

Working Group on Physics, Modern IT and Artificial Intelligence Arbeitskreis Physik, moderne Informationstechnologie und Künstliche Intelligenz (AKPIK)

Tim Ruhe
TU Dortmund
Otto Hahn-Straße 4a
44227 Dortmund
tim.ruhe@tu-dortmund.de

Overview of Posters and Sessions (Poster P)

Sessions

AKPIK 1.1–1.4	Tue	11:15–12:45	H1	RDM I: NFDI consortia (joint session AGI/AKPIK)
AKPIK 2.1–2.5	Tue	13:30–16:30	H1	RDM II: Perspectives in Research Data Management (joint session AGI/AKPIK)
AKPIK 3.1–3.7	Thu	13:30–15:30	P	AKPIK Postersession

AKPIK 1: RDM I: NFDI consortia (joint session AGI/AKPIK)

Time: Tuesday 11:15–12:45

Location: H1

Invited Talk AKPIK 1.1 Tue 11:15 H1
Challenges in data preservation in high energy physics —
 ●ULRICH SCHWICKERATH — CERN, CH-1211 Genf 23

We preserve our data to extend the scientific reach of our experiments. In high energy physics it is cost-efficient to warehouse data from completed experiments on the tape archives of our national and international laboratories. To use data archived in such a way we must also preserve our ability of use the data, specifically the documentation, computing environment and software of the experiments and analyses. Successful data preservation thus requires careful planning and ongoing effort. The contribution illustrates the challenges of long-term data preservation with experience especially from LEP, and will give a brief overview over the ongoing efforts in the LHC experiments at CERN.

Invited Talk AKPIK 1.2 Tue 11:45 H1
The PUNCH4NFDI Consortium in the NFDI — ●THOMAS SCHÖRNER — Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany

With the "Nationale Forschungsdateninfrastruktur" (NFDI, national research data infrastructure), a massive effort is undertaken in Germany to provide a coherent research data management, to make research data sustainably utilisable and to implement the FAIR data principles.

PUNCH4NFDI is the consortium of particle, astro- and astroparticle, as well as hadron and nuclear physics within the NFDI. It aims for a FAIR future of the data management of its community and at harnessing its massive experience not least in "big data" and "open data" for the benefit of "PUNCH" sciences (Particles, Universe, NuClei and Hadrons) as well as for physics in general and the entire NFDI.

In this presentation, we will introduce the work programme of PUNCH4NFDI, its connection to everyday work in the physical sciences and beyond, and in particular the idea of digital research products and the PUNCH science data platform.

Invited Talk AKPIK 1.3 Tue 12:05 H1
DAPHNE4NFDI - Daten aus Photonen und Neutronenexperimenten — ANTON BARTY¹, BRIDGET MURPHY², ASTRID

SCHNEIDEWIND³, WIEBE LOHSTROH⁴ und ●CHRISTIAN GUTT⁵ —
¹DESY, Hamburg — ²CAU Kiel — ³FZ Jülich — ⁴TU München —
⁵Universität Siegen

Die Methoden der Synchrotron- und Neutronenstreuung werden in einer großen, interdisziplinären Bandbreite von Wissenschaftsfeldern angewendet. Die Nutzer repräsentieren dabei verschiedene Fachbereiche in den Naturwissenschaften, die sich dem gemeinsamen Bedarf an anspruchsvoller, schneller und tiefer Datenanalyse sowie den Herausforderungen der Implementierung eines qualifizierten Forschungsdatenmanagements gegenübersehen.

Ziel von DAPHNE4NFDI ist es, eine übergreifende Infrastruktur zu schaffen, welche die Forschungsdaten entsprechend den FAIR-Prinzipien verarbeitet. DAPHNE4NFDI bringt dazu Großforschungseinrichtungen und Nutzer/innen aus den wichtigsten Anwendungsbereichen zusammen, um das Datenmanagement im Sinne der FAIR-Kriterien voranzutreiben.

Invited Talk AKPIK 1.4 Tue 12:25 H1
FAIRmat – Making Materials Data Findable and AI Ready — CLAUDIA DRAXL¹ and ●FAIRMAT TEAM² — ¹Institut für Physik, Humboldt-Universität zu Berlin — ²<https://www.fair-di.eu/fairmat/fairmatteam>

The enormous amounts of research data produced every day in the field of condensed matter physics and the chemical physics of solids represent a gold mine of the 21st century. This gold mine is, however, of little value, if these data are not comprehensively characterized and made available. How can we refine this feedstock, i.e., turn data into knowledge and value? Here, a FAIR (Findable, Accessible, Interoperable, and Re-usable) data infrastructure plays a decisive role. Only then, data can be readily shared and explored by data analytics and artificial-intelligence (AI) methods. Making data Findable and AI Ready (a forward-looking interpretation of the acronym) will change the way how science is done today.

In this talk, we discuss how the NFDI consortium FAIRmat (<https://fair-di.eu/fairmat>) is approaching these goals, and how researchers can profit from our first steps already now.

AKPIK 2: RDM II: Perspectives in Research Data Management (joint session AGI/AKPIK)

Time: Tuesday 13:30–16:30

Location: H1

Invited Talk AKPIK 2.1 Tue 13:30 H1
NFDI4Phys research data management for the next decades — ●HANS-GÜNTHER DÖBEREINER — Institut für Biophysik, Universität Bremen

NFDI4Phys.de is applying to become part of the NFDI.de process. We are working towards digital transformation of academia from the viewpoint of physics with an emphasis on disciplinary and transdisciplinary research. For a list of our domains, see <https://nfdi4phys.de/domains/>. Data need to become FAIR (Findable, Accessible, Interoperable, Reusable). We promote FAIR Digital Objects (FDOs). These are, e.g., digital twins of objects in real life with a unique identifier. For a detailed list of our task areas and strategy, see <https://nfdi4phys.de/task-areas/>. We strive to categorise data according to their structure based on the hierarchical emerge of levels in nature, see <https://nfdi4phys.de/emergence/>. This implies building a bridge between natural and social sciences, engineering, and the humanities. Key to these efforts is to develop physics of complex systems further. We need to structure the qualitative by quantifying it with semantic metrics. Generally, the techniques to do so are mainly available, scattered across various disciplines, waiting to be picked up. We need to overcome the complexity barrier in our minds to make progress. Quantum supremacy provided by quantum computing will eventually provide another level of computing power. Finally, faced with tremendous opportunities, we urgently need to develop ethics of information in order to guide us in judging the impact of technology on our society.

Invited Talk AKPIK 2.2 Tue 14:00 H1

Semantic Research Data Management in the National Research Data Initiative (NFDI) — ●SÖREN AUER — TIB, Welfengarten 1b, 30167 Hannover, Germany

In this talk, we will give an overview of the concepts and implementation of semantic Research Data Management for the National Research Data Initiative (NFDI). We will introduce vocabularies and ontologies for establishing a common understanding of research data and showcase their use in the context of the NFDI initiatives NFDI4Ing, NFDI4Chem, and NFDI4DataScience. We give an overview of three open technology components, ready to be used: - Terminology service for the collaborative creation of terminologies, vocabularies, and ontologies: <https://service.tib.eu/ts4tib/index> - Open Research Knowledge Graph (ORKG) for organizing scientific contributions in a knowledge graph: <https://www.orkg.org> - Leibniz Data Manager as a meta-data repository for research data: <https://labs.tib.eu/info/projekt/leibniz-data-manager/>

Invited Talk AKPIK 2.3 Tue 14:30 H1
NFDI, EOSC, Gaia-X: Three Data Clouds - One Goal? — ●KLAUS TOCHTERMANN — Düsternbrooker Weg 120, 24105 Kiel

Currently, there are three data clouds having an influence on how data management will be shaped in Europe in the future: NFDI, Gaia-X and EOSC. This talk will focus on the synergies of and difference between these three major projects.

The talk will explore how the alignment of these Open Science infrastructures will significantly shape (open) science system of the future. It will shed light on the following questions: Which cooperations are necessary to successfully organise the exchange of scientific resources?

How do they complement each other to advance the implementation of the FAIR principles as a whole? Which aspects are crucial for future engagement?

break

Invited Talk AKPIK 2.4 Tue 15:30 H1
Research Data Management and Higher Education in Physics — ●JANICE BODE¹ and PHILIPP JAEGER² — ¹Westfälische Wilhelms-Universität Münster, Germany — ²University of Manitoba, Canada and Bergische Universität Wuppertal, Germany

This year, funding has been awarded to three NFDI consortia from the physics community. DPG is involved in all of them, aiming to provide a platform for collaboration and exchange of ideas. As its students organization, jDPG also got involved as well as representatives of the Federal Conference of Physics Student Councils (ZaPF - Zusammenkunft aller deutschsprachigen Physikfachschaften).

If research data management (RDM) is to be promoted throughout the physics community, there is no way around reaching out to physi-

cists in their early stages of education. This should be achieved in close alignment between the consortia and the stakeholders in higher education in physics to avoid redundant work and incompatible developments. To this end, we discuss possible routes towards implementing RDM in the physics curricula at low expenses in terms of both effort and funds.

Topical Talk AKPIK 2.5 Tue 16:00 H1
Discussion — ●PHILIPP JÄGER¹, UWE KAHLERT², and TIM RUHE³ — ¹University of Manitoba, Canada and Bergische Universität Wuppertal, Germany — ²RWTH Aachen University — ³Technische Universität Dortmund

In order to provide data obtained in physics experiments to the respective communities and to society as a whole, these data need not only to be stored in a sustainable way, but also be prepared and maintained. Such an effective research data management (RDM) does not only require unified solutions with respect to the anticipated resources, but also common standards across various scientific disciplines. We will discuss current and future initiative of RDM with the speakers of the session.

AKPIK 3: AKPIK Postersession

Time: Thursday 13:30–15:30

Location: P

AKPIK 3.1 Thu 13:30 P
Towards optical neuromorphic hardware at Cesium wavelength — ●ELIZABETH ROBERTSON^{1,2}, MINGWEI YANG¹, LUISA ESGUERRA¹, LEON MESSNER¹, LINA JAURIGUE², GUILLERMO GALLEGO^{2,3}, KATHY LÜDGE², and JANIK WOLTERS^{1,2,3} — ¹Deutsches Zentrum für Luft- und Raumfahrt, Institute for Optical Sensor Systems, Rutherfordstraße 2, 12489 Berlin, Germany — ²Technische Universität Berlin, Str. des 17. Junis 135, 10623 Berlin, Germany — ³Einstein Center Digital Future Robert-Koch-Forum, Wilhelmstraße 67, 10117 Berlin, Germany

With the exponential growth of research in Machine Learning, so too has the demand for fast, energy efficient neuromorphic hardware, to execute such algorithms. Optics provides an attractive tool for implementing neuromorphic hardware due to its speed, low crosstalk and high parallelism. In particular, convolutional- and recurrent neural networks benefit from an optics implementation; by using passive optics to carry out convolution [1], and time division multiplexing to demonstrate a reservoir computer [2]. We present our results towards realizing an optical convolutional neural network with an atomic non-linearity [3] and an all-optical reservoir computer based on a delay loop with an additional optical random-access memory (RAM)[4].

- [1] Miscuglio, M. et al. *Optica* 7, 1812-1819 (2020).
- [2] Brunner, D., et al. *Nat Commun* 4, 1364 (2013).
- [3] Yang, M. et al. CLEO/Europe-EQEC, poster JSIV-P.4 (2021).
- [4] Wolters, J., et al. *Phys. Rev. Lett.* 119, 060502 (2017).

AKPIK 3.2 Thu 13:30 P
Optical Convolutional Neural Network with Atomic Non linearity — ●MINGWEI YANG^{1,2}, ELIZABETH ROBERTSON^{2,3}, LUISA ESGUERRA^{2,3}, and JANIK WOLTERS^{2,3} — ¹Humboldt Universität zu Berlin, Newtonstr.15, D 12489 Berlin, Germany — ²Deutsches Zentrum für Luft und Raumfahrt e.V. (DLR), Rutherfordstraße 2, D 12489, Berlin, Germany — ³Technische Universität Berlin, Straße des 17. Juni 135, D 10623, Berlin, Germany

An optical convolutional neural network is demonstrated in which linear operations are implemented by lenses and spatial light modulators (SLMs), while an optical nonlinearity is realized by a cesium vapor cell as a saturable absorber. We use this network to demonstrate nonlinear image processing, as well as to improve the classification of the MNIST data by a single layer fully connected.

AKPIK 3.3 Thu 13:30 P
Convolutional Neural Network Framework for the Analysis of X-ray photoelectron spectra — ●LUKAS PIELSTICKER¹, RACHEL L. NICHOLLS¹, GUDRUN KLIHM¹, ROBERT SCHLÖGL^{1,2}, and MARK GREINER¹ — ¹Department Heterogeneous Reactions, Max Planck Institute for Chemical Energy Conversion, Mülheim an der Ruhr — ²Department Inorganic Chemistry, Fritz Haber Institute of the Max Planck Society, Berlin

X-ray photoelectron spectroscopy (XPS) enables studying the electronic structure and chemical state of solid materials and their surfaces. Quantitative analysis of the phases present in XP spectra is typically performed by manual peak fitting. However, such elemental quantification often suffers from, among other things, superposition of core-levels of different elements, incorrect instrument calibration, poor choice of backgrounds and lineshapes, as well as from noise in the data. Moreover, as XPS instruments are becoming increasingly automated and capable of producing large amounts of data, an equally automated approach to elemental quantification is desirable.

Here, a scalable automation framework for XPS analysis using Convolutional Neural Networks (CNNs) is presented. For model training, synthetic mixed metal-oxide spectra were generated based on known reference spectra. CNNs are shown to be capable of quantitatively determining the presence of metallic and oxide phases, as well as identifying morphological features such as over- and sublayers, exhibiting more reliable performance than standard XPS users. The use of Bayesian CNNs for the determination of quantification uncertainty is illustrated.

AKPIK 3.4 Thu 13:30 P
Conversion of Molecular Dynamics (MD) simulations to Neutron and X-Ray Scattering Data using High Performance Computing — ●ARNAB MAJUMDAR¹, SEBASTIAN BUSCH¹, and MARTIN MÜLLER² — ¹Lichtenberg Straße 1, 85747, Garching b. München — ²Leibnitz Straße 19, 24098, Kiel

Sassena is one of the software solutions to convert molecular dynamics (MD) simulations into elastic and quasi-/inelastic neutron and X-ray scattering curves. Current work makes an effort to introduce different strategies of parallel computing into sassena. Parallel computing can be a huge leap in the journey of reducing the computing time within sassena. It consists of different strategies like distributed memory parallelization (MPI), shared memory parallelization (OpenMP) and vectorization. Sassena inherits distributed memory parallelization from its previous version. This work further augments vectorization and shared memory parallelization into it. Through vectorization, this work bolsters the computing speed of sassena to a new height of up to an order of magnitude faster than its previous version. On the other hand, shared memory parallelization introduces a possibility of doing hybrid parallelization within sassena. Furthermore, this work plans to benefit from the achieved performance gain by validating simulations of hydrogen storage materials with neutron scattering data.

AKPIK 3.5 Thu 13:30 P
Making computation material science data FAIR — ●JENS BRÖDER^{1,2}, VOLKER HOFMANN^{1,2}, DANIEL WORTMANN³, STEFAN BLÜGEL³, and STEFAN SANDFELD^{1,2} — ¹Institute for Advanced Simulation, Forschungszentrum Jülich, D-52425 Jülich, Germany — ²Helmholtz Metadata Collaboration, Hub Information — ³Peter Grünberg Institut and Institute for Advanced Simulation, Forschungszentrum Jülich, D-52425 Jülich, Germany

For research data to be reusable by scientists or machines, the gathered research data and metadata should comply with the so-called "FAIR principles", i.e. it should be findable, accessible, interoperable, and reusable [1], a task which is not straightforward. In computational materials science, workflows often encompass many different simulation steps. The enrichment of data with detailed metadata is often only feasible close to the data creation process. Therefore, designated software, workflows, tools and standards will be needed throughout the community. Using an exemplary research project, we show in detail how to reconcile data from simulations with FAIR principles. The project contains data from a high-throughput simulation performed with the program FLEUR (www.flapw.de) using the AiiDA framework (<https://aiida.net>) on over 5000 different materials. All data and software is openly available and FAIR through the materialscloud archive [2]. We also discuss challenges for the domain of material science and how the Helmholtz Metadata Collaboration (HMC) tries to address these issues. [1] Wilkinson, M.D. et al. *Sci Data* 3, 160018 (2016) [2] L. Talirz et al., *Sci Data* 7, 299 (2020)

AKPIK 3.6 Thu 13:30 P

Design and validation of a Digital Twin for prostate cancer from a physics point of view — ●CARLOS ANDRES BRANDL¹, ANNA NITSCHKE¹, and MATTHIAS WEIDEMÜLLER^{1,2} — ¹Physikalisches Institut, Ruprecht-Karls Universität Heidelberg, Im Neuenheimer Feld 226, 69120 Heidelberg, Germany — ²National Laboratory for Physical Sciences at Microscale and Department of Modern Physics, and CAS Center for Excellence and Synergetic Innovation Center in Quantum Information and Quantum Physics, Shanghai Branch, University of Science and Technology of China, Shanghai 201315, China

Digital Twins (DT) are virtual representations of physical assets and e.g. promise improved decision making. DT can help to personalize healthcare for complex diseases like prostate cancer by combining large amount of clinical parameters and answering the questions of tumor risk, tumor stage and optimal treatment. In medical applications interpretability and uncertainty quantification are crucial.

Combining data-driven approaches with interpretable machine-

learning models and evidence based clinical guidelines will lead to more reliable and confidential outcomes. Correlation analysis gives an insight to the data and enables to unravel the dependencies of important clinical parameters like prostate specific antigen (PSA) with others and determine their distributions in the Heidelberg patient cohort. The backpropagation of outcome errors to the found distributions of the input parameters with Bayesian methods allows to determine the impact of the input parameter uncertainties on the twin predictions and helps clinicians to interpret the results accordingly.

AKPIK 3.7 Thu 13:30 P

Comparison of structural representations for machine learning-accelerated *ab initio* calculations — ●JOHANNES WASMER, PHILIPP RÜSSMANN, and STEFAN BLÜGEL — Forschungszentrum Jülich, Germany

Quantum mechanical calculations based on density functional theory (DFT) are the workhorse in today's computational materials design. Here we explore the possibility to accelerate the DFT calculations with potentials generated from a surrogate machine learning model. Finding a better starting potential could drastically reduce the number of required self-consistency steps during the convergence of DFT calculations. The juKKR code (jukkr.fz-juelich.de) allows high-throughput *ab initio* impurity embedding calculations which we use to generate a training dataset of 10'000 impurities from most elements of the periodic table embedded into elemental crystals with the help of the workflow engine AiiDA. The choice of a structural representation of the atomic environment which a machine learning model can understand has been identified as a crucial step. We compare a variety of such representations as training input for our surrogate model. Finally, we benchmark results for the converged impurity potential from DFT calculations against the output of the trained surrogate model.

We acknowledge support by the Joint Lab Virtual Materials Design, by the DFG under Germany's Excellence Strategy Cluster of Excellence ML4Q, by the AIDAS2 virtual lab, and thank for computing time provided on the JARA Partition part of the supercomputer CLAX at RWTH Aachen University.