Nationale Forschungsdateninfrastruktur

für

disziplinäre und transdisziplinäre Physik

Letter of Intent

August 15, 2020

1 Binding letter of intent as advance notification or non-binding letter of intent

	Binding letter of intent (required as advance notification for proposals in 2020)
x	Non-binding letter of intent (anticipated submission in 2021)

2 Formal details

- Planned name of the consortium
  Nationale Forschungsdateninfrastruktur für disziplinäre und transdisziplinäre Physik
- Acronym of the planned consortium
  NFDI4Phys

Applicant institution

Universität Bremen, Bibliothekstraße 1, 28359 Bremen
Prof. Dr. Bernd Scholz-Reiter

Spokesperson

Prof. Dr. Hans-Günther Döbereiner, hgd@uni-bremen.de, Institut für Biophysik

Participants

Prof. Dr. Uwe Engel, uengel@uni-bremen.de, Institut für Soziologie
Prof. Dr. Rolf Drechsler, drechsler@uni-bremen.de, Data Science Center
Dr. Lena Steinmann, lena.steinmann@uni-bremen.de, Data Science Center
Prof. Dr. Frank Kirchner, frank.kirchner@dfki.de, FB 03, Informatik

Co-applicant institution

Physikalisch-Technische Bundesanstalt (PTB)
Bundesallee 100, 38116 Braunschweig
President: Prof. Dr. Joachim Ullrich

Co-spokesperson

Dr. Holger Israel (TBC), holger.israel@ptb.de, PTB

Co-applicant institution

Technische Informationsbibliothek (TIB), Welfengarten 1B, 30167 Hannover
Director: Prof. Dr. Sören Auer

Co-spokesperson

Dr. Esther Tobschall, esther.tobschall@tib.eu, TIB

Participants

Dr. Irina Sens, irina.sens@tib.eu, TIB
Dr. Angelina Kraft, angelina.kraft@tib.eu, TIB

Co-applicant institution

Leibniz Institute for Plasma Science and Technology (INP)
Felix-Hausdorff-Straße 2, 17489 Greifswald
Director: Prof. Dr. Klaus-Dieter Weltmann

Co-spokesperson
Prof. Dr. Dirk Uhrlandt, uhrlandt@inp-greifswald.de, INP
Participant
Dr. Markus Becker, markus.becker@inp-greifswald.de, INP

Co-applicant institution
Karlsruhe Institute of Technology (KIT), Kaiserstraße 12, 76131 Karlsruhe
President: Prof. Dr.-Ing. Holger Hanselka
Co-spokesperson
Frank Tristram, frank.tristram@kit.edu, Cluster of Excellence 3DMM2O

Co-applicant institution
FIZ Karlsruhe - Leibniz Institut für Informationsinfrastruktur GmbH (FIZ)
Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen
CEO: Sabine Brünger-Weilandt
Co-spokesperson
Prof. Dr. Harald Sack, harald.sack@fiz-karlsruhe.de, FIZ

Co-applicant institution
Gottfried Wilhelm Leibniz Universität Hannover
Welfengarten 1, 20167 Hannover
President: Prof. Dr. Volker Epping
Co-spokesperson:
Prof. Dr. Holger Frahm, frahm@itp.uni-hannover.de, Institut für Theoretische Physik

Co-applicant institution
Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, 52428 Jülich,
Chairman of the Board: Wolfgang Marquardt
Co-spokesperson
Dr. Stefan Krieg, s.krieg@fz-juelich.de, Jülich Supercomputing Centre

Participant institution
Deutsche Physikalische Gesellschaft (DPG), Hauptstrasse 5, 53604 Bad Honnef

Besides the DPG, there is a large number of other participant institutions and individuals which will be added to the consortium over the next 12 months. We are in the process of building and extending our user base.

3 Objectives, work program and research environment

- Research area of the proposed consortium

Physics is structured in several disciplines driven by communities relying on shared large facilities and research collaborations and by communities organized in independent research groups and laboratory infrastructure. These independent groups are the community for NFDI4Phys, and most of them are not covered by any other consortium. Naturally, these groups need substantial activation energy to join the NFDI, but they individually maintain and improve open solutions as a community for the NFDI. The majority of these groups work in subfields of physics, like atomic and molecular physics, optics, (cold) plasma physics, and soft matter physics. These fields found their home in our consortium and form our disciplinary user group. Further, starting in the middle of last century, but especially during the last few decades, a quite remarkable process took place. Slowly, but steadily, physics research extended into a number of other disciplines, e.g., biology, neurobiology, medicine, sociology, and economics, forming the corresponding subfields, biological physics, neurophysics, medical physics, sociophysics, and econophysics along the way. We may call all these subfields transdisciplinary physics, which build our transdisciplinary user group, where biological physics spearheaded the interdisciplinary process and in fact together with soft matter physics is now considered disciplinary in DFG classification.

Due to the wide range of disciplines addressed by conceptually physical methods, including social sciences and the humanities, we form a NFDI *en miniature*. This provides us with the unique opportunity to attempt an integration of research data management (RDM) across disciplines right from the start. Research conducted by user groups of NFDI4Phys spans across a wide range of the DFG classification scheme. We give a selected choice of main fields below. Of those, plasma physics and biological physics share a common physical concept: Both are concerned with reactive processes out of thermal equilibrium. Likewise, quantum optics and biological physics both acknowledge the importance of information content and processing. Finally, all disciplinary and transdisciplinary fields are joined by a hierarchical sequence of emergent properties and behavior typical for complex systems. The physics of complex systems searches for universality within and in between hierarchical levels. The NFDI process will bring together disciplines and foster new unexpected collaborations.

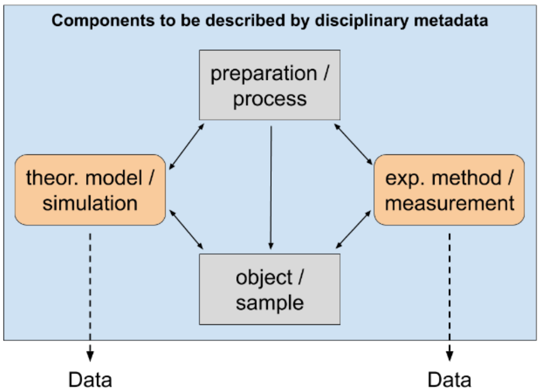
Disciplinary physics fields
307: Physik der kondensierten Materie
308: Optik, Quantenoptik, Physik der Atome, Moleküle und Plasmen
310: Statistische Physik, Weiche Materie, Biologische Physik, Nichtlineare Dynamik

Transdisciplinary fields
108: Philosophie
110: Psychologie
111: Sozialwissenschaften
112: Wirtschaftswissenschaften
201: Grundlagen der Biologie und Medizin

- Concise summary of the planned consortium's main objectives and task areas

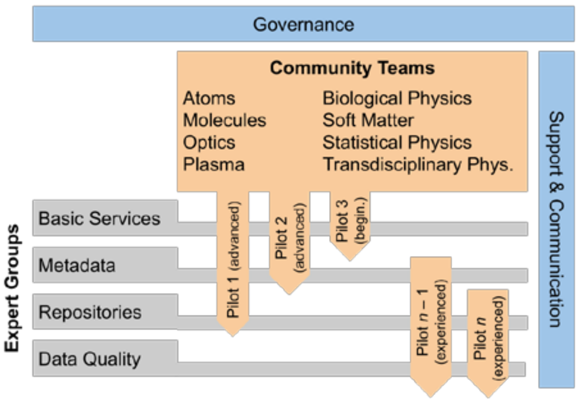
The sub-areas of physics addressed in NFDI4Phys have in common that for their studies, be it theoretical or experimental, samples or simulation series are generated or prepared with a large set of parameters. The processes of generating or preparing samples play a major role, yet in this consortium, self-developed devices, models and methods are used for both, the set-up of the studies as well as for the analysis of the results. This commonality and challenge form the fundamental basis for the FAIRification (to make research data Findable, Accessible, Interoperable, Reusable) of research data in NFDI4Phys. The generalized schema shown in Figure 1 illustrates that objects/ samples, processes for their generation or preparation, as well as the experimental methods or theoretical models used in scientific studies are considered equally important when documenting data by means of disciplinary metadata. Each of these components might be the focus of a scientific study. For example, a biophysical study may focus on the analysis of a cell structure (i.e. of a sample), a medical study monitors the tumor of a patient (i.e. an object), or the development of a novel method to investigate a small interaction force (theoretical or experimental method). Only if all relevant components of a study are sufficiently described, the corresponding data can really be FAIR. The modular approach used in NFDI4Phys particularly supports the inter- and transdisciplinary interoperability and reusability of the data by linking the metadata modules, vocabularies and ontologies to be developed within the framework of NFDI4Phys with systems already available or being developed, e.g., within other NFDI consortia.

Figure 1: Common structure of metadata in NFDI4Phys



The starting point for the development of services and tools for research data management are user groups (pilots), which initially come primarily from the fields of atomic and molecular physics, (quantum) optics and photonics, and plasma physics, dynamics and statistical physics, soft matter, biological physics, and other transdisciplinary areas of physics. The basic procedure in the project will be to guide the user groups systematically through a process of building research data management structures. This process will be adapted to beginning pilots on the one hand and advanced pilots on the other. At the same time, a continuous access for new, not yet specified user groups will be enabled. As shown in Figure 2, this approach determines the structure of the project in the form of a matrix. Both the pilots and the expert

Figure 2: Structure of the NFDI4Phys Consortium



groups have their own working tasks. Joint work packages are formed at the interface points of the matrix to define the individual user journey. To give an example: The expert group ‘metadata’ will develop specific metadata modules together with the pilot user ‘quantum optics’. Similarly, individual metadata modules for other disciplines of physics are developed by the same expert group together with the respective pilot users. This includes the deduction of common schemes and interfaces between the modules, e.g., network models describing the topology and dynamics are ubiquitous from socioeconomic to metabolic networks. Therefore structured data and the integration of network data and metadata formats are important in bridging the disciplines.

Community Teams are the initial point of contact for new use cases. To address different fields, these teams must understand the specific ways of dealing with the particular objects of research and, in the case of transdisciplinary areas of physics, to speak the language of the other discipline. The community teams articulate the needs of a given field and support onboarding by recommending a sequence of tasks for the building of FAIR processes based on an assessment of the users' maturity in RDM.

Several expert groups are set up to help the pilots through the individual tasks:
Basic Services are directed at beginning pilots: They include a first characterization of use cases, measures for raising data literacy and provision of RDM tools suiting the users’ needs.

Metadata standards and ontology development are the key for almost all tasks and users within NFDI4Phys, as they are essential to drive extensive digitization of research processes and documentation in physics by ontology development and harmonizing metadata. We will focus on flexible modular solutions for specific use cases and an ontology service for individual and dynamically changing usage. In particular, many use cases related to physical methods can be described in terms of structural or numerical factual (network) data (triplets of numerical value, unit, uncertainty range) which we plan to underpin with a common, harmonized ontology and metadata system.

Repositories play a critical role in linking data and software repositories for physics accessible from a single portal, addressing the current gap of dedicated infrastructure, long-term storage and the construction of a knowledge graph for physics.

Data Quality, located at the most advanced stage of the use case flow, seeks to establish quality criteria for the data and metadata relevant to our communities.

Assistance with generic tasks within NFDI4Phys and communicating with external partners:
Support & Communication coordinates our collaborative efforts with other NFDI consortia to identify and address cross-cutting topics. With DPG and other learned and professional societies as well as local stakeholders at universities and research institutions we attract new users. With journals, funding agencies and industry we are going to promote FAIR principles for data publication and the use of open standards. As a service to the expert groups we calibrate questionnaires, e.g., for metadata, conduct surveys and evaluate user satisfaction.

Governance will monitor the development of NFDI4Phys and the progress of work in the different task areas. It ensures efficient financial and organizational management.

- <u>Brief description of the proposed use of existing infrastructures, tools and services that are essential in order to fulfill the planned consortium's objectives.</u>

Harmonizing metadata systems for *numerical factual data* is crucial for digitizing and improving the workflows for many researchers in physics and related fields. The Physikalisch-Technische Bundesanstalt (PTB) as Germany's National Metrology Institute will play a prominent role in this process, building on its longstanding experience in harmonizing metrology in the analogue world, and its high degree of European and international connectivity. Within NFDI4Phys, *FAIR data and metadata services* will be established by using and extending state-of-the-art technologies along the research data lifecycle and process workflows of the targeted communities. NFDI4Phys will support the harmonization and development of desired metadata standards including the use of existing and new PID services together with Technische Informationsbibliothek (TIB) and DataCite. *PID services* will include the DOI service, the connection of metadata from various PIDs (e.g. ORCID for researchers, ROR for institutions and DOI for data) using, e.g., the *PID Graph* infrastructure, and the provision of PIDs for instruments within a device database. NFDI4Phys will also significantly increase the visibility and access of relevant NFDI4Phys research data and research software in existing repositories, using, e.g, the Leibniz Data Manager (LDM) software (based on CKAN, provided by TIB). NFDI4Phys also recognizes the importance of ontology development and the use of semantic technologies to maximize the findability and machine interoperability of produced research data. This will be addressed by establishing a *Knowledge Graph infrastructure* (FIZ Karlsruhe) and by the deployment of a *Terminology Service* (TIB). This will enhance the built up of user initiated mutually linked repositories as spear-headed by FIZ.

The new Data Science Center (DSC) at University Bremen fosters collaborations between specialists from computer, natural, life, economic, social, engineering sciences and mathematics, thereby pushing the consortium's transdisciplinary agenda. The DSC in Bremen provides a high-performance IT-infrastructure to employ complex, data-intensive workloads. The DSC will expand its infrastructure to also include modules for hosting repositories and archiving. By introducing machine learning into user processes, we provide novel possibilities to analyze data and extract information, thereby increasing data usability and adding extra value.

Jülich Supercomputing Centre (JSC) operates supercomputers of the highest performance class in Europe, enabling scientists and engineers to explore some of the most complex grand challenges facing society. JSC research is performed through collaborative infrastructures exploiting extreme-scale supercomputing and federated data services.

As representatives of many others we would like to mention here that Leibniz Universität Hannover has established a research data support team with TIB and operates a research data repository with REST-API to connect external applications. INP Greifswald develops a plasma knowledge graph with FIZ and operates INPTDAT, a data platform for plasma technology. Our users are well connected internationally and have established national centers and international initiatives or are members thereof. We note the Bremen center for decision research. Our consortium benefits from the Cluster of Excellence *3DMM2O* at KIT and University of Heidelberg as well as CRC 1032 in Munich. Our Members organize international conference series, e.g., *Physics of Cancer*, CellPhysics, and *PhysNet*.

- Interfaces to other proposed NFDI consortia: brief description of existing agreements for collaboration and/or plans for future collaboration

In the NFDI, physics is covered by four candidate consortia: DAPHNE, FAIRmat, PUNCH4NFDI, and NFDI4Phys. Coordinated by the Deutsche Physikalische Gesellschaft (DPG), the consortia have agreed on common definitions of their respective communities they intend to represent. The profile of NFDI4Phys states:

“Our use cases are drawn especially from the fields of optics, quantum optics and photonics, atomic, molecular and plasma physics, biological physics, dynamics and statistical physics, soft matter, and the physics of socio-economic systems. NFDI4Phys addresses the highly dynamic turn towards new areas of research and technologies and the corresponding shift of classic disciplinary boundaries by adopting a transdisciplinary approach with the goals of digitizing workflows and harmonizing RDM through quality standards.”

On this basis, we have defined our interfaces with DAPHNE, FAIRmat, and PUNCH4NFDI, and are planning for work packages for issues relevant to all four physics consortia. Shared physics cross-cutting topics will be defined by the shared physical methods and their resulting data types and structures.

Inherent to the structure of our consortium is an open attitude towards all disciplines. PUNCH4NFDI and NFDI4Phys share a number of participants, and there is going to be an exchange led by FAU Erlangen-Nuremberg in NFDI4Phys, dedicated to the interoperability of the NFDI4Phys-device ontologies with the “big data” datasets commonly found in astronomy and particle physics. Similar test cases linking our metadata systems to photon and neutron beamline experiments are going to be suggested to DAPHNE, e.g., based on the experiments of PTB researchers.

An inter-consortium working group with, e.g., FAIRmat, DAPHNE, NFDI4Chem, and MatWerk should seek common RDM problems and exploit synergies in their solution. Likewise, NFDI4Phys, NFDI4Ing, and NFDI4BIMP are all concerned with applicative processes, and we advocate a harmonized approach to their RDM in a suitable collaboration format. NFDI4Chem and LifesciencesUmbrella will provide alignment with other initiatives in the natural sciences. In order to foster our trans-disciplinary approach, we seek collaboration agreements with consortia from the Social Science and the Humanities.

Finally, there are several consortia focusing on methodologies that are crucial for sciences driven by physical methods. We are going to approach MaRDI, NFDI4DS, NFDI4RSE, and NFDI4HPC in order to bring together their expertise in mathematical, computational, and data science methods with our use cases from disciplinary and transdisciplinary physics.

4 Cross-cutting topics

NFDI4Phys has signed the “Leipzig Berlin Declaration on NFDI cross-cutting topics” and will collaborate with other NFDI consortia on these topics.

Ontology Development: Outside of big data communities, poor or missing (meta)data standards and the lack of harmonized workflows are major obstacles and a special challenge in physics as well as for transdisciplinary researchers using physical methods. A major problem in transdisciplinary research are central key expressions shared by different fields with differing connotations and semantic content. Since our vision is to eventually arrive at universally accepted metadata systems for physics-related data, while ensuring interoperability across disciplines, we will coordinate with NFDI consortia from physics and beyond.

RDM Education: We consider it crucial to expose students to RDM best practices early in their education. Physics departments are particularly well positioned for this task since they offer physics lab courses not only for students of physics but also for those studying chemistry, biology, and engineering. NFDI4Phys plans to design a concept for RDM in introductory physics labs, in collaboration with lecturers responsible for the labs as well as student representatives.

Data Storage & Federated Repositories form the backbone of the NFDI. Connecting existing infrastructures and thus forming a federated system of repositories, research data should be stored as locally as possible and as centrally as necessary. Hence, local computing centers and university libraries will handle datasets from all NFDI consortia.

Software Development & Management will be at the heart of implementing the NFDI. Consortia must not see themselves competing over limited human resources but follow the calls for building an integrated, federated infrastructure by cooperating and exploiting synergies.

Data Quality & Review for discipline-specific data are a notable aspect of NFDI4Phys that we offer to other physics consortia and the wider NFDI.

Authorization & Authentication Infrastructure should be consistent throughout the NFDI, requiring a high level of expertise regarding data protection and information security.

Self-Assessment of the NFDI as an institution aiming at FAIR access to data is warranted. It is important to investigate the sustainability of solutions found to reach such a goal. Further, we need to ask what impact FAIR access to data will have on to society. Could public FAIRness be unfair to single individuals? Should RDM really do all that what technology allows? Accordingly, we propose inter-consortium working groups for **technology assessment** and process **philosophy of NFDI**. The former should reflect causal influences of the NFDI process of amassing large amounts of data for German society. The latter should ask the question how the digitization process affects German sciences and NFDI in particular.