, Qatar

[f]Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar

The need for good research data management (RDM) practices is becoming more recognized as a critical part of research. This may be attributed to the 5V challenge in Big Data, including volume, variety, velocity, veracity, and value. The materials science community is no exception to these challenges as it heralds its new paradigm of data-driven science, which uses artificial intelligence to accelerate materials discovery but requires massive datasets to perform effectively. Hence, there are efforts to standardize, curate, preserve, and disseminate these data in a way that is Findable, Accessible, Interoperable, and Reusable (FAIR). To understand the current state of data-driven materials science and learn about the challenges faced with RDM, we gather user stories of researchers from small and large-scale projects. This enables us to provide relevant recommendations within the data-driven research life cycle to develop and/or procure an effective RDM system following the FAIR guiding principles.

Bringing Scientific Communities and Technology Partners Together

Lukas Pielsticker[1,2], Florian Dobener[2], Rubel Mozumder[2], Laurenz Rettig[3], Walid Hetaba[1,2], Sandor Brockhauser[2], Heiko Weber[2,5]
[1]Max Planck Institute for Chemical Energy Conversion, Mülheim an der Ruhr, Germany
[2]FAIRmat, Humboldt-Universität zu Berlin, Berlin, Germany
[3]Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Germany
[5]Friedrich-Alexander-Universität, Erlangen-Nürnberg, Germany

The complexity and size of photoemission data is rapidly increasing as new technological breakthroughs have enabled multidimensional parallel acquisition. However, most of the community is currently using heterogeneous data formats and workflows. Therefore, there is an urgent need to develop standardized exchange and storage formats describing photoelectron spectroscopy techniques and their numerical data and metadata. We propose a new data format based on NeXuS, a community standardisation platform with a controlled vocabulary and defined relations between hierarchically organised entries. The proposed format for multidimensional photoemission spectroscopy is designed to allow high-performance automated access, enabling experimental databases for high-throughput material search. We recently started to implement our approach in several labs and techniques, e.g., for ARPES, XPS, and PEEM. In order to to establish a community standard, we have started to collaborate with and include not only the scientific community, but also reached out to vendors of photoelectron spectroscopy instruments - our technology partners - to ensure that the standard developed together will be supported by state-of-the-art measurement systems. In a first kick-off workshop, we discussed with the technology partners about our approach and started a process of cooperation to eventually reach full support of FAIR data generated by commercially available scientific instruments. Here, we report on the results from this workshop and show it as an example of answering the need for interaction with the wider scientific and industrial community in order to make data FAIR also in other experimental domains.

How are we making Matter data FAIR?

Gabriel Preuß, Helmholtz-Zentrum Berlin für Materialien und Energie GmbH

Creating data according to the FAIR principles is already part of good scientific practice but the FAIR principles are high level guidelines rather than concrete implementations. That is why one of the key missions of HMC is to support the Helmholtz community in making their data FAIR in an easy and comparable way. Here, we present concrete approaches of HMC Hub Matter to make data of the BESSY II Light Source FAIR, ranging from uses cases at beamlines over training of the instrument staff to policy review and information tools.

Towards Intelligible Metadata

Gerrit Günther, Helmholtz Metadata Collaboration

The process of annotating research data with bibliographic metadata when publishing research results is widely understood. Less well understood is how to annotate research data with domain-specific metadata in an interoperable manner. Here, we address machine-readable definitions of metadata that are suitable for the Helmholtz Hub Matter, ranging from data format standards such as NeXus to the implementation of web-proven semantics like RDFa through concrete examples and use cases.

An Ontology for Sample Synthesis by Physical Vapor Deposition

Hampus Näsström[1], José A. Márquez[1], Michael Götte[2], Andrea Albino[1], Sebastian Brückner[1], Holger von Wenckstern[3]

[1] Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin, Germany

[2]Helmholtz-Zentrum Berlin

[3] Universität Leipzig

A very common thin-film deposition technique used throughout the full range of technology readiness levels is that of Physical Vapor Deposition (PVD). This deposition method is deployed from fundamental materials research all the way to large scale industry production lines. Although most PVD techniques are well described in literature, there is, as far as we can tell, no comprehensive ontology for these processes. As part of the work done by the NFDI consortium FAIRmat (https://fairmat-nfdi.eu), we therefore want to construct an ontology for PVD techniques starting with thermal evaporation, pulsed laser deposition (PLD) and sputtering. The ontology will be based on the Basic Formal Ontology (BFO) and more specifically on the taxonomy defined in the Chemical Methods Ontology (CHMO). In this poster we present our first effort in generalizing the three methods mentioned above and open up for input from the community.

FID Materials Science - Your Open Pathway to the World of Materials

Henriette Wiehl[1] & Hoa Nguyen[2]
[1]Saxon State and University Library Dresden (SLUB)
[2]German National Library of Science and Technology (TIB).

The DFG-funded Fachinformationsdienst für Materialwissenschaft und Werkstofftechnik (Specialised Information Service for Materials Science), or FID Materials Science for short, is aimed primarily at researchers and institutions in the fields of materials science and materials engineering. The aim of the project is to provide rapid and effective information and networking to support research and initiate cooperations. For this purpose a couple of services are provided: FID Portal, Material Hub, Connection to TIB's catalog, Materials Guide and Materials Map. The portal of FID Materials Science (materials-science.info) gives an overview of the tools and services to find specialised information. Material Hub is a data base for finding characteristic materials properties and it combines material information from industry and research and thus enables intuitive and efficient research. It was developed at the SLUB in the ERDF-funded project of the same name and will be further developed in the FID. Materials Guide lists external specialist information databases in materials science. This way, required materials properties or further material class-related information can be found even if they are not yet available in the Material Hub. Using Materials Map it is easy to find experts in the interdisciplinary field of materials science. Based on publicly

available information, the Materials Map provides an overview of the research community, its actors and their expertise and publications. The connection to the TIB's catalog holdings enables access to scientific information and thus supports further research. Scientific publications and textbooks as well as audiovisual media, research data, patents and standards can be searched via the TIB portal. The FID Materials Science is jointly operated and further developed by the Saxon State and University Library Dresden (SLUB) and the German National Library of Science and Technology (TIB).

A2 Use case: Acquisition of data and metadata during pulsed laser deposition

Holger von Wenckstern, Felix-Bloch-Institut, Universität Leipzig

We present an application case for the handling, storage, retrieval, analysis and visualization of data and metadata generated during thin film growth by pulsed laser deposition. We have implemented a custom PLD control software that collects all process-relevant data and metadata (equipment related, source materials, process parameters, etc.) and transfers them to an electronic lab book. In addition, we implemented a self-developed front-end that allows a very efficient screening of the database as well as analysis and visualization of the stored data.

NOMAD-CAMELS: A Configurable Application for Measurements, Experiments and Laboratory Systems

A. D. Fuchs[1,2], J. A. F. Lehmeyer[1,2], P. Oppermann[3], H. Junkes[3], H. B. Weber[2], M. Krieger[2]

[1] Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin, Germany

[2]Lehrstuhl für Angewandte Physik, Department Physik, Friedrich-Alexander-Universität Erlangen-Nürnberg

[3]Fritz-Haber-Institut der Max-Planck-Gesellschaft (FHI), Berlin

We are currently developing an open-source configurable measurement software, especially targeted towards the rather heterogeneous requirements of experimental solid-state physics. Here many experiments utilize a great multitude of measurement instruments, which are used in dynamically changing ad-hoc setups. The software NOMAD-CAMELS [1,2] allows the researcher to define instrument control and measurement protocols using a graphical user interface (GUI). We aim for a low entry threshold enabling the creation of new measurement protocols without programming knowledge or deeper understanding of device communication. The GUI compiles the configured measurement and process protocols into fully transparent python code that interfaces with instruments and actually performs the required tasks, thus allowing the researcher to review the code or even use it as a template for further development. While NOMAD-CAMELS is lightweight, it is scalable and can readily connect to large-scale, distributed systems using e.g. the EPICS framework. NOMAD-CAMELS generates FAIR-

compliant output data including rich metadata (instruments settings, measurement protocol, python code, timestamps, default plots, etc.). NeXus standards, immediate NOMAD integration and hence a FAIRmat-compliant data pipeline can be readily implemented. The overall goal is to create a software, which enables and supports the complete experimental workflow from envisioning a new experiment including the design of the experimental protocol to finally generating FAIR data.

[1] NOMAD-CAMELS documentation: https://fau-lap.github.io/NOMAD-CAMELS/ [2] NOMAD-CAMELS GitHub repository: https://github.com/FAU-LAP/NOMAD-CAMELS

LumiSpy - Luminescence Spectroscopy Data Analysis in Python Using HyperSpy

Jonas Lähnemann[1], Mikel Gomez[1], and the HyperSpy and LumiSpy developers
[1] Paul-Drude-Institut für Festkörperelektronik

Python is a versatile programming language that has found widespread use in data analysis boasting a broad range of specialized scientific libraries. HyperSpy [1] is an open source python library dedicated to the interactive analysis of multi-dimensional datasets, in particular spectrum images (one or two-dimensional arrays of spectra), in an easy and reproducible fashion. HyperSpy facilitates the application of analytical procedures operating on individual spectra to a multi-dimensional dataset and gives easy access to tools that exploit the multi-

dimensionality of the dataset. LumiSpy [2] is an extension to HyperSpy specifically for luminescence spectroscopy data. It is aimed at helping with the analysis of luminescence spectroscopy data, in particular continuous as well as time-resolved catho-, and photoluminescence spectral imaging, but also electroluminescence, and Raman spectroscopy. Import from a range of scientific data formats, including the relevant metadata, is supported through the RosettaSciIO [3] library. Through the integration with HyperSpy, LumiSpy is particularly well suited not only for the analysis of single spectra, but also for the treatment of spectral maps, which includes the fitting of complex models to the data. The poster will give an introduction to the HyperSpy and LumiSpy libraries and where the development is heading.
References: [1] https://hyperspy.org, DOI: 10.5281/zenodo.592838 [2] https://lumispy.org, DOI: 10.5281/zenodo.4640445 [3] https://hyperspy.org/rosettasciio

Integration of the Labfolder and elabFTW ELNs in NOMAD

Jonathan Noky[1], Amir Golparvar[2], Sebastian Brückner[3]
[1]Max-Planck-Institut für chemische Physik fester Stoffe Dresden
[2] Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin, Germany
[3]Leibniz-Institut für Kristallzüchtung Berlin

Here, we present an approach to integrate the LabFolder and elabFTW electronic lab notebooks (ELNs) into NOMAD. This provides a good solution for groups where either the ELNs are

already established or more accepted by the users for its different user interface and experience. In LabFolder, we achieve this by making use of the template functionality in combination with the data element feature of the ELN to ensure structured entries. Knowing the structure, the entries are exported via the API and imported into NOMAD in an adapted form. This whole process can be done from the NOMAD server. elabFTW provides an API endpoint to export the (experiment) data in ELNFileFormat (a standard file format for ELNs - https://github.com/TheELNConsortium/). NOMAD uses this functionality and integrates the elabFTW data by importing and mapping the ELN file into a built-in NOMAD schema. With the integration we show, that it is possible to manage (parts of) the user input via ELNs while still maintaining the capabilities of NOMAD. In the future, similar integrations should be possible for other ELNs providing sufficient API functionality.

Development of a FAIR Data Repository Infrastructure for Catalysis Data

J. Schumann[1,2], A. Trunschke[1,2]

[1] Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin, Germany

[2]Department of Inorganic Chemistry, Fritz-Haber-Institut der Max-Planck-Gesellschaft

Catalysis research generates a vast amount of complex data, which can be difficult to manage, share, and reuse effectively. To address these challenges and enable future AI-assisted catalysis

development, a FAIR data infrastructure tailored to catalysis research has become an urgent need. This infrastructure should ensure that data is Findable, Accessible, Interoperable, and Reusable, according to the FAIR data principles. The development of such a successful infrastructure is a complex task which needs to be a carried by the catalysis community. In this contribution, we want to focus on two important aspects: Firstly, on the individual research laboratory level, there is a need to transition from manual, paper-based documentation, to automated experiments and digitalization of laboratory documentation. This involves using standardized data formats and complete metadata. This shift will enhance efficiency, reproducibility, and streamline data management processes. Secondly, on the catalysis community level, it is essential to develop controlled vocabulary ideally recorded in a standardized ontology, and establish agreement on reporting standards and mandatory metadata to accompany every publication containing catalysis data. This concerted effort will ensure consistency and facilitate data integration and comparison. We will showcase an existing solution for local data acquisition and storage at the Fritz-Haber-Institute (FHI) and display a first blueprint for a data entry schema designed specifically for catalysis data in NOMAD. By addressing these aspects, we aim to contribute to the development of a robust and unified data infrastructure for catalysis research, enabling efficient data sharing and application of new machine learning tools for faster and more successful catalyst development in the future.

FAIRifying Computational Materials-Science Data: Workflows, Ontologies, and Data

Luca Ghiringhelli[1], Nathan Daelman[1], Joseph Rudzinski[1], José Maria Pizarro[1]

[1] Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin, Germany

The NOMAD Laboratory maintains since 2014 a data infrastructure for computational materials-science data: the NOMAD Repository & Archive. Data from more than 50 atomistic simulation codes, spanning the overwhelming majority of the data produced in the field, are stored and translated into the NOMAD Metainfo, a modular, hierarchical metadata schema that allows for data Findability and Accessibility, which are the first two requirements of FAIR data. For a full Interoperability and Reusability, however, more powerful technologies need to be developed: extension of the current Metainfo to represent complex computational workflows, ontologies for establishing semantic relationship among data and metadata, and data quality assessments. In FAIRmat, we are developing tools in these three directions. In this contribution, we present recent progress in: 1) Annotation of precision (towards data-quality) for electronic structure calculations; 2) Semantically rich representation of exchange-correlation functionals of DFT, in order to find similarity amongst functionals; 3) Hierarchical representation of atomistic structures (atoms build moieties that build molecules that build super-molecules, ...); 4) Support of a variety of excited-state and advanced many-body methods: GW, BSE, DMFT; 5) Complex

workflows and visualizations of electronic properties. The audience will be prompted to experience on their device of choice updated demos for all recent developments, reachable via a suitable QR code. We also welcome users to sit with us with questions, feedback, and suggestions.

Writing a research data management plan

Ahmed E. Mansour[1], Lucia Alessandra Rotheray[1], Kerstin Helbig[2],Silvana Botti[3], Heiko B. Weber[4], Martin Aeschlimann[5] and Claudia Draxl[1]

[1] Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin, Germany

[2]Computer and Media Service, Humboldt-Universität zu Berlin

[3]Institut für Festkörpertheorie und -optik, Friedrich-Schiller-Universität Jena, Germany

[4]Department of Physics, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany

[5]Department of Physics and Research Center OPTIMAS, Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau, Germany

Good research data management has always been crucial, but is now becoming a more urgent topic due to the increased volume and variety of data produced by scientific research. As well as an increased awareness of data issues in the community, this is being driven by an increased demand from funders and journals that researchers manage their data FAIRly and sustainably. In particular many funders now require a comprehensive data

management plan (DMP). While all researchers need to be involved in preparing and executing a DMP, many are still uncertain how to approach writing one. FAIRmat offers practical advice and resources, which we summarize here.

Automating NOMAD equation of state workflow for High throughput DFT calculations

Mariano Forti, ICAMS, Ruhr Universität Bochum

High Throughput (HT) DFT calculations of the so-called energy-volume curves are common practice in state of the art materials discovery. They are essential to determine materials stability and equilibrium properties, and requires high amount of cpu-time and fine-tuning calculation parameters. This step usually generates high volume of information and its preservation is essential not only for the validation of the calculations but also for allowing the re-usage of the data. In this work, we will show the integration of NOMAD into our tailored HT manager to automate the publication of energy-volume curves for bulk intermetallics.

Steps Taken Towards Reproducible and FAIR Atom Probe Research

Markus Kühbach[1], Sandor Brockhauser[1], Sarath Menon[2], Alaukik Saxena[2], Mariano Forti[3], Baptiste Gault[2,4], Tilmann Hickel[2,5], Thomas Hammerschmidt[3]
[1]Humboldt-Universität zu Berlin, Department of Physics,Germany

[2]Max-Planck-Institut fÃ¼r Eisenforschung GmbH, Düsseldorf, Germany
[3]ICAMS Ruhr-Universität Bochum, Department of Atomistic Modelling and Simulation, Bochum, Germany
[4]Royal School of Mines, Imperial College London
[5]Materials Informatics, Bundesanstalt für Materialforschung und -prüfung (BAM), Berlin, Germany

Atom probe tomography and related field-ion microscopy became mature nanoscale structure and composition characterization methods. There are several possibilities for improving the current state-of-the-art in terms of how software for analyzing atom probe data gets developed, how such software is being used, and how we can support scientists who would like to share datasets and results more frequently between laboratories. These efforts are guided by the FAIR and FAIR4RS principles. With this poster, we deliver an update of the steps and software tools developed reporting domain-specific additions to NOMAD supporting the atom probe research community. Specifically, we will report about a set of NeXus data schemas whereby atom probe experiments can be described for all relevant instruments including about a dozen computational-geometry-focused data analysis steps whereby reconstructed atom position are transformed into engineering-relevant descriptors of the materials' micro- and nanostructure. Domain-specific parsers and electronic lab notebook templates as well as customized containerized applications for atom probe practitioners will also be covered. The

implementation of these tools profits from having members of the FAIRmat and the NFDI-MatWerk consortia working together.

Piloting a Local Research Data Infrastructure for Heterogeneous Catalysis

Mohammad Khatamirad, Technical University of Berlin

NFDI4Cat, a consortium within the NFDI initiative, aims at advancing the digitalization of research data within the German catalysis community enabling scientists to utilize research data throughout the whole research data lifecycle. This initiative is specifically important in modern catalysis research, where the connection to data-driven methods is rapidly growing. This connection supports the linking of different disciplines within the catalysis research community through the help of data scientists and mathematicians. To enable an interdisciplinary exchange of knowledge, data and tools, a data infrastructure and a fundamental base is needed. A local pilot was set up by BasCat and Fraunhofer FOKUS as a framework to promote the required interdisciplinary work for digital catalysis research. This pilot serves as framework to implement tools for data cleaning, visualization, exploration, statistical investigation, correlation analysis as well as quality assurance. These tools are scripted in a generic form enabling reuse for different reactions while also tracing changes in ontology/meta data supporting reproducibility of analysis outputs. Synthesis and test of the catalysts is carried out at BasCat for Rh-based syngas hydrogenation catalysts. The base of the pilot is formed by a Dataverse repository storing data

in a structured manner, enabling exploration, sharing and publishing of data. It also supports the integration of external applications providing tools enabling data-driven research. A web application hosting a set of visualization tools was connected to the repository as well as a JupyterHub (JupyterLab) to increase flexibility enabling users to develop and operate their personal models and algorithms on the data provided by the repository. On top of those systems a meta portal is deployed harvesting the metadata provided by the Dataverse repository and serving as an open access point for this metadata.

HMC Dashboard on Open and FAIR Data in Helmholtz

Mojeeb R. Sedeqi[1,2], Alexander Schmidt[1], Gabriel Preuß[1,2], Astrid Gilein[1], Tempest Glodowski[1], Vivien Serve[1,2], Oonagh Mannix[1,2], Markus Kubin[1,2]
[1]Helmholtz-Zentrum Berlin für Materialien und Energie
[2]Helmholtz Metadata Collaboration

As part of our work in the Helmholtz Metadata Collaboration (HMC), we present an integrated approach to monitoring and assessing the state of open and FAIR data in the Helmholtz Association through harvesting and interactive visualization methods. The results are published as the "HMC Dashboard on Open and FAIR Data" under https://fairdashboard.helmholtz-metadaten.de. Literature metadata is harvested from several research centers in the Helmholtz Association. Data publications linked to that literature are identified using the SCHOLIX API. In a first approach to an automated FAIR assessment of the data

identified we adopted the F-UJI framework [1] (originally developed by the FAIRSFAIR consortium). The collected results are presented in an interactive dashboard, which allows users to monitor, analyze and explore statistics on open and FAIR data of data publications found in Helmholtz- and non-Helmholtz repositories. All plots are interactive, allowing users to filter and focus on the data they are interested in. The HMC dashboard targets different groups such as Helmholtz researchers, data-managers, libraries, and repositories within Helmholtz and beyond; as well as management and provides target group-specific information. For example, it provides information supporting Helmholtz researchers to improve the FAIRness of their data. It provides information on the repositories where Helmholtz data is deposited and helps to identify gaps that can be addressed towards improving the FAIRness of that data. The dashboard is now publicly available. All source code is open for contributions and reuse by all research communities. Keywords: OPEN DATA, FAIR DATA, METADATA HARVESTING, DASHBOARD
References: [1] A Devaraju and R Huber: F-UJI - An Automated FAIR Data Assessment Tool (2020). doi: 10.5281/zenodo.4063720

Development and Implementation of Data Model for Scanning Tunneling Microscopy and Scanning Tunneling Spectroscopy

Yichen Jin [1], Rubel Mozumder[1], Yan Wang[2] ,Jürgen P. Rabe[1], Heiko B. Weber[3], Tamás Haraszti[4], Sabine Maier[3], Carlos-Andres Palma[1,2], Sandor Brockhauser[1], Claudia Draxl[1]

[1] Department of Physics & IRIS Adlershof, Humboldt-Universität zu Berlin, Germany

[2]Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing , People's Republic of China

[3]Department of Physics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

[4]Leibniz-Institut für Interaktive Materialien e.V., RWTH Aachen, Germany

Since the first realization of scanning tunneling microscopy (STM) and scanning tunneling spectroscopy (STS), a wide range of applications has been developed by using the technique of quantum tunneling effect. With the complexity of this technique, it has become possible to analyze to molecular structural image of surface as well as electronic properties of metal, semiconductor , single molecule, and even supermolecular structures. In materials science, the community aims at organize experimental data along FAIR principles, thereby developing or using a standard vocabulary of metadata. NeXus allows to generate such a sort of data modeling and with that we have developed specific data

model called NXiv_sweep2 for STM/STS experiment. In FAIRmat (Area-B, associated with experimental area in materials science), beside the STM/STS we also cover several experimental fields (e.g. APM - Atom Probe Microscopy, XPS - X-ray Photoelectron Spectroscopy, ARPES - Angle-Resolved Photoemission Spectroscopy). Our work focuses not only on storing data and metadata (via NOMAD not discussed here) from STM/STS experiments in terms of the NXiv_sweep2 vocabulary but also supports handling data-analysis results (e.g. topography, I/V, and dI/dV). In this presentation, we will show how experimental data can be converted into the NeXus data format. We will detail the basic software architecture of reader that transforms experimental data file into STM/STS specific data model which in terns follows NeXus data format concepts.

BAM Data Store: An Institutional Approach to Research Data Management for MSE

Angela Ariza[1], Rukeia El-Athman[1], Jörg Rädler[1], Thilo Muth[1]
[1]Bundesanstalt für Materialforschung und -prüfung (BAM)

The Bundesanstalt für Materialforschung und -prüfung (BAM) participates as an active partner in several NFDI consortia and contributes to the development of guidelines and metadata standardization efforts to implement the FAIR data principles in the field of materials science and engineering (MSE). In the eScience and IT sections of BAM, we aim to facilitate these efforts by providing a centralized framework for RDM and electronic lab notebooks that enable researchers at BAM to represent

heterogenous workflows in a standardized manner while still allowing for domain-specific requirements. To this end, we conducted a pilot project to test the implementation of a central BAM Data Store based on the open-source software openBIS. During the pilot phase, five groups from different MSE disciplines successfully used the framework to digitally document research processes enriched with metadata and to store associated data in a central and structured form in line with RDM best practices. By offering interfaces to other systems, openBIS further allowed researchers to automate (meta)data import and data-driven analysis. Based on the positive experience of the pilot project, the rollout of the system begins in May 2023 with the aim of establishing the Data Store as the backbone infrastructure for RDM at BAM. Despite the success of the project, we noticed that the sustainable adoption of RDM tools and software by researchers remains a challenge because the initial setup and customization require significant effort by the researchers which often remains invisible in the academic context. To ensure researchers' continued engagement in enriching these tools with domain-specific knowledge, funding bodies must provide resources for the continuous development of RDM standards as well as for the underlying software.

Data governance framework for a sustainable change of research data management

Sebastian Brückner, Department of Physics & IRIS Adlershof, Humboldt-Universität zu Berlin & Leibniz-Institut für Kristallzüchtung Berlin, Germany

Implementation of data governance policies and frameworks is necessary for a sustainable change of research data management (RDM) at a research organization. In particular, introducing and applying new RDM tools and the FAIR principles can be challenging at institute level if no adequate data governance policy is at place. Data governance policies and frameworks help to work out topics such as data strategy & vision, data access, data management & infrastructure, data skills and training, definition of data quality, definition of roles and responsibilities of the different user groups, management of interests of all stakeholders. Here, we will show our measures regarding data governance at the Leibniz Institut für Kristallzüchtung (IKZ Berlin) to support the digital transformation of the institute.

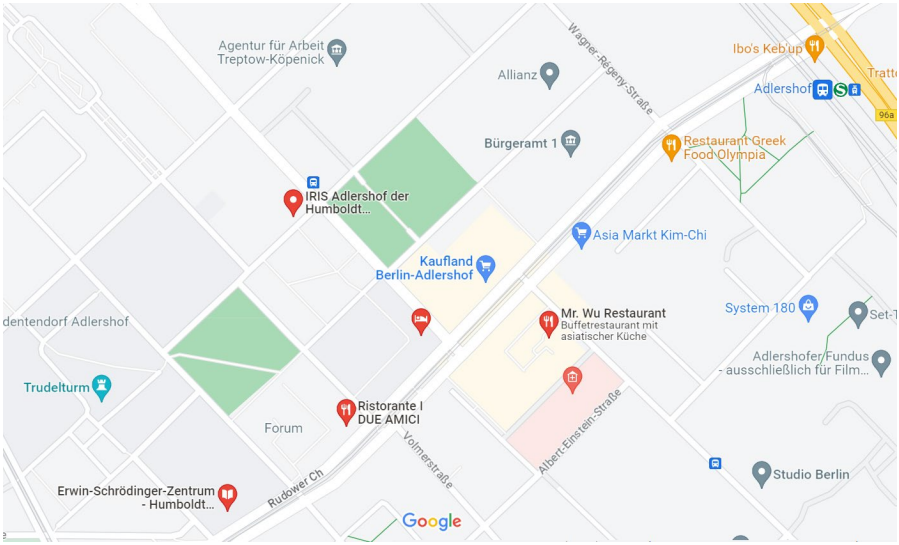
Towards Standard Data Structures and Tools for FAIR Synthesis Data

Sebastian Brückner[1,2], Andrea Albino[1], Hampus Näsström[1], José Márquez[1], Florian Dobener[1], Claudia Draxl[1], Martin Albrecht[1,2]
[1]Department of Physics & IRIS Adlershof, Humboldt-Universität zu Berlin, Germany
[2]Leibniz-Institut für Kristallzüchtung Berlin, Germany

The area A team FAIRmat, focuses on the development of structured data tools and standards and their application to FAIR (Findable, Accessible, Interoperable, and Reusable) synthesis data. We highlight the importance of structured data for semantic interoperability, machine-readability, classification, AI integration, searchability, and shareability, and the potential for data-driven

decisions in solid-state physics research. We introduce our work on FAIRmat, following a bottom-up approach. Here, we discuss the ongoing development of an experimental data model and encourage input from existing ontologies. We showcase the Research Data Management structure that we are implementing in NOMAD, a web-based software for FAIR research data management in materials science, which we are extending for synthesis data through FAIRmat.
We discuss the challenges and goals related to materials synthesis data, including the harmonization of metadata schemas of synthesis processes and experimental characterization. This includes introducing schema and template concepts using the NOMAD platform, emphasizing the importance of base sections for modularity, flexibility, and interoperability. Finally, we present the concept of inheritance and composition in the context of these base sections, demonstrating how these can enhance searchability and processing capabilities in NOMAD.

Venues



IRIS Adlershof

Zum Großen Windkanal 2, 12489 Berlin

Erwin-Schrödinger-Zentrum (ESZ)

Rudower Ch. 26, 12489 Berlin

Contact

Please send all enquiries to fairmat@physik.hu-berlin.de.