

# National Research Data Infrastructure for Adaptive Immune Receptor Repertoires (NFDI4AIRR)

Sprecher: C. Busse, DKFZ, [christian.busse@dkfz-heidelberg.de](mailto:christian.busse@dkfz-heidelberg.de)

## Angaben zum Konsortium

### Forschungsgebiet

Immunologie (204-05<sup>1</sup>, 205-18) und damit zusammenhängende Fächer

- Hämatologie / Onkologie (205-14)
- Mikrobiologie / Virologie (204-xx)
- Neuropathologie (206-07)
- Pathologie (205-06)
- Tiermedizin (207-11, 207-12)
- Theoretische Biologie (201-07)

### Datentypen

NFDI4AIRR strebt langfristig an, alle für die Immunologie relevanten Datentypen in der NFDI abbilden zu können. Im Rahmen der Beantragung planen wir jedoch uns vorerst auf die folgenden Datentypen und Methodiken zu fokussieren:

- AIRR-seq<sup>2</sup>: Ein sequenzierungs-basiertes Verfahren, das die adaptiven Immunrezeptoren (Immunglobuline & T-Zellrezeptoren) untersucht. Damit lassen sich über längere Zeiträume Proliferations-, Selektions- und Mutationsprozesse innerhalb des Immunsystems verfolgen, wobei die individuellen Rezeptoren als Barcodes dienen. Zusätzlich ist diese Information kritisch zur Rekonstruktion und zur potenziellen Vorhersage der Reaktivitäten dieser Rezeptoren (z.B. Antikörper/Antigen-Interaktion).
- Durchflusszytometrie: Ein weit verbreitetes Standardverfahren zur Phänotypisierung von Immunzellen, welches inhärent Daten in Einzelzell-Auflösung liefert und zusätzlich die Isolation und Weiterverarbeitung von Zellen ermöglicht. Nahezu alle relevanten Zellpopulationen des adaptiven Immunsystems können hierüber charakterisiert und quantifiziert werden.
- Mikroskopie: Ein für die Immunologie zunehmend wichtig werdender Datentyp, der auf Grund der hohen Ortsauflösung innerhalb des Gewebes und langen Beobachtungszeiträumen komplementär zur Durchflusszytometrie ist.

Die Auswahl dieser Datentypen beruht darauf, dass sie (1) inhaltlich reichhaltige Information über immunologische Prozesse liefern, (2) sich wechselseitig ergänzen und (3) auf verschiedenen Ebenen bereits einer gewissen Standardisierung unterzogen wurden. Damit werden sie es ermöglichen, ein umfassendes Datenmodell für die Immunologie zu entwickeln und zu mplementieren, in das später weitere Datentypen eingefügt werden können.

1 Fach-Nr. gemäß [DFG Fachsystematik](#) vom 2020-02-18

2 Repertoiresequenzierung von adaptiven Immunrezeptoren (Adaptive Immune Receptor Repertoire sequencing)

## **Domänen-spezifische FDM-Anforderungen**

Immunologische Forschung erstreckt sich über alle lebenswissenschaftlichen Betrachtungsebenen, das heißt von den molekularen Mechanismen bis zu den Auswirkungen auf der Populationsebene des Wirtsorganismus. Ein sinnvolles Forschungsdatenmanagement (FDM) muss daher dieser starken "vertikalen" Komponente Rechnung tragen und die einzelnen Ebenen so verknüpfen, dass domänen-spezifische Entitäten und Prozesse abgebildet werden können. Um auf der anderen Seite die "horizontale" Interoperabilität mit anderen Domänen zu gewährleisten, müssen jedoch innerhalb einer Ebene domänenübergreifende Standards und Ontologien identifiziert und verwendet werden. Hinzu kommen, bedingt durch das weite Spektrum der von der Immunologie untersuchten Spezies (Mensch, Maus, andere Primaten, Nutztiere, etc.), Unterschiede in der Komplexität der vorhandenen Metadaten auf den einzelnen Ebenen. Hier ist es daher wichtig zwischen einem obligatorischen Kern- und einem optionalen Erweiterungsset von Metadaten zu unterscheiden.

## **Geplante FDM-Maßnahmen und Dienste**

NFDI4AIRR plant Aktivitäten in den folgenden Bereichen: (1) Den Aufbau, Ausbau und Betrieb von verteilten Repositorien für die oben genannten Datentypen, teilweise auch in Kooperation mit anderen Konsortien. Dies umfasst auch ein gemeinsames Register aller Datensätze innerhalb der Repositorien um einfache Auffindbarkeit zu gewährleisten. (2) Die Etablierung und Weiterentwicklung von Community-Standards für immunologische Metadaten, sowohl auf Ebene der Inhalte (*was* wird annotiert) als auch der Repräsentation (*wie* wird die Information codiert). (3) Die Entwicklung von Qualitätsrichtlinien für Daten und Metadaten und eine darauf basierende Auditierung der unter (1) genannten Repositorien. (4) Kuratierung und FAIRifizierung von existierenden öffentlichen Datensätzen mit hoher Relevanz für die Community. (5) Training und Unterstützung von Wissenschaftler/innen bei der Benutzung und Einreichung von Daten aus bzw. in Repositorien und eine Förderung von *Open Science* Praktiken.

## **Erfahrungen und Hintergrund im FDM**

Mehrere der Partner in NFDI4AIRR haben langjährige Erfahrungen in der Entwicklung und Harmonisierung von Metadatenstandards durch internationale Communities. Zusätzlich sind Kompetenzen im Bereich der Entwicklung und des Betriebs lokaler Datenbanken, dem Aufbau föderierter Repositorien, Big Data Analysen und maschinellem Lernen vorhanden.

## **Partnerorganisationen und vorhandenen Infrastrukturen**

NFDI4AIRR strebt eine enge Interaktion mit den relevanten Organisationen und Infrastrukturen an. Hierzu gehören auf nationaler Ebene die Deutsche Gesellschaft für Immunologie (DGfI), die Deutsche Gesellschaft für Zytometrie (DGfZ), und die German Biobank Alliance (GBA). Auf europäischer Ebene stellen EOSC Life, BBMRI ERIC, EuroClonalityNGS und das European B Cell Network relevante Stakeholder dar. International bestehen enge Verbindungen zur AIRR Community und zum iReceptor Plus Konsortium. Zusätzlich bestehen Kontakte zum Human Immunology Project Consortium (HIPC) des NIAID (USA) und verschiedenen Arbeitsgruppen der Research Data Alliance (RDA).

## **Schnittstellen innerhalb der NFDI**

Als Konsortium im Bereich der Lebenswissenschaften bestehen für NFDI4AIRR wichtige inhaltliche Schnittstellen zu NFDI4Microbiota, NFDI-Neuro, NFDI4Agri und NFDI4Health. Wir

unterstützen die Bemühungen dieser und andere Konsortien im Rahmen der NFDI4Life Umbrella Initiative auf eine gemeinsame Bearbeitung spezifischer Querschnittsthemen der Lebenswissenschaften hinzuwirken (z.B. Harmonisierung der eingesetzten Ontologien). Außerdem bestehen Schnittstellen bezüglich der Methodiken und Datentypen zu GHGA (DNA-Sequenzierung), Advanced Material Safety (Durchflusszytometrie) und NFDI4BIMP (Mikroskopiedaten). Zusätzlich teilen wir das Interesse der NFDI4RSE Querschnittsinitiative zur Steigerung der Nachnutzbarkeit und Nachhaltigkeit von Forschungssoftware.

## Querschnittsthemen

NFDI4AIRR betrachtet die in der Leipzig-Berlin-Erklärung<sup>3</sup> identifizierten Querschnittsthemen als gemeinsamen Arbeitsauftrag an die gesamte NFDI, dessen Durchführung einen wichtigen strukturbildenden Prozess darstellt. Wir unterstützen daher ausdrücklich die gemeinschaftliche Bearbeitung aller in der Erklärung aufgeführten Handlungsfelder und die vorgeschlagenen Umsetzungsmöglichkeiten. Als besonders relevant für unser eigenes Vorhaben haben wir die folgenden Querschnittsthemen/-dienste identifiziert:

- Offene Daten- und Metadatenstandards
- Harmonisierung und Standardisierung von Ontologien
- Förderung von FAIRen und *Open Science* Praktiken in den Fachcommunities
- *Data literacy* und FDM Curricula für Wissenschaftler/innen
- Qualitätssicherungsrichtlinien
- Freie, quelloffene und nachhaltige Forschungssoftware ("Public Money, Public Code")
- Umgang mit personenbezogenen und vertraulichen Daten
- Identifizierungs- und Zugriffsverwaltung
- Integration der NFDI in europäische FDM-Strukturen (EOSC)

Unsere Erfahrungen in der Community-geleiteten Entwicklung und Harmonisierung von Metadatenstandards, Datenmodellen und föderierten Datenrepositorien planen wir über die entsprechenden Gremien in die NFDI einzubringen.

## Erwartungen an die NFDI-Konferenz

Von der Teilnahme an der NFDI-Konferenz erhoffen wir uns Einblicke in die geplanten Aktivitäten der bis dahin bewilligten Konsortien der ersten Runde und Updates von den Initiativen der zweiten Runde. Des Weiteren würden wir uns gerne mit potentiell neuen Akteuren im Bereich der NFDI vernetzen. Daher freuen wir uns auf die Gelegenheit, unsere Ziele und Planungen präsentieren zu können.

<sup>3</sup> Bierwirth M *et al.* Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung. DOI: [10.5281/zenodo.3895209](https://doi.org/10.5281/zenodo.3895209)

## Vorgesehene Mitglieder des Konsortiums

(Co-Sprecherinnen/Co-Sprecher und die weiteren, beteiligten Institutionen)

Co-Sprecher/in	Zugehörige Institution
Dr. Christian Busse Projektleiter <a href="mailto:christian.busse@dkfz-heidelberg.de">christian.busse@dkfz-heidelberg.de</a>	Deutsches Krebsforschungszentrum (DKFZ) Im Neuenheimer Feld 280 69120 Heidelberg
Prof. Dr. Anca Dorhoi Institutsleiterin <a href="mailto:anca.dorhoi@fli.de">anca.dorhoi@fli.de</a>	Friedrich-Loeffler-Institut (FLI) Südufer 10 17493 Greifswald - Insel Riems
Prof. Dr. Michael Hummel Abteilungsleiter <a href="mailto:michael.hummel@charite.de">michael.hummel@charite.de</a>	Charité - Universitätsmedizin Berlin Charitéplatz 1 10117 Berlin
Prof. Dr. Ralf Küppers Univ.-Professor <a href="mailto:ralf.kueppers@uk-essen.de">ralf.kueppers@uk-essen.de</a>	Universitätsklinikum Essen Hufelandstr. 55 45147 Essen
Dr. Dr. Alexander Schmidt Chief Medical Officer <a href="mailto:gf-buero@dkms.de">gf-buero@dkms.de</a>	DKMS gGmbH Kressbach 1 72072 Tübingen
Prof. Dr. Nicholas Schwab Univ.-Professor <a href="mailto:nicholas.schwab@ukmuenster.de">nicholas.schwab@ukmuenster.de</a>	Universitätsklinikum Münster Albert-Schweitzer-Campus 1 48149 Münster
Prof. Dr. Michael Meyer-Hermann Abteilungsleiter <a href="mailto:mmh@theoretical-biology.de">mmh@theoretical-biology.de</a>	Helmholtz-Zentrum für Infektionsforschung (HZI) Rebenring 56 38106 Braunschweig
Dr. Anne Eugster Scientist <a href="mailto:anne.eugster@tu-dresden.de">anne.eugster@tu-dresden.de</a>	Technische Universität Dresden Center for Regenerative Therapies Dresden Faculty of Medicine Fetscherstrasse 105 01307 Dresden
Dr. Sven Nahnsen Direktor <a href="mailto:sven.nahnsen@qbic.uni-tuebingen.de">sven.nahnsen@qbic.uni-tuebingen.de</a>	Eberhard Karls Universität Tübingen Quantitative Biology Center Auf der Morgenstelle 10 72076 Tübingen