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 Immunology

- *Acronym of the planned consortium*
  NFDI4Immuno

- *Applicant institution*
  Deutsches Krebsforschungszentrum (DKFZ)
  Stiftung des öffentlichen Rechts
  Im Neuenheimer Feld 280
  69120 Heidelberg

  Vorstand: Prof. Dr. Michael Baumann, Ursula Weyrich

- *Spokesperson*
  Dr. Christian Busse
  Deutsches Krebsforschungszentrum
  christian.busse@dkfz-heidelberg.de

- *Co-applicant institution*
  Ruhr-Universität Bochum (RUB)
  Universitätsstraße 150
  44801 Bochum
  Dekan: Professor Dr. Ralf Gold

- *Co-spokesperson*
  Prof. Dr. Nina Babel
  nina.babel@elisabethgruppe.de
  Centrum für Translationale Medizin
  Universitätsklinikum der Ruhr-Universität Bochum

- *Co-applicant institution*
  Deutsches Rheuma-Forschungszentrum (DRFZ)
  Charitéplatz 1
  10117 Berlin
  Kaufmännische Direktorin: Petra Starke

- *Co-spokesperson*
  Prof. Dr. Hyun-Dong Chang
  chang@drfz.de
  DRFZ

- *Co-applicant institution*
  Friedrich-Loeffler-Institut (FLI)
  Südufer 10
  17493 Greifswald - Insel Riems
  Leiter der Hauptverwaltung: Dietmar Nobis

- *Co-spokesperson*
  Prof. Dr. Anca Dorhoi
  anca.dorhoi@fli.de
  FLI Riems

- *Co-applicant institution*
  Zentrum für Regenerative Therapien der TU Dresden (CRTD)
  Fetscherstraße 105
  01307 Dresden
  Direktor: Prof. Michael Brand

- *Co-spokesperson*
  Dr. Anne Eugster
  anne.eugster@tu-dresden.de
  CRTD

- *Co-applicant institution*
  Charité - Universitätsmedizin Berlin
  Charitéplatz 1
  10117 Berlin
  Dekan: Prof. Dr. Axel Radlach Pries

- *Co-spokesperson*
  Prof. Dr. Michael Hummel
  michael.hummel@charite.de
  Institut für Pathologie - Molekulare Diagnostik

- *Co-applicant institution*
  Universitätsklinikum Essen
  Hufelandstr. 55
  45147 Essen
  Kaufmännnischer Direktor: Thorsten Kaatze

- *Co-spokesperson*
  Prof. Dr. Ralf Küppers
  ralf.kueppers@uk-essen.de
  Institut für Zellbiologie (Tumorforschung)

- *Co-applicant institution*
  Helmholtz-Zentrum für Infektionsforschung (HZI)
  Inhoffenstraße 7
  38124 Braunschweig
  Wissenschaftlicher Geschäftsführer: Prof. Dr. Dirk Heinz

- *Co-spokesperson*
  Prof. Dr. Michael Meyer-Hermann
  mmh@theoretical-biology.de
  HZI

- *Co-applicant institution*
  Eberhard Karls Universität Tübingen
  Geschwister-Scholl-Platz
  72074 Tübingen
  Rektor: Prof. Dr. Bernd Engler

- *Co-spokesperson*
  Dr. Sven Nahnsen
  sven.nahnsen@qbic.uni-tuebingen.de
  Quantitative Biology Center (QBiC)

- *Co-applicant institution*
  DKMS gGmbH
  Kressbach 1
  72072 Tübingen
  Chief Executive Officer: Dr. Elke Neujahr

- *Co-spokesperson*
  Prof. Dr. Johannes Schetelig
  schetelig@dkms.de
  Clinical Trials Unit

- *Co-applicant institution*
  Universitätsklinikum Münster
  Albert-Schweitzer-Campus 1
  48149 Münster
  Dekan: Prof. Dr. Frank Ulrich Müller

- *Co-spokesperson*
  Prof. Dr. Nicholas Schwab
  nicholas.schwab@ukmuenster.de
  Klinik für Neurologie mit Institut für Translationale Neurologie

- *Participant*

  Prof. Barbara M. Bröker
  barbara.broeker@med.uni-greifswald.de
  Institut für Immunologie
  Universität Greifswald
  Ferdinand-Sauerbruch-Straße, DZ7
  17475 Greifswald

- *Participant*

  PD Dr. Cornelia Eckert
  cornelia.eckert@charite.de
  Klinik für Pädiatrie mit Schwerpunkt Onkologie und Hämatologie
  Charité - Universitätsmedizin Berlin
  Augustenburger Platz 1
  13353 Berlin

- *Participant*

  Dr. Katharina Imkeller
  imkeller@embl.de
  Neurologisches Institut / Edinger Institut
  Universitätsklinikum Frankfurt
  Theodor-Stern-Kai 7
  60590 Frankfurt am Main

- *Participant*

  Dr. Felix Meissner
  meissner@biochem.mpg.de
  Experimentelle Systemimmunologie
  Max-Planck-Institut für Biochemie
  Am Klopferspitz 18
  82152 Martinsried

- *Participant*

  Dr. Katherina Siewert
  katherina.siewert@bfr.bund.de
  Studienzentrum Dermatotoxikologie
  Bundesinstitut für Risikobewertung
  Max-Dohrn-Str. 8-10
  10589 Berlin

- *Participant*

  Prof. Dietmar Zehn
  dietmar.zehn@tum.de
  Lehrstuhl für Tierphysiologie und Immunologie
  Wissenschaftszentrum Weihenstephan
  Technische Universität München (TUM)
  Weihenstephaner Berg 3
  85356 Freising

3 Objectives, work programme and research environment

- *Research area of the proposed consortium (according to the DFG classification system)*

  Immunology (204-05) and related research areas:
  - Hematology, Oncology / Rheumatology (205-14, 205-18)
  - Microbiology / Virology / Parasitology (204-03, 204-04, 204-06)
  - Pathology / Neuropathology (205-06, 206-07)
  - Veterinary medicine (207-08)
  - Theoretical biology (201-07)

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

The immune system plays a fundamental role in health and disease and efficiently protects vertebrate hosts from infections and cancer. However, failures in its regulation can cause autoimmunity, allergy, immunodeficiencies and lymphoid malignancies. To perform the critical task of self/non-self recognition, the adaptive immune system utilizes billions of randomly generated immunoglobulins/antibodies and T-cell receptors (hereafter: Ig/TCR), which are expressed by cells of the B-cell and T-cell lineage, respectively. The phenotype, activation state, secretion pattern and histo-anatomic location of these cells, as well as the entireness of the Ig/TCR of an individual host (commonly referred to as Adaptive Immune Receptor Repertoire or "AIRR"), are reflective of the key processes within the immune system: Diversification, selection, antigen recognition and clonal expansion. A comprehensive understanding of these processes will facilitate mechanistic insights and allow the development of diagnostic markers and novel therapeutic strategies. To this end, it is necessary to obtain the capability to combine data and metadata from diverse experimental technologies, e.g., sequencing, cytometry, affinity measurements, concentration assays and imaging, which provide complementary observations of these processes. Thus, **the main objective of NFDI4Immuno is to build a network of federated repositories for data describing the state of the immune system in close cooperation with the German immunological community and to provide tools and services that will facilitate integrated data analyses across these repositories.** NFDI4Immuno will start by building a federated infrastructure for AIRR-seq and cytometry data and by expanding existing data models to cover reactivity, cytokine and spatial information while initiating the integration of additional data types, as determined by the needs of the community. The detailed steps are as follows:

Build a network of federated AIRR-seq and cytometry data repositories: AIRR-seq is a heterogeneous set of NGS-based technologies that provide information on the AIRR, i.e., the highly variable regions of Ig/TCR. Several partners in the consortium have a long-standing expertise in the generation and analysis of this data type and are among the founding members of the international AIRR Community. As such they have an established involvement in the community's efforts to provide standards and promote FAIR practices for AIRR-seq data. Building on this expertise and the existing software stack (iReceptor Turnkey) for sharing AIRR-seq data, we plan to roll-out the initial set of data repositories within the first year.
Cytometry (i.e., both flow and mass cytometry) is the current gold standard in immunology to describe, define and isolate immune cells and provides a rich phenotypic description at single-cell resolution. In spite of its pervasiveness and importance, there is currently no generally accepted way to share these often large and high-dimensional data sets. To fully utilize the wealth of information in cytometric data, we will expand existing software stacks to facilitate the storage, annotation and analysis of cytometry data, while providing for close integration with associated AIRR-seq data sets. This is considered to be critical activity, as the advent of commercially available platforms that use DNA barcoding of cells and surface markers to combine cytometry with AIRR-seq (e.g., 10X Chromium), has made the borders between these two technologies become increasingly fluid.
Throughout the funding period, AIRR-seq and cytometry data - both from the consortium members and third parties - will be curated and made publicly available via our repositories. We plan to give special attention to data from species other than human and mouse, e.g., livestock and non-human primates, which are often not appropriately covered by other biomedical

infrastructures. The resulting network of repositories will constitute the seed infrastructure for the further expansion of NFDI4Immuno to cover more data types.

Expanding existing data models: Information on Ig/TCR reactivity, cytokine secretion, gene expression and temporo-spatial location is generated and used by immunologists on a daily basis. However, due to the variety of methods involved, there is a large spectrum of distinct data types that need to be integrated into an ecosystem of FAIR data. Therefore, NFDI4Immuno will lead the extension of existing immunological data models to facilitate the annotation, sharing and re-use of these data types, which is a precondition for their integration into the repository infrastructure described above.

Initiating the integration of additional data types: The needs of the community will be continuously assessed over the funding period. For data types that are not yet covered by ongoing efforts, we will develop, circulate and request feedback on roadmaps to their integration. To support data types that cannot be readily integrated into our data models, NFDI4Immuno will build and maintain an infrastructure for third-party annotation of data sets based on the W3C Web Annotation standard. This will also facilitate the annotation of data sets stored by general purpose repositories with immunological metadata.

These data-centric activities will be translated into the following five task areas:

- *Data curation* - Provide consistent metadata annotation to existing data sets and make them available via the consortium's repositories. Additionally, organize the consortium's Quality Assurance (QA) Panel, which will develop and maintain guidelines for data and metadata quality and perform periodic audits among the federated repositories.
- *Interoperability* - Harmonize metadata descriptors and ontologies with other NFDI consortia to facilitate fast and easy cross-consortia queries for data sets. Develop and standardize programmatic interfaces (APIs) for advanced computational workflows and exchange and interconnection of data within and beyond the NFDI.
- *Community involvement and outreach* - Disseminate and support the adoption of these and related strategies by the immunological community, to promote the cultural change towards Open Science. Engage stakeholders outside of the national scope of NFDI with the long-term goal to connect NFDI4Immuno at the international level, e.g., with EOSC.
- *Repository DevOps* - Develop and operate the consortium's federated repository infrastructure.
- *Application development and support* - Build novel user-friendly applications both for data analysis utilizing NFDI4Immuno's comprehensive federated infrastructure, as well as simplified (meta-)data submission to NFDI4Immuno repositories. Provide support and training to users with a primarily experimental background enabling them to appropriately utilize the resources provided by the consortium.

The planned governance structure will consist of the Executive Board (Speaker + two co-Speakers), the General Assembly (all partners) and the Community Council, a committee representing the diverse NFDI4Immuno user communities and their needs. The Executive Board will manage the organizational, financial and legal concerns of the consortium. It will implement a unified strategy to ensure the long-term availability and usability of the consortium's data and code.

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

Hardware and network infrastructure: In general terms, NFDI4Immuno is designed as a federated repository infrastructure, which will substantially lower the requirements for storage and compute capacity at the individual partner sites. Therefore, it is by default expected that partners can provide these resources as part of their contribution to the consortium. However, especially for large third-party data sets, we are planning to utilize centralized storage services like the Helmholtz Data Federation (HDF). These data sets might also require additional network bandwidth as well as some compute resources. Therefore, the consortium is currently in the process of estimating the potential resource requirements and is in contact with the Helmholtz Infrastructure for Federated ICT Services (HIFIS) for further input regarding this topic.

Software: NFDI4Immuno will build on the freely available iReceptor Turnkey software stack, the core of which is licensed under LGPL3 (GNU Lesser General Public License). Notably, the consortium partners at DKFZ are currently part of the Horizon 2020-funded *iReceptor Plus* consortium, which aims to substantially enhance the capabilities of the software stack until the end of 2022. Therefore, the necessary know-how for further development and operation of the platform will already be present in NFDI4Immuno.

Biobanking: Many pathological processes involve the immune system, thus biobanked samples of diseased tissues are an attractive resource for immunological analysis. Therefore, we consider it valuable to link immunological data at NFDI4Immuno that derives from samples of the German Biobank Alliance (GBA) with its respective metadata stored by the GBA. To this end, we will work together with the German Biobank Node (GBN), which coordinates the GBA and is part of BBMRI-ERIC, to develop and implement such linkage on the sample level.

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

NFDI4Immuno's "inwards" activities aim to provide deep integration of immunological data sets with rich, domain-specific metadata. To embed this information within the efforts of other NFDI consortia, we are planning the following "outwards" activities:

NFDI4Microbiota: The continuous interaction between the immune system of vertebrate hosts and their microbiota makes this consortium a natural partner. To facilitate the parallel analysis of both, we plan on harmonizing metadata and ontologies describing the host, as well as identifying and enhancing formalized descriptions of sampling procedures.

InnoMatSafety: InnoMatSafety is the other consortium within the NFDI that plans to manage a significant amount of cytometry data. Therefore, we aim to establish a platform for regular knowledge exchange in respect to the technical aspects of quality control, data management and storage. In the best case, this will develop into a shared backend storage infrastructure with domain-specific metadata layers on top of it. In addition, we will evaluate the potential shared requirements for domain-specific (i.e., immunological) metadata.

NFDI4BIOIMAGE: Microscopy data is one of the key data types in immunological research. As NFDI4BIOIMAGE aims to provide generic and domain-spanning tools and services for the storage and management of microscopy and photonics-based imaging data, NFDI4Immuno is interested in utilizing these resources. To ensure the findability of imaging data describing the immunological processes that are the central focus of NFDI4Immuno, we will cooperate in developing and harmonizing metadata structures and ontologies.

NFDI-Neuro: Our cooperation with NFDI-Neuro will revolve around creating semantic interoperability between data sets from our respective research domains. To this end, we will focus on the joint use case of neurological autoimmune disorders.

NFDI4Patho: With its expertise regarding data from and about patients, NFDI4Patho is an important partner especially for use cases in translational immunology. We plan to work together on such use cases aiming to combine and leverage data sets from lymphoid malignancies as well as autoimmune disorders.

GHGA: We consider GHGA as the NFDI's central infrastructure for generic sequencing data from human subjects and will interact closely with regard to the deposition of such data (e.g., transcriptome) within their repositories and cross-referencing and accessing information between our consortia. Generic sequencing data outside of the scope of GHGA (i.e., majority of studies with non-human subjects) will be linked via ENA and we aim to develop the necessary programmatic abstraction layers together with GHGA.

NFDI4Health: The centralized metadata storage and catalog proposed by NFDI4Health could be a complementary top-layer to NFDI4Immuno's federated data repositories. To this end, we plan to develop the required interfaces for data exchange and will make our expertise in domain-specific metadata and ontologies available to the NFDI4Health consortium.

4 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.*

  We consider the cross-cutting topics identified in the Leipzig-Berlin Declaration[1] as a common work assignment for the whole NFDI, which will also constitute an important process of organizational development. Therefore, NFDI4Immuno explicitly supports the task areas described in the Declaration and their suggested implementations. We identified the following cross-cutting topics and services as being especially relevant for our consortium:

  - Open data and metadata standards
  - Harmonization and standardization of ontologies
  - Supporting FAIR and Open Science practices within the communities
  - Data literacy and RDM curricula for scientists
  - Guidelines for quality assurance
  - Free and Open Source research software
  - Management of personal and/or confidential data
  - Common authentication and authorization infrastructure
  - Integration of the NFDI into European RDM infrastructures (EOSC)

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

  NFDI4Immuno can contribute its expertise regarding the development of open community data standards, as well as the extension and harmonization of domain-spanning ontologies. Several members of the planned consortium have already engaged in similar activities in the context of data standards development by the AIRR Community.

1Bierwirth M *et al*. Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung. DOI: 10.5281/zenodo.3895209