NFDI4Patho

National Research Data Infrastructure for Pathology

Binding letter of intent as advance notification of a full proposal

Letter of intent

1. Binding letter of intent as advance notification of a full proposal

X	Binding letter of intent (required as advance notification for proposals in 2021)

2. Formal details

Planned name of the consortium

National Research Data Infrastructure for Pathology

Acronym of the planned consortium

NFDI4Patho

Applicant institution

University Hospital Aachen (UKA) of the RWTH Aachen University

Pauwelsstr. 30, 52074 Aachen, Germany

Univ.-Prof. Dr. rer. nat. Dr hc mult. Ulrich Rüdiger

Spokesperson

Univ.-Prof. Dr. med. Peter Boor, Ph.D., pboor@ukaachen.de, UKA, Institute of Pathology

Co-spokespersons

Univ.-Prof. Dr. med. Rainer Röhrig, rroehrig@ukaachen.de, UKA, Institute for Medical Informatics

Univ.-Prof. Dr. Ing. Ulrike Meyer, meyer@itsec.rwth-aachen.de, RWTH, IT-Security

Co-applicant institution

Charité - University Medicine Berlin,

Charitéplatz 1, 10117 Berlin, Germany

Prof. Dr. Axel Radlach Pries

Co-spokespersons:

Prof. Dr. Peter Hufnagl, peter.hufnagl@charite.de, Institute of Pathology

Norman Zerbe, norman.zerbe@charite.de, Institute of Pathology

Co-applicant institution

Institute for Information Technology (OFFIS)

Escherweg 2, 26121 Oldenburg, Germany

Prof. Dr.-Ing. Andreas Hein

Co-spokesperson

Dr. rer. nat. Marco Eichelberg, marco.eichelberg@offis.de, Automation and Integration Technology in the Healthcare Research and Development Division

Co-applicant institution

Technology, Methods, and Infrastructure for Networked Medical Research (TMF)

Charlottenstraße 42/Ecke Dorotheenstraße, 10117 Berlin, Germany

Sebastian C. Semler

Co-spokesperson

Sebastian C. Semler, Sebastian.Semler@TMF-ev.de, TMF General Manager

Co-applicant institution

Johann Wolfgang Goethe-University Frankfurt

Theodor-W.-Adorno-Platz 1, 60323 Frankfurt am Main, Germany

Univ.-Prof. Dr. Enrico Schleiff

Co-spokesperson

Univ.-Prof. Dr. med. Peter J. Wild, peter.wild@kgu.de, Dr. Senckenberg Institute of Pathology (SIP), University Hospital Frankfurt

Co-applicant institution

Fraunhofer Institute for Digital Medicine MEVIS

Max-von-Laue-Str. 2, 28359 Bremen, Germany

Prof. Dr.-Ing. Horst K. Hahn

Co-spokesperson

Prof. Dr.-Ing. Horst K. Hahn, horst.hahn@mevis.fraunhofer.de, MEVIS Managing Director

Participants

- Christian-Albrechts-University of Kiel, Prof. Dr. med. Christoph Röcken
- Commission Digital Pathology of the Association of German Pathologists, Prof. Dr. med. Gunter Haroske
- Frankfurt Institute for Advanced Studies (FIAS), Prof. Dr. Volker Lindenstruth
- German Society for Neuropathology and Neuroanatomy (DGNN), Prof. Dr. med. Till Acker
- German Society of Pathology e. V., Dipl.-Politologe Joerg Maas
- Göttingen University Medical Center, Prof. Dr. med. Phillip Ströbel & Prof. Dr. med. Otto Rienhoff
- Hannover Medical School, Univ.-Prof. Dr. med. Danny D. Jonigk
- Ludwig-Maximilians-University Munich, Prof. Dr. med. Frederick Klauschen
- University Hospital Carl Gustav Carus Dresden, Prof. Dr. med. Gustavo Baretton
- University Hospital Erlangen, Prof. Dr. Arndt Hartmann & Prof. Dr. Kerstin Amann
- University Medical Center Hamburg-Eppendorf, Prof. Dr. Thorsten Wiech
- University of Giessen, Prof. Dr. med. Till Acker
- University of Regensburg, Prof. Dr. Christoph Brochhausen-Delius

1. Objectives, work programme and research environment

Research area of the proposed consortium: 205 Medicine / 205-06 Pathology

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

Pathology, including neuropathology and all other specific pathology subdisciplines, represents a cross-sectional medical field that focuses on tissue diagnostics. One of the main focuses of pathology is the multiscale morphological tissue analysis. This brings together the results of macroscopy, microscopy and ultrastructural examinations together with analyses at the molecular level and clinical data. As a cross-sectional research area, pathology works with all types of tissues and plays a key position in almost all areas of applied medical research, e.g. oncology, cardiology, neurology, gastroenterology, nephrology, immunology etc. Recent technological developments have enabled effective digitization of pathology, particularly generating novel types of large and unique image data, the so-called whole-slide-images, which exceed other current imaging approaches in terms of the level of detail and volume of generated data. This catalyzes unprecedented research opportunities towards automated, highly accurate and reproducible pathology research. Digitization enables new insights to be gained through automated cell-level data analysis using artificial intelligence techniques, particularly using deep learning approaches. This new area of pathology referred to as digital (computational) pathology, will foreseeably transform and improve pathology research and diagnostics. The main bottleneck in current preclinical and clinical research in pathology is the lack of access to relevant, well-curated, heterogeneous, multicentric, (sufficiently) large datasets according to the FAIR principles (findable, accessible, interoperable, and reusable).

The main objective of NFDI4Patho is to tackle this unmet need by developing a secure, federated national research data infrastructure and ecosystem making pathology data FAIR and available to the entire research community.

The goal is to build and operate a service and ecosystem for FAIR data that the community will want to use, that will facilitate research and development, and, in the broader outlook, accelerate the translation of medical research into patient applications. Particularly, the latter is essential in all medical research, given the immense problem of the "translational valley of death", in which many if not most research endeavors fail to translate for the benefit of the patient. NFDI4Patho provides the right instrument, at the right time, and the right partners for building a unique FAIR-data infrastructure and shaping the digital transformation & translational research for the entire community. NFDI4Patho also provides highly complementary & interdisciplinary expertise for the NFDI and integrates closely with several other projects (see below). While NFDI4Patho does not aim to encompass the whole "medical" research field, with its focused approach it will serve as a use-case addressing all relevant issues in detail and providing a potential blueprint for other use-cases.

The type of data tackled in NFDI4Patho will include all pathology data, with particular focus on the novel image data, but also metadata, preanalytical processing data, and the important routine clinical data (complementary to other NFDIs), and molecular data (in close cooperation with GHGA). NFDI4Patho will not only focus on basic research but also translational research. This will require comprehensive data provenance, including the regulatory basis of data processing, to assess whether and how which data from NFDI4Patho can be used for the development and validation of diagnostics.

To tackle these goals, NFDI4Patho will provide the following main infrastructures & services:

- central point for requests for research data, coordination of provisioning
- middleware for queries on availability of data and samples, data merging and transfer
- infrastructure for performing distributed data analysis (distributed computing)
- supporting the data-providing centers in setting up the required local infrastructure, including processes to be standardized

- information, education and training of the connected data providing and data processing centers and the data using researchers.

We have identified eight task areas that are required for the development, implementation, and operation of NFDI4Patho:

Governance: Will coordinate the overall progress of the project and, in alignment and contribution of all partners in an open and participative governance structure, will develop consensus on rules, processes, sample documents, and standards for operations of NFDI4Patho.This includes rules for data use and access.

IT-Infrastructure: Will design and implement a decentralized IT architecture with a central platform (office) for data requests & transfer using existing IT systems and components. The platform will handle the research requests, coordinate the review by the Data Use & Access Committee (DUAC), and provide the link to the data owners. After consenting by requesting researcher and data owners, the platform will merge and transfer the data. The processes will be based on experiences and best practices from the Medical Informatics Initiative (MII), German Biobank Alliance (GBA), AKTIN emergency department registry, the German Center for Lung Research (DZL), and the German Registry of COVID-19 Autopsies (DeRegCOVID).

Data Stewardship: Will develop concepts for standardized and improved data collection and curation in the data generating centers and establish standard operating procedures (SOPs) for standardization and harmonized data acquisition and preanalytics, focusing on all relevant technical and medical aspects and metadata. The concepts and SOPs will be made available to the whole community, aiming for broad dissemination of standardization and improving data availability and quality while at the same time optimizing the data collection effort. The coordinated processes are core elements in setting up data provenance. We will first focus on selected use-cases, representing the full spectrum of pathology datasets and lighthouse research areas of the involved centers.

Interoperability: Interoperability will build upon the necessary communication standards DICOM and HL7 FHIR. These will be consented in the corresponding standardization organizations and will support the standardization effort in the field of digital pathology. These developments will be done in close coordination with other initiatives, in particular EMPAIA, MII, BIGPICTURE, other NFDI consortia, and the German Centers for Health Research (DZG).

AI & Analytics: Will create tools & infrastructure to apply analytics and AI to pathology data and enable distributed processing of data across medical centers. We will demonstrate the value of NFDI4Patho by implementing high-profile use cases in distributed data analytics, computational pathology studies, and federated learning.

Ethics & Regulations: Will tackle ethical and legal requirements when dealing with medical data. The developed solutions will be based on numerous preliminary works and experiences within the central, national platform (TMF - Technology, Methods, and Infrastructure for Networked Medical Research). All relevant regulations will be included, i.e. the WMA Declaration of Helsinki, supplemented by the WMA Declaration of Taipei and the data protection legislation. A major goal of NFDI4Patho is to provide solutions for data generation applicable for the development of computer-aided decision support systems (CDSS), which fall under the medical device legislation regulated by the EU regulation 2071/745 (MDR).

Education & Dissemination: Will disseminate the developed solutions, provide the essential education and training for all researchers to develop the required competencies, and tackle public relations.

Security: To develop a secure infrastructure, we will define systematic security measures and enable privacy by design by automatic pseudonymization/anonymization of personal data. We will carry out penetration tests and threat modeling to identify potential vulnerabilities and perform audits to identify weaknesses in our system.

Brief description of the proposed use of existing infrastructures, tools and services that are essential in order to fulfil the planned consortium's objectives

The specific technical and medical aspects of NFDI4Patho are fully addressed by the **complementary expertise of consortium partners**. **NFDI4Patho builds upon existing infrastructures and close coordination with major research initiatives** within the NFDI, particularly GHGA, NFDI4Health, but also NFDI Neuro, NFDI4Immuno and NFDI4Bioimage, and outside of the NFDI, i.e. MII, EMPAIA, BIGPICTURE, DZG, GBA/GBN and the Network of University Medicine (NUM; particularly the projects DEFEAT PANDEMIcs, AKTIN, CODEX and RACOON). The consortium partners are **directly involved in most of these initiatives, allowing close and direct, coordinated interactions**. The NFDI4Patho concept builds on horizontal networking across the individual locations to achieve vertical networking within the university locations through interoperability with the MII data integration centers and other networks. This will make the **data available for broad interdisciplinary research projects**. On the international level, close interaction with the majority (if not all) of international initiatives in digital pathology has been initiated (e.g. IC3R, piCC, BigPicture, PathLake, Imaging Data Commons), with several consortium partners being directly involved in most.

The consortium involves **all major German professional societies** involved in pathology, i.e., the Association of German Pathologists (BDP), the German Society of Pathology (DGP), and the German Society of Neuropathology and Neuroanatomy (DGNN). This will enable a very high, in fact potentially a complete coverage and broad education and implementation of NFDI4Patho. **NFDI4Patho will follow the principle of open and participative governance, in which the inclusion of additional interested centers is anticipated and welcome.** NFDI4Patho will provide support and assistance in aligning additional partners with the NFDI4Patho infrastructure. The consortium already involves several **centers strongly involved in digital pathology**, all of which have expertise in quality management and assurance and are accredited according to the DIN EN ISO/IEC 17020:2012 (DAkkS). Even though digital pathology is still an emerging field, these centers have existing infrastructures allowing for a digital workflow. The consortium partners are also involved as experts in (digital) Pathology in the World Health Organization (WHO). NFDI4Patho will also profit from the existing high-performance computing (HPC) clusters on various sites, enabling advanced computational pathology approaches (e.g., JARA, FIAS, BIFOLD, MDC, etc.). NFDI4Patho will also align with the GAIA-X and EOSC initiatives, establishing transparent digital ecosystems on the European level. The consortium has a long-standing experience with relevant interoperability standards (DICOM, HL7, IHE, FHIR), development of the DICOM toolkit DCMTK, and experience in standardization work. Extensive experience also exists in the field of dissemination and education and the development and analysis of security architectures and protocols, the development of privacy-preserving distributed applications, and attack detection with machine learning methods.

All the "Pathology" partners and all the above-mentioned professional societies participate in the internationally unique national **COVID-19 Autopsy Registry** (DeRegCOVID, www.DeRegCOVID.ukaachen.de) and the NUM project DEFEAT PANDEMIcs, focusing on collecting and analyzing data of COVID-19 autopsies in Germany. These large projects (with >35 national pathologies involved) will enable access to autopsy datasets and already provides strong evidence of a **highly collaborative, nationwide work of the NFDI4Patho community**. We will closely collaborate with the "Ecosystem for Pathology Diagnostics with AI Assistance" (EMPAIA) project, which has complementary goals but does not have the access to research data that NFDI4Patho will provide. This joined effort might uniquely facilitate the development and translation of AI into diagnostic pathology. NFDI4Patho has strong **expertise in Medical Informatics**, e.g., in software engineering, education, and development and maintaining research data infrastructures for various collaborative research projects (e.g. AKTIN and DeRegCOVID). The Technology and Methods Platform for Networked Medical Research (TMF), as an umbrella organization for medical collaborative research projects and networks in Germany, provides, among others, extensive legal and ethical expertise for NFDI4Patho.

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

NFDI4Patho will put a strong focus on integration and interoperability, working closely with other NFDI consortia, other infrastructure projects as well as standardization organizations nationally and internationally. We will either integrate or utilize already available solutions or develop, support, or align novel solutions with the NFDI and respective communities. The added value of NFDI4Patho lies in supporting the very specific requirements of pathology, representing a medical cross-sectional field supporting many other medical specialties not tackled by other NFDI consortia. NFDI4Patho also aims to serve as a use case for the development of research data infrastructure for other medical specialties. Importantly, NFDI4Patho tackles an emerging and future-oriented research field of digital pathology, in which no comparable infrastructures exist yet. This thereby provides a unique opportunity to potentially provide solutions and standards not only nationally, but also internationally. It also provides a unique opportunity to develop standards and processes right from the beginning.
Potential interaction with many NFDI consortia could be easily envisaged and we also plan to integrate and if possible, help to shape the NFDI beyond the medical domain. In the medical field, NFDI4Patho will focus on intensive collaboration with consortia that provide direct, relevant, and bilateral complementarity and value, and that have already been outlined and planned. A key added value provided by NFDI4Patho is the expertise, (meta-)data and standards from the pathology domain, while NFDI4Patho will benefit from the developed standards from the fields of genetics and omics (GHGA), personalized health data (NFDI4Health), neurology (NFDI4Neuro) and immunology (NFDI4Immuno). This will facilitate and enable the development of common and shared solutions across the domains and the communities. Together with NFDI4Health, GHGA and NFDI Neuro we will align our efforts in tackling and handling sensitive patient data regarding ethical, legal, privacy and security issues, within the framework of NFDI4Patho, drawing on the extensive experience of the umbrella organization TMF.
The linkage of molecular and pathology data provides a very strong added value and is an essential approach not only in cancer but also in other diseases, particularly rare diseases. Therefore, with GHGA we will work on specific use-cases aligning and making digital pathology and -omics data available, particularly addressing record linkage. With NFDI4Health, we will work on standards, particularly for the chronic disease use cases addressed by both consortia and for mortality registries (NFDI4Health) aligned with autopsies as the ultimate method for assessing mortality (NFDI4Patho). We will also jointly address the interoperability and use of structured (NFDI4Health) and unstructured (NFDI4Patho) medical data. Since MII and NFDI4Health themselves cannot cover the complex and rapidly growing field of digital pathology, together we aim to align and support the convergence to a single research inquiry site, i.e., the national data-sharing platform (formerly ZARS) of the MII. With NFDI Neuro, we will address multi-scalar datasets which are generated in both consortia, as well as text data mining, specific use cases in neurology and potential approaches for record linkage. With NFDI4Immuno, we will address selected use-cases, e.g., in the field of immune-pathology and lymphoid malignancies and autoimmune diseases and provide expertise in working with routine unstructured routine medical data as well as imaging and pathology data. With NFDI4Bioimage, we will address connections between clinical imaging (NFDI4Patho) and pre-clinical/biological imaging (NFDI4Bioimage), we will provide expertise and solutions on patient data handling and record linkage, and together address the interoperability between medical imaging standards (e.g., DICOM) and bioimaging data formats (TIFF, ZARR).

3. Cross-cutting topics

Please identify cross-cutting topics that are relevant for your consortium and that need to be designed and developed by several or all NFDI consortia.

Of the 16 topics identified in the "NFDI Cross-cutting Topics Workshop Report" and the "Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung", NFDI4Patho has a strong interest in topics related to sensible medical data (together with GHGA, NFDI4Health, NFDI Neuro), biological data (e.g. with NFDI4Bioimage) and management of large data sets. In particular, we are interested in

- Topic 1[1]: Research Data Commons (cloud infrastructure, identity management, distributed AI, …)
- Topic 2: (Meta) Data, Findability (e.g., record linkage)
- Topic 4: Provenance (e.g., reproducibility of experimental data, data quality, long-term storage (e.g. readable data formats))
- Topic 5: Infrastructure/Interoperability/Interfaces (e.g., standardized data management plans, record linkage, …)
- Topic 6: Quality Management and Assurance Legal and ethical aspects (e.g., enabling use of the data to build medical devices)
- Topic 8: Ethical-legal Aspects, person-related (e.g., consent to use medical records, legal frameworks for frictionless data sharing)
- Topic 10: Training and Education (e.g., to enable adoption in as many centers as possible).
- Topic 11: Common Vision and Strategy
- Topic 14: HR
- Topic 15: Internationalization (e.g., compatibility to related international projects)

[1] The topic numbers refer to Ebert, B., Fluck, J., Glöckner, F. O., Koepler, O., Miller, B., Schmitt, R., Schrade, T., Stegle, O., Steinbeck, C., von Suchodoletz, D., Wagemann, K., Knebes, J., Kraft, S., Seitz-Moskaliuk, H., Sure-Vetter, Y., & Wössner, E. (2021). **NFDI Cross-cutting Topics Workshop Report.** https://doi.org/10.5281/zenodo.4593770

Please indicate which of these cross-cutting topics your consortium could contribute to and how

NFDI4Patho will tackle similar major topics and required task areas as other NFDI consortia or other infrastructure projects, especially (but not only) in the biomedical field, and we will be happy to contribute to common cross-cutting developments in these areas.
Our focus will particularly be on data-related cross-cutting topics, as our unique selling point is the use of routine medical pathology data. Also, another unique focus of NFDI4Patho is on the regulatory aspects enabling translational medical research, i.e. bringing research results from the bench to the bedside and patient benefit. In addition to these data-related issues, NFDI4Patho partners have strong expertise in education, dissemination and providing ongoing research support. We also aim to contribute to these cross-cutting areas as well.
Infrastructure: ethic and legal restrictions and consent management (Topics 6, 8). Taking the patient perspective into account, we will contribute to legal and ethical questions. Possible contributions range from standardized legal consent frameworks which need to be compatible with other initiatives in the medical context (e.g., MII, GHGA) to re-usable agreements for data-sharing and data analysis in trusted environments. Another important point are the regulatory requirements for data use in the context of the development of diagnostic systems (IVDR, EU-V 746/2017, MDR EU-V745/2017), where we will contribute to open, quality-assured software systems. Regulatory aspects are essential in medical research and still a large hurdle on the way to bring findings and insights to clinical use. Universal, re-usable building blocks could facilitate future developments in this area.
Infrastructure: long-term storage and access, decentralized computing (Topics 1, 4, 5). The size of histological image data requires special concepts for decentralized storage, decentralized use and demand-driven merging, streaming of partial data and decentralized data evaluation. These concepts will be further developed and implemented in NFDI4Patho in a modular architecture such that it is easy to adapt them to other research data. Wherever possible, existing components will be used or built upon. We will ensure long-term usability by aligning and integrating all developments with existing interoperability standards (especially DICOM) in coordination with other consortia.
Infrastructure: privacy, protection and record linkage (Topics 2, 8). One key element of our work will be the pseudonymization or anonymization of image data (along with their metadata) to ensure reliable research results based on secondary use of the routine clinical data. We will coordinate with other consortia to ensure the compatibility of unique identifiers and metadata formats to enable a record linkage across projects.
Community and collaborative governance: vision and (international) dissemination (Topics 10, 11, 14, 15): As a federated research data infrastructure project, international collaboration is the logical next step to maximize user benefits. We will support collaboration between international projects. Our goal is to connect and link international projects with similar goals, ideally leading to cross-project search and database connectivity. At the national level, NFDI4Patho aims to involve as many centers as possible to maximize its impact. User-focused information and training programs are an important step toward achieving broad coverage of clinical centers. We will align and contribute to existing formats for training and user engagement.