

# Gesamtvorhabensbeschreibung

zur Förderung durch das BMBF

Thema: FAIR Data Spaces – Aufbau eines gemeinsamen Cloud-basierten Datenraums für Wirtschaft und Wissenschaft

Akronym: FAIR-DS

Gesamtkosten: 6.313.712 €

Datum: 12.04.2021

Projektdauer: 36 Monate: 17.05.2021–16.05.2024

## Kontaktdaten des Antragstellers

|   |   |
|---|---|
| Organisation: <b>Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V. (FhG)</b><br><b>Postfach 20 07 33, 80007 München</b>  |   |
| Einrichtungen: Fraunhofer-Institut für Angewandte Informationstechnik FIT<br>Schloss Birlinghoven, 53757 Sankt Augustin<br>Fraunhofer-Institut für Software- und Systemtechnik ISST<br>Emil-Figge-Str. 91, 44227 Dortmund |   |
| Anrede: Herr  | Vorname: Christoph                                |
| Titel: Dr.  | Nachname: Lange-Bever                             |
| Telefon: <a href="tel:+49224114">+49 2241 14-****</a>   | E-Mail: Vorname.Nachname[at]fit.fraunhofer[dot]de |
| <u>Weitere Projektpartner:</u>  |   |
| <b>Nationale Forschungsdateninfrastruktur (NFDI) e.V.</b><br>Prof. Dr. York Sure-Vetter<br>Albert-Nestler-Str. 13<br>76131 Karlsruhe  |   |
| <b>Justus-Liebig-Universität Gießen (JLU)</b><br>Prof. Dr. Alexander Goesmann<br>Heinrich-Buff-Ring 58<br>35392 Gießen  |   |
| <b>Technische Informationsbibliothek (TIB)</b><br>Prof. Dr. Sören Auer<br>Welfengarten 1B<br>30167 Hannover   |   |

**Universität zu Köln (UzK)**

Prof. Dr. Oya Beyan  
Albertus-Magnus-Platz  
50923 Köln

**Universität Leipzig (UL)**

Prof. Dr.-Ing. Toralf Kirsten  
Härtelstr. 16-18  
04107 Leipzig

**RWTH Aachen University**

Dr. rer. nat. Marius Politze  
Templergraben 55  
52062 Aachen

**FIZ Karlsruhe - Leibniz-Institut für Informationsinfrastruktur GmbH**

Prof. Dr. Franziska Boehm  
Hermann-von-Helmholtz Platz 1  
76344 Eggenstein-Leopoldshafen

**Universität Heidelberg****Universitätsklinikum Heidelberg (UKHD),** Sektion Translationale Medizinethik

Nationales Centrum für Tumorerkrankungen  
Prof. Dr. Dr. Eva Winkler  
Im Neuenheimer Feld 460

**Eberhard Karls Universität Tübingen (EKUT)**

Prof. Dr.-Ing. Oliver Kohlbacher  
Sand 14  
72076 Tübingen

**Heidelberger Akademie der Wissenschaften (HAdW)<sup>1</sup>**

Dr. iur. Fruzsina Molnar-Gabor  
BioQuant-Zentrum der Univ. Heidelberg (BQ 049)  
INF 267, 69120 Heidelberg

**European Molecular Biology Laboratory (EMBL)**

Dr. rer. nat. Jan Korb  
Meyerhofstraße 1  
69117 Heidelberg

---

<sup>1</sup> Die Heidelberger Akademie der Wissenschaften schied im Februar 2022 als Verbundpartner aus. Ebenfalls im Februar 2022 wurde die Universität Heidelberg als Verbundpartner in FAIR-DS aufgenommen

**Philipps-Universität Marburg (UMR)**

Prof. Dr. Bernhard Seeger  
Fachbereich Mathematik und Informatik  
Hans-Meerwein-Str. 6

**Universität Münster (WWU)**

Prof. Dr. Thomas Hoeren  
Institut für Informations-, Telekommunikations- und Medienrecht  
– Zivilrechtliche Abteilung –  
Leonardo-Campus 9, Raum 110.118  
48149 Münster

**ZBW – Leibniz Informationszentrum Wirtschaft**

Prof. Dr. Klaus Tochtermann  
Düsternbrooker Weg 120  
24105 Kiel

**Atos Information Technology GmbH**

Klaus Ottradovetz  
Otto-Hahn-Ring 6  
81739 München

# 1. Ziele

## 1.1. Gesamtziel des Vorhabens

Das Projekt „FAIR Data Spaces“ strebt den Aufbau eines gemeinsamen Cloud-basierten Datenraums für Wirtschaft und Wissenschaft durch Zusammenführung der beiden Initiativen Gaia-X und NFDI an. „FAIR“ steht dabei für Leitprinzipien für die verantwortliche Verwaltung von Forschungsdaten – diese sollen auffindbar, zugreifbar, interoperabel und wiederverwendbar (findable, accessible, interoperable, reusable) sein (Wilkinson et al. 2016).

Gaia-X<sup>2</sup> ist eine föderierte und sichere Dateninfrastruktur für Europa. Ihr Aufbau wird koordiniert von der gemeinnützigen Gaia-X European Association for Data and Cloud, mit bisher über 300 Mitgliedern, überwiegend Unternehmen. Die Nationale Forschungsdateninfrastruktur (NFDI<sup>3</sup>) für das deutsche Wissenschaftssystem wird unter Koordination des NFDI-Vereins aufgebaut von – im Endausbau – etwa 30 Konsortien, an denen überwiegend Hochschulen und Forschungsinstitute beteiligt sind.

Das Projekt „FAIR Data Spaces“ hat als **Gesamtziele**,

Z1: Synergien der Zusammenarbeit der beiden Initiativen Gaia-X und NFDI zu identifizieren und zu heben,

Z2: durch die Definition und Etablierung eines Vorgehensmodells die Klärung rechtlicher und ethischer Fragestellungen und die Bereitstellung technischer Grundlagen eine inhaltliche Verschränkung der Initiativen zu erreichen, und

Z3: den souveränen Austausch von Daten zwischen Wirtschaft und Wissenschaft sowohl national als auch in der EU in konkreten Anwendungen und Arbeitsgebieten zu fördern.

Aus den Gesamtzielen leiten sich folgende **Unterziele** ab, um Prinzipien für die Entwicklung von gemeinsamen Diensten herauszuarbeiten:

Z1.1: zu Beginn des Projekts die Zusammenarbeit der beiden Initiativen Gaia-X und NFDI in einem Roadmapping-Prozess zu konzeptionieren,

Z1.2: sodann den Wirkungsbereich der Initiativen durch den Aufbau einer gemeinsamen Community zu vergrößern und zu stärken,

Z2.1: die Entwicklung und das Deployment der Dienste und Demonstratoren von vornherein auf ein Multi-Cloud-Umfeld zu konzeptionieren,

Z2.2: die Kompatibilität von Diensten anhand von ersten Beispielen zu demonstrieren,

Z2.3: den rechtssicheren Datenaustausch gemäß der FAIR-Prinzipien zu erproben,

Z2.4: die Einhaltung von Compliance-Fragen prototypisch umzusetzen,

Z3.1: Mehrwerte des Datenaustauschs für eine große Bandbreite von Akteuren insbesondere aus der Wirtschaft zu schaffen,

---

<sup>2</sup> Siehe <https://gaia-x.eu/>

<sup>3</sup> Siehe <https://www.nfdi.de/>

Z3.2: den Austausch von Daten zwischen Wissenschaft und Wirtschaft sowohl technisch durch Demonstratoren zeigen als auch durch Aufbau von Kompetenzen an der Schnittstelle beider Welten fördern,

Z3.3: die Empfehlung des EOSC Interoperability Frameworks<sup>4</sup> der European Open Science Cloud in diesem Kontext auf Anwendbarkeit zu überprüfen, sowie

Z3.4: sich mit den 9 EU-weiten gemeinsamen Datenräumen zu verlinken, die als Teil der Europäischen Datenstrategie (Europäische Kommission 2020) umgesetzt werden, und gleichzeitig ihre Entwicklung zu unterstützen.

Schließlich soll über die Projektlaufzeit hinaus eine Grundlage für einen gegenseitigen Austausch von Forschungsdaten und für gemeinsame Analysemöglichkeiten von Daten zwischen Wirtschaft und Wissenschaft geschaffen werden.

## 1.2. Bezug des Vorhabens zu den förderpolitischen Zielen

Seit 2014 berät der von der Gemeinsamen Wissenschaftskonferenz (GWK) eingesetzte Rat für Informationsinfrastrukturen (RfII) die Politik bei der koordinierten Weiterentwicklung des Informationsinfrastrukturaufbaus. Eine zentrale Empfehlung des RfII<sup>5</sup> war die Einrichtung einer Nationalen Forschungsdateninfrastruktur (NFDI). Die mögliche Ausgestaltung und Struktur von NFDI wurde vom RfII auch unter Einbeziehung internationaler Erfahrungen in einer Reihe von Diskussionspapieren beleuchtet. Eines der wichtigsten Ergebnisse dieser RfII-Analyse war, dass NFDI eine wissenschaftsgeleitete Initiative sein soll.

Gaia-X wurde 2019 von den Wirtschaftsministerien von Deutschland und Frankreich im Rahmen einer gemeinsamen Roadmap für Künstliche Intelligenz<sup>6</sup> initiiert, von vornherein mit dem Ziel, weitere Akteure in Europa zu gewinnen. Mit der bereits erwähnten Europäischen Datenstrategie im Blick (Europäische Kommission 2020) soll insbesondere ein Gegenentwurf zu den großen Cloud-Plattformbetreibern (Hyperscalers) aus den USA und China geschaffen werden, der europäischen Werten folgt. Eine wichtige Grundlage bildet die in Deutschland mit Fördermitteln des BMBF seit 2015 angeschobene Technologie „International Data Spaces“<sup>7</sup>, die ursprünglich unter dem Titel „Industrial Data Space“ einen Datenraum für die Industrie 4.0 schaffen sollte, inzwischen aber auf zahlreichen weiteren Gebieten zum Einsatz kommt.

Schon die KI-Strategie der Bundesregierung von 2018 verspricht, „Anreize und Rahmenbedingungen für das freiwillige, datenschutzkonforme Teilen von Daten [zu] verbessern sowie den Aufbau einer vertrauenswürdigen Daten- und Analyseinfrastruktur einschließlich des Aufbaus einer zugrundeliegenden Cloud-Plattform mit skalierbarer Speicher- und Rechenkapazität voran[zum]treiben“; insbesondere werden auch „Datenpartnerschaften zwischen Unternehmen sowie mit Forschungseinrichtungen“ genannt (Bundesregierung 2018). Um die Datensouveränität, Datenkompetenz und Dateninfrastrukturen in

---

<sup>4</sup> Siehe <https://www.eoscsecretariat.eu/news-opinion/achieving-interoperability-eosc-interoperability-framework>

<sup>5</sup> Siehe <http://www.rfii.de/?p=1998>

<sup>6</sup> Roadmap for a Research and Innovation Network on Artificial Intelligence between the Governments of the French Republic and the Federal Republic of Germany. [https://www.data-infrastructure.eu/GAIX/Redaktion/EN/Downloads/roadmap-research-and-innovation-network-on-artificial-intelligence.pdf?\\_blob=publicationFile&v=2](https://www.data-infrastructure.eu/GAIX/Redaktion/EN/Downloads/roadmap-research-and-innovation-network-on-artificial-intelligence.pdf?_blob=publicationFile&v=2)

<sup>7</sup> Siehe <https://www.internationaldataspaces.org/>

Deutschland und der EU weiter zu stärken, wurde im Januar 2021 die Datenstrategie<sup>8</sup> der Bundesregierung verabschiedet. Die Datenstrategie zielt durch die Nachnutzbarkeit von Daten einerseits auf einen Innovationsschub in der Wirtschaft, andererseits auf eine Stärkung des Wissenschaftsstandorts Deutschland. Gaia-X und NFDI sind zentrale Teile dieser Datenstrategie der Bundesregierung. Weitere Bereiche sind die Einrichtung und Erprobung von Datentreuhandmodellen, der Aktionsplan Forschungsdaten zur Förderung des Teilens von Forschungsdaten sowie die Nationale Digitale Bildungsoffensive, um die Datenkompetenz in der breiten Bevölkerung zu erhöhen.

Der Wissenschaftsrat gibt in seinem Positionspapier „Zum Wandel in den Wissenschaften durch datenintensive Forschung“<sup>9</sup> Beispiele für datenintensive Forschungen, untersucht die Dimensionen des Wandels und gibt schließlich Empfehlungen in Form von 8 Leitlinien. NFDI adressiert diese Leitlinien, indem z.B. der Kulturwandel hin zum Data Sharing befördert wird, gemeinsame Standards für den Austausch von Forschungsdaten vereinbart werden und Kompetenzen für den Umgang mit Daten und Methoden geschärft werden.

Die Deutsche Forschungsgemeinschaft (DFG) befasst sich in einem Impulspapier mit dem Thema „Digitaler Wandel in den Wissenschaften“<sup>10</sup>. Das Papier beschreibt die Merkmale des digitalen Wandels in den Wissenschaften, benennt die Auswirkungen auf die Wissenschaften und gibt einen Überblick zu den Handlungsfeldern der DFG zur Gestaltung des digitalen Wandels. Als wesentliches Gestaltungsmerkmal wird hervorgehoben, dass Daten aus der Forschung, wo immer möglich, nach den FAIR-Prinzipien (Wilkinson et al. 2016) zugänglich gemacht werden sollen.

### 1.3. Wissenschaftliche und technische Arbeitsziele des Vorhabens

Das Vorhaben befasst sich mit dem Datenaustausch aus mehreren Perspektiven: aus der ethischen und rechtlichen Perspektive, aus der technischen Perspektive von Architektur und Implementation sowie aus der praktischen Perspektive der Umsetzung in konkreten Szenarien und des Aufbaus einer Community.

Schon der Aufbau von Dateninfrastrukturen für Wissenschaft oder Wirtschaft allein stellt eine erst teilweise gelöste Herausforderung dar (Otto & Jarke 2019). Wir haben also herausfordernde wissenschaftliche Fragestellungen zu bewältigen und werden diese konkret in den in Abschnitt 3 erläuterten Arbeitspaketen angehen:

- Wie muss das Design einer gemeinsamen Dateninfrastruktur für Wissenschaft und Wirtschaft aussehen, um Netzwerkeffekte systematisch auszunutzen, so dass exponentielles Wachstum gefördert wird? (AP1 und AP3)
- Wie sollte bei datenübergreifender Nutzung von forschungsorientierten Dateninfrastrukturen, wie NFDI, und wirtschaftsorientierten Dateninfrastrukturen, wie Gaia-X, der Begriff Forschung ausgehend von der datenschutzrechtlichen Perspektive verstanden werden? (AP2)
- Wie behandeln die Fachbereiche Datenschutz und Ethik das Thema der Forschungsdaten und wie sollte mit möglichen Unterschieden (aber auch Gemeinsamkeiten) in der Beurteilung umgegangen werden? Wie sollte man vorgehen, wenn der Spielraum im Datenschutz größer ist als der in der Ethik bzw. umgekehrt? (AP2)

---

<sup>8</sup> Siehe <https://www.bildung-forschung.digital/de/datenstrategie-der-bundesregierung-3616.html>

<sup>9</sup> Siehe [https://www.wissenschaftsrat.de/download/2020/8667-20.pdf?\\_\\_blob=publicationFile&v=5](https://www.wissenschaftsrat.de/download/2020/8667-20.pdf?__blob=publicationFile&v=5)

<sup>10</sup>

[https://zenodo.org/record/4191345/files/20201028\\_Digitaler\\_Wandel\\_in\\_den\\_Wissenschaften\\_DFG-Impulspapier\\_de.pdf?download=1](https://zenodo.org/record/4191345/files/20201028_Digitaler_Wandel_in_den_Wissenschaften_DFG-Impulspapier_de.pdf?download=1)

- Wie kann man die Daten, die übergreifend mittels Verknüpfung von NFDI und Gaia-X ausgetauscht werden, immaterialgüterrechtlich einordnen? Welche verschiedenen Schutzsysteme kommen in Bezug auf diese Daten zur Anwendung (Urheberrecht, Datenbankherstellerrecht, Geheimnisschutz-Gesetz, Patentrecht)? Bestehen beispielsweise Datenbankrechte nach § 87a Urheberrechtsgesetz? (AP2)
- Wie kann eine kombinierte Dateninfrastruktur (bestehend aus NFDI- und Gaia-X-Diensten) für europäische und außereuropäische Gaia-X-Anwender zugänglich gemacht werden und wie verhält sich diese Infrastruktur zu den (weiteren) europäischen Datenräumen? Welchen Einfluss haben die neue unionale Regelungsstruktur und die neuen europäischen Infrastrukturvorgaben auf die Verknüpfung und Vernetzung? (AP2)
- Wie können die zentral definierte Architektur von Gaia-X und die einzeln pro NFDI-Konsortium definierten Architekturen zu einer gemeinsamen Referenzarchitektur kombiniert werden? (AP3)
- Wie können vorhandene Dienste aus unterschiedlichen NFDI-Konsortien möglichst wiederverwendbar für den Einsatz in einer gemeinsamen Dateninfrastruktur für Wissenschaft und Wirtschaft bereitgestellt werden? (AP3)
- Wie können Werkzeuge zur Nachnutzung der Daten so angepasst und erweitert werden, dass eine skalierbare Anwendbarkeit auf großen heterogenen Datenbeständen in unterschiedlichen Cloud-Infrastrukturen und Multi-Cloud-Infrastrukturen gewährleistet wird. (AP4)
- Wie kann die Gaia-X Infrastruktur genutzt werden, um Qualitätssicherungs- und Analyseworkflows zu kuratieren, wiederzuverwenden und für Nutzende zur Verfügung zu stellen und wie kann ein wissenschaftsgeleitetes Bewirtschaftungskonzept für das Scale-Out dafür realisiert werden? (AP4)
- Wie können verteilte Analysen durchgeführt werden, welche verschiedene Dienste von NFDI-Konsortien und Gaia-X kombinieren? (AP4)

Technisch betrachtet, geht es überwiegend darum, Dienste, die in einzelnen NFDI-Konsortien schon ansatzweise realisiert wurden, für den Einsatz über NFDI hinaus zu verallgemeinern und nutzbar zu machen. Die Nutzung solcher Dienste in wirtschaftsnahen Szenarien soll insbesondere dadurch demonstriert werden, dass sie mit Gaia-X-konformen Schnittstellen versehen und somit mit den Gaia-X Federation Services für Identity & Trust, Federated Catalogue, Sovereign Data Exchange und Compliance interoperabel werden, die den Kern eines jeden auf Gaia-X-Technologie basierenden Datenraums bilden. Dazu wird auf vorhandenen Infrastrukturen für grundlegende Cloud-Technologie aufgesetzt. Dazu zählen insbesondere de.NBI (Deutsches Netzwerk für Bioinformatik-Infrastruktur<sup>11</sup>) als Angebot für die Wissenschaft und die Open Telekom Cloud<sup>12</sup> als Angebot für die Wirtschaft. Die Erprobung wird in mehreren Schichten eines Technologie-Stacks realisiert: Aufbauend auf einer Storage-Infrastruktur werden Dienste bereitgestellt für Compliance & FAIRness Monitoring, FAIR Identity Spaces (d.h. Authentifizierung, Autorisierung und Delegation), (Meta)data Terminologies & Validation sowie für Verteilte Analysen (vgl. Abbildung 1). Abbildung 2 zeigt, an welcher Stelle in ihrem Lebenszyklus Daten von einem bestimmten Dienst verarbeitet werden.

---

<sup>11</sup> Siehe <https://www.denbi.de/cloud>

<sup>12</sup> Siehe <https://open-telekom-cloud.com/>

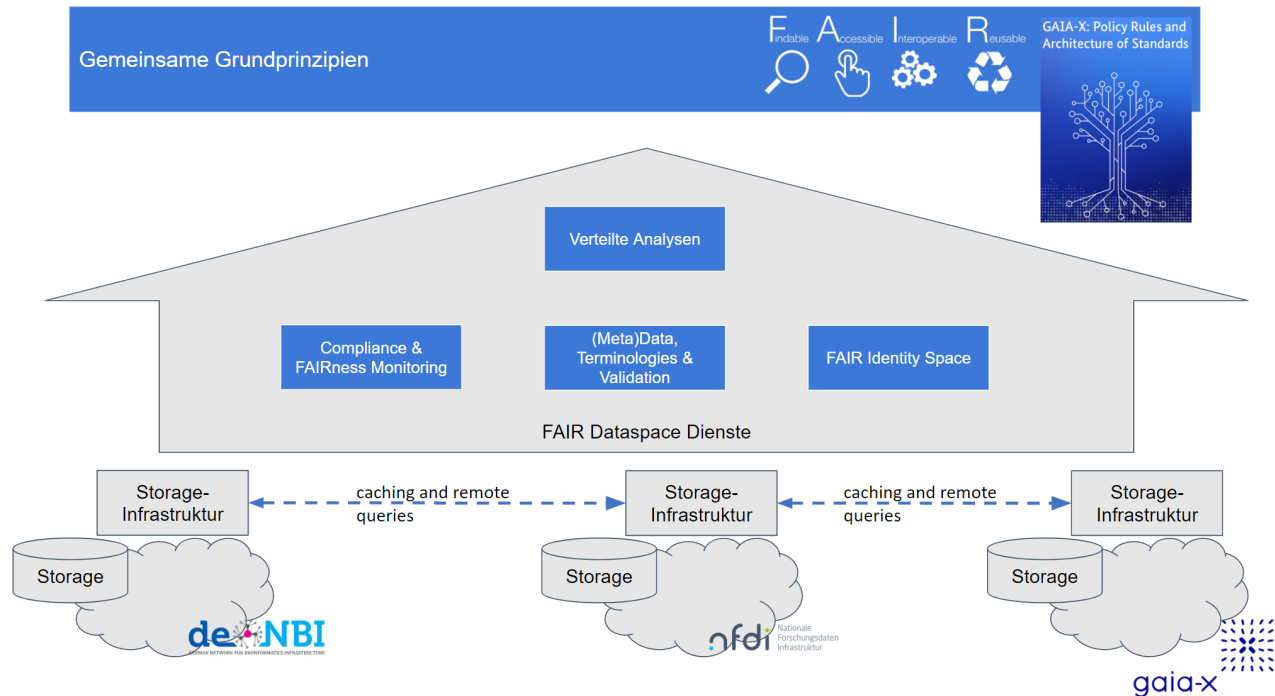


Abbildung 1: Schematische Architektur eines FAIR Data Space (vgl. AP3)

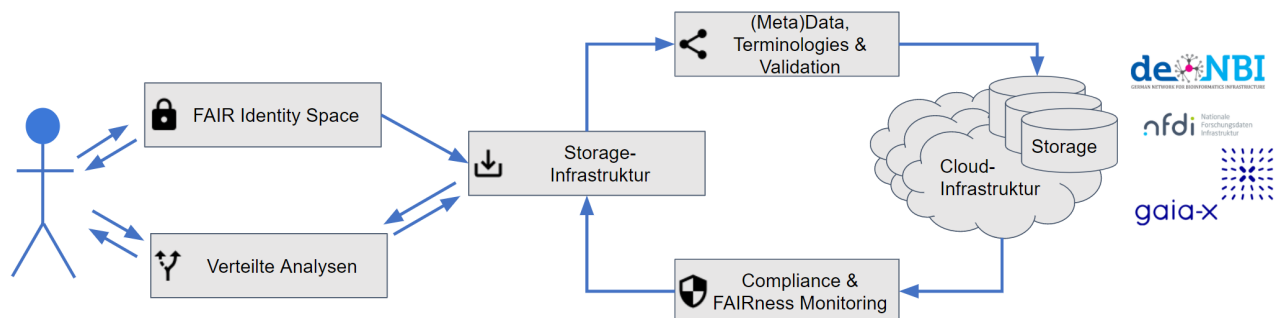


Abbildung 2: Zusammenspiel der Dienste entlang des Daten-Lebenszyklus

Die in AP4 erarbeiteten Demonstratoren für konkrete Datenaustausch-Szenarien sollen zum Ende der ersten Projektphase bis Ende 2021 die grundsätzliche Machbarkeit des Ansatzes zeigen. Dazu werden im NFDI-Umfeld bereits vorhandene Dienste in Gaia-X-konforme Schnittstellen verpackt. Als Grundlage auf Gaia-X-Seite dienen hierbei die bis Mitte 2021 dort veröffentlichten Dokumente: die dann aktuelle Neuausgabe der Technical Architecture (BMWi 2020a) und die erste Spezifikation der Federation Services<sup>13</sup>. Deren erste Implementation wird jedoch erst Ende 2021 fertig und fließt somit in der zweiten Projektphase in die Demonstratoren ein. Ebenso werden die Demonstratoren in der zweiten Phase die Arbeiten zu den rechtlichen Aspekten berücksichtigen, die in der ersten Phase noch nicht technisch umgesetzt werden.

Abbildung 3 ordnet schematisch das Zusammenspiel von Infrastruktur- und Datendiensten im Falle des NFDI4Biodiversity-Demonstrators in die Architektur eines Gaia-X-basierten Datenraums ein, der ein

<sup>13</sup> Siehe <https://www.eco.de/presse/gaia-x-federation-services-eco-uebernimmt-projektmanagement/>



Ökosystem für Speicher- und Netzwerkinfrastruktur umfasst und eines für Daten und datenbasierte Dienste.

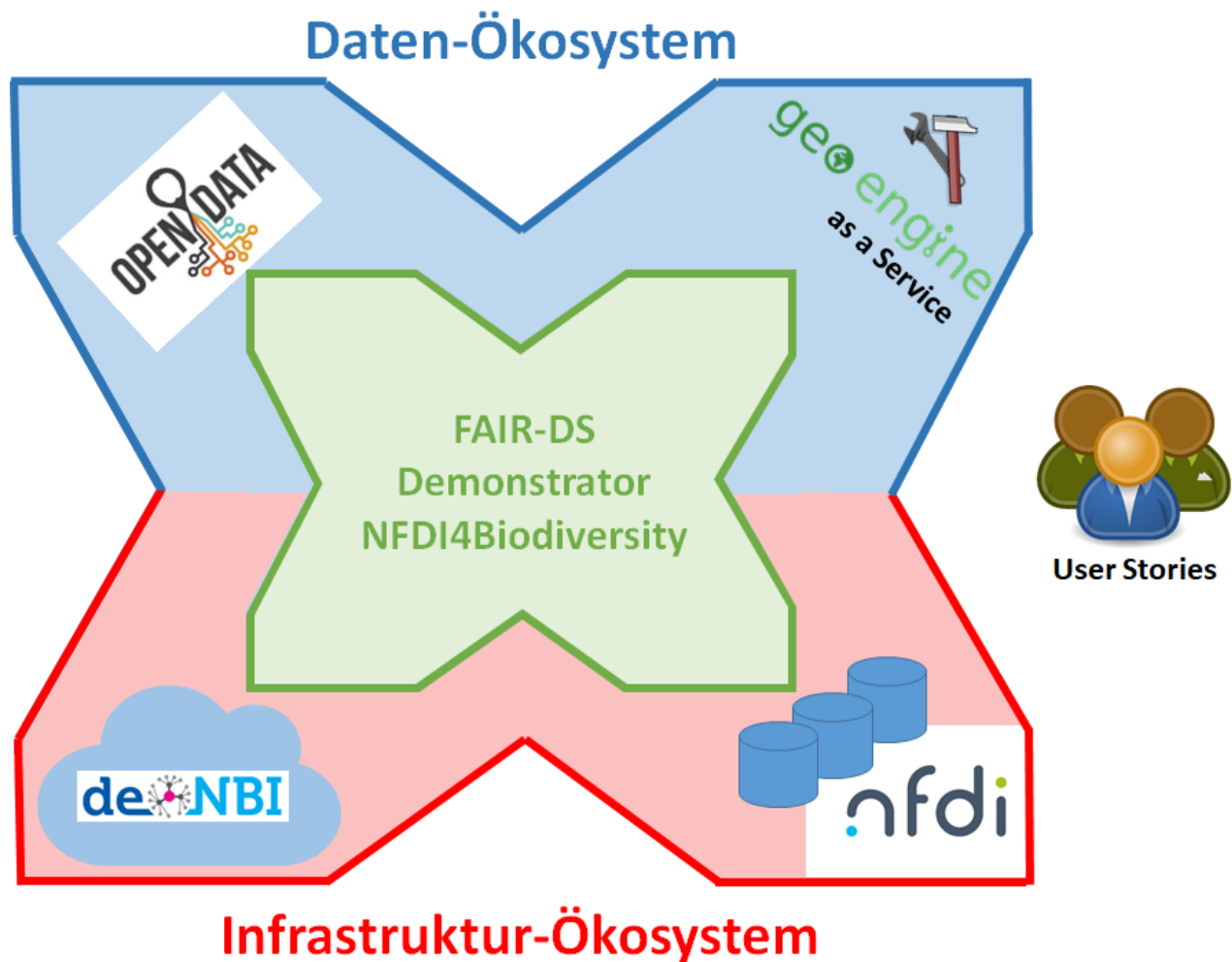


Abbildung 3: Zusammenspiel von Infrastruktur- und Daten-Diensten im Falle des NFDI4Biodiversity-Demonstrators (AP4.1)

## 2. Stand der Wissenschaft und Technik; bisherige Arbeiten

### 2.1. Stand der Wissenschaft und Technik (einschließlich alternativer Lösungen, der Ergebnisverwertung entgegenstehende Rechte, Informationsrecherchen)

Die Wissenschaft verändert sich durch den digitalen Wandel grundlegend und mit einer rasanten Geschwindigkeit. Forschungsdaten digital zu erheben, zu analysieren und langfristig verfügbar zu machen ist ins Zentrum vieler Einzelvorhaben gerückt. Es gibt bereits eine Reihe von nationalen (z.B. Länderinitiativen zum Forschungsdatenmanagement, DINI, Nestor, Projekte des DFG-Programms Wissenschaftliche Literaturversorgungs- und Informationssysteme<sup>14</sup>) und von internationalen (z.B.

<sup>14</sup> Siehe <https://www.dfg.de/foerderung/programme/infrastruktur/lis/>

European Open Science Cloud EOSC, Research Data Alliance RDA, GO FAIR, EUDAT) Projekten und Initiativen im Bereich des Forschungsdatenmanagements.

Der Rat für Informationsinfrastrukturen (RfII) hat die Landschaft 2016 untersucht und hat sich auf Basis der Analyse für die Empfehlung der Einrichtung einer Nationalen Forschungsdateninfrastruktur ausgesprochen. Wie der RfII in seiner Empfehlung heraus arbeitete, gibt es zwei Seiten der Medaille: positiv gesprochen existiert eine „Vielfalt“, von der durch Erfahrungsaustausch profitiert werden kann, negativ gesprochen existiert eine „Zersplitterung“, die eine gemeinsame Nachnutzung von Forschungsdaten massiv erschwert. Erst mit dem Aufbau des Vereins Nationale Forschungsdateninfrastruktur (NFDI) e.V. wird dem rasanten Wandel strukturell angemessen auf nationaler Ebene Rechnung getragen, wird die Vielfalt als positiv verstärkendes Element aufgegriffen und wird der Zersplitterung entgegengewirkt.

Zu Gaia-X sind bisher Konzeptpapiere und erste Demonstratoren<sup>15</sup> entstanden. Weitere Implementierungen folgen ab 2021. Mit wissenschaftlichen Arbeiten ist erst im Zuge der 2021 beginnenden Forschungsprojekte mit Gaia-X-Bezug zu rechnen. Wissenschaftliche und technische Ergebnisse, die in weiten Teilen eine Grundlage für Gaia-X bilden, sind im Rahmen der International Data Spaces (IDS) entstanden, überwiegend unter Beteiligung des Verbundpartners Fraunhofer-Gesellschaft (siehe deshalb auch Abschnitt 2.3). Gaia-X strebt nach dem Prinzip der „Architecture of Standards“ (BMW 2020) an, zuvor etablierte Standards zu unterstützen bzw. auf ihnen aufzubauen. Dazu zählen insbesondere das IDS-Referenzarchitekturmodell (Otto et al. 2019) und die ihm zugrunde liegenden Standards etwa des World Wide Web Consortium W3C. Ein aktuelles Positionspapier zeigt die Gemeinsamkeiten von Gaia-X und IDS aus IDS-Sicht auf (Otto et al. 2021). Die meisten Kernkomponenten des IDS, zu großen Teilen von der Fraunhofer-Gesellschaft mit BMBF-Förderung in den Projekten InDaSpace (Förderkennzeichen 01IS15054) und InDaSpacePlus (01IS17031) entwickelt, sind inzwischen unter dem Dach der IDS Association als Open Source veröffentlicht worden<sup>16</sup>. Die Implementation der Gaia-X Federation Services wird ebenfalls Open Source sein.

## 2.2. Bestehende Schutzrechte

In einzelnen NFDI-Konsortien und konkreten auf Gaia-X- oder IDS-Prinzipien basierenden Datenräumen kommt Technologie zum Einsatz, die Schutzrechten Dritter unterliegt. Diese Komponenten fließen jedoch nicht in die Grundlagen und Demonstratoren des FAIR Data Space ein. Stattdessen kommen Open-Source-Implementierungen der Verbundpartner oder Dritter zum Einsatz.

## 2.3. Bisherige Arbeiten der Antragsteller

Die nachfolgende Tabelle gibt einen Überblick, an welchen Konsortien bzw. übergreifenden Initiativen auf Seiten von NFDI bzw. Gaia-X die Verbundpartner beteiligt sind. Die weiteren Absätze stellen zunächst NFDI-Konsortien und weitere Initiativen vor, an denen mehrere Verbundpartner beteiligt sind und deren bisherige Arbeiten zentral in das Vorhaben einfließen; schließlich werden einschlägige Vorarbeiten weiterer einzelner Partner vorgestellt.

---

<sup>15</sup> Siehe <https://dev-portal.gaia-x-demonstrator.eu/>

<sup>16</sup> Siehe <https://github.com/International-Data-Spaces-Association>

| Konsortium/<br>Initiative | Beteiligte Partner |                   |     |     |     |    |      |     |      |      |      |     |     |      |     |      |
|---------------------------|--------------------|-------------------|-----|-----|-----|----|------|-----|------|------|------|-----|-----|------|-----|------|
|                           | FhG                | NFDI              | JLU | TIB | UzK | UL | RWTH | FIZ | UKHD | HAdW | EMBL | UMR | WWU | EKUT | ZBW | Atos |
| NFDI4Ing                  |                    | (x) <sup>17</sup> | X   | X   |     |    | X    |     |      |      |      |     |     |      |     |      |
| NFDI4Chem                 |                    | (x)               |     | X   |     |    | X    | X   |      |      |      |     |     |      |     |      |
| NFDI4Health               | X                  | (x)               |     |     | X   | X  |      |     |      |      |      |     |     |      |     |      |
| NFDI4Culture              |                    | (x)               | X   | X   |     |    |      | X   |      |      |      | X   |     |      |     |      |
| GHGA                      |                    | (x)               |     |     | X   |    |      |     | X    | X    | X    |     |     | X    |     |      |
| NFDI4BioDiversity         |                    | (x)               | X   |     |     |    |      |     |      |      |      | X   |     |      |     |      |
| KonsortSWD                |                    | (x)               |     |     |     |    |      |     |      |      |      |     |     |      | X   |      |
| MII                       | X                  |                   |     |     | X   | X  | X    |     |      |      |      |     |     | X    |     |      |
| Gaia-X Association        | X                  |                   |     |     |     |    |      |     |      |      |      |     |     |      |     | X    |
| IDS Association           | X                  |                   |     |     |     |    |      |     |      |      |      |     |     |      |     | X    |

## NFDI4Ing (vertreten durch RWTH, TIB, JLU)

NFDI4Ing<sup>18</sup> nutzt den methoden- und nutzerorientierten Ansatz um Forschungsdaten aus den Ingenieurwissenschaften FAIR zu machen. Dafür definiert das Konsortium Archetypen, die Forschende nach ihren Forschungsmethoden unterscheiden. Auf die Anforderungen der Archetypen kann so individuell innerhalb der Basisdienste eingegangen werden. Diese Arbeit wird durch Community Cluster unterstützt, die die Blickwinkel der ingenieurwissenschaftlichen Fachdisziplinen einnehmen. Seit 2017 konnten innerhalb des Konsortiums mehrere IT-Dienste, Arbeitsgruppen und Workshops etabliert werden um auf die vielseitigen Bedarfe und Anforderungen zu reagieren.

Innerhalb der RDA setzt sich NFDI4Ing für die deutsche Beteiligung an der „Interest Group Research Data Management in Engineering“ ein, um diese offene Plattform für Koordination und Kommunikation mit der weltweit vernetzten ingenieurwissenschaftlichen Community zu nutzen.

Mit dem Ziel Wege zu finden, um die Interoperabilität zwischen disziplinspezifischen Metadaten und Terminologien sicherzustellen hat die Arbeitsgruppe Metadata4Ing<sup>19</sup> Workshops für die ingenieurwissenschaftliche Community organisiert und über NFDI auch die Abstimmung mit anderen Fachdisziplinen vorangetrieben. Die Arbeitsgruppe besteht aus Ingenieuren und Experten für Metadaten. Metadata4Ing kann auf Erfahrungen und Ideen zurückgreifen, die in Projekten bei den Konsortialpartnern entwickelt wurden. Um den gewünschten Grad an Interoperabilität zu erreichen, umfasst die Arbeitsgruppe auch Teilnehmer aus Institutionen außerhalb von NFDI4Ing und ist noch offen für weitere Partner.

Im Rahmen des Konsortiums wird ein zentrales Versionierungssystem, GitLab, zur Verfügung gestellt und darauf aufbauend gemeinsame Schulungsangebote und Nutzungskonzepte für Forschende entwickelt. Grund für diese Entscheidung war die Beobachtung, dass in verschiedenen Abteilungen eine Vielzahl von

<sup>17</sup> NFDI e.V. ist an keinem NFDI-Konsortium als Partner beteiligt, hat aber die Aufgabe, alle Konsortien zu koordinieren.

<sup>18</sup> Siehe <https://nfdi4ing.de/>

<sup>19</sup> Siehe <https://nfdi4ing.de/projects/metadata4ing/>

auf die Quellcode- und Dokumentenverwaltung zugeschnittenen Installationen existieren. Als zentralen Service bietet NFDI4Ing daher eine GitLab-Instanz, die allen Nutzenden innerhalb des Konsortiums über die DFN-AAI Föderation zugänglich ist.<sup>20</sup>

### **Medizininformatik Initiative (MII; vertreten durch UL, UzK, EKUT, FhG, RWTH)**

Im Rahmen der Medizininformatik-Initiative fördert das BMBF mehrere Konsortien, zu denen sich mehrere Universitätskliniken mit weiteren Partnern wie Forschungsinstituten, Hochschulen, Unternehmen oder nicht-universitären Krankenhäusern zusammengeschlossen haben. Die Konsortien arbeiten gemeinsam daran, die Voraussetzungen zu schaffen, um Daten aus Forschung und Patientenversorgung untereinander austauschen zu können. Im Fokus stehen zunächst die Universitätskliniken, weil hier die engste Verbindung zwischen Krankenversorgung und klinischer Forschung besteht. An den Konsortien sind jeweils mehrere Universitätskliniken sowie weitere Partner wie Forschungsinstitute, Hochschulen, Unternehmen oder nicht-universitäre Krankenhäuser beteiligt.

In der ersten Phase wurden vier Konsortien ausgewählt und gefördert: SMITH, DIFUTURE, HIGHMED und MIRACUM. Ausgewählte Partner dieses Antrags sind in diese Konsortien involviert. Innerhalb dieser Konsortien werden vorerst an jedem Universitätsklinikum sogenannte Datenintegrationszentren aufgebaut und miteinander vernetzt. In diesen Zentren werden die technischen und organisatorischen Voraussetzungen für einen standortübergreifenden Datenaustausch zwischen Krankenversorgung und klinischer/biomedizinischer Forschung geschaffen. Dabei kommt den Aspekten der Interoperabilität, dem Data Sharing und dem Datenschutz eine Bedeutung zu. Aus Sicht der Interoperabilität werden aktuelle Standards aus dem medizinischen Bereich verwendet. Dazu zählen bspw. Integrated Healthcare Environment (IHE), das die Interaktion zwischen verschiedenen Informationssystemen in einer und zwischen Organisationen in Form von Profilen beschreibt. Darüber hinaus wurde von der MII das hierarchisch organisierte FHIR (Fast Healthcare Interoperable Resource) Format für den Datenaustausch präferiert. Es bietet gleichsam die Möglichkeit der semantischen Interoperabilität, z.B. mit der Nutzung von Terminologien. Dazu steht SNOMED CT für die medizinische Forschung in Deutschland zur Verfügung. Darüber hinaus etabliert die MII mit der Zentralen Antrags- und Registerstelle (ZARS) eine Infrastruktur, die einen zentralen Zugang auf die Daten ermöglicht. Dem Zugang geht eine Beantragung voraus. Der Zugriff kann auch unter Nutzung verteilter Analyseinfrastrukturen, wie z.B. dem Personal Health Train und DataSHIELD erfolgen.

### **GHGA – German Human Genome Phenome Archive (vertreten durch UKHD, HAdW, EMBL, EKUT)**

Ziel des NFDI-Konsortiums GHGA<sup>21</sup> ist der Aufbau eines Genomarchivs als nationaler, förderter Knoten des Europäischen Genom-Phänom-Archiv (EGA). GHGA wird die sichere Speicherung, den Zugriff und die Analyse menschlicher Omics-Daten (z.B. Genome, Transkriptome) in einem einheitlichen ethisch-rechtlichen Rahmen ermöglichen. Dies erfolgt vor dem Hintergrund, dass die momentane zentrale EGA-Infrastruktur die spezifischen nationalen Regelungen zum Datenschutz nur ungenügend abbilden kann. Daher ist eine Umwandlung in eine förderierte Infrastruktur aus nationalen Knoten („förderiertes

<sup>20</sup> Siehe <https://git.rwth-aachen.de/nfdi4ing/>

<sup>21</sup> Siehe <https://ghga.dkfz.de/>

EGA“) geplant. Als Teil der Nationalen Forschungsdateninfrastruktur integriert GHGA dabei breite Teile der Genomforschungsgemeinde in Deutschland inklusive der großen Genomsequenzierzentren. GHGA wird Forschenden in Deutschland in die Lage versetzen, humane Genomdaten rechtskonform entsprechend der FAIR-Richtlinien auszutauschen und für die Forschung zu nutzen. Zudem kann GHGA dabei internationale Standards zum Datenaustausch stärker mitgestalten. GHGA ist dabei eingebunden in flankierende internationale Forschungsnetzwerke wie etwa die europäische B1MG-Initiative.

Das UKHD bringt mit der Sektion Translationale Medizin-Ethik jahrelang gewachsene Expertise zu ethischen und rechtlichen Begleitung von Daten-Initiativen mit einem Fokus auf die Förderung von Data-Sharing von personenbezogenen und sensiblen Daten ein. Die Leiterin, Prof. Dr. Eva Winkler, ist im Direktorium von GHGA für ELSA-Fragen (Winkler & Knoppers 2020), in der AG Consent des Nationalen Steuerungsgremiums der Medizininformatik Initiative und Sprecherin des EURAT-Konsortiums zu ethischen und rechtlichen Fragen der Genomsequenzierung (Bartram et al. 2013).

Die Heidelberger Akademie der Wissenschaften (HAdW) bringt rechtswissenschaftliche Expertise zur Verarbeitung (Pictor et al. 2019, Molnar-Gabor 2021, Molnar-Gabor 2021a) und zum Austausch sensibler personenbezogener Daten (Molnar-Gabor 2018, Molnar-Gabor & Korbelt 2020), insbesondere zu den grenzüberschreitenden Aspekten auf unionaler und internationaler Ebene (Molnar-Gabor 2019a).

Die Arbeitsgruppe von Dr. Molnar-Gabor an der HAdW zeichnet sich durch langjährige Erfahrung in der juristischen Bearbeitung von Dateninfrastruktur-Vorhaben vor allem in der Omics-Forschung aus, die unter Beteiligung unterschiedlicher Länder und Akteure aufgebaut wurden (Philips et al. 2020).

Dr. Jan Korbelt vom Europäischen Laboratorium für Molekularbiologie (EMBL) ist im Direktorium von GHGA und bringt langjährige Expertise im nationalen und internationalen Datenaustausch und Omics-Forschung (z.B. als Gründer und Mitglied des Lenkungsausschusses des internationalen Pan-Cancer Analysis of Whole Genomes Projektes) in das beantragte Konsortium ein.

Prof. Dr. Oliver Kohlbacher ist einer der Mitinitiatoren von GHGA und Direktor des Instituts für Biomedizinische Informatik an der Universität Tübingen. Seine Arbeitsgruppe hat langjährige Erfahrung mit der Implementierung von Dateninfrastrukturen und der verteilten Analyse von (Multi-)Omics-Daten insbesondere im biomedizinischen Kontext.

## **NFDI4Health (vertreten durch UL, RWTH, FhG)**

Das Ziel von NFDI4Health<sup>22</sup> ist die Verschmelzung von epidemiologischer, Public Health- und klinischer Forschung: Ein multidisziplinäres Team aus Wissenschaftlerinnen und Wissenschaftlern wird in Deutschland eine Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten aufbauen. Die Mission ist die Wertsteigerung der Forschung in den Bereichen Epidemiologie, Gesundheitswesen und klinische Studien. Dazu werden hochwertige Daten nach den FAIR-Prinzipien international zugänglich gemacht. NFDI4Health bietet eine vollständige Abdeckung großer epidemiologischer Studien, der Public-Health-Forschung und von Prüfarzten und -ärztinnen initiiertes klinischer Studien in Deutschland sowie die gemeinsame Entwicklung von NFDI4Health mit der Nutzergemeinde.

Aufbauend auf den Entwicklungen und Erfahrungen von vorangegangenen Projekten der beteiligten Partner, z.B. Universität Leipzig, RWTH, HITS gGmbH, BIH und BIPS, wird eine verteilte Infrastruktur anvisiert, die die einzelnen Datengeber (z.B. Studienzentren), sogenannte *local access points*, miteinander

---

<sup>22</sup> Siehe <https://www.nfdi4health.de/>

verbindet. Das Netzwerk kann um neue Datengeber an einer beliebigen Stelle erweitert werden. Ein zentraler Zugriffsknoten (sog. *central access point*) ermöglicht, den Zugang zu den Daten zu beantragen. Dazu soll die ZARS Infrastruktur der Medizininformatik Initiative mitgenutzt werden. Datenstrukturen und -formate sind an aktuellen Standards (z.B. HL7 FHIR) ausgerichtet, die übergreifend Verwendung finden. Die verteilte Datenanalyse wird mit Hilfe des Personal Health Train (Beyan et al. 2020) und DataSHIELD abgedeckt.

## **NFDI4BioDiversity (vertreten durch JLU, UMR)**

Biodiversität ist mehr als nur die Betrachtung und Erfassung der Vielfalt der Arten auf unserem Planeten. Biodiversität umfasst die genotypische, phänotypische, funktionelle Diversität und Interaktion von Arten, Populationen und Ökosystemen. Der erschütternde Bericht des Weltbiodiversitätsrates (IPBES) hat jüngst bestätigt, dass das Artensterben mit bisher unabsehbaren Konsequenzen für die Menschheit rasant voranschreitet. Die drängenden wissenschaftlichen und gesellschaftlichen Fragen zu den Veränderungen in der Biodiversität können nur durch das Zusammenführen von fundierten Daten und Wissen aus verschiedenen Bereichen beantwortet werden. NFDI4BioDiversity hat sich sechs Projektziele gesetzt, um den Zugang zu dezentralen Daten zu ermöglichen: 1. Forschungsdatenmanagement als integralen Bestandteil der Biodiversitätsforschung fördern, 2. Daten-FAIRness ermöglichen, 3. FAIRness um Qualität erweitern, 4. NFDI4BioDiversity in die nationalen und internationalen Aktivitäten integrieren, 5. bei NFDI-weiten Querschnittsthemen zusammenarbeiten und 6. NFDI als Service-Infrastruktur gemeinsam zu steuern und diese dabei dynamisch und nachhaltig zu gestalten. Zu den Schwerpunkten im Bereich der Software-Entwicklung in NFDI4BioDiversity gehört insbesondere die Etablierung einer generischen Architektur, die sogenannte Research Data Commons (RDC). Die RDC-Architektur beinhaltet zum einen eine Cloud-basierte Speicherarchitektur zur systematischen Erfassung von Daten und Metadaten, die in diesem Projekt als eine wichtige Schnittstelle zu Gaia-X ausgebaut und weiterentwickelt werden soll. Zum anderen werden im RDC basierend auf einer semantischen Speicherebene Werkzeuge, wie z. B. die aus der Universität Marburg ausgegründete Geo Engine<sup>23</sup>, angeboten, die als Demonstrator in diesem Projekt für andere Cloud- und Multi-Cloud Infrastrukturen erweitert und adaptiert werden sollen.

## **KonsortSWD (vertreten durch ZBW)**

Der Bereich Forschungsdatenmanagement bildet bereits seit mehreren Jahren einen wichtigen Innovationsschwerpunkt in der ZBW. Dies lässt sich u.a. anhand der folgenden Aktivitäten veranschaulichen: Das international agierende Unterstützungs- und Koordinierungsbüro der GO FAIR Initiative in Deutschland ist an der ZBW eingerichtet. Es hat die Aufgabe, weltweit ein Netzwerk aus wissenschaftlichen Fachgemeinschaften und Forschungsinfrastrukturen aufzubauen, die Akzeptanz und Anwendung der FAIR-Prinzipien voranzubringen und damit den Weg für die EOSC zu ebnen. Durch ihr Engagement in der DataTogether-Initiative ist die ZBW mit den internationalen Akteuren, d.h. CODATA, WDS und RDA, im Bereich Forschungsdatenmanagement bestens vernetzt. Prof. Dr. Klaus Tochtermann, der Direktor der ZBW, ist gewähltes Mitglied des Board of Directors der EOSC Association und gestaltet in dieser Rolle die Weiterentwicklung der EOSC aktiv mit. Dort ist er insbesondere zuständig für die Identifikation von Synergien zwischen Gaia-X und der EOSC. Im Kontext von NFDI ist die ZBW für die

---

<sup>23</sup> Siehe <https://www.geoengine.de/>

Technologieentwicklung im Rahmen des Konsortiums KonsortSWD verantwortlich. Dieselbe Rolle nimmt sie in dem derzeit in Begutachtung befindlichen Konsortium BERD@NFDI ein, in dem der Fokus auf non-Standard Daten, wie z.B. Daten aus Geschäftsberichten oder Social Media Daten, gelegt wird. Mit Bezug zum Thema Forschungsdatenmanagement betreibt die ZBW aktiv Forschung im Bereich Open Science. Sie koordiniert den multidisziplinären Leibniz-Forschungsverbund *Open Science*, in dem Phänomene im Zusammenhang mit der Digitalisierung der Wissenschaft erforscht werden. Über diese Expertise ist die ZBW in zahlreiche forschungspolitische Arbeitsgruppen zum Thema eingebunden.

## **Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V. (FhG)**

Ein Verbund mehrerer Fraunhofer-Institute unter Leitung des **Fraunhofer-Instituts für Software- und Systemtechnik ISST** mit Beteiligung des **Fraunhofer-Instituts für Angewandte Informationstechnik FIT** entwickelt seit 2014 den Kern der IDS-Technologie, zum großen Teil in den BMBF-Projekten InDaSpace (2015–2018) und InDaSpacePlus (seit 2018). Einschlägige Vorarbeiten dieser beiden Institute umfassen die IDS-Referenzarchitektur (Otto et al. 2019; Otto & Jarke 2019), die Spezifikation und Implementation der für den Datenaustausch erforderlichen Connector-Schnittstelle<sup>24</sup> und des Informationsmodells, mit dem in einem Data Space Teilnehmende sich selbst sowie die von ihnen angebotenen Daten und die von ihnen betriebenen Infrastruktur-Komponenten beschreiben (Bader et al. 2020, Spiekermann et al. 2018), sowie ein 2020 begonnenes Arbeitspaket zu FAIR Data in InDaSpacePlus. ISST-Leiter Prof. Dr. Boris Otto ist stellvertretender Vorsitzender des Vorstands der IDS Association; Dr. Christoph Lange-Bever vom FIT leitet die Arbeitsgruppe „Informationsmodell“.

Fraunhofer ist Gründungsmitglied der Gaia-X Association. ISST-Leiter Prof. Dr. Boris Otto war in der Aufbauphase 2020–2021 deren CTO; Dr. Christoph Lange-Bever vom FIT leitet das zum IDS-Informationsmodell analoge Open Work Package „Self-Description“. Beide Institute sind an der Spezifikation der Gaia-X Federation Services beteiligt.

Das Fraunhofer FIT leitet im Rahmen der GO-FAIR-Initiative die Entwicklung des Personal-Health-Train-Ansatzes zur verteilten Datenanalyse (Beyan et al. 2020) und bringt diese in NFDI4Health und – über seine Kooperation mit der RWTH Aachen – in mehrere Projekte der vom BMBF geförderten Medizininformatik-Initiative (MII) ein.

## **Nationale Forschungsdateninfrastruktur (NFDI) e.V.**

Seit Gründung von Nationale Forschungsdateninfrastruktur (NFDI) e.V. im Oktober 2020 wurden erste Strukturen zur Koordination von gemeinsamen Arbeiten von Organisationen, die in DFG-geförderten NFDI-Konsortien mitwirken, geschaffen und erste gemeinsame Aktivitäten wurden durchgeführt. Dazu zählen beispielsweise die Durchführung von gemeinsamen Abstimmungs- und Strategietreffen, die Bereitstellung einer technischen Kommunikationsinfrastruktur (z.B. Mailinglisten), die Außendarstellung des Vereins durch eine Webseite sowie die Erzeugung von Reichweite für den Verein in sozialen Medien. Gleichzeitig wurden erste Verwaltungsstrukturen aufgebaut, so dass der Verein seit März 2021 als Arbeitgeber wirken kann.

---

<sup>24</sup> Siehe <https://www.dataspace-connector.io/>

## **Justus-Liebig-Universität Gießen (JLU)**

Seit August 2013 ist Alexander Goesmann als Professor für Systembiologie mit dem Schwerpunkt Genomik, Proteomik und Transkriptomik an der Justus-Liebig-Universität (JLU) Gießen tätig. Zu seiner Gruppe gehört auch die Bioinformatics Core Facility (BCF), die eine leistungsfähige IT-Bioinformatik-Infrastruktur bereitstellt. Im deutschen Netzwerk für Bioinformatik Infrastruktur (de.NBI) leitet er das Zentrum für Mikrobielle Bioinformatik (BiGi), welches an den beiden Standorten Bielefeld und Gießen verortet ist. Im Rahmen dieses Großprojekts wurde die JLU als einer von sechs Standorten zur Etablierung einer Cloud-Computing-Infrastruktur ausgewählt und es flossen aus Sondermitteln des Bundes bereits mehr als 7 Mio. Euro in den Aufbau der erforderlichen Rechnersysteme. Im Fokus der wissenschaftlichen Arbeiten steht die Bereitstellung und Weiterentwicklung von bioinformatischen Software-Lösungen und automatisierten Workflows zur Analyse des Erbguts von medizinisch und biotechnologisch relevanten Mikroorganismen, aber auch diverse eukaryotische Datensätze werden in der Professur verarbeitet und analysiert. Darüber hinaus ist die Gruppe beteiligt am Deutschen Zentrum für Lungenforschung (DZL), am Deutschen Zentrum für Infektionsforschung (DZIF), am LOEWE-Zentrum für Insektenbiotechnologie, sowie an verschiedenen Verbundprojekten der beiden Universitäten in Gießen und Marburg (u.a. LOEWE-Schwerpunkt Medical RNOMICS, SFB TRR81, GRK 2355, KFO309). Zu den weiteren Tätigkeiten gehört die aktive Beteiligung als Area Leader in die Gestaltung der Bioinformatik-Plattform (Translational Hub 2: Cardio-Pulmonary Systems Biology and Medicine) des Exzellenzclusters „Cardio-Pulmonary Institute (CPI)“. In diesem Verbundprojekt steht zumeist die Analytik eukaryotischer und insbesondere auch humaner omics-Datensätze (u.a. aus Einzelzell-Sequenzierungen) im Vordergrund. In all den vorgenannten Projekten ist die Gruppe verantwortlich für die strukturierte und standardisierte Datenerfassung gemäß der FAIR-Prinzipien, sowie für die automatische und reproduzierbare Datenauswertung. Dabei kommen zunehmende Cloud-basierte Technologien (Object Storage, Open Stack, Kubernetes, Software-Container, etc.) zum Einsatz.

A. Goesmann hat mehr als 247 Publikationen in zum Teil sehr hochrangigen Zeitschriften als Autor und Koautor veröffentlicht und an der JLU Gießen bislang 15,8 Mio. Euro an Drittmitteln eingeworben.

## **TIB Leibniz Informationszentrum Technik und Naturwissenschaften**

Die TIB ist eine nationale Forschungsinfrastruktur für die Bereitstellung von wissenschaftlichen Informationen aus den Bereichen Technik, Informatik, Mathematik, Physik, Chemie und Architektur. Als weltweit größtes Fachinformationszentrum in ihren Fachgebieten verfügt die TIB über herausragende Expertise bei der Entwicklung, Verwaltung und Erhaltung von Wissen, insbesondere in Schlüsselbereichen wie großen Forschungsdaten, Vokabularen und Ontologien, sowie Patenten und Standards. Die TIB betreibt weltweit anerkannte Forschung mit dem Ziel, den Informations-, Daten- und Wissensaustausch im digitalen Zeitalter voranzutreiben, z.B. mit dem Open Research Knowledge Graph<sup>25</sup>. Die TIB hat auf dem Gebiet der Ontologieentwicklung umfangreiche Erfahrungen. Hierzu gehören die Entwicklung der Ontologie STO, die Standards zur Realisierung von Industrie 4.0 beschreibt (Grangel-González et al. 2017), oder die Entwicklung von VoCol, einer integrierten Umgebung zur Entwicklung von Vokabularen (Halilaj et al. 2016). Im derzeit laufenden STREAM-Projekt (BMBF, 16QK11A–16QK11E, 2019–2022) werden eine semantische Repräsentation und Qualitätssicherungsmerkmale aus den Materialwissenschaften erarbeitet. Diese Ergebnisse werden ebenfalls in die hier dargestellte Ontologie-Entwicklung einfließen.

---

<sup>25</sup>Siehe <https://www.orkg.org/>



Seit 2017 baut die TIB zusammen mit der Leibniz Universität Hannover ein Datenmanagementsystem (Leibniz Data Manager) für die semantische Erschließung und Visualisierung von heterogenen Datensammlungen auf. Dieses beinhaltet eine semantische Beschreibung und Vernetzung von Forschungsdaten und anderen Portalen, sowie Vorschaufunktionen (Screening) für unterschiedliche Datentypen und assistiert Forschenden bei der Auswahl relevanter Datensätze für ihre jeweilige Fachdisziplin. Das Screening umfasst bisher u. a. die Visualisierungen von AutoCAD-Dateien, Video-Dateien und eine Live-Coding Funktion mittels Jupyter-Notebooks.

## **Universität zu Köln (UzK)**

UzK als Partner im Projekt HighMED Medical Informatics Initiative arbeitet an innovativen, interoperablen Lösungen in der medizinischen Informatik mit dem Ziel, medizinische Patientendaten für die klinische Forschung und Ausbildung zugänglich zu machen, was wiederum die Patientenversorgung verbessern wird. Die UzK hat das Datenintegrationszentrum MeDIC gegründet, das Daten aus der realen Welt sammelt, die von Leistungserbringern und Forschern generiert werden. Das MeDIC stellt diese Daten der medizinischen Forschung zur Verfügung. Darüber hinaus entwickelt die UzK Technologieplattformen, die es Kliniken ermöglichen, datenbasierte und patientenzentrierte Entscheidungen zu treffen, indem sie die im MeDIC gesammelten Daten wiederverwenden. Die UzK ist an insgesamt zwei NFDI-Konsortien beteiligt: NFDI4Health und GHGA und hat eine führende Rolle im GO FAIR Personal Health Train Implementierungsnetzwerk (Beyan et al. 2020).

## **Universität Leipzig (UL)**

Die Universität Leipzig hat mit dem LIFE Forschungszentrum für Zivilisationserkrankungen sowie dem Institut für Medizinische Informatik, Statistik und Epidemiologie zwei Institutionen der Medizininformatik, an denen vorwiegend klinische und große epidemiologische Studien begleitet und durchgeführt werden. Beide Institutionen haben sich mit der Teilnahme an großen Forschungsverbänden und der Publikation von Studienergebnissen einen Namen gemacht. Daneben sind zahlreiche technischen und organisatorischen Infrastrukturen in Bezug auf Data Sharing entstanden. Dazu zählen solche zur Metadaten-gesteuerten Datenintegration, insbesondere im Kontext großer epidemiologischer Studien, aber auch zur Datenspeicherung mit dedizierten FHIR Servern (z.B. BLAZE FHIR Server) und Portallösungen (Leipziger Data Portal). Am LIFE Zentrum wurde an der Universität umfangreich Data Sharing personenbezogener (medizinischer) Daten vollzogen und durchgesetzt. Darüber hinaus ist die Universität Leipzig an verschiedenen Initiativen und Data Sharing Projekten, wie der MII und der NFDI4Health, beteiligt und hat Projekte in diesem Bereich (z.B. Leipzig Health Atlas<sup>26</sup>) erfolgreich umgesetzt. Insbesondere in der MII ist die UL als Konsortialführer des SMITH Konsortium federführend beteiligt. Weitere Infrastrukturen zu den Themen Privacy Preserving Record Linkage und zu verteilten Analysen sind entstanden; letzteres, der Personal Health Train, in enger Zusammenarbeit mit der RWTH, FhG und EKUT.

## **RWTH Aachen University (RWTH)**

Als IT-Dienstleister der RWTH ist das IT Center eine führende Einrichtung der Informationstechnologie in der RWTH Aachen und in der europäischen Hochschullandschaft, die aktiv mit Partnern aus Forschung,

---

<sup>26</sup> Siehe <https://www.health-atlas.de/>

Lehre und Industrie kooperiert. Im Bereich Forschungsdatenmanagement bietet das IT Center disziplinübergreifende (ProjektRepository, AIMS; Politze & Decker 2016), und disziplinspezifische (u.a. SFB985, SFB1382; Politze, Schwarz et al. 2020), IT-Dienste und Beratung an. Seit 2010 entwickelt das IT Center entsprechende Systeme für das Forschungsdatenmanagement, die sich aktuell in zwei Schlüsselfelder aufteilen: die wissenschaftsgeleitete und projektbasierte Bewirtschaftung und Provisionierung von IT-Ressourcen in den Bereichen Hochleistungsrechnen und Forschungsdatenspeichersystemen (u.a. JARDS; Politze, Claus et al. 2020) sowie das FAIR Management und die disziplinspezifische Validierung von (Meta-)Daten (Politze et al. 2019) in verteilten Speichersystemen (Coscine; Schmitz & Politze 2018). Seit 2019 erprobt das IT Center zudem die Nutzung der Cloud-Plattformen Microsoft Azure und Open Telekom Cloud für Hybrid-Cloud und Scale-Out Szenarien.

## **FIZ Karlsruhe**

Die IGR-Abteilung am FIZ Karlsruhe verfügt über juristische Expertise im Bereich des Datenschutzrechts. Ein datenschutzrechtlicher Schwerpunkt stellt hierbei die Verarbeitung personenbezogener Daten in der digitalen Wissenschaft einschließlich deren Infrastruktureinrichtungen dar. Neben der rechtlichen Begleitung von technischen Projekten und Dienstleistungen, erstellt IGR Empfehlungen zum datenschutzkonformen und privatsphärefreundlichen Design von neuen Technologien (Boehm 2016) – beispielsweise in mehreren NFDI-Konsortien. IGR veröffentlicht regelmäßig wissenschaftliche Aufsätze zum Thema Datenschutzrecht (Hallinan 2021, Hallinan 2021a, Hallinan 2020).

## **Universitätsklinikum Heidelberg (UKHD → siehe oben GHGA)**

## **Heidelberger Akademie der Wissenschaften (HADW → siehe oben GHGA)**

## **European Molecular Biology Laboratory (EMBL → siehe oben GHGA)**

## **Philipps-Universität Marburg (UMR → siehe oben NFDI4Biodiversity)**

## **WWU Münster**

Als Institut für Informations-, Telekommunikations- und Medienrecht verfügt das ITM über besondere Expertise im Bereich des Immaterialgüterrechts. Fragen des Patentrechts, des Geheimnisschutzrechts oder des Urheberrechts sind langjährige Forschungsschwerpunkte des Instituts. Das ITM hat hierzu auch zahlreiche Schriften zum Datenrecht, insbesondere im Hinblick auf urheberrechtliche und geheimnisschutzrechtliche Fragen, veröffentlicht (Hoeren et al. 2021, Hoeren & Pinelli 2020, Hoeren 2020, Hoeren & Pinelli 2020a, Hoeren & Pinelli 2020b, Hoeren 2020a, Hoeren et al. 2020, Hoeren & Pinelli 2020c). Die WWU hat über das ITM schon eine Reihe von Vorarbeiten juristischer Natur erstellt. In mehreren Forschungsprojekten des BMBF hat das ITM auch über immaterialgüterrechtliche Probleme von künstlicher Intelligenz erforscht (so etwa in dem Projekt ABIDA oder Goal).

## Eberhard Karls Universität Tübingen (EKUT → siehe oben GHGA)

## ZBW (→ siehe oben KonsortSWD)

## Atos Information Technology GmbH

Atos ist Gründungsmitglied der Gaia-X Foundation und hat eine führende Rolle in der Architektur-Arbeitsgruppe von Gaia-X. Insbesondere ist Atos-Projektleiter Klaus Ottradovetz Hauptautor der Gaia-X Architecture of Standards (BMW 2020), d.h. des Prozesses, mit dem die Architektur von Gaia-X auf vorhandenen Standards aufgebaut wird, und Co-Autor der Technical Architecture (BMW 2020a). In das Roadmapping und die Definition der FAIR-DS-Architektur bringt Atos weitere Erfahrung aus seiner Arbeit in der International Data Spaces Association ein wie auch seine Plattformexpertise aus Projekten und Lösungen wie ApeX, Venecom, Atos Market Place, Travel Platform, Mundi, Atos Intelligent Transportation System mit einem Fokus auf nutzbare Open-Source Technologien für Data-Sharing und Governance Funktionalitäten und schließlich seine langjährige Erfahrung in der Umsetzung von Sicherheits- und Authentifizierungslösungen u.a. unter Einbezug von DLT-Technologien und Self-Sovereign Identity (SSI) im europäischen Kontext sowohl aus dem kommerziellen Umfeld wie auch aus aktuellen europäischen Förderprogrammen (z.B. EU-Projekt FISHY). Weitere Erfahrungen werden durch Zusammenarbeit mit oneM2M, ETSI, FIWARE in den Bereichen Standards, Architektur und Umsetzung interoperabler Datenplattformen und Konnektoren eingebracht.

### 3. Ausführliche Beschreibung des Arbeitsplans

#### 3.1. Vorhabenbezogene Ressourcenplanung

Der Arbeitsplan ist in fünf Arbeitspakete aufgeteilt. Die nachfolgende Tabelle zeigt, welches Arbeitspaket welche Projektziele adressiert. Abbildung 4 zeigt, wie die Arbeitspakete zusammenwirken.

| AP | Inhalt  | Adressierte Ziele                        |
|----|---|--|
| 1  | Aufbau der gemeinsamen Community, mit der gemeinsame Ziele und eine Roadmap erarbeitet werden                                   | Z1 (Z1.1, Z1.2)<br>Z3 (Z3.4)             |
| 2  | Erarbeitung ethischer und rechtlicher Rahmenbedingungen für den Datenaustausch zwischen Wissenschaft und Wirtschaft             | Z2 (Z2.3, Z2.4)<br>Z3 (Z3.4)             |
| 3  | Schaffung technischer Grundlagen: konzeptionell durch gemeinsame Referenzarchitektur, praktisch durch wiederverwendbare Dienste | Z2 (Z2.1, Z2.2, Z2.3, Z2.4)<br>Z3 (Z3.3) |
| 4  | Demonstration der Machbarkeit des Datenaustauschs zwischen Wissenschaft und Wirtschaft in konkreten Einsatzgebieten             | Z3 (Z3.1, Z3.2)                          |
| 5  | Projektmanagement   | (alle)                                   |

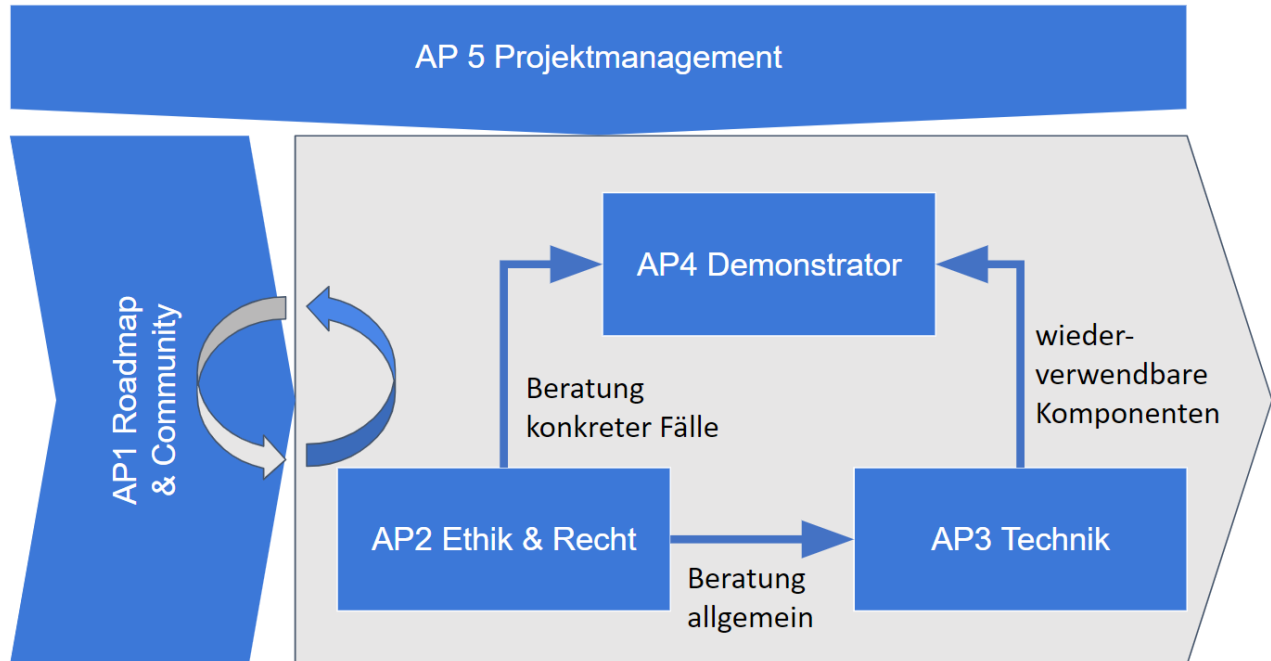


Abbildung 4: Zusammenwirken der Arbeitspakete

Wir arbeiten insbesondere durch AP1 iterativ und berücksichtigen dabei, dass es Vorarbeiten aus den NFDI-Konsortien gibt sowie parallel zu unserem Vorhaben weitere Implementierungen u.a. zu Gaia-X entstehen werden. Separat beigefügt ist eine Tabelle, welche die Arbeitsteilung der Partner in den AP zusammenfasst und durch ein Gantt-Chart den zeitlichen Ablauf der Arbeitsschritte zeigt.

Jedes Unter-Arbeitspaket ist in Form einer Tabelle beschrieben, die über Dauer, beteiligte Partner und deren Ressourcen (Personenmonate) informiert, sowie über

- Ziele
- Arbeitsschritte zur Erreichung der Ziele und daran beteiligte Partner
- Ergebnisse, die aus den Arbeitsschritten hervorgehen, deren Natur (Bericht, Software<sup>27</sup>, Website) und Lieferdatum und ggf. der Meilenstein (vgl. Abschnitt 3.2), zu dem das Ergebnis beiträgt.

Die Partner FhG, NFDI und UKHD beantragen Personalmittel für studentische Hilfskräfte. Da dies wegen der günstigen Kostensätze und der eigenen Definition von „Vollzeit“ (64, 60 bzw. 167,4 Stunden pro Monat) teilweise zu hohen Personenmonats-Zahlen führt, sind Aufwände für über Personalkosten finanzierte studentische Hilfskräfte separat ausgewiesen in der Form ## + ## (Personenmonate wiss. Mitarbeitende + Personenmonate Hilfskräfte).

Es wird darauf geachtet, dass das Projekt zum Ende des Jahres 2021 („Phase 1“) ein in sich abgeschlossenes Teilergebnis liefert. Dies wird dadurch sichergestellt, dass der Meilenstein 1 sich am Ende des Jahres 2021 befindet. Alle Arbeitsschritte sind eindeutig diesem Meilenstein zugeordnet oder beginnen frühestens im Januar 2022 („Phase 2“), d.h., es gibt keinen Arbeitsschritt (wohl aber Unter-Arbeitspakete und Arbeitspakete), der die Jahresgrenze 2021/2022 überschreitet.

<sup>27</sup> Neben Software im engeren Sinne zählen wir hierzu auch Demonstratoren sowie Ontologien.

## AP1: Roadmapping und Community (Leitung: FhG)

Die Herausforderung der Zusammenarbeit zwischen Gaia-X und NFDI besteht in der Schaffung einer gemeinsamen Vision, die es ermöglicht Daten und Wissen zwischen den verschiedenen Communities und Anwendungsgruppen auszutauschen. Dabei müssen mit Gaia-X und NFDI kurzfristig die verschiedenen Perspektiven der Industrie bzw. der Forschung abgeglichen werden. In einer ersten Phase muss daher eine gemeinsamer Dialog geschaffen werden, der auf der Basis der FAIR-Prinzipien die Definition einer gemeinsamen Infrastruktur ermöglicht. Daran beteiligen sich alle Partner unter Leitung der Fraunhofer-Gesellschaft, die zugleich zusammen mit Atos Gaia-X vertritt. Die NFDI-Seite wird in Phase 1 von der RWTH Aachen (als Teil von NFDI4Ing) koordiniert, wobei der NFDI-Verein projektbegleitend unterstützt. In Phase 2 bringt der NFDI-Verein eigene Ressourcen in das Projekt ein und übernimmt die Leitung.

Datenaustausch ist notwendigerweise eine europaweite und globale Herausforderung. Das langfristige Ziel besteht in der Erweiterung der nationalen Architektur zu einer europaweiten (bzw. globalen) Infrastruktur zum Datenaustausch und Verwendung basierend auf europäischen Normen und Werten. Mittelfristig stoßen die europäischen Datenräume (Europäische Kommission 2020) zu NFDI/EOSC bzw. Gaia-X hinzu. Diese sollten ebenfalls interoperabel gestaltet werden um Datensilos zu vermeiden und die gestellten Anforderungen zu realisieren. Dabei hilft, dass die europäischen Datenräume ebenfalls auf den FAIR-Prinzipien aufbauen – daher der Titel des Vorhabens: „FAIR-Data Spaces“ – FAIR-DS. Die FAIR-Prinzipien können daher über NFDI und Gaia-X hinaus als Grundlage zum Aufbau einer Community, gemeinsamen Basisstandards sowie auf den Standards beruhender Infrastruktur dienen. Die beiden Unter-AP initiieren auf dieser gemeinsamen Basis in zwei Schritten eine Roadmap sowie Community und entwickeln diese sukzessive.

| AP1.1 Roadmapping und Kommunikation |                   |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |              |
|-------------------------------------|-------------------|------|-----|-----|-----|----|------|-----|------|------|------|-----|-----|------|-----|------|--------------|
| Dauer                               | Mai–Dezember 2021 |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |              |
| Leitung                             | FhG               |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |              |
| Partner                             | FhG               | NFDI | JLU | TIB | UzK | UL | RWTH | FIZ | UKHD | HAdW | EMBL | UMR | WWU | EKUT | ZBW | Atos | Σ            |
| PM                                  | 8,5<br>+3,5       |      | 1   | 1   | 1   | 1  | 4    | 1   | 1    | 1    | 1    | 1   | 1   | 1    | 3   | 2    | 27,5<br>+3,5 |

### Ziele:

In der Anfangsphase intensiviert dieses AP die Kontakte zwischen NFDI, den Communities der Gaia-X Foundation, des Gaia-X Hub Germany und der International Data Spaces Association (IDSA). Durch die Schaffung einer Roadmap werden gemeinsame Ziele identifiziert und eine Vision entwickelt, die beiträgt, die Integration voranzutreiben. Dies dient zunächst einer breiten Erhebung möglicher Einsatz-Szenarien für souveränen Datenaustausch in der Schnittmenge zwischen Wirtschaft und Wissenschaft sowie dem Erheben von Anforderungen an technische Lösungen dafür. Im weiteren Verlauf werden über das hier geschaffene Netzwerk die Ergebnisse dieses Projekts zurück in die jeweiligen Einzel-Initiativen verbreitet.

Die Aufbauphase von NFDI ist noch nicht abgeschlossen. Bisher wurden 9 Konsortien in einem wissenschaftsgeleiteten Verfahren unter Leitung der DFG ausgewählt. Insgesamt soll es bis zum Jahr 2022 bis zu 30 Konsortien geben. Beim Roadmapping und darauf aufbauenden Aktivitäten ist daher sicherzustellen, dass auch die Expertise und die Anforderungen von zukünftigen Konsortien mit

berücksichtigt werden können.

Dies liefert dann die Basis für die in AP1.2 definierte Erweiterungen auf die European Open Science Cloud (EOSC) sowie die europäischen Data Spaces. Dabei werden auch die Entwicklungen im Kontext der Global Open Science Cloud im Blick behalten, d.h. der Vernetzung globaler Initiativen, wie der New Digital Research Infrastructure Organization in Kanada, the CSTCloud der Chinese Academy of Sciences, der ARDC e-infrastructure in Australien usw.

**Arbeitsschritte:**

**AP1.1.1: Setup und Erstbefüllung der Themen-Landkarte** (Mai–Dezember 2021; **FhG**, RWTH, ZBW). FhG richtet eine öffentliche Website mit einem geeigneten Content Management System und einem geeigneten Datenmodell ein. FhG und NFDI erfassen darin den kurz nach Projektbeginn aktuellen Überblick über Kernthemen, Konsortien, Communities, Anwendungsdomänen, jeweils dafür Verantwortliche, Strukturen usw. sowohl auf Seiten von NFDI als auch auf Seite von Gaia-X; ZBW ergänzt die internationale Perspektive. Die so gebildete Landkarte fasst das Wissen zusammen, mit dem anschließend die gemeinsame Roadmap erarbeitet wird. In der Roadmap werden wieder zu verwendende Technologie und einzubindende Akteure identifiziert und ausgewählt. (4+1 PM)

**AP1.1.2: Roadmapping von Zielen und Vision (Phase 1)** (Mai–Dezember 2021; **FhG**, JLU, TIB, UzK, UL, RWTH, FIZ, UKHD, HAdW, EMBL, UMR, WWU, EKUT, ZBW, Atos). Erarbeitung der gemeinsamen Ziele und Vision für NFDI (RWTH, alle Partner in NFDI-Konsortien sowie WWU) und Gaia-X (FhG, Atos) insbesondere mit den bisher etablierten Konsortien von NFDI und Gaia-X-Partnern. (11,5+0,5 PM)

**AP1.1.3: Erhebung von Einsatz-Szenarien und Anforderungen** (Mai–Dezember 2021; **FhG**, RWTH). Aufbauend auf Erfahrungen aus den domänenübergreifenden Lastenheft-Workshops im Gaia-X Hub Germany (ehemals „Workstream 1“; FhG – Unterstützungsschreiben des Hub liegt vor) sowie aus ersten Erfahrungen aus konsortialübergreifenden Workshops von NFDI (RWTH) werden als Vorbereitung des Roadmappings kurze Online-Workshops zur Erhebung von Einsatz-Szenarien und Anforderungen durchgeführt. (3+1 PM)

**AP1.1.4: Gemeinsame Kommunikationsstrategie und deren Realisierung (Website, Social Media, Veranstaltungen, Vorträge)** (Mai–Dezember 2021; **FhG**, JLU, TIB, UzK, UL, RWTH, FIZ, UKHD, HAdW, EMBL, UMR, WWU, EKUT, ZBW). Die Partner erarbeiten gemeinsam eine Strategie für die Kommunikation und Verbreitung der gemeinsamen Ziele und Visionen (z.B. an Unternehmen, wissenschaftliche Partner und die Öffentlichkeit). RWTH koordiniert die NFDI-Sicht; ZBW ergänzt diese international. Die Strategie wird realisiert durch den Aufbau einer Website, der Schaffung von Reichweite in Social-Media-Kanälen, gemeinsamen Veranstaltungen und Vorträgen (9+1 PM)

**Ergebnisse:**

**E1.1.1: Struktur der Themen-Landkarte mit ersten Inhalten** (Website, September 2021, MS1)

**E1.1.2: Erster Bericht Roadmapping** (Bericht, Dezember 2021, MS1)

**E1.1.3: Erster Bericht Kommunikation** (Bericht, Dezember 2021)

| AP1.2   | Aufbau der Community für FAIR Data Spaces |           |     |     |     |     |      |     |      |      |      |     |     |      |     |      |           |
|---------|---|-----------|-----|-----|-----|-----|------|-----|------|------|------|-----|-----|------|-----|------|-----------|
| Dauer   | Januar 2022–April 2024                    |           |     |     |     |     |      |     |      |      |      |     |     |      |     |      |           |
| Leitung | NFDI                                      |           |     |     |     |     |      |     |      |      |      |     |     |      |     |      |           |
| Partner | FhG                                       | NFDI      | JLU | TIB | UzK | UL  | RWTH | FIZ | UKHD | HAdW | EMBL | UMR | WWU | EKUT | ZBW | Atos | Σ         |
| PM      | 22<br>+16                                 | 20<br>+20 | 0,5 | 0,5 | 0,5 | 0,5 | 0,5  | 0,5 | 0,5  | 0,5  | 0,5  | 0,5 | 0,5 | 0,5  | 14  | 5    | 71<br>+32 |

**Ziele:**

- Aufbau der FAIR Data Spaces Community in Deutschland
- Ausbau der FAIR Data Spaces Community in Europa
- Internationalisierung der FAIR Data Spaces Community
- FAIR-Data-Spaces-Themen sichtbar machen für Wirtschaft und Wissenschaft
- Ergebnisse des Projekts zurück in die einzelnen NFDI-Konsortien und die Gaia-X-Community kommunizieren

**Arbeitsschritte:**

**AP1.2.1: Fortschreibung der Themen-Landkarte** (Januar 2022–Dezember 2023; FhG, **NFDI**, ZBW). Aufgrund der aktuell hohen Dynamik bei der Entwicklung von Themen (national und international) ist eine regelmäßige Überarbeitung erforderlich. (7+5 PM)

**AP1.2.2: Roadmapping von Zielen und Vision** (Januar 2022–Dezember 2023; FhG, **NFDI**, JLU, TIB, UzK, UL, RWTH, FIZ, UKHD, HAdw, EMBL, UMR, WWU, EKUT, ZBW, Atos). Dieses AP umfasst die agile Fortschreibung der gemeinsamen Ziele und Visionen für NFDI, Gaia-X, und eventuelle weiterer Dataspaces (z.B., die EU Common Dataspaces) in einem lebenden Dokument. Alle Partner treffen sich zu gemeinsamen Roadmapping-Workshops. Ein wichtiger Aspekt ist die zukünftige Einbeziehung von Organisationen aus noch kommenden NFDI-Konsortien. Diese werden zukünftig ausgewählt in den von der DFG organisierten Begutachtungsrunden 2 und 3 in den Jahren 2020–21 und 2021–22. (16+4 PM)

**AP1.2.3: FAIR-DS-DE Workshop-Serie** (Januar 2022–Juni 2023; FhG, **NFDI**, ZBW). Erarbeitung eines Konzepts für den Aufbau einer FAIR Data Spaces Community in Deutschland. Wesentliches Ziel ist die Etablierung und Stärkung einer gemeinsamen Community für Datenräume, in der sich neben NFDI- und Gaia-X-Partnern auch weitere Akteure wiederfinden und beteiligen sollen. Umsetzung des Konzepts durch die Organisation, Durchführung und Etablierung einer Workshop-Serie. Dazu wird untersucht, welche Datenräume in Deutschland etabliert werden. Dies sind insbesondere die Datenräume, zu denen im Rahmen des Gaia-X Hub Germany Use Cases identifiziert wurden (energy, health, industry 4.0/SME, agriculture, mobility, public sector, smart living, finance). An diesen Datenräumen beteiligte Organisationen werden jeweils angesprochen und dazu eingeladen, sich inhaltlich an der Gestaltung von Workshops zu beteiligen. (10+6 PM)

**AP1.2.4: FAIR-DS-EU Workshop-Serie** (August 2022–April 2024; FhG, **NFDI**, ZBW). Erarbeitung eines Konzepts für den Aufbau einer FAIR Data Spaces Community in Europa. Wesentliches Ziel ist die Ausweitung der gemeinsamen Community auf die europäische Ebene. Umsetzung des Konzepts durch die Organisation, Durchführung und Etablierung einer Workshop-Serie. Dazu wird untersucht, welche Datenräume in Europa etabliert werden. Dies sind insbesondere die Datenräume der EU Datenstrategie (industrial, green deal, mobility, health, financial, energy, agriculture, public administration, skills). An diesen Datenräumen beteiligte Organisationen werden gezielt angesprochen und dazu eingeladen, sich inhaltlich an der Gestaltung von Workshops zu beteiligen. (10+6 PM)

**AP1.2.5: FAIR-DS Konferenz** (Juli 2023–April 2024; FhG, **NFDI**, ZBW). Erarbeitung eines Konzepts für die Internationalisierung der FAIR Data Spaces Community. Umsetzung des Konzepts durch die Organisation und Durchführung einer Konferenz, voraussichtlich 2023 oder 2024, mit internationaler Beteiligung und Partnern wie z.B. BDVA/DAIRO. Die Konferenz soll mindestens einen digitalen Anteil haben (Streaming von Vorträgen) und präferiert als hybride Veranstaltung durchgeführt werden. Dabei ist zu berücksichtigen, dass hybride Veranstaltungsformate nach ersten, 2020 gewonnen Erfahrungen besonders anspruchsvoll sind. (10+6 PM)

**AP1.2.6: Messeauftritt** (Januar–Juni 2023; **FhG**). Organisation eines Messeauftritts (Ziel: Hannover-Messe 2023) mit Wirkung auf die nationale und internationale Industrie. Präsentiert werden die Demonstratoren aus AP4. (6+3 PM)

**AP1.2.7: Kommunikation der Ergebnisse in die NFDI-Konsortien und Gaia-X-Arbeitsgruppen** (Januar

2022–Mai 2024; FhG, **NFDI**). Die Ergebnisse des Projekts werden durch das NFDI-Direktorat über die Leitungen der am Projekt beteiligten NFDI-Konsortien zurück in die einzelnen NFDI-Konsortien kommuniziert, mit dem Ziel, dort ein tieferes Bewusstsein zu schaffen für den Datenaustausch mit der Wirtschaft. Ebenso werden die Ergebnisse durch die FhG in die Arbeitsgruppen von Gaia-X kommuniziert (die der gesamten Gaia-X Foundation und die des German Hub) sowie in die Arbeitsgruppen der International Data Spaces Association. (8+6 PM)

**Ergebnisse:**

**E1.2.1: Bericht Community-Aufbau in Deutschland** (Dezember 2022, MS2)

**E1.2.2: Bericht Community-Ausbau in Europa** (Dezember 2023)

**E1.2.3: Bericht Konferenz** (Mai 2024, MS3)

**E1.2.4: Fertige Themen-Landkarte** (Website, April 2024, MS3)

**E1.2.5: Messestand und Dokumentation Messe-Auftritt** (Bericht, Juni 2023)

**AP2: Rechtliche und ethische Rahmenbedingungen (Leitung: FIZ)**

Dieses AP definiert rechtliche und ethische Rahmenbedingungen für die Bereitstellung, Nutzung und Verwertung von Industrie- und Forschungsdaten in einer Verknüpfung von NFDI und Gaia-X. In der ersten Phase des Projekts – bis Dezember 2021 – werden die allgemeinen rechtlichen Rahmenbedingungen erarbeitet (AP2.1.1–2.1.3). Vor allem werden Fragen des Immaterialgüterrechts und des Datenschutzrechts identifiziert und geklärt. In der zweiten Phase des Projekts – bis Mai 2024 – werden konkrete Rechtsfragen der dann erarbeiteten Use Cases und Demonstratoren beantwortet (AP2.2). In diese Analyse werden die Auswirkungen aktueller Gesetzesvorhaben, wie dem EU Data Governance Act oder der Datenstrategie des Bundes (z.B. Datennutzungsgesetz/E-Government) mit einbezogen. In AP2.4 wird die Verbindung zwischen den verknüpften Infrastrukturen NFDI und Gaia-X und den weiteren europäischen Dateninfrastrukturen im Bereich Forschung und Gesundheitsdatenverarbeitung geklärt. Aus ethischer Sicht gilt liegt der Schwerpunkt auf der Problematik der informierten Einwilligung bei personenbezogenen Daten sowie auf Fragestellungen, die sich im Spannungsfeld wissenschaftliche/industrielle Daten ergeben (AP2.3). Die rechtliche und ethische Analyse sind miteinander verknüpft. In AP2.2.10 werden die Erkenntnisse aus AP2.2 und 2.3 zusammengeführt. Die Ergebnisse von AP2 sollen als Input für die Bildung eines ELSA-Trainingscurriculums zur Unterstützung der ethischen Nutzung von Forschungsdaten in AP4.5 zur Verfügung gestellt werden.

|                |  |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
|----------------|--|------|-----|-----|-----|----|------|-----|------|------|------|-----|-----|------|-----|------|----------|
| <b>AP2.1</b>   | <b>Erarbeitung des rechtlichen und ethischen Rahmenwerks</b> |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Dauer</b>   | Mai 2021–Dezember 2022                                       |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Leitung</b> | WWU  |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Partner</b> | FhG  | NFDI | JLU | TIB | UzK | UL | RWTH | FIZ | UKHD | HAdW | EMBL | UMR | WWU | EKUT | ZBW | Atos | $\Sigma$ |
| <b>PM</b>      |  |      |     |     |     |    |      | 6   |      | 6    |      |     | 6   |      |     |      | 18       |

**Ziele:**

- Immaterialgüterrechtliche Einordnung der Daten in Gaia-X und NFDI
- Datenschutzrechtliche Einordnung der Daten in Gaia-X und NFDI
- Datenschutzrechtliche Einordnung eines Zugangs zu den Daten von europäischen und internationalen Gaia-X-Anwendern

**Arbeitsschritte:**



**AP2.1.1: Immaterialgüterrechtliche Einordnung der Daten in Gaia-X und NFDI** (Mai 2021–Dezember 2021; **WWU**). Es ist zu klären, welche verschiedenen Schutzsysteme im Bezug auf die Daten zur Anwendung kommen (Urheberrecht, Datenbankherstellerrecht, Geheimnisschutz-Gesetz, Patentrecht). Wie sind die Daten immaterialgüterrechtlich geschützt? Bestehen Datenbankrechte nach § 87a Urheberrechtsgesetz? Oder sind sie sogar Datenbankwerke nach § 4 Urheberrechtsgesetz? Bestehen Geheimhaltungsvereinbarungen? Patentrecht? Wer ist Inhaber der Schutzrechte? (6 PM)

**AP2.1.2: Datenschutzrechtliche Einordnung der Daten in Gaia-X und NFDI** (Mai 2021–Dezember 2021; **FIZ**). Beim Zusammenbringen von „Industrie- und Forschungsdaten“ ist zu klären, welche Verarbeitungstätigkeiten als „Forschung“ nach EU-Datenschutzrecht gelten und was die Reichweite des Begriffes für die Verknüpfung von NFDI und Gaia-X bedeutet. Unter welchen Bedingungen können personenbezogene Daten für Forschungszwecke in einer Verknüpfung von NFDI und Gaia-X verwendet werden? Unter welchen Bedingungen können personenbezogene Daten für andere Nicht-Forschungszwecke in einer Verknüpfung von NFDI und Gaia-X verwendet werden? (6 PM)

**AP2.1.3: Datenschutzrechtliche Einordnung eines Zugangs zu den Daten von europäischen und internationalen Gaia-X-Anwendern, Phase 1** (Dezember 2021; **HAdW**). Des Weiteren muss geklärt werden, unter welchen Bedingungen personenbezogene Daten von NFDI-Infrastrukturen an Gaia-X Anwender in der Europäischen Union und außerhalb der EU transferiert oder ihnen zugänglich gemacht werden können. In Phase 1 werden die betroffenen Akteure identifiziert und datenschutzrechtlich eingeordnet. (1 PM)

**AP2.1.4: Datenschutzrechtliche Einordnung eines Zugangs zu den Daten von europäischen und internationalen Gaia-X-Anwendern, Phase 2** (Januar 2022–Dezember 2022; **HAdW**).

Bei der Bearbeitung der Frage, nach welchen Modalitäten künftig grenzüberschreitende Übermittlungen zu Gaia-X-Anwender in der EU und in Drittstaaten stattfinden können oder ihnen Zugriff auf die Daten erlaubt werden kann, wird die rechtliche Analyse des Zugangs und der Zurverfügungstellung in Phase 2 akteurs(typ)bezogen durchgeführt. Ein besonderes Augenmerk wird hierbei auf den Schutz sensibler Daten gelegt. (5 PM)

**Ergebnisse:**

**E2.1.1: Bericht immaterialgüterrechtliche und datenschutzrechtliche Einordnung** (Dezember 2021, MS1)

**E2.1.2: Bericht datenschutzrechtliche Einordnung eines Zugangs zu den Daten von europäischen und internationalen Gaia-X-Anwendern** (Dezember 2022, MS2)

|                |   |      |     |     |     |    |      |      |      |      |      |     |     |      |     |      |          |
|----------------|---|------|-----|-----|-----|----|------|------|------|------|------|-----|-----|------|-----|------|----------|
| <b>AP2.2</b>   | <b>Rechtliche Überprüfung der Anwendungsfälle</b> |      |     |     |     |    |      |      |      |      |      |     |     |      |     |      |          |
| <b>Dauer</b>   | Januar 2022–April 2024                            |      |     |     |     |    |      |      |      |      |      |     |     |      |     |      |          |
| <b>Leitung</b> | FIZ   |      |     |     |     |    |      |      |      |      |      |     |     |      |     |      |          |
| <b>Partner</b> | FhG   | NFDI | JLU | TIB | UzK | UL | RWTH | FIZ  | UKHD | HAdW | EMBL | UMR | WWU | EKUT | ZBW | Atos | $\Sigma$ |
| <b>PM</b>      |   |      |     |     |     |    |      | 28,5 |      |      |      |     | 24  |      |     |      | 52,5     |

**Ziele:**

Dieses Arbeitspaket leistet die rechtliche Beratung der in AP3 zu entwickelnden Komponenten und der Datenflüsse der in AP4 zu entwickelnden Demonstratoren und berücksichtigt dabei Folgendes:

- Vertragsrechtliche Überprüfung
- Immaterialgüter- und datenschutzrechtliche Compliance-Überprüfung
- Entwurf Musterverträge/Codes of Conduct
- Zweckbindung der Daten

- Haftung
- Verbindung Datenschutzrecht/Ethik

### **Arbeitsschritte:**

**AP2.2.1: Vertragsrechtliche Überprüfung Immaterialgüterrecht** (Januar 2022–Juli 2023; **WWU**). Sind die Verträge hinsichtlich der Demonstratoren, die in Bezug auf die Einbringung von Daten abgeschlossen worden, immaterialgüterrechtlich zulässig? Auswertung der (bis dahin) abgeschlossen Verträge. (6 PM)

**AP2.2.2: Vertragsrechtliche Überprüfung Datenschutzrecht** (Januar 2022–Juli 2023; **FIZ**). Sind die Verträge hinsichtlich der Demonstratoren, die in Bezug auf die Einbringung von Daten abgeschlossen worden, datenschutzrechtlich zulässig? Auswertung der (bis dahin) abgeschlossen Verträge. (6 PM)

**AP2.2.3: Compliance-Überprüfung Immaterialgüterrecht** (Januar 2022–Juli 2023; **WWU**). Wie sind die gesetzlichen und vertragsrechtlichen Restriktionen für die eingehenden Daten in Compliance-Strukturen in den Demonstratoren umgesetzt? Immaterialgüterrechtliche Betrachtung. Welche juristischen Beziehungen haben die ethischen FAIR-Prinzipien mit dem Immaterialgüterrecht (etwa über FRAND)? (6 PM)

**AP2.2.4: Compliance-Überprüfung Datenschutzrecht** (Januar 2022–Juli 2023; **FIZ**). Wie sind die gesetzlichen und vertragsrechtlichen Restriktionen für die eingehenden Daten in Compliance-Strukturen in den Demonstratoren umgesetzt? Datenschutzrechtliche Betrachtung. (6 PM)

**AP2.2.5: Musterverträge** (Mai 2022–Dezember 2023; **WWU**). Welche unterschiedlichen Rechtskulturen in Bezug auf IP treffen zusammen bei der Zusammenführung beider Großprojekte? Wo sind die Kulturen unterschiedlich und wo ähnlich? Wie kann man diese Unterschiede durch Musterverträge in den erarbeiteten Use Cases/Demonstratoren durchbrechen? (4 PM)

**AP2.2.6 Codes of Conduct** (Mai 2022–Dezember 2023; **WWU**). Die Musterverträge werden ergänzt durch konkrete Codes of Conduct etwa zu notwendigen Geheimhaltungsmaßnahmen (technisch-organisatorisch, z.B. für Datenzugriffsrechte, Passwortkontrolle, Datensicherheit). In enger Rückkopplung zu den Ergebnissen aus AP3.5 wird auch geprüft, wie man die rechtliche Verbindlichkeit solcher Codes erhöhen kann. (2 PM)

**AP2.2.7: Zweckbindung der Daten** (Januar–Dezember 2022; **FIZ**) Wie können die auftretenden Probleme bei der Nachnutzung und ggf. bei der Zweckänderung konkret in den Use Cases/Demonstratoren gelöst werden? Wie kann der Zweckbindungsgrundsatz bei der Zusammenführung der Daten eingehalten werden? Welche technischen Maßnahmen (Anonymisierung/Pseudonymisierung) müssen dafür vorliegen? (6 PM)

**AP2.2.8: Haftung Immaterialgüterrecht** (Januar 2023–April 2024; **WWU**). Es ist zu klären, wie die Rechtsmängelhaftung sich auf bestehende Lücken bei der Nutzung eingebrachter Daten auswirkt. Dabei gehen wir vor allem auch der kollisionsrechtlichen Frage nach, die sich aus dem Einfluss verschiedener Rechtsordnungen ergibt. Immaterialgüterrechtliche Betrachtung (6 PM)

**AP2.2.9: Haftung Datenschutzrecht** (Januar 2023–April 2024; **FIZ**). Es ist zu klären, wie die Rechtsmängelhaftung sich auf bestehende Lücken bei der Nutzung eingebrachter Daten auswirkt. Dabei gehen wir vor allem auch der kollisionsrechtlichen Frage nach, die sich aus dem Einfluss verschiedener Rechtsordnungen ergibt. Datenschutzrechtliche Betrachtung (6 PM)

**AP2.2.10: Verbindung Secrecy/Datenschutzrecht/Ethik** (Januar 2023–April 2024; **FIZ**). Wie können die ethischen und rechtlichen Anforderungen bezüglich der Verarbeitung insbesondere personenbezogener Daten in der Verknüpfung von NFDI und Gaia-X vereinbart werden? Welche Fragestellungen ergeben sich aus dem Zusammenschluss und wie können etwaige Widersprüche aufgelöst werden? (Zusammenarbeit mit AP2.1 und AP2.3 Ethik. Der rechtliche Spielraum wird mit dem ethischen abgeglichen. Wenn dabei Ungleichheiten gefunden werden, werden Lücken identifiziert und analysiert.) (4,5 PM)

### **Ergebnisse:**

- E2.2.1: Bericht Vertragsrechtliche Überprüfung und Compliance** (Juli 2023)  
**E2.2.2: Bericht Musterverträge und Codes of Conduct** (Dezember 2023)  
**E2.2.3: Bericht Zweckbindung der Daten** (Dezember 2022, MS2)  
**E2.2.4: Bericht Haftung** (Mai 2024, MS3)  
**E2.2.5: Bericht Verbindung Secrecy/Datenschutzrecht/Ethik** (Mai 2024, MS3)

|                 |                     |      |     |     |     |    |      |     |              |      |      |     |     |      |     |      |              |
|-----------------|---------------------|------|-----|-----|-----|----|------|-----|--------------|------|------|-----|-----|------|-----|------|--------------|
| <b>AP2.3</b>    | <b>Ethik</b>        |      |     |     |     |    |      |     |              |      |      |     |     |      |     |      |              |
| <b>Zeitraum</b> | Mai 2021–April 2024 |      |     |     |     |    |      |     |              |      |      |     |     |      |     |      |              |
| <b>Leitung</b>  | UKHD                |      |     |     |     |    |      |     |              |      |      |     |     |      |     |      |              |
| <b>Partner</b>  | FhG                 | NFDI | JLU | TIB | UzK | UL | RWTH | FIZ | UKHD         | HAdW | EMBL | UMR | WWU | EKUT | ZBW | Atos | Σ            |
| <b>PM</b>       |                     |      |     |     |     |    |      |     | 18,5<br>+4,3 |      |      |     |     |      |     |      | 18,5<br>+4,3 |

**Ziele:**

- Empfehlungen für ein Informed-Consent-Modul für die Sekundärnutzung personenbezogener Daten in Gaia-X und in NFDI
- gerechte Kriterien für den Datenzugang im Spannungsfeld zwischen akademischer und privatwirtschaftlicher Sekundärdatennutzung
- Absicherung von Vertrauen in eine gemeinwohlfördernde Forschung durch gute Governance und Rechenschaftspflichten (accountability)
- Empfehlungen zu den Pflichten mit Blick auf Data-Sharing der relevanten Stakeholder

**Arbeitsschritte:**

**AP2.3.1: Informed Consent Modul für prospektive Datenerhebung** (Mai–Dezember 2021; UKHD). Mit Blick auf personenbezogene Daten stellt sich die Frage nach der Legitimierung einer Sekundärnutzung im Gaia-X-/NFDI-Austausch. Wir wissen aus empirischen Untersuchungen in Deutschland, dass die Zustimmungquote für die Sekundärnutzung für die akademische Forschung hoch ist, aber deutlich geringer ist, wenn nach um die Datennutzung für kommerzielle Unternehmen gefragt wird (Voigt et al. 2020). Nur 23% der repräsentativ Befragten in Deutschland würden z.B. an eine private Firma ihre genetischen Patientendaten spenden (Richter et al. 2020). Wie sollte für die Sekundärnutzung personenbezogener Daten aufgeklärt werden? (3,5+1,3 PM)

**AP2.3.2: Gerechtigkeit in Datennutzung und Datenzugang: Welche Kriterien und Policies sollen für Datenzugang und Datennutzung im Spannungsfeld zwischen Open Science und privatwirtschaftlicher Gewinnmaximierung gelten?** (Januar–Dezember 2022; UKHD). Die Anreizsysteme der akademischen und privatwirtschaftlichen Welt unterscheiden sich. Forschung im öffentlichen Interesse ist nicht abhängig von der Gewinnmaximierung für Aktionäre bspw. durch den Schutz der Eigentumsrechte. Vielmehr wurzeln die Anreize von Open Science überwiegend im Wunsch neues Wissen zu schaffen, Anerkennung durch Peers und beruflichen Aufstieg. Die beiden Systeme des Wissensgewinns, privatwirtschaftlich und akademisch, sind damit im besten Fall komplementär zu denken mit einer Durchlässigkeit bezogen auf Ideen, Daten und Personal. Aus der Unterschiedlichkeit der beiden Systeme und ihrer Anreizstrukturen folgt, dass es eines Aushandlungsrahmens bedarf, der einerseits die Abgrenzung der Rechte an geistigem Eigentum berücksichtigen muss und andererseits den Schutz des öffentlichen Datenraums von wissenschaftlichen Daten. Was sind relevante Eckpunkte für diesen Aushandlungsrahmen und wie beeinflusst er die Zugangskriterien und die Vertragsgestaltung (Zusammenarbeit mit AP2.2.10 Immaterialgüterrecht und Datenschutz. Der ethische Spielraum wird mit dem rechtlichen abgeglichen. Wenn dabei Ungleichheiten gefunden werden, werden Lücken identifiziert und analysiert.) (5,5+1 PM)

**AP2.3.3: Vertrauen & Governance: Wie kann Vertrauen in eine gemeinwohlfördernde Forschung von Daten aus Industrie und Akademia aufgebaut und gefördert werden?** (Januar–Dezember 2023; UKHD). Datensouveränität wird vom Deutschen Ethikrat mit Blick auf personenbezogene Daten als das Recht auf Kontroll- und Einsichtsmöglichkeiten in die Nutzung der Daten definiert. Wie kann mit Blick auf die verbreitete Skepsis bzgl. der Sekundärnutzung durch kommerzielle Unternehmen die Zusammenarbeit von Gaia-X und NFDI im Rahmen einer Governance so gestaltet werden, dass z.B. durch Transparenz und Repräsentation legitimer Interessen von Betroffenen (auch z.B. durch Patientenbeteiligung) Vertrauen aufgebaut und durch Accountability-Kriterien erhalten wird? (5,5+1 PM)

**AP2.3.4: Moralische Pflichten zum Data-Sharing (Industrie und Akademie)** (Januar–April 2024; UKHD). Auch hier geht es um die Aushandlung von Anreizsystemen (Geheimniserhalt und notwendige andere Anreize) und die (moralischen) Pflichten zur Maximierung des Datennutzens im Sinne des Gemeinwohls, die mit Blick auf die Unternehmensethik selbst oder soziale Verantwortung begründbar wären. Haben privatwirtschaftliche Unternehmen die (moralische) Pflicht, ihre Forschungsdaten für die Sekundärnutzung in der Forschung zur Verfügung zu stellen, und, wenn ja, unter welchen Bedingungen? (4+1 PM)

**Ergebnisse:**

**E2.3.1: Generische Mustertexte für die Einwilligung in die Sekundärnutzung personenbezogener Daten in Gaia-X/NFDI** (Bericht, Dezember 2021, MS1)

**E2.3.2: Bericht zum Spannungsfeld zwischen akademischer und privatwirtschaftlicher Sekundärdatennutzung** (Dezember 2022, MS2)

**E2.3.3: Prozedurale und substantielle Kriterien für die Governance von Gaia-X/NFDI** (Bericht, Dezember 2023, MS3)

**E2.3.4: Positionspapier zu den (moralischen) Pflichten zum Data-Sharing** (Bericht, Mai 2024, MS3)

|                |   |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |    |
|----------------|---|------|-----|-----|-----|----|------|-----|------|------|------|-----|-----|------|-----|------|----|
| <b>AP2.4</b>   | <b>Rechtliche Anschlussfähigkeit des Datenraums</b> |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |    |
| <b>Dauer</b>   | November 2021–Mai 2024                              |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |    |
| <b>Leitung</b> | EMBL  |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |    |
| <b>Partner</b> | FhG   | NFDI | JLU | TIB | UzK | UL | RWTH | FIZ | UKHD | HAdW | EMBL | UMR | WWU | EKUT | ZBW | Atos | Σ  |
| <b>PM</b>      |   | 0,5  |     |     |     |    |      |     |      | 19   | 5,5  |     |     | 5    |     |      | 30 |

**Ziele:** Ziel des Arbeitspaketes ist es, die rechtliche Interoperabilität zwischen den verknüpften Dateninfrastrukturen (NFDI und Gaia-X als FAIR Data Space), den bestehenden sowie den zu entwickelnden europäischen und grenzüberschreitenden Infrastrukturen (z.B. verschiedene European Data Spaces, 1+ Million Genomes, EGA) zu gewährleisten.

**Arbeitsschritte:**

**AP2.4.1: Datentypen, sensible Daten, Rechtsgrundlagen, Phase 1** (November 2021–Dezember 2021; EKUT). Die Datentypen im Zusammenschluss von Gaia-X und NFDI, insbesondere die sensiblen Daten, beeinflussen die rechtliche Interoperabilität mit (anderen) europäischen Infrastrukturen. Es wird der Frage nachgegangen, ob die Verarbeitung dieser Daten an der Schnittstelle verschiedener europäischer Infrastrukturen einen Einfluss auf die Rechtfertigungsgründe ihrer Verarbeitung hat. (1 PM)

**AP2.4.2: Datentypen, sensible Daten, Rechtsgrundlagen, Phase 2** (Februar–September 2022; HAdW, EKUT). Die Datentypen im Zusammenschluss Gaia-X und NFDI, insbes. die sensiblen Daten, beeinflussen die rechtliche Interoperabilität mit (anderen) europäischen Infrastrukturen. Auf Grundlage des

Ergebnisse in AP2.4.1, Phase 1, wird die Frage eruiert, auf welche Rechtsgrundlagen die Verarbeitungen gestützt werden können und wie sich die verschiedenen Rechtsgrundlagen vor diesem Hintergrund zueinander verhalten. (5 PM)

**AP2.4.3: Erarbeitung der Anschlussmöglichkeiten an die europäischen Forschungsdateninfrastrukturen, Phase 1** (Dezember 2021; EMBL). Dazu wird untersucht, welchen datenschutzrechtlichen Einfluss die Verknüpfung der Infrastrukturen in Deutschland in Form des FAIR Data Space auf ihren Anschluss an weitere europäische (auch föderierte) Forschungs- und Industriedateninfrastrukturen hat. Zunächst werden die verschiedenen Anwendungsszenarien bedacht. (0,5 PM)

**AP2.4.4: Erarbeitung der Anschlussmöglichkeiten an die europäischen Forschungsdateninfrastrukturen, Phase 2** (März 2022–Januar 2023; HAdW, EMBL, EKUT). Aufbauend auf den Ergebnissen in AP2.4.3, Phase 1, wird untersucht, welchen Einfluss die fortschreitende Entwicklung europäischer Infrastrukturen auf die angestoßenen nationalen Verknüpfungen hat. Es werden die Modalitäten für die Berücksichtigung der entstehenden europäischen Vorgaben bei der nationalen Infrastrukturentwicklung bestimmt. (5,5 PM)

**AP2.4.5: Erarbeitung der Anschlussmöglichkeiten an die europäischen Gesundheitsdateninfrastrukturen** (Oktober 2022–Juni 2023; HadW, EMBL, EKUT). Bestimmte NFDI-Konsortien werden von Beginn an in enger Abstimmung mit europäischen Partnern aufgebaut, und zukünftig als Deutschlands Knotenpunkt in föderierten europäischen Infrastrukturen fungieren (z.B. GHGA). Neben den NFDI-Konsortien wird Gaia-X die kommerzielle Nutzung von sensiblen Daten ermöglichen. Wenn an einer nationalen Dateninfrastruktur europäisch eingebettete Partner beteiligt sind, die personenbezogene und sensible Daten verarbeiten, bedarf es der Klärung, welchen Einfluss dies auf die rechtliche Interoperabilität der vereinten Infrastruktur nach außen hat und wie diese Interoperabilität rechtlich im Einzelnen geleistet werden kann. (6 PM)

**AP2.4.6: Eruiierung datenschutzrechtlich bestimmter Optionen des Verhältnisses zu den europäischen Dateninfrastrukturen mit europäischen Partnern** (Februar–Oktober 2023; NFDI, HAdW, EMBL, EKUT). Mit diesem Arbeitsschritt soll erreicht werden, dass die nationale Dateninfrastruktur als FAIR Data Space die datenschutzkonformen Möglichkeiten ihres Anschlusses an die grenzüberschreitende Forschung mit personenbezogenen und sensiblen Daten in größtmöglichem Umfang wahrnimmt und diese mitgestalten und -implementieren kann. Hierbei ist besonders zu beachten, dass Gaia-X als offene Dateninfrastruktur, die auf europäischen Werten basiert, schon von vornherein datenschutzkonform entworfen wird, deshalb also als Grundlage dienen kann. Das Arbeitspaket kann parallel und in enger Abstimmung mit dem Aufbau der Communities für FAIR Data Spaces (AP1.2) laufen und wichtige Impulse hierfür geben bzw. aufnehmen. (6 PM)

**AP2.4.7: Anschlussmöglichkeiten an international verteilte Forschungsnetzwerke von Repositorien zur gemeinsamen Nutzung von personenbezogenen und sensiblen Daten** (August 2023–April 2024; HAdW, EMBL). Dazu wird untersucht, wie die erweiterte nationale Dateninfrastruktur die Beteiligung an internationalen Forschungsnetzwerken beeinflusst. Weiterhin wird untersucht, wie die angestrebte Interoperabilität zwischen Gaia-X und NFDI zur Nutzung von Forschungsdaten für kommerzielle Zwecke auf internationaler Ebene beitragen kann und umgekehrt. Es werden die Modalitäten einer Freigabe der Daten in international verteilten Repositorien aus einer verknüpften nationalen Infrastruktur für die Forschung definiert sowie die Zugangsformen zu diesen Daten bestimmt. Dabei wird die Rolle verschiedener Akteure berücksichtigt (z.B. internationaler Organisationen). (6 PM)

**Ergebnisse:**

**E2.4.1: Bericht zur daten(typ)basierten rechtlichen Interoperabilität an der Schnittstelle diverser europäischer Dateninfrastrukturen** (Februar 2023, MS2)

**E2.4.2: Handreichung für die nationalen Entscheidungsträger über die Vor- und Nachteile unterschiedlicher Anschlussoptionen an europäische Dateninfrastrukturen** (Bericht, Oktober 2023, MS3)

**E2.4.3: Handreichung für die nationalen Entscheidungsträger über die Vor- und Nachteile unterschiedlicher Anschlussoptionen an international verteilte Forschungsnetzwerke** (Bericht, Mai 2024, MS3)

### AP3: Technische Grundlagen (Leitung: JLU)

In AP3 werden basierend auf bereits vorhandenen Vorarbeiten und geplanten Arbeiten verschiedene Basisdienste an die Gaia-X-Spezifikationen angepasst (AP 3.3–3.7) und mit den Gaia-X Federation Services verbunden, die Ende 2021 in einer Grundversion, Ende 2022 in einer erweiterten Version als Open Source Software (OSS) vorliegen sollen und im Folgenden bedarfsgerecht weiterentwickelt werden. Als Grundlage für diese Anpassungen werden in AP3.1 die Gaia-X-Anforderungen zentral erfasst und aufgearbeitet. In AP3.2 werden gemeinsam mit den einzelnen Entwicklern der Basisdienste die Architekturen derselben an die Gaia-X-Anforderungen adaptiert. Hierbei werden auch generelle nicht Gaia-X-spezifische Überlegungen zu Cloud-nativen Architekturen einbezogen, insbesondere im Hinblick auf die Skalierbarkeit der Systeme. Die Basisdienste werden bei ausgewählten und entsprechend kompatiblen Cloud-Providern deployed. Basierend auf diesen Deployments nehmen die Demonstrator-Projekte diese Dienste in Anspruch und belegen somit deren Anwendbarkeit und die Kompatibilität zwischen Gaia-X und NFDI. Für Entwicklung und Deployment der Dienste und Demonstratoren wird daher von vornherein auf ein Multi-Cloud-Umfeld gesetzt. Dabei sind de.NBI für einen wissenschaftlichen Cloud-Provider und die Open Telekom Cloud, das kommerzielle Pendant, als mögliche Beispiele zu nennen. Eine Übertragbarkeit zwischen wissenschaftlichen und kommerziellen Infrastrukturen ist somit sichergestellt.

Aufgrund der hohen Zahl gegenseitiger Abhängigkeiten einzelner Services werden im Rahmen des Entwicklungsprozesses DevOps best-practices genutzt, um die einzelnen Teilkomponenten zu einem frühen Zeitpunkt auf Kompatibilität zu testen und gegebenenfalls auftretende Probleme frühzeitig zu erkennen und zu lösen. Weiterhin erleichtert dies die Implementierung der Anforderungen von Gaia-X aus AP3.1. Zur Unterstützung dieser Herangehensweise, die eine weit verbreitete Praxis bei der Software-Entwicklung in Cloud-Infrastrukturen darstellt, dienen Teile der allozierten Ressourcen in AP3.2 sowie die beantragte Hardware.

Die entwickelten Services werden schließlich als Produktivsysteme und Referenzimplementierung genutzt.

|                |                                  |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |            |
|----------------|----------------------------------|------|-----|-----|-----|----|------|-----|------|------|------|-----|-----|------|-----|------|------------|
| <b>AP3.1</b>   | <b>Konzeptionelle Grundlagen</b> |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |            |
| <b>Dauer</b>   | Mai 2021–Mai 2024                |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |            |
| <b>Leitung</b> | FhG                              |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |            |
| <b>Partner</b> | FhG                              | NFDI | JLU | TIB | UzK | UL | RWTH | FIZ | UKHD | HAdW | EMBL | UMR | WWU | EKUT | ZBW | Atos | Σ          |
| <b>PM</b>      | 16<br>+9                         |      | 6,5 |     |     |    |      |     |      |      |      | 2   |     |      |     | 5    | 29,5<br>+9 |
| <b>Ziele:</b>  |                                  |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |            |

- Bezug zwischen den Architekturen von Gaia-X und den NFDI-Konsortien herstellen und in gemeinsames Konzept überführen
- Überblick über die verfügbare Landschaft der verfügbaren Technologien in den NFDI-Konsortien gewinnen, entlang der Entwicklung in den NFDI-Konsortien fortschreiben und um die externe Forschungsinfrastruktur-Landschaft ergänzen
- Gaia-X-konforme Selbstbeschreibung von Daten- und Dienstangeboten in einer Forschungsinfrastruktur ermöglichen

#### **Arbeitsschritte:**

**AP3.1.1: Initiales Architektur-Konzept** (Mai–August 2021; **FhG**, JLU, UMR, Atos). Abbildung der NFDI-Architekturen auf die Gaia-X-Architektur (die Version der Technical Architecture, die für Juni 2021 erwartet wird, und die erste Spezifikation der Federation Services, die für Mai 2021 erwartet wird). (4+1 PM)

**AP3.1.2: Referenzarchitektur-Modell** (Januar 2022–Mai 2024; **FhG**, JLU, UMR, Atos). Durchgängige Weiterentwicklung des initialen Konzepts aus AP3.1.1 unter Berücksichtigung der bis dahin stattgefundenen Architektur-Entwicklungen bei Gaia-X (weitere Releases der Technical Architecture und der Federation-Services-Spezifikation) und NFDI. Zwischenstände werden passend zu den Phasen der Entwicklung der Komponenten und Demonstratoren geliefert. (12+1 PM)

**AP3.1.3: Dokumentation der Gaia-X-konformen Selbstbeschreibung von Forschungsdaten- und -dienstangeboten** (Juni–Oktober 2021; **FhG**, NFDI). Initial wird die auf Gaia-X-Seite vorhandene Dokumentation zur semantischen Selbstbeschreibung von Daten und Diensten ergänzt um eine Dokumentation, wie entsprechende Selbstbeschreibungen in Forschungsinfrastrukturen verfasst werden können. (1,5+1,5 PM)

**AP3.1.4: Entwicklung einer Selbstbeschreibungs-Ontologie für Forschungsdaten- und -dienstangebote** (Januar 2022–Dezember 2023; **FhG**, JLU, UMR) – als domänenspezifische Erweiterung der Gaia-X-Selbstbeschreibungs-Ontologie, ergänzt um eine Anbindung an Konzepte aus spezifischen Ontologien der NFDI und wissenschaftlicher Fach-Communities, wie sie in Form der Terminologie in AP3.6 integriert worden sind. Hierzu bauen wir auf den Vorarbeiten der NFDI-übergreifenden Metadaten-Initiative und des Gaia-X Open Work Package „Self-Description“ auf und vernetzen uns mit diesen. (6+4 PM)

**AP3.1.5: Erste Sammlung der verfügbaren technischen Grundlagen** (Mai–August 2021; **FhG**, JLU). Für alle in den Diensten der ersten Phase (erste Phase von AP3.2–3.7) benötigten technischen Grundlagen schaffen wir in Form einer Technologie-orientierten Erweiterung der Themenlandkarte aus AP1 einen Überblick: Was (z.B. welcher PID-Dienst) ist implementiert worden, auf welchem Reifegrad (TRL), welche Schnittstellen hat es, unter welcher Lizenz ist es verfügbar, wer betreibt eine Instanz? (2+1 PM)

**AP3.1.6: Fortschreibung der Technologie-Landkarte** (Januar 2022–Mai 2024; **FhG**) um weitere Entwicklungen in NFDI, EOSC und Gaia-X (4+0,5 PM)

#### **Ergebnisse:**

**E3.1.1: Initiales Architektur-Konzept** (Bericht, August 2021, MS1)

**E3.1.2: Referenzarchitektur-Modell v1** (Bericht, März 2022)

**E3.1.3: Referenzarchitektur-Modell v2** (Bericht, Februar 2023, MS2)

**E3.1.4: Referenzarchitektur-Modell v3** (Bericht, Mai 2024, MS3)

**E3.1.5: Dokumentation der Gaia-X-konformen Selbstbeschreibung von Forschungsdaten- und -dienstangeboten** (Bericht, Oktober 2021, MS1)

**E3.1.6: Selbstbeschreibungs-Ontologie für Forschungsdaten und -dienstangebote v1** (Software, Juni 2022, MS2)

**E3.1.7: Selbstbeschreibungs-Ontologie für Forschungsdaten und -dienstangebote v2** (Software, Juli 2023, MS3)

**E3.1.8: Technologie-Landkarte v1** (Website, Februar 2023, MS2)

**E3.1.9: Technologie-Landkarte v2** (Website, Mai 2024, MS3)

|                |                                 |      |     |     |     |    |      |     |      |       |      |     |     |      |     |      |    |
|----------------|---------------------------------|------|-----|-----|-----|----|------|-----|------|-------|------|-----|-----|------|-----|------|----|
| <b>AP3.2</b>   | <b>Cloud-native Architektur</b> |      |     |     |     |    |      |     |      |       |      |     |     |      |     |      |    |
| <b>Dauer</b>   | Mai 2021–Mai 2024               |      |     |     |     |    |      |     |      |       |      |     |     |      |     |      |    |
| <b>Leitung</b> | JLU                             |      |     |     |     |    |      |     |      |       |      |     |     |      |     |      |    |
| <b>Partner</b> | FhG                             | NFDI | JLU | TIB | UzK | UL | RWTH | FIZ | UKHD | HAAdW | EMBL | UMR | WWU | EKUT | ZBW | Atos | Σ  |
| <b>PM</b>      | 13<br>+2                        |      | 21  |     |     |    |      |     |      |       |      |     |     |      |     |      | 36 |

**Ziele:**

- Bereitstellung von Querschnittsinfrastrukturen und -kompetenzen (DevOps, Kubernetes, skalierbare Architekturen, etc.) für die Implementierung der Services und Demonstrator-Projekte
- Sicherstellung der Umsetzung der Anforderungen von AP 3.1 in den Services
- Gewährleistung des Betriebs der Demonstrator-Projekte
- Bereitstellung der Cloud-Umgebung

Um einen möglichst großen Nutzen aus den zur Verfügung stehenden Cloud-Infrastrukturen zu ziehen und angesichts der heterogenen Entwicklungskulturen für Software in den NFDI-Konsortien und bei Gaia-X eine Einhaltung etablierter Best Practices bei der Software-Entwicklung zu erreichen, leistet dieses Arbeitspaket zentral Unterstützung für die anderen APs. Dies umfasst die Etablierung von DevOps best-practices bei der Entwicklung der individuellen Services, sowie Unterstützung bei der Einrichtung der benötigten Kubernetes Cluster und dem Deployment der Services und Demonstratoren. Da die Verfügbarkeit der technischen Infrastruktur in den NFDI-Konsortien sehr unterschiedlich ausgeprägt ist, kommt dem Multi-Cloud-Aspekt ein besonderer Stellenwert zu. So wird die Kompatibilität zwischen und innerhalb der wissenschaftlichen und kommerziellen Cloud-Umgebungen sichergestellt und die Übertragbarkeit für die Nutzung der Services im wirtschaftlichen Kontext ermöglicht.

Dementsprechend werden in diesem AP die in AP3.1 entwickelten Architekturkonzepte gemeinsam mit den Services und unter Berücksichtigung der relevanten Vorarbeiten in einer technischen Architektur verwirklicht. Für ein eigenständiges Deployment wird in geringem Umfang Hardware beschafft. Ausfallsicherheit und Skalierbarkeit werden darüber hinaus durch die Anbindung bzw. Teilintegration in die bereits vorhandene de.NBI-Cloud-Infrastruktur am Standort Gießen sichergestellt. Dies beinhaltet insbesondere die mehrfach redundante Datenspeicherung in einem leistungsfähigen Ceph Storage-Cluster. Die auf der Hardware aufbauenden Infrastruktur-Services (Objects Storage und Kubernetes) werden vollständig Gaia-X-kompatibel deployed. Die hierbei gewonnenen Erkenntnisse fließen dabei auch in die zukünftige Ausrichtung der de.NBI Cloud ein, um hier ebenfalls eine entsprechende Kompatibilität zu Gaia-X außerhalb dieses Projekts zu erreichen. Die Hardware kann über die Projektlaufzeit hinaus zur Bereitstellung und Weiterentwicklung der Services mit eigenen Mitteln betrieben und zur Verfügung gestellt werden.

**Arbeitsschritte:**

**AP3.2.1: DevOps Phase 1** (Mai–Dezember 2021; FhG, JLU). Etablierung von CI/CD-Pipelines und Deployment der Demonstratoren. (6 PM)

**AP3.2.2: Übersetzung der konzeptionellen Anforderungen in Spezifikationen für die Services und Demonstratoren** (Januar 2022–April 2024; FhG, JLU). Die in AP3.1 entwickelte konzeptionelle Architektur



wird hier in eine technische Architektur weiterentwickelt und gemeinsam mit AP 3.2.3 in den einzelnen Services umgesetzt. (10+2 PM)

**AP3.2.3: IT-Architekturberatung** (Januar 2022–April 2024; **FhG, JLU**). Consulting für die einzelnen Services und Demonstrator-Projekte, um eine Cloud-native Architektur zu implementieren basierend auf den Gaia-X-Standards und damit einhergehend auf der in AP 3.1 entwickelten konzeptionellen Architektur. (6,5 PM)

**AP3.2.4: DevOps Phase 2** (Januar 2022–April 2024; **FhG, JLU**). Einrichtung von CI/CD-Pipelines und notwendiger Infrastruktur für die Phase 2-Demonstratoren und die entwickelten Services. (8,5 PM)

**AP3.2.5: Cluster-Provisioning Phase 1** (Mai–Dezember 2021, **JLU**). Integration der Hardware sowie Deployment und Management des Kubernetes-Clusters. (1 PM)

**AP3.2.6: Cluster-Provisioning Phase 2** (Januar 2022–April 2024, **JLU**). Betrieb und Management des Kubernetes-Clusters. Weiterentwicklung der Kubernetes- und Storage-Infrastruktur im Hinblick auf Gaia-X-Kompatibilität. (2 PM)

**Ergebnisse:**

**E3.2.1: Deployment der Phase-1-Demonstratoren abgeschlossen** (Software, Dezember 2021, MS1)

**E3.2.2: Deployment der Phase-2-Demonstratoren abgeschlossen** (Software, Februar 2024, MS3)

**E3.2.3: Erste Einrichtung des Kubernetes-Clusters** (Software, September 2021, MS1)

|                |                              |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
|----------------|------------------------------|------|-----|-----|-----|----|------|-----|------|------|------|-----|-----|------|-----|------|----------|
| <b>AP3.3</b>   | <b>Storage-Infrastruktur</b> |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Dauer</b>   | Mai 2021–Mai 2024            |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Leitung</b> | JLU                          |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Partner</b> | FhG                          | NFDI | JLU | TIB | UzK | UL | RWTH | FIZ | UKHD | HAdW | EMBL | UMR | WWU | EKUT | ZBW | Atos | $\Sigma$ |
| <b>PM</b>      | 6,5<br>+2,5                  |      | 31  |     |     |    |      |     |      |      |      |     |     |      |     |      | 40       |

**Ziele:**

- Umsetzung von Standards bezogen auf Gaia-X und NFDI und technische Anforderungen aus den Demonstrator-Projekten bei der Entwicklung einer Datenspeicher-Architektur
- Implementierung einer Datenspeicher-Architektur als versatile Cloud-native Lösung zur strukturierten Speicherung, Nutzung und Weitergabe von Daten im Sinne der FAIR-Prinzipien.

In diesem Arbeitspaket werden die technischen Voraussetzungen geschaffen, um den Datenaustausch zwischen verschiedenen NFDI- und Gaia-X-Datenräumen zu ermöglichen. Dazu wird die in NFDI4Biodiversity vorgesehene Datenspeicher-Architektur (**COre stoRage Engine - CORE**) an die in AP3.1 und AP3.2 entworfene Referenzarchitektur und die dort als relevant identifizierten Standards angepasst und entsprechend weiterentwickelt. CORE ist ein hoch skalierbares System für die systematische, performante und effiziente Speicherung großer heterogener Daten. Es basiert auf etablierten Cloud-Standards wie Objektspeicher und Kubernetes und wird im Laufe dieses Projekts um zusätzliche Funktionalitäten erweitert.

Dazu werden die Schnittstellen der in AP3.5–3.7 entwickelten Services in die Speicherarchitektur integriert, um den vollen Umfang der Funktionalitäten den in AP4 skizzierten Demonstrator-Projekten zur Verfügung zu stellen. AP3.3.1 wird in der ersten Projektphase umgesetzt, alle weiteren Arbeitspakete werden in der zweiten Phase bearbeitet.

**Arbeitsschritte:**

**AP3.3.1: Grundlagen für Demonstrator-Projekte** (Mai 2021–Dezember 2021; FhG, **JLU**). In der ersten

Projektphase werden zunächst die notwendigen technischen Grundlagen gelegt (insbesondere durch Support, Consulting und Schulungen, sowie durch initiale Deployments), um die Speicherarchitektur in die Demonstrator-Projekte einbinden zu können. Hierzu gehört insbesondere das CORE-Deployment in mindestens zwei Datenzentren, wobei jeweils eines im NFDI- und eines im Gaia-X-Kontext angesiedelt ist. Es werden erforderliche technische Anpassungen vorgenommen, um die Operabilität zu gewährleisten. Die zentrale Speicherarchitektur wird an die unmittelbaren Bedürfnisse der Demonstrator-Projekte unter Berücksichtigung der in AP3.1 gesammelten Grundlagen angepasst und es werden benötigte Funktionalitäten hinzugefügt. (6,5 PM)

**AP3.3.2: Allgemeine Weiterentwicklung der Speicherarchitektur CORE** (Januar 2022–Mai 2024; FhG, JLU). In der zweiten Projektphase werden weiterhin Bedarfe und Anforderungen der Demonstrator-Projekte gesammelt und entsprechende Funktionalitäten implementiert, sofern dies nicht bereits unter den weiteren Arbeitsschritten in diesem Arbeitspaket geschieht. Weiterhin wird die Gesamtarchitektur basierend auf den in AP3.2 gesammelten Grundlagen weiterentwickelt. (7 PM)

**AP3.3.3: Integration der Authentifizierungs- und Autorisierungsinfrastruktur** (Juni 2022–Mai 2024; FhG, JLU). Im Verlauf der zweiten Projektphase wird die in AP3.5 entwickelte Authentifizierungs- und Autorisierungsinfrastruktur als Zugriffsschutzmechanismus integriert und die Interoperabilität mit anderen Komponenten, insbesondere den Demonstrator-Projekten getestet. (7 PM)

**AP3.3.4: Integration des (Meta-)Daten-Validierungsdienstes** (Juni 2022–Mai 2024; JLU). Der in AP3.6 entwickelte Validierungsdienst wird in die Speicherarchitektur integriert. Hierbei sollen Daten und Metadaten manuell und/oder automatisch gegen vom Benutzer definierte und ausgewählte Ontologien getestet werden können. Darüber hinaus kann die Freigabe von sensiblen Daten bei Bedarf zurückgehalten werden. Weiterhin wird die Implementierung so ausgestaltet, dass weitere Validierungsdienste eingebunden werden können. Zusätzlich wird es möglich sein, die FAIRness einzelner Datensätze zu überprüfen. (5 PM)

**AP3.3.5: Integration der Interfaces für verteilte Analysen** (Juni 2022–April 2024; JLU). In diesem Teilpaket werden die Schnittstellen für eine verteilte Datenanalyse implementiert. Diese basieren maßgeblich auf den in AP3.7 für NFDI4Health entwickelten Services. Die Schnittstellen sollen nicht nur für diesen speziellen Service, sondern auch für weitere Implementierungen nutzbar sein. Hierbei liegt der Schwerpunkt auf zwei Nutzungsszenarien. Zum einen sollen mit Hilfe der verteilten Analyse personenbezogene Daten ohne direkte Weitergabe dieser Daten analysiert werden können, zum anderen soll es ermöglicht werden, Datensätze, die zu groß sind, um sie sinnvoll übertragen zu können, lokal am Ort der Datenaufbewahrung zu analysieren. (6 PM)

**AP3.3.6: PID Service** (Juni 2022–April 2024; FhG, JLU). Der CORE-Storage enthält bereits eine Komponente, um abgespeicherten Objekten individuelle IDs zuzuweisen. Die Zuweisung dieser IDs wird in diesem Unter-AP weiterentwickelt, um mit dem in AP3.4.2 entwickelten FAIR (Meta-)PID-Resolver kompatibel zu sein. Initial wird die interne Vergabe von einzigartigen IDs für Datenobjekte implementiert basierend auf den Vorgaben von AP3.4.2. Die Einbindung der Meta-PID-bezogenen Komponenten wird im Folgenden realisiert. (6+2,5 PM)

#### **Ergebnisse:**

**E3.3.1: Stabiles Deployment der Speicherarchitektur für Demonstrator-Projekte** (Software, November 2021, MS1)

**E3.3.2: Implementierung des internen PID Service verfügbar** (Software, Januar 2023, MS2)

**E3.3.3: Erstes stabiles Release mit vollständiger Integration aller Services** (Software, Dezember 2023, MS3)

|                |   |             |            |            |            |           |             |            |             |             |             |            |            |             |            |             |            |
|----------------|---|-------------|------------|------------|------------|-----------|-------------|------------|-------------|-------------|-------------|------------|------------|-------------|------------|-------------|------------|
| <b>AP3.4</b>   | <b>Compliance &amp; FAIRness Monitoring</b> |             |            |            |            |           |             |            |             |             |             |            |            |             |            |             |            |
| <b>Dauer</b>   | Januar 2022–Mai 2024                        |             |            |            |            |           |             |            |             |             |             |            |            |             |            |             |            |
| <b>Leitung</b> | JLU   |             |            |            |            |           |             |            |             |             |             |            |            |             |            |             |            |
| <b>Partner</b> | <b>FhG</b>                                  | <b>NFDI</b> | <b>JLU</b> | <b>TIB</b> | <b>UzK</b> | <b>UL</b> | <b>RWTH</b> | <b>FIZ</b> | <b>UKHD</b> | <b>HAdW</b> | <b>EMBL</b> | <b>UMR</b> | <b>WWU</b> | <b>EKUT</b> | <b>ZBW</b> | <b>Atos</b> | <b>Σ</b>   |
| <b>PM</b>      | 6<br>+2                                     |             | 12         |            | 4          |           | 8,5         |            |             |             |             |            |            |             |            |             | 30,5<br>+2 |

**Ziele:**

- Umsetzung und Überwachung der Einhaltung der FAIR-Prinzipien
- Implementierung und Betrieb eines Services zur Überwachung der in AP2 definierten Compliance-Kriterien

Dieses Arbeitspaket adressiert die Entwicklung der benötigten Software-Services, um Daten rechtssicher und gemäß der FAIR-Prinzipien auszutauschen und die Konformität mit diesen Regeln zu überwachen.

**Arbeitsschritte:**

**AP3.4.1: Compliance Monitoring** (Juni 2022–Mai 2024; **JLU**) implementiert einen Service, der basierend auf den in AP2 identifizierten Kriterien zum Datenaustausch zwischen Gaia-X und NFDI die Einhaltung derselben beim Austausch von Daten garantiert. Insbesondere findet eine Differenzierung zwischen verschiedenen Aspekten in Bezug auf Datenhaltung und Zugriffsmuster statt. Diese umfassen beispielsweise die Überwachung der Einhaltung von Richtlinien bei der Speicherung und Nutzung von sensiblen bzw. personenbezogenen Daten oder die Beachtung von rechtlichen Rahmenbedingungen zum Austausch von nicht-personenbezogenen Daten. (10 PM)

**AP3.4.2: FAIR Data Registries** (Juni 2022–Mai 2024; **JLU, UzK, RWTH**) liefern eine Schnittstellendefinition für die Abfrage von Daten und Metadaten auf Basis eines persistenten Identifiers auf Grundlage der FAIR-Prinzipien. Entlang der Empfehlungen des EOSC Interoperability Framework werden dafür Schnittstellen bestehender Meta-PID-Resolver, wie N2T, erweitert. Diese Schnittstellen erlauben auf Basis von PID Information Types eine Automatisierung der Abfrage von Daten-„Bit-Streams“ und Metadaten auf Basis verschiedener PID Systeme, sowie eine Abbildung in eine einheitliche Repräsentation auf Basis des W3C Data Catalog Vocabulary DCAT, auf dem auch das Gaia-X Self Description Schema aufbaut. (8,5 PM)

**AP3.4.3: FAIR Validation Services** (Juni 2022–November 2024; **FhG, JLU, UzK, RWTH**) evaluiert die Konformität von Daten mit den FAIR-Prinzipien. In der ersten Phase werden die Validierungsdienste das F-UJI Automated FAIR Data Assessment Tool mit einem IDS-Connector demonstriert. Die FAIR Data Object Assessment Metrics als abgeschlossene Menge werden benutzt, um die über den IDS-Connector bereitgestellten Metadaten zu prüfen. In der zweiten Phase werden die FAIR-Validierungsdienste auf der Basis der Ergebnisse der Durchführung von Roadmapping-Aktivitäten des AP1 weiterentwickelt. Dabei werden Schnittstellen zu verschiedenen NFDI-Konsortien untersucht und daraus abgeleitet weitestgehend standardisierte Schnittstellen entwickelt. (12+2 PM)

**Ergebnisse:**

**E3.4.1: Demonstration der ersten Version der FAIR Validation Services mit IDS-Konnektoren** (Software, Dezember 2021, MS1)

**E3.4.2: PoC für das Compliance Monitoring** (Software, Februar 2023, MS2)

**E3.4.3: Erste stabile Version des Compliance-Monitoring Service** (Software, November 2023, MS3)

|              |                            |
|--------------|----------------------------|
| <b>AP3.5</b> | <b>FAIR Identity Space</b> |
|--------------|----------------------------|

|                |                   |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |    |
|----------------|-------------------|------|-----|-----|-----|----|------|-----|------|------|------|-----|-----|------|-----|------|----|
| <b>Dauer</b>   | Mai 2021–Mai 2024 |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |    |
| <b>Leitung</b> | RWTH              |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |    |
| <b>Partner</b> | FhG               | NFDI | JLU | TIB | UzK | UL | RWTH | FIZ | UKHD | HAdW | EMBL | UMR | WWU | EKUT | ZBW | Atos | Σ  |
| <b>PM</b>      | 6                 |      |     |     |     |    | 31   |     |      |      |      |     |     |      |     |      | 37 |

**Ziele:**

- Synchronisierung verschiedener Konzepte zur Autorisierung in NFDI und Gaia-X
- Verknüpfung bestehender Authentifizierungsinfrastrukturen
- Schaffung eines interoperablen, föderierten „Identity Space“ für NFDI

Dezentrale und föderierte Authentifizierungstechnologien sind in wirtschaftlichen sowie in wissenschaftlichen Anwendungen weit verbreitet. AP3.5 setzt auf die bestehenden Entwicklungen und Infrastrukturen um einen einheitlichen „Identity Space“ für Anwendungen von NFDI zu schaffen, die auf den Gaia-X Federation Services für Identity & Trust aufbauen. Wie der gesamte FAIR Dataspace sollen auch Identitäten den FAIR-Prinzipien entsprechen.

**Arbeitsschritte:**

**AP3.5.1: Authentifizierungsframeworks, Phase 1** (Mai–Dezember 2021; FhG, RWTH). Frameworks, die bereits in Services von NFDI Anwendung finden, werden zunächst katalogisiert und gemeinsame Schnittstellen herausgearbeitet. Dabei wird erfasst, welchen Anforderungen die aktuellen Implementierungen genügen müssen. Auf Basis des Katalogs können dann Brückentechnologien bereitgestellt werden um die bestehenden Implementierungen zu verbinden. Hierbei ist die transparente Verknüpfung verschiedener Authentifizierungsprotokolle, insbesondere SAML2 im akademischen und OAuth2/OIDC im kommerziellen (sog. „Social Sign On“) Umfeld, zu planen und die notwendigen Rahmenbedingungen für eine solche zu schaffen. (5 PM)

**AP3.5.2: (Meta-)Authentifizierungsinfrastrukturen (AAls), Phase 2** (Januar 2022–April 2024; RWTH). Auf Basis der Erkenntnisse aus AP3.5.1 und der Einbindung von AAls, die bereits in verschiedenen Cloud-Szenarien in Deutschland und im europäischen Kontext etabliert sind, wird ein Kern-Dienst für die Nutzerauthentifizierung implementiert. So sind im Kontext der EOSC bereits technologische Rahmenbedingungen abgesteckt, wie z.B. durch die „AARC II Blueprint Architecture“, oder Meta-Identity-Provider wie „Unity“, um die Basissoftware zu liefern, mit denen entsprechende AAls als Service im Kontext von NFDI aufgebaut werden können. Dabei sind insbesondere die bestehenden Entwicklungen für föderiertes „Single-Sign-On“ wie die eduGAIN-Föderation zu integrieren und eine Nutzung innerhalb der Gaia-X-Infrastruktur zu ermöglichen. (6PM)

**AP3.5.3: Schnittstellenspezifikation für FAIRe Autorisierungsservices** (Mai 2022–Oktober 2022; FhG, RWTH). Ein weiter Grundpfeiler ist der nahtlose Austausch von virtuellen Identitäten und Berechtigungen. Dabei ist zwischen bestehenden, zentralen sog. „Secure Token Services“ oder dezentralen Services auf Basis kryptografischer Signaturen oder Claims zu unterscheiden. Ziel ist es insbesondere die Schnittstellen zum Austausch der Autorisierungs- und Delegationsinformationen abzustimmen und zu vereinheitlichen und entsprechende Services für das Verwalten von Zugriffsrechten bereitzustellen. Dabei dient die „Web Access Control“ Spezifikation als Grundlage. (7 PM)

**AP3.5.4: Dezentrale Autorisierungs- und Delegationsservices** (Oktober 2022–Dezember 2024; FhG, RWTH). Auf Basis der Schnittstellendefinition liefern dezentrale Autorisierungsservices im Sinne eines „Web of trust“ gesicherte Zugriffsinformationen für die digitalen Objekte oder Ressourcen. Insbesondere das, von NFDI geforderte, Multi-Cloud Umfeld stellt hier besondere Herausforderungen für die Dezentralität, sodass die Durchsetzung der Berechtigungen dabei zumeist in den datenhaltenden Systemen verbleibt. (7 PM)

**AP3.5.5: Gruppen und Eigenschaften** (Dezember 2022–Mai 2023; RWTH). Ziel des AP ist die Identifikation von Individuen und Organisationen als Entitäten und die Übertragung von Informationen (sog. Claims) zu diesen Entitäten zwischen den beteiligten Systemen. Für die Identifikation lassen sich bestehende PID services wie ORCID nutzen. Jedoch definieren z.B. eduPersonSchema und OIDC unterschiedliche Claims für Personen, die ineinander überführt und im Sinne der FAIR-Prinzipien mit der ORCID verbunden werden müssen. (6PM)

**AP3.5.6: FAIRe Organisationen und Organisationszuordnung** (Mai 2023–April 2024; RWTH).

Ähnlich wie im Falle von Personendaten können Organisationen mit der ROR-ID, internen Organisationsverzeichnissen und eduGAIN entityIDs beschrieben werden. Der Kern-Service auf Basis der „Organization Ontology“ stellt diese Verbindung her und liefert Schnittstellen für die Definition und Abfrage von Organisationsstrukturen und Zugehörigkeiten. (6 PM)

**Ergebnisse:**

**E3.5.1: Katalog von Authentifizierungsframeworks in Gaia-X und NFDI** (Bericht, Dezember 2021, MS1)

**E3.5.2: Authentifizierungsservice auf Basis der „AARC II BPA“ (beta)** (Software, August 2022)

**E3.5.3: Schnittstellendefinition für den Autorisierungsservice auf Basis von „Web Access Control“** (Bericht, Oktober 2022)

**E3.5.4: Autorisierungsservice (beta)** (Software, Dezember 2022, MS2)

**E3.5.5: Schnittstellendefinition für Organisationsmitgliedschaften auf Basis der „Organization Ontology“** (Bericht, Mai 2023)

**E3.5.6: Organisationsservice (beta)** (Bericht, September 2023)

**E3.5.7: Authentifizierungsservice, Autorisierungsservice und Organisationsservice (final)** (Software, Februar 2024, MS3)

**E3.5.8: Dokumentation der Schnittstellen und Implementierung** (Bericht, April 2024, MS3)

|                |  |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
|----------------|--|------|-----|-----|-----|----|------|-----|------|------|------|-----|-----|------|-----|------|----------|
| <b>AP3.6</b>   | <b>(Meta)data Terminologies &amp; Validation</b> |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Dauer</b>   | Januar 2022–Mai 2024                             |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Leitung</b> | TIB  |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Partner</b> | FhG  | NFDI | JLU | TIB | UzK | UL | RWTH | FIZ | UKHD | HAdW | EMBL | UMR | WWU | EKUT | ZBW | Atos | $\Sigma$ |
| <b>PM</b>      | 6<br>+6  | 2    |     | 34  |     |    | 6    |     |      |      |      |     |     |      |     |      | 46<br>+6 |

**Ziele:**

- (Meta)daten-Interoperabilität und Nachnutzbarkeit ermöglichen, indem (Meta)daten mit formaler Semantik beschrieben werden
- Aufbau Terminology Service für Zugang zu und Kuratierung von Terminologie und deren Semantik
- Einbindung Terminology Service in mehrwertschaffenden Diensten, z.B. Datenkuratierung in Datenrepositorien, (Meta)daten-Validierung, etc.

**Arbeitsschritte:**

**AP3.6.1: Terminologie Landscaping** (Januar–März 2022; TIB). Aufbau eines Katalogs mit relevanten existierenden Terminologien (kontrolliertes Vokabular, Taxonomien, Thesauri, Ontologien) und Ermittlung, ob diese Ressourcen die Terminologie-Bedarfe für FAIR Data Spaces decken, resp. welche Lücken bestehen (z.B. fehlende Terme, Duplikate, fehlende/ungenau Semantik, etc.) und mit welcher Priorität diese adressiert werden müssen um die FAIR Data Space Demonstratoren zu unterstützen. Es wird aufgebaut auf den Vorarbeiten der NFDI-Metadaten-Initiative. (1 PM)

**AP3.6.2: Terminologie Qualitätssicherung** (Februar 2022–Juli 2022; TIB). Auf Basis der Erkenntnisse aus AP3.6.1 soll der Terminologie-Katalog so bereinigt werden, dass die wichtigsten Bedarfslücken adressiert werden. Ziel ist es die Terminologie-Qualität zu sichern und die praktische Nutzung von Terminologie in FAIR Data Spaces, insb. in den Demonstratoren, zu ermöglichen. (5 PM)

**AP3.6.3: Terminology Service** (April 2022–Dezember 2023; TIB). Der bereinigte Terminologie-Katalog wird mittels eines Terminology Service über das Web Menschen und Maschinen zugänglich gemacht. Der Dienst soll den (maschinellen) Terminologie-Lookup sowie die Abfrage der Terminologie-Semantik unterstützen. Außerdem, soll der Dienst die nachhaltige, kollaborative Terminologie-Kuratierung ermöglichen. Er soll auf den verwandten Entwicklungen in diversen NFDI-Konsortien (insb. NFDI4Ing, NFDI4Chem, NFDI4BioDiversity) aufbauen, diese nachnutzen und für FAIR Data Spaces anpassen. Im Vergleich zu NFDI, wo Interoperabilität und Nachnutzung über Disziplinen hinweg eine wichtige Bedeutung hat, ist in Gaia-X Interoperabilität und Nachnutzbarkeit insbesondere innerhalb einer Organisation wichtig. Dieser Unterschied zwischen Wirtschaft und Wissenschaft wird berücksichtigt. Die Basisinfrastruktur bildet dabei insb. EBI-OLS für den Zugang, WebProtégé für die Kuratierung, Git für die Versionierung, WebVOWL für die Visualisierung und wird für die Bedarfe des FAIR Data Space angepasst. (11 PM)

**AP3.6.4: Terminology Service Einbindung in Datenzentren** (Juni 2022–Mai 2024; TIB, RWTH). Ziel ist es den Terminology Service und die über den Dienst zugänglichen Terminologien in die Datenkuratierung an Datenzentren/-repositorien/-archiven einzubinden. Erprobt wird die Einbindung zunächst im Rahmen der FAIR Data Space Demonstratoren. (8 PM)

**AP3.6.5: Terminology Service Einbindung in Schema- und Profildefinition** (November 2022–Mai 2024; TIB, RWTH, FhG). Die Auffindung, Interoperabilität und Nachnutzung von Daten und deren Selbstbeschreibung (Metadaten) beruhen auf anerkannten Schema- und Profildefinitionen (z.B. schema.org) und deren breiter Anwendung in der Praxis. Diese beschreiben (Meta)daten Struktur (Attribute) und Datentypen. Dabei werden Wertebereiche nur selten mittels kontrollierter Terminologie definiert. Ziel ist diese Terminology-Service-Anwendung im Rahmen der FAIR Data Space Demonstratoren zu erproben. (8+6 PM)

**AP3.6.6: Terminology Service Einbindung in (Meta)daten Validierung** (Februar 2023–Mai 2024; FhG, TIB, RWTH). Auf Basis der Einbindung von Terminologie in Schema- und Profildefinition (AP3.6.5) untersucht AP3.6.6 den Mehrwert dieser Einbindung für die (Meta)daten Validierung mittels Schemata in, z.B., XSD oder SHACL wie sie z.B. von NFDI4Ing im Projekt AIMS erprobt werden. Dafür wird eine API entwickelt, die (Meta)daten als Bytestream mit einer Typdefinition (z.B. JSON) erwartet, diese validiert und eine Zusammenfassung der Validierung retourniert. Darüber hinaus, erprobt dieses AP auch die Einbindung des Dienstes in (Meta)daten Validierung entsprechend den FAIR Daten-Prinzipien und in entsprechenden Validierungs-Tools, z.B. FAIR Maturity Model Validation (F-UJI) und die CI-Pipeline des Gaia-X-Self-Description-Arbeitspakets gemeinsam mit AP 3.4.3. (7+6 PM)

**Ergebnisse:**

**E3.6.1: Terminology Service, Alpha-Version** (Software, Dezember 2022, MS2)

**E3.6.2: Terminology Service, Beta-Version** (Software, Juli 2023)

**E3.6.3: Terminology Service, finale Version** (Software, Dezember 2023, MS3)

**E3.6.4: Terminology-Service-Einbindungen** (Bericht, April 2024, MS3)

|                |                           |
|----------------|---------------------------|
| <b>AP3.7</b>   | <b>Verteilte Analysen</b> |
| <b>Dauer</b>   | Mai 2021–November 2023    |
| <b>Leitung</b> | FhG                       |

| Partner | FhG       | NFDI | JLU | TIB | UzK | UL | RWTH | FIZ | UKHD | HAdW | EMBL | UMR | WWU | EKUT | ZBW | Atos | Σ         |
|---------|-----------|------|-----|-----|-----|----|------|-----|------|------|------|-----|-----|------|-----|------|-----------|
| PM      | 18<br>+18 |      |     |     | 11  | 11 |      |     |      |      |      |     |     |      |     |      | 40<br>+18 |

### Ziele:

Die verteilte Analyseplattform Personal Health Train (PHT) führt Data-Science- und Machine-Learning-Aufgaben auf föderierten Datenquellen aus, ohne die Daten von ihrem ursprünglichen Standort zu entfernen. Dies inkludiert:

- Erweiterung und Anpassung des PHT um Anforderungen und Standards aus Gaia-X und NFDI
- Implementierung von Schnittstellen mit Gaia-X- und NFDI-Infrastrukturkomponenten

### Arbeitsschritte:

**AP3.7.1: Proof-of-Concept und Anforderungsanalyse** (Mai–Dezember 2021; **FhG**): In der ersten Phase wird die verteilte Analyseplattform Personal Health Train (PHT) eingesetzt und mit den bestehenden Kerndiensten integriert. Der PHT wird im AP4 NFDI4Health-Demonstrator verwendet, um Algorithmen an die verteilten Datenquellen heranzuführen und Lernen zu ermöglichen, ohne Daten physisch zu teilen. Die möglichen Schnittstellen zwischen bestehenden Komponenten wie Speicherarchitektur CORE (AP3.2), Gaia-X Federated Catalogue oder IDS-Connectoren werden analysiert; wenn möglich werden bestehende Dienste und Komponenten integriert. Ein Anwendungsfall für verteilte Analytik mit synthetischen Daten wird mit lokalen Datenpunkten von NFDI4Health demonstriert. Rückmeldungen und Validierungsergebnisse werden dokumentiert, zukünftige Integrationsanforderungen werden identifiziert (6+6 PM).

**AP3.7.2: Weiterentwicklung und Integration mit Gaia-X- und NFDI-Services** (Januar 2022–November 2023; **FhG**, **UzK**, **UL**): In der zweiten Phase werden plattform- und konsortiumsübergreifende Anwendungsfälle entwickelt und die verteilten Analysedienste um benötigte Schnittstellen und zusätzliche Funktionalitäten erweitert: Das Architekturkonzept und die verfügbaren Komponenten und Standards, die in AP3.1 identifiziert wurden, werden mit den bestehenden Komponenten und der Architektur des PHT verglichen. Zusammen mit den Ergebnissen von AP3.7.1 werden erforderliche Schnittstellen und notwendige Verbesserungen identifiziert und umgesetzt. (12+6 PM)

**AP3.7.3: FAIR data access** (März 2022–September 2023; **UL**, **UzK**): Es werden PHT-Datenkonnectoren für die Cloud-Dienste und NFDI-Datenzugriffspunkte implementiert. Die Konnectoren werden mit den in AP3.4.3 entwickelten Datenregistern kommunizieren und die in AP3.1 und 3.6 bereitgestellten Terminologien und Ontologien nutzen, um Metadaten im Zusammenhang mit verteilten Analyseaufgaben bereitzustellen. PHT-Stationen werden an Datenzugriffspunkten installiert. (11 PM)

**AP3.7.4: Trust-Framework-Implementierung** (Januar 2022–Juni 2023; **UzK**, **FhG**, **UL**): Es wird ein Trust-Framework entwickelt, das die Nachvollziehbarkeit und Transparenz der Ausführung von Algorithmen ermöglicht. Die in AP3.5 implementierten Dienste Authentifizierung & Autorisierung werden genutzt. FAIRness der Analyseaufgaben wird erzielt, indem der in AP3.3.6 entwickelte PID-Dienst verwendet wird und umfangreiche Metadaten über die ausgeführten Analyseaufgaben generiert sowie verwaltet werden. (11+6 PM)

### Ergebnisse:

**E3.7.1: Demonstration des Prototyps einer funktionalen, verteilten Analyseplattform mit Use Cases für die Gesundheitsversorgung** (Software, November 2021, MS1)

**E3.7.2: Anforderungsanalyse** (Bericht, Dezember 2021, MS1)

**E3.7.3: Integration PHT-Plattform: erste Version** (Software, November 2022, MS2)

**E3.7.4: Integration PHT-Plattform finale Version** (Software, November 2023, MS3)

**E3.7.5: Roll-out und Installation von PHT-Stationen Version 1** (Software, September 2022, MS2)

**E3.7.6: Roll-out und Installation von PHT-Stationen Version 2** (Software, September 2023, MS3)

**E3.7.7: Sicherheitskonzept für verteilte Analysen** (Bericht, Juni 2023)

## AP4: FAIR Data Space Demonstratoren (Leitung: UMR)

Die in AP3 erarbeiteten Grundlagen sollen in diesem Arbeitspaket genutzt werden, um die aus verschiedenen NFDI-Konsortien stammenden Demonstratoren zu erweitern und anzupassen, sodass spezifische Anforderungen von Gaia-X erfüllt werden. Insbesondere sollen Demonstratoren durch ein PoC nachweisen, wie die Multi-Cloud-Fähigkeit unterstützt, unter Verwendung einer Gaia-X-konformen Selbstbeschreibung auf Forschungsdaten und Services zugegriffen und ein datenschutzkonformer Umgang mit verteilten Forschungsdaten erzielt werden kann.

Die Demonstratoren werden in zwei Phasen entwickelt. In Phase 1 wird auf bereits in den NFDI-Konsortien vorhandenen und geplanten Projekten aufgebaut. Zunächst wird die Konzeption der Demonstratoren an die Gaia-X-Standards sowie die neu entwickelten Gaia-X kompatiblen Services angepasst. Außerdem werden Mock-up oder PoC Demonstratoren basierend auf den bereits vorhandenen Vorarbeiten entwickelt. Die Demonstratoren werden am Ende von Phase 1 auf verschiedene Gaia-X kompatiblen Infrastrukturen deployed um die Kompatibilität zu demonstrieren. In Phase 2 werden die in AP3 entwickelten Komponenten in die Demonstratoren integriert und durchgängig deployed um die Kompatibilität zu gewährleisten.

Über einen Open Call sollen weitere Demonstratoren gewonnen werden, insbesondere aus der Industrie, um das Zusammenspiel von wissenschaftlichen und industriellen Services aufzuzeigen.

Zur Demonstration der Erkenntnisse aus AP2, und um zusätzlich zur Bereitstellung technischer Lösungen auch Kompetenzen aufzubauen, wird ein Schulungsangebot zu ELSA-Themen erstellt.

|   |  |      |     |     |     |    |      |     |      |      |      |      |     |      |     |      |      |
|---|--|------|-----|-----|-----|----|------|-----|------|------|------|------|-----|------|-----|------|------|
| <b>AP4.1</b>  | <b>FAIR-DS Demonstrator NFDI4Biodiversity and Gaia-X</b> |      |     |     |     |    |      |     |      |      |      |      |     |      |     |      |      |
| <b>Dauer</b>  | Mai 2021–Mai 2024  |      |     |     |     |    |      |     |      |      |      |      |     |      |     |      |      |
| <b>Leitung</b>  | UMR  |      |     |     |     |    |      |     |      |      |      |      |     |      |     |      |      |
| <b>Partner</b>  | FhG  | NFDI | JLU | TIB | UzK | UL | RWTH | FIZ | UKHD | HAdW | EMBL | UMR  | WWU | EKUT | ZBW | Atos | Σ    |
| <b>PM</b>   |  |      |     |     |     |    |      |     |      |      |      | 32,5 |     |      |     |      | 32,5 |
| * Es werden zusätzlich 7 Personenmonate über einen Unterauftrag an Geo Engine eingebracht.  |  |      |     |     |     |    |      |     |      |      |      |      |     |      |     |      |      |
| <b>Ziele:</b>   |  |      |     |     |     |    |      |     |      |      |      |      |     |      |     |      |      |
| Phase 1:  |  |      |     |     |     |    |      |     |      |      |      |      |     |      |     |      |      |
| <ul style="list-style-type: none"> <li>• Bereitstellung eines One-Click Kubernetes-Deployment der Geo Engine für die de.NBI-Cloud und Gaia-X.</li> <li>• Umsetzung von User Stories für die Verknüpfung und Verarbeitung von zwei Datenräumen (aus NFDI und Industrie).</li> </ul>  |  |      |     |     |     |    |      |     |      |      |      |      |     |      |     |      |      |
| Phase 2   |  |      |     |     |     |    |      |     |      |      |      |      |     |      |     |      |      |
| <ul style="list-style-type: none"> <li>• Erweiterung des initialen Demonstrators um die vier bei Gaia-X zentralen Federation Services: Identity &amp; Trust, Federated Catalogue, Sovereign Data Exchange und Compliance.</li> <li>• Fortschreibung von User-Stories und Anbindung weiterer Daten und Services.</li> </ul>  |  |      |     |     |     |    |      |     |      |      |      |      |     |      |     |      |      |
| Die Geo Engine ist eine Plattform für eine interaktive Geodatenanalyse mit einer Web-basierten Benutzerschnittstelle. Mit ihr können Daten bereitgestellt werden, sodass Benutzer diese kombinieren und in einem Datenportal oder einer Analyse-App visuell explorieren können. So wird auf Basis der Geo Engine in den Projekten GFBio und NFDI4BioDiversity das Geoportal VAT betrieben. Das Start-up mit dem |  |      |     |     |     |    |      |     |      |      |      |      |     |      |     |      |      |



Namen Geo Engine sorgt für die Weiterentwicklung der Geo Engine als Open Source und die Verbreitung der Plattform als Daten- und Analyseportal in unterschiedlichen Projekten.

Im Rahmen dieses Showcase soll unter Mitwirkung des Start-ups die Geo Engine in der Gaia-X-Cloud und VAT-ähnliche Dienste sowohl der Forschung als auch der Industrie bereitgestellt werden. Die Geo Engine kann dann direkt auf Cloud-gehostete Daten zugreifen, die von NFDI-Konsortien in Repositorien zur Verfügung gestellt werden. Insbesondere sollen hierbei die Datenströme aus dem Ammod-Projekt eingebracht werden. Die Elastizität der Cloud erlaubt eine Skalierung der Geo Engine gemäß Anzahl und Komplexität der Nutzung. Über das föderierte Identitätsmanagement können die Daten rechtskonform verwaltet und zugänglich gemacht werden. Dadurch ergibt sich durch die Verwendung der Geo Engine für Benutzer aus Industrie und Wissenschaft ein echter Mehrwert.

#### **Arbeitsschritte:**

**AP4.1.1: Anforderungsanalyse** (Mai 2021; **UMR**). Es werden zunächst Anforderungen für den Demonstrator definiert, die in der ersten Phase umgesetzt werden sollen. Dabei werden sowohl Datensätze aus dem Kontext von NFDI4Biodiversity und dem industriellen Umfeld ausgewählt als auch die darauf basierenden Workflows definiert. (1 PM)

**AP4.1.2: Datenkonnektor** (Juli 2021; **UMR**). Für den Zugriff auf die NFDI-Daten wird ein erster Konnektor im Demonstrator bereitgestellt, der sich mit einer Gaia-X-konformen Selbstbeschreibung beschreibt. (1 PM)

**AP4.1.3: Automatisches Deployment** (September 2021; **UMR**). Es wird ein automatisches Deployment (one-click) von Geo-Engine-Instanzen auf einer Gaia-X-Cloud bereitgestellt, in der die Unternehmensdaten verfügbar sind. (1 PM)

**AP4.1.4: User Stories auf Basis einer Verknüpfung von Datenräumen** (Oktober–Dezember 2021; **UMR**). Es sollen User Stories für den Demonstrator entwickelt werden, bei denen öffentliche Daten (z.B. denen aus Ammod) und Unternehmensdaten kombiniert und analysiert werden. (3 PM)

**AP4.1.5: Evaluation** (Januar–März 2022; **UMR**). Auf Basis der Rückmeldung potentieller Nutzer auf die User Stories werden Schwächen und Stärken des Demonstrators identifiziert, die in die Konzeption der zweiten Phase eingehen. (2,5 PM)

**AP4.1.6: Autorisierung** (April–September 2022; **UMR**). Die in der ersten Phase prototypische SSO-Authentifizierung wird verfeinert und eine Autorisierung gemäß Gaia-X-Spezifikation realisiert. (6 PM)

**AP4.1.7: Umsetzung und Nutzung eines Federated Catalogue** (November 2022–April 2023; **UMR**). Es wird ein Federated Catalogue für Daten und Services im Demonstrator angebunden. Damit werden die Ressourcen (Daten und Workflows) über Standardschnittstellen bereitgestellt. Es wird gezeigt, wie andere Instanzen der Geo-Engine und Softwarekomponenten von Gaia-X über den Katalog auf diese Ressourcen zugreifen. (6 PM)

**AP4.1.8: Datenhoheit & Compliance** (August 2023– Januar 2024; **UMR**). Auf Basis der Gaia-X-Autorisierung im Rahmen des Sovereign Data Exchanges wird ein Ownership-Modell umgesetzt, welches Datenhoheit sowie Rechtevergabe (z.B. Rollenmanagement in Firmen) anbietet. Bezüglich Compliance wird die Registrierung als Gaia-X-Provider angestrebt sowie EU-Regularien (DSGVO) umgesetzt. (6 PM)

**AP4.1.9: Erweiterte User Stories** (Oktober 2022–April 2024; **UMR**). Auf Grundlage der erweiterten Funktionalität und der Verfügbarkeit weiterer Daten aus NFDI und Industrie werden die User Stories aus der ersten Phase überarbeitet und erweitert. Dabei wird das Querschnittsthema Skalierbarkeit und die Multi-Cloud-Fähigkeit adressiert. (6 PM)

**Ergebnisse:****E4.1.1: Vollständige Spezifikation der Anforderungen, Phase 1** (Bericht, August 2021)**E4.1.2: Demonstrator und User Stories, Phase 1** (Software, Dezember 2021, MS1)**E4.1.3: Evaluation erste Phase und Spezifikation der Aufgaben für die zweite Phase** (Bericht, März 2022)**E4.1.4: Autorisierung, Phase 2** (Software, September 2022)**E4.1.5: Federated Catalogue, Phase 2** (Software, Februar 2023, MS2)**E4.1.6: Ownership-Modell & Compliance** (Software, Oktober 2023, MS3)**E4.1.7: Erweiterter Demonstrator und User Stories, Phase 2** (Software, Januar 2024, MS3)**E4.1.8: Evaluation und Dokumentation, Phase 2** (Bericht, April 2024, MS3)

|                |   |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
|----------------|---|------|-----|-----|-----|----|------|-----|------|------|------|-----|-----|------|-----|------|----------|
| <b>AP4.2</b>   | <b>FAIR Research Data Quality Assurance and Workflows</b> |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Dauer</b>   | Mai 2021–Mai 2024   |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Leitung</b> | RWTH  |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Partner</b> | FhG   | NFDI | JLU | TIB | UzK | UL | RWTH | FIZ | UKHD | HAdW | EMBL | UMR | WWU | EKUT | ZBW | Atos | $\Sigma$ |
| <b>PM</b>      |   |      |     | 6   |     |    | 25   |     |      |      |      |     |     |      |     |      | 31       |

**Ziele:**

- Phase 1: Bereitstellung eines Demonstrators zur Qualitätssicherung von Forschungsdaten auf Basis der Open-Source-Software GitLab und der Gaia-X-Infrastruktur
- Phase 1: Demonstration der Verknüpfung von Private und Public Cloud („Hybrid-Cloud“) als Scale-Out Szenario
- Phase 2: Erweiterung des Demonstrators zur Anbindung an die in AP3 erarbeiteten Infrastrukturkomponenten, insbesondere Synchronisation von Gruppen und Projektmitgliedern
- Phase 2: Bereitstellung eines Katalogs für erprobte Qualitätssicherungs- und Analyseworkflows
- Phase 2: Spezifikation der Schnittstellen und Übertragbarkeit auf ein Multi-Cloud-Umfeld
- Phase 2: Erarbeitung eines Bewirtschaftungskonzepts für die genutzten Ressourcen in einem Public-Cloud- oder Multi-Cloud-Umfeld

Der Demonstrator in Phase 1 zeigt auf Basis der Open-Source-Software GitLab, wie weitgehend händisch erhobene Datensätze (z.B. Messdaten oder Software-Quellcodes) automatisiert qualitätsgesichert werden können. Der Demonstrator nutzt dafür die von GitLab bereitgestellte Workflow-Engine „Runner“ um Jobs zu definieren und in einer Kubernetes Infrastruktur zu orchestrieren.

Mit der Infrastruktur aus Phase 1 sollen in Phase 2 die Möglichkeiten des Demonstrators für Nutzende einfacher zugänglich gemacht werden. Dafür soll insbesondere ein Katalog für speziell kuratierte Workflows etabliert werden, auf den Nutzende zurückgreifen können. Die damit einhergehenden, erhöhten Anforderungen an die Workflow-Infrastruktur soll zusätzlich durch ein entsprechendes wissenschaftsgeleitetes Bewirtschaftungskonzept für das Scale-Out begleitet werden.

**Arbeitsschritte:**

**AP4.2.1: Anforderungsanalyse** (Mai–Juni 2021; RWTH). Zunächst werden geeignete Datensätze und Workflows aus dem Wirkungsbereich von NFDI4Ing für den Demonstrator ausgewählt. Auf Basis der Auswahl werden die Anforderungen an die Infrastruktur definiert. (3 PM)

**AP4.2.2: Aufbau Demonstrator Technische Grundlagen** (Juli–September 2021; RWTH). Implementierung der automatischen Allokation der entsprechenden Ressourcen („Scale-Out“) und Verbindung mit der Cloud-Infrastruktur über Gaia-X-Schnittstellen z.B. auf Basis von Terraform und Kubernetes. (4 PM)

**AP4.2.3: Aufbau Demonstrator Inhaltliche Grundlagen** (Oktober–Dezember 2021; RWTH, TIB).

Definition von Demo-Abläufen für Qualitätssicherung in verschiedenen Phasen des Forschungsdaten-Life-Cycle: Datenerfassung, Analyse und Archivierung. (6 PM)

**AP4.2.4: Wrap-Up Phase 1 & Ergebnissicherung** (Januar–März 2022; RWTH). Zusammenfassen der Ergebnisse und „lessons-learned“ auf Basis des technischen Scale-Out-Szenarios. Reflexion und Übertragung auf AP4.2.5–4.2.7. (2 PM)

**AP4.2.5: Katalog für Qualitätssicherungs- und Analyseworkflows** (November 2022–September 2023; RWTH, TIB). Um die Verfügbarkeit von Workflows zu erhöhen soll zunächst ein Katalog von verbreiteten Definitionen bereitgestellt werden. Der Katalog soll in einem Open-Source-Modell von Freiwilligen, sog. Maintainern, kuratiert werden und Workflows sollen als Docker-Registry bereitgestellt werden. Die im Katalog angebotenen Workflows sollen dabei den FAIR-Prinzipien entsprechen. (9 PM)

**AP4.2.6: Erweiterung des Demonstrators** (Juni 2023–März 2024; RWTH). Anbindung des Demonstrators an weitere Kern-Services, insbesondere in den Bereichen AAI zur Synchronisation von Projektmitgliedern (AP 3.5), sowie Sicherstellen der Verbindung mit dem in AP4.2.2 entwickelten Katalog der Analyseworkflows. Damit verbunden Erweiterung des Scale-Out-Prinzips zur Berücksichtigung von Workflows mit hohem Ressourcenbedarf (z.B. GPUs für KI-Trainings). (5 PM)

**AP4.2.7: Dokumentation und Bewirtschaftungskonzept** (September 2023–Mai 2024; RWTH). Ergebnissicherung der zwei Phasen des Demonstrators, sowie Dokumentation der etablierten Prozesse zur Übergabe im Sinne einer langfristigen Sicherstellung des Weiterbetriebs im Rahmen eines wissenschaftsgeleiteten Bewirtschaftungskonzepts. (2 PM)

**Ergebnisse:**

**E4.2.1: Evaluation der notwendigen Schnittstellen** (Bericht, August 2021)

**E4.2.2: Anbindung der GitLab-Infrastruktur von NFDI4Ing an Kubernetes-Infrastruktur** (Demonstrator, Dezember 2021, MS1)

**E4.2.3: Spezifikation „Blaupause“ für zukünftige Anbindungen** (Bericht, März 2022)

**E4.2.4: Infrastruktur für Katalog von Qualitätssicherungs- und Analyseworkflows** (Software, September 2023, MS2)

**E4.2.5: Erweiterter Demonstrator: Gruppensynchronisation** (Software, November 2023, MS3)

**E4.2.6: Erweiterter Demonstrator: Analyseworkflows mit hohem Ressourcenbedarf** (Software, Februar 2024, MS3)

**E4.2.7: Dokumentation und Bewirtschaftungskonzept** (Bericht, April 2024, MS3)

|                |  |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
|----------------|--|------|-----|-----|-----|----|------|-----|------|------|------|-----|-----|------|-----|------|----------|
| <b>AP4.3</b>   | <b>Cross-Plattform FAIR Datenanalyse</b> |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Dauer</b>   | Mai 2021–Mai 2024                        |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Leitung</b> | UzK                                      |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Partner</b> | FhG                                      | NFDI | JLU | TIB | UzK | UL | RWTH | FIZ | UKHD | HAdW | EMBL | UMR | WWU | EKUT | ZBW | Atos | Σ        |
| <b>PM</b>      | 4<br>+4                                  |      |     |     | 18  | 18 |      |     |      |      |      |     |     |      |     |      | 40<br>+4 |

**Ziele:**

- Demonstration der plattformübergreifenden datenschutzkonformen Datenanalyse auf verteilten Datensätzen, ohne die Daten selbst zu teilen
- Phase 1 wird die Interoperabilität zwischen den lokalen NFDI4Health-Datenpunkten (Local Access Points) und Gaia-X-konformen Komponenten demonstrieren. Da die Daten sensibel sind, werden wir verteilte Analysedienste anwenden, um die Daten zu analysieren. In der ersten Phase wird ein Anwendungsfall der Bildanalyse demonstriert.

- In Phase 2 werden neue Anwendungsfälle mit zusätzlichen Gaia-X-Komponenten inkl. Federation Services spezifiziert. Die an NFDI beteiligten Datenzentren werden die erforderlichen Daten kuratieren. Die PHT-Plattform (Personal Health Train) wird um zusätzliche, in der ersten Phase identifizierte Anforderungen erweitert, um Analyseaufgaben zwischen Gaia-X- und NFDI-Partnern durchzuführen.

#### **Arbeitsschritte:**

**AP4.3.1: Demonstrator-Entwurf** (Mai–August 2021; **UL**). Im Fokus steht ein Anwendungsfall zur Analyse medizinischer Bilddaten als Proof of Concept. Dazu wird der Anwendungsfall spezifiziert, die Bilddaten ausgewählt und um weitere realitätsnahe aber synthetisch generierte Daten erweitert. Darauf aufbauend werden die auszuführenden Analysen geplant und relevante vorhandene Algorithmen zur Datenanalyse ausgewählt. Abschließend werden die Erfolgskriterien spezifiziert. (3 PM)

**AP4.3.2: Demonstrator-Ausführung, Phase 1** (August 2021–Dezember 2021; **UzK, UL**). Installation von Services auf Seiten der Datengeber (Local Access Points). Ausführung der Analysen und Evaluierung der Ergebnisse. (5 PM)

**AP4.3.3: Identifikation weiterer Anwendungsfälle** (Januar 2022–Mai 2023; **FhG, UzK, UL**). Der Proof-of-Concept-Demonstrator wird den Gaia-X- und NFDI-Partnern präsentiert, um Feedback einzuholen. Neue Anwendungsfälle für die zweite Phase werden definiert und ihre Umsetzung entworfen und implementiert. Durch Studium der Ergebnisse der ersten Runde des Open Call aus AP4.4 werden weitere Szenarien für Anwendungsfälle eröffnet. (6+2 PM)

**AP4.3.4: Datenkuratierung** (März 2022–August 2023; **UzK, UL**). Die Datenzentren in Leipzig und Köln identifizieren verfügbare Datenquellen und Formate. Ethikanträge werden gestellt. Die benötigten Datenkuratierungsschritte werden ausgeführt. Dabei werden die FAIR-Data-Prinzipien beachtet, insbesondere dadurch, dass Metadaten sowohl die Daten als auch die durchlaufenen Prozessschritte beschreiben. (6 PM)

**AP4.3.5: Implementation von Algorithmen zur Datenanalyse** (Juli 2022–Oktober 2023; **UL, UzK**). Für die definierten Anwendungsfälle werden Machine-Learning- und weitere KI-Algorithmen entwickelt. In Abhängigkeit von den Anforderungen der Anwendungsfälle werden die Algorithmen für inkrementelles Lernen oder für föderiertes Lernen entwickelt. (6 PM)

**AP4.3.6: Installation von PHT-Stationen** (August 2022–November 2023; **UzK, UL**). PHT-Stationen werden in den Datenintegrationszentren installiert und eine sichere Verbindung zu den lokalen Datenquellen aufgebaut. Die Anbahnungs- und Ausführungsprozesse werden entwickelt und aufgebaut. (6 PM)

**AP4.3.7: Aufbau der zentralen PHT-Services** (März 2022–Dezember 2023; **FhG**). Aufbau, Installation und Etablierung von zentralen PHT-Services werden vorgenommen. (2+2 PM)

**AP4.3.8: Evaluierung der Ergebnisse der verteilten Analysen** (Oktober 2022–März 2024; **UL, UzK**). Die Ergebnisse der ausgeführten verteilten Analysen werden mit zuvor spezifizierten Metriken evaluiert. Zu diesen Metriken zählen bspw. Performanz (z.B. Laufzeit und Nutzung von / Anforderungen an Systemressourcen), Qualität der erzielten Ergebnisse (Accuracy) und Sicherheit. (6 PM)

#### **Ergebnisse:**

**E4.3.1: Entwurf des Demonstrators** (Bericht, August 2021)

**E4.3.2: Demonstrator, Phase 1** (Software, Dezember 2021, MS1)

**E4.3.3: Föderierter Datenkatalog v1** (Software, August 2022)

**E4.3.4: ML-Algorithmen v1** (Software, Oktober 2022, MS2)

**E4.3.5: ML-Algorithmen final** (Software, Oktober 2023)

**E4.3.6: Erweiterter Demonstrator v1** (Software, Januar 2023, MS2)

**E4.3.7: Erste weitere PHT Einsatzszenarien aufgenommen und beschrieben** (Bericht, Januar 2023, MS2)

**E4.3.8: Erweiterter Demonstrator v2** (Software, Februar 2024, MS3)

**E4.3.9: Weitere PHT Einsatzszenarien final beschrieben** (Bericht, Februar 2024, MS3)

**E4.3.10: Evaluation und Dokumentation** (Bericht, April 2024, MS3)

| AP4.4 Open call (2 Runden)   |                      |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
|--|----------------------|------|-----|-----|-----|----|------|-----|------|------|------|-----|-----|------|-----|------|----------|
| <b>Dauer</b>   | Januar 2022–Mai 2024 |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Leitung</b>   | FhG                  |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Partner</b>   | FhG                  | NFDI | JLU | TIB | UzK | UL | RWTH | FIZ | UKHD | HAdW | EMBL | UMR | WWU | EKUT | ZBW | Atos | Σ        |
| <b>PM</b>  | 12<br>+9             | 2    |     |     |     |    |      |     |      |      |      |     |     |      |     | 4    | 18<br>+9 |
| <p><b>Ziele:</b> Durch die Struktur des Verbunds ist die Wissenschaft im Projekt stark vertreten. Nun sollen Unteraufträge an Unternehmen vergeben werden, die sich an die im Projekt selbst geschaffenen Demonstratoren andocken. Ziele im Einzelnen sind:</p> <ul style="list-style-type: none"> <li>• Verstärkte Einbindung von Akteuren aus der Wirtschaft in das Projekt, insbesondere auch KMU</li> <li>• Breitere Nutzung der Demonstratoren durch weitere Akteure aus der Wissenschaft, insbesondere jedoch aus der Wirtschaft</li> <li>• Demonstration weiterer Szenarien des Datenaustauschs zwischen Wirtschaft und Wissenschaft</li> <li>• Breitere Abdeckung der Anwendungsdomänen, insbesondere auch solcher, die möglicherweise bei Gaia-X bis 2022/2023 weniger stark vertreten waren</li> <li>• Faire und rechtssichere Vergabe der Unteraufträge an diejenigen Unternehmer, die die beste Erfüllung der definierten Kriterien versprechen</li> </ul>   |                      |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <p><b>Arbeitsschritte:</b></p> <p>Die drei Schritte des Open Call werden in zwei Runden ausgeführt, einmal nach Fertigstellung von Phase 1 der Demonstratoren, ein zweites Mal so, dass zu Beginn der Bearbeitung der Unteraufträge die Demonstratoren der Phase 2 zur Verfügung stehen.</p> <p>Die in diesem Unter-AP kalkulierten Aufwände decken die Organisation des Verfahrens ab. Die inhaltlichen Kriterien werden in den Phasen „Definition“, „Bewertung“ und „Auswahl“ mit den weiteren NFDI-Partnern abgestimmt, denen jedoch hierfür nur geringe Aufwände im Rahmen ihrer eigentlichen AP entstehen. Die Gaia-X-Vertreter im Verbund (FhG, Atos) ziehen über den acatech organisierten Gaia-X Hub Germany und die internationalen Gaia-X-Arbeitsgruppen weitere Expert*innen hinzu.</p> <p><b>AP4.4.1: Definition und Ausschreibung, Phase 1</b> (Januar–Februar 2022; FhG, Atos): Es werden fachliche Kriterien für die Bewertung von Bewerbungen definiert und der vergaberechtliche Rahmen für eine Ausschreibung geklärt. Die Kriterien sollen Anreize für Lösungen schaffen, die sich an die in AP4.1–4.3 entwickelten Demonstratoren andocken und demonstrieren,</p> <ol style="list-style-type: none"> <li>1. wie Unternehmen Daten aus der Wissenschaft nutzen,</li> <li>2. wie sie Daten der Wissenschaft zur Verfügung stellen,</li> <li>3. oder wie neue, eigene Demonstratoren weitere Probleme innovativ lösen.</li> </ol> <p>Weitere Kriterien können sein:</p> <ul style="list-style-type: none"> <li>• Einfügen in die gegebene Technologie aus AP3 und 4.</li> <li>• Mögliche rechtliche Fragestellungen werden benannt.</li> <li>• Mehrwert durch den Austausch von Daten zwischen Wirtschaft und Wissenschaft wird gezeigt.</li> </ul> <p>Gleichzeitig sollen Budget-Grenzen und weitere inhaltliche Kriterien sicherstellen, dass ein möglichst ausgewogenes Portfolio von Lösungen der Arten (1.), (2.) oder (3.) entsteht. Eine mögliche Budget-Grenze für eine Bewerbung ist „max. 1/3 des insgesamt verfügbaren Budgets für die Grundfunktionalität, max. 1/2 des verfügbaren Budgets für erweiterte Funktionalität durch optionale</p> |                      |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |

Zusatzleistungen“.

Die Ausschreibung wird über alle in AP1 erschlossenen Kommunikationskanäle der Communities in Wirtschaft und Wissenschaft verbreitet sowie über etablierte Inkubatorenprogramme wie z.B. BMBF EXIST. Es wird ein System zum Einreichen der Bewerbungen und zum Klären von Bieterfragen eingerichtet. (3,5+2 PM)

**AP4.4.2: Bewerbung und Auswahl, Phase 1** (März–Mai 2022; FhG, NFDI, Atos): Während der Teilnahmefrist werden Bieterfragen beantwortet. Die eingereichten Bewerbungen werden nach den Kriterien bewertet und priorisiert; es wird eine Auswahl getroffen und eine rechtssichere Benachrichtigung durchgeführt. Ausgewählt wird, sofern es die eingegangenen Bewerbungen ermöglichen,

- mindestens ein Antrag für ein Projekt, in dem ein oder mehrere Wirtschaftsunternehmen die verfügbaren Daten aus der Wissenschaft nutzt und
- mindestens ein Antrag für ein Projekt, in dem ein oder mehrere Wirtschaftsunternehmen der Wissenschaft Daten zur Nutzung zur Verfügung stellt. (4+1 PM)

**AP4.4.3: Abschluss und Abwicklung der Unteraufträge, Phase 1** (Juni–September 2022; FhG): Mit den schließlich ausgewählten Bietern werden Unteraufträge abgeschlossen. Die Bieter führen diese Aufträge durch und erfahren dabei im Rahmen von AP2.2 rechtliche Beratung. Die Ergebnisse werden abgenommen und gemeinsam mit den Unterauftragnehmern zur Demonstration vorbereitet. (2+2 PM)

**AP4.4.4: Definition und Ausschreibung, Phase 2** (Oktober–November 2023; FhG, Atos): siehe AP4.4.1. Die Definition wird auf Grundlage der Erfahrungen aus Phase 1 überarbeitet. (3+1 PM)

**AP4.4.5: Bewerbung und Auswahl, Phase 2** (Dezember–Januar 2024; FhG, NFDI, Atos): siehe AP4.4.2 (3,5+1 PM)

**AP4.4.6: Abschluss und Abwicklung der Unteraufträge, Phase 2** (Februar–Mai 2024; FhG): siehe AP4.4.3 (2+2 PM)

**Ergebnisse:**

**E4.4.1: Website mit Links zu den Demonstratoren aus Phase 1** (Website/Software, September 2022, MS2)

**E4.4.2: Bericht Open Call Phase 1** (Bericht, September 2022, MS2)

**E4.4.3: Website mit Links zu den Demonstratoren aus Phase 2** (Website/Software, Mai 2024, MS3)

**E4.4.4: Bericht Open Call Phase 2** (Bericht, Mai 2024, MS3)

|                |  |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
|----------------|--|------|-----|-----|-----|----|------|-----|------|------|------|-----|-----|------|-----|------|----------|
| <b>AP4.5</b>   | <b>ELSA Training für Data Scientists</b> |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Dauer</b>   | Mai 2021–Mai 2024                        |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Leitung</b> | UzK                                      |      |     |     |     |    |      |     |      |      |      |     |     |      |     |      |          |
| <b>Partner</b> | FhG                                      | NFDI | JLU | TIB | UzK | UL | RWTH | FIZ | UKHD | HAdW | EMBL | UMR | WWU | EKUT | ZBW | Atos | <b>Σ</b> |
| <b>PM</b>      | 2<br>+3                                  |      |     |     | 16  |    |      |     |      |      |      |     |     |      |     |      | 18<br>+3 |

**Ziele:**

- Förderung der Kenntnisse ethischer und rechtlicher Rahmenbedingungen und Sensibilisierung für ethische Fragen bei Datenexperten aus der Informatik, Gesundheitsforschung und aus dem Wirtschaftsbereich, durch Training
- Ein ELSA-Trainingscurriculum vorbereiten, um die rechtliche und ethische Nutzung von Forschungsdaten durch Gaia-X-Innovatoren zu unterstützen
- Verwendung der Ergebnisse von AP2, um das Wissen um die rechtlichen und ethischen Aspekte

der Zusammenarbeit von NFDI und Gaia-X an die Datenwissenschaftler\*innen weiterzugeben.

**Arbeitsschritte:**

**AP4.5.1: Landschaftsanalyse** (September–Dezember 2021; **UzK**). Überprüfung bestehender Grundsätze, Leitlinien und Codes of Conduct sowie bestehender ELSA-Ethiktraining-Curricula und relevanter geschäftsethischer Konzepte. Aufbau einer Verbindung zu internationalen Initiativen wie der Ethics and Social Aspects of Data (ESAD) Interest Group der RDA und interdisziplinären Wissenschaftler\*innen in diesem Bereich. (2 PM)

**AP4.5.2: Workshop-Serie „ELSA in Data Science“** (Juni 2022–September 2023; **UzK, FhG**). Organisation einer Workshop-Serie mit dem Titel „Towards an ELSA training curriculum“ unter Beteiligung von interdisziplinären Experten aus Forschung, Wissenschaft und ggf. Industrie mit dem Ziel, Beiträge zur Gestaltung eines ELSA-Curriculums sowie Feedback für die erste Version desselben zu sammeln. (7+2 PM)

**AP4.5.3: Entwicklung eines ELSA-Trainingscurriculums** (Januar 2022–März 2024; **UzK**). Ausgehend von den Beiträgen der Workshops, des AP2 und der Demonstratoren aus AP4.1–4.3 als Use Cases in drei kritischen Bereichen (Geodaten, Datenqualität, Gesundheitsdaten) werden zwei Versionen eines ELSA-Trainingscurriculums erstellt. Die erste Version wird in einem zweiten Teil der Workshop-Serie (AP4.5.2) sowie bei den Gaia-X Industriepartnern (AP4.5.4) vorgestellt und diskutiert, deren Feedback zu einer zweiten Version führen wird. (8 PM)

**AP4.5.4: Outreach und Kontaktaufnahme mit Gaia-X-Partnern** (Juni 2023–März 2024; **FhG**). In diesem AP werden wir interessierte Gaia-X-Partnern suchen und Feedback von ihnen einholen, um die Nachhaltigkeit des Trainingscurriculums zu gewährleisten. Die Diskussionsgrundlage für den Outreach wird die Präsentation der Version 1 des ELSA-Trainingscurriculums sein, und die Ergebnisse des Prozesses werden für die endgültige Version berücksichtigt. (1+1 PM)

**Ergebnisse:**

**E4.5.1: Bewertung der bestehenden Ansätze** (Bericht, Dezember 2021, MS1)

**E4.5.2: Workshop-Bericht** (September 2022)

**E4.5.3: ELSA-Trainingscurriculum Version 1** (Bericht, Februar 2023, MS2)

**E4.5.4: Workshop-Bericht** (September 2023)

**E4.5.5: ELSA-Trainingscurriculum Version 2** (Bericht, Mai 2024, MS3)

**AP5: Projektmanagement (Leitung: FhG)**

Alle Arbeiten werden dezentral von den einzelnen beteiligten Institutionen innerhalb der AP1–4 ausgeführt und zentral durch dieses AP koordiniert. Dazu zählen die organisatorische und administrative Begleitung sowie die Kommunikation mit dem BMBF. Darüber hinaus obliegt dem AP das Projekt-Controlling.

|                |                          |             |            |            |            |           |             |            |             |             |             |            |            |             |            |             |          |
|----------------|--------------------------|-------------|------------|------------|------------|-----------|-------------|------------|-------------|-------------|-------------|------------|------------|-------------|------------|-------------|----------|
| <b>AP5.1</b>   | <b>Projektmanagement</b> |             |            |            |            |           |             |            |             |             |             |            |            |             |            |             |          |
| <b>Dauer</b>   | Mai 2021–Mai 2024        |             |            |            |            |           |             |            |             |             |             |            |            |             |            |             |          |
| <b>Leitung</b> | FhG                      |             |            |            |            |           |             |            |             |             |             |            |            |             |            |             |          |
| <b>Partner</b> | <b>FhG</b>               | <b>NFDI</b> | <b>JLU</b> | <b>TIB</b> | <b>UzK</b> | <b>UL</b> | <b>RWTH</b> | <b>FIZ</b> | <b>UKHD</b> | <b>HAdW</b> | <b>EMBL</b> | <b>UMR</b> | <b>WWU</b> | <b>EKUT</b> | <b>ZBW</b> | <b>Atos</b> | <b>Σ</b> |
| <b>PM</b>      | 8<br>+5                  | 8           |            |            |            |           |             |            |             |             |             |            |            |             |            |             | 16<br>+5 |

**Ziele:** Einhaltung von Deadlines und Budget überwachen, Kommunikation unter den Partnern sicherstellen, reibungsloser, kontrollierter und organisierter Projektablauf.

**Arbeitsschritte:**

**AP5.1.1: Kommunikation zwischen den Projektpartnern, Phase 1** (Mai–Dezember 2021; **FhG**). Start des Projekts; Initiierung, Vorbereitung und Durchführung eines Kick-Off-Workshops; administrative und organisatorische Begleitung des Projekts inklusive Meilenstein- und Risikoplanung sowie Controlling; finale Abstimmung und Unterzeichnung des Kooperationsvertrages zwischen allen Projektpartnern. (1 PM)

**AP5.1.2: Kommunikation zwischen den Projektpartnern, Phase 2** (Januar 2022–Mai 2024; **FhG, NFDI**). Administrative und organisatorische Begleitung des Projekts inklusive Meilenstein- und Risikoplanung sowie Controlling; Initiierung, Vorbereitung und Durchführung von Konsortial-Workshops. (10+1 PM)

**AP5.1.3: Kommunikation mit dem BMBF / Berichte, Phase 1** (Mai–Dezember 2021; **FhG**). Kommunikation und Absprachen mit dem BMBF; Berichtswesen ggü. dem BMBF. (0,5 PM)

**AP5.1.4: Kommunikation mit dem BMBF / Berichte, Phase 2** (Januar 2022–Mai 2024; **FhG, NFDI**). Kommunikation und Absprachen mit dem BMBF; Berichtswesen ggü. dem BMBF. (2,5+1,5 PM)

**AP5.1.5: Einrichtung der Kollaborations-Infrastruktur** (Mai 2021–Dezember 2021; **FhG**). Initiierung und Aufbau einer Kommunikations- und Kollaborationsinfrastruktur. Hierzu gehört auch ein grundlegendes Repository für Software-Quelltexte, welches im Rahmen von AP3.2 um weitere Infrastruktur ergänzt wird. (1+0,5 PM)

**AP5.1.6: Pflege der Kollaborations-Infrastruktur** (Januar 2022–Mai 2024; **FhG**). Pflege der Kommunikations- und Kollaborationsinfrastruktur. (1,5+1,5 PM)

**Ergebnisse:**

**E5.1.1: Kollaborations-Infrastruktur eingerichtet: Dokumenten-Ablage, Mailingliste, Chat, Code-Repository** (Bericht, Juni 2021)

**E5.1.2: Bericht Kick-Off-Workshop** (Juni 2021)

**E5.1.3: 1. Zwischenbericht nach Ende von Phase 1 = Erreichen von MS1** (Januar 2022)

**E5.1.4: 2. Zwischenbericht** (August 2022)

**E5.1.5: 3. Zwischenbericht nach Erreichen von MS2** (März 2023)

**E5.1.6: 4. Zwischenbericht** (Oktober 2023)

**E5.1.7: Abschlussbericht zum Ende von Phase 2 = Erreichen von MS2** (Mai 2024)

## 3.2. Meilenstein- und Risikoplanung

### 3.2.1. Meilensteine

Wir planen einen Meilenstein (MS1) zum Ende des Jahres 2021, wo insbesondere einige Demonstratoren in einer ersten Version fertiggestellt sein werden. Ein weiterer Meilenstein (MS2) ist im Februar 2023 geplant, also kurz nach der Mitte der Projektlaufzeit. Schließlich definiert die Fertigstellung der endgültigen Demonstratoren den letzten Meilenstein MS3. Die drei Meilensteine sind im beigefügten Gantt-Chart als rote Linien eingezeichnet.

| #   | Monat         | Definition   |
|-----|---------------|--|
| MS1 | Dezember 2021 | Erste Demonstratoren fertiggestellt. Erste weitere Ergebnisse fertig, auf denen die nachfolgenden Arbeiten aufbauen.<br><br>E1.1.1: Struktur der Themen-Landkarte mit ersten Inhalten (Website, September 2021)<br>E1.1.2: Erster Bericht Roadmapping (Bericht, Dezember 2021)<br>E2.1.1: Bericht immaterialgüterrechtliche und datenschutzrechtliche Einordnung |



|     |              |   |
|-----|--------------|---|
|     |              | <p>(Dezember 2021)</p> <p>E2.3.1: Generische Mustertexte für die Einwilligung in die Sekundärnutzung personenbezogener Daten in Gaia-X/NFDI (Bericht, Dezember 2021)</p> <p>E3.1.1: Initiales Architektur-Konzept (Bericht, August 2021)</p> <p>E3.1.5: Dokumentation der Gaia-X-konformen Selbstbeschreibung von Forschungsdaten- und -dienstangeboten (Bericht, Oktober 2021)</p> <p>E3.2.1: Deployment der Phase-1-Demonstratoren abgeschlossen (Software, Dezember 2021)</p> <p>E3.2.3: Erste Einrichtung des Kubernetes-Clusters (Software, September 2021)</p> <p>E3.3.1: Stabiles Deployment der Speicherarchitektur für Demonstrator-Projekte (Software, November 2021)</p> <p>E3.4.1: Demonstration der ersten Version der FAIR Validation Services mit IDS-Konnektoren (Software, Dezember 2021)</p> <p>E3.5.1: Katalog von Authentifizierungsframeworks in Gaia-X und NFDI (Bericht, Dezember 2021)</p> <p>E3.7.1: Demonstration des Prototyps einer funktionalen, verteilten Analyseplattform mit Use Cases für die Gesundheitsversorgung (Software, November 2021)</p> <p>E3.7.2: Anforderungsanalyse (Bericht, Dezember 2021)</p> <p>E4.1.2: Demonstrator und User Stories, Phase 1 (Software, Dezember 2021)</p> <p>E4.2.2: Anbindung der GitLab-Infrastruktur von NFDI4Ing an Kubernetes-Infrastruktur (Demonstrator, Dezember 2021)</p> <p>E4.3.2: Demonstrator, Phase 1 (Software, Dezember 2021)</p> <p>E4.5.1: Bewertung der bestehenden Ansätze (Bericht, Dezember 2021)</p>   |
| MS2 | Februar 2023 | <p>Bedeutende Zwischenergebnisse sind erzielt, von denen insbesondere die Fertigstellung der endgültigen Version der Demonstratoren abhängt.</p> <p>E1.2.1: Bericht Community-Aufbau in Deutschland (Dezember 2022, MS2)</p> <p>E2.1.2: Bericht datenschutzrechtliche Einordnung eines Zugangs zu den Daten von europäischen und internationalen Gaia-X-Anwendern (Dezember 2022)</p> <p>E2.2.3: Bericht Zweckbindung der Daten (Dezember 2022)</p> <p>E2.3.2: Bericht zum Spannungsfeld zwischen akademischer und privatwirtschaftlicher Sekundärdatennutzung (Dezember 2022)</p> <p>E2.4.1: Bericht zur daten(typ)basierten rechtlichen Interoperabilität an der Schnittstelle diverser europäischer Dateninfrastrukturen. (Februar 2023)</p> <p>E3.1.3: Referenzarchitektur-Modell v2 (Bericht, Februar 2023)</p> <p>E3.1.6: Selbstbeschreibungs-Ontologie für Forschungsdaten und -dienstangebote v1 (Software, Juni 2022)</p> <p>E3.1.8: Technologie-Landkarte v1 (Website, Februar 2023)</p> <p>E3.3.2: Implementierung des internen PID Service verfügbar (Software, Januar 2023)</p> <p>E3.4.2: PoC für das Compliance Monitoring (Software, Februar 2023)</p> <p>E3.5.2: Authentifizierungsservice auf Basis der „AARC II BPA“ (beta) (Software, August 2022)</p> <p>E3.5.3: Schnittstellendefinition für den Autorisierungsservice auf Basis von „Web Access Control“ (Bericht, Oktober 2022)</p> <p>E3.5.4: Autorisierungsservice (beta) (Software, Dezember 2022)</p> <p>E3.6.1: Terminology Service, Alpha-Version (Software, Dezember 2022)</p> <p>E3.7.3: Integration PHT-Plattform: erste Version (Software, November 2022)</p> |

|     |          |   |
|-----|----------|---|
|     |          | <p>E3.7.5: Roll-out und Installation von PHT-Stationen Version 1 (Software, September 2022)</p> <p>E4.1.4: Autorisierung, Phase 2 (Software, September 2022)</p> <p>E4.1.5: Federated Catalogue, Phase 2 (Software, Februar 2023)</p> <p>E4.2.4: Infrastruktur für Katalog von Qualitätssicherungs- und Analyseworkflows (Software, September 2023)</p> <p>E4.3.4: ML-Algorithmen v1 (Software, Oktober 2022)</p> <p>E4.3.6: Erweiterter Demonstrator v1 (Software, Januar 2023)</p> <p>E4.3.7: Erste weitere PHT Einsatzszenarien aufgenommen und beschrieben (Dokument, Januar 2023)</p> <p>E4.4.1: Website mit Links zu den Demonstratoren aus Phase 1 (Website/Software, September 2022)</p> <p>E4.4.2: Bericht Open Call Phase 1 (Bericht, September 2022)</p> <p>E4.5.3: ELSA-Trainingscurriculum Version 1 (Bericht, Februar 2023)</p>   |
| MS3 | Mai 2024 | <p>Die Demonstratoren der Phase 2 und weitere Ergebnisse von Bedeutung über das Projekt hinaus sind fertiggestellt.</p> <p>E1.2.3: Bericht Konferenz (Mai 2024)</p> <p>E1.2.4: Fertige Themen-Landkarte (Website, April 2024)</p> <p>E2.2.4: Bericht Haftung (Mai 2024)</p> <p>E2.2.5: Bericht Verbindung Secrecy/Datenschutzrecht/Ethik (Mai 2024)</p> <p>E2.3.3: Prozedurale und substantielle Kriterien für die Governance von Gaia-X/NFDI (Bericht, Dezember 2023, MS3)</p> <p>E2.3.4: Positionspapier zu den (moralischen) Pflichten zum Data-Sharing (Bericht, Mai 2024, MS3)</p> <p>E2.4.2: Handreichung für die nationalen Entscheidungsträger über die Vor- und Nachteile unterschiedlicher Anschlussoptionen an europäische Dateninfrastrukturen. (Bericht, Oktober 2023)</p> <p>E2.4.3: Handreichung für die nationalen Entscheidungsträger über die Vor- und Nachteile unterschiedlicher Anschlussoptionen an international verteilte Forschungsnetzwerke. (Bericht, Mai 2024)</p> <p>E3.1.4: Referenzarchitektur-Modell v3 (Bericht, Mai 2024)</p> <p>E3.1.7: Selbstbeschreibungs-Ontologie für Forschungsdaten und -dienstangebote v2 (Software, Juli 2023)</p> <p>E3.1.9: Technologie-Landkarte v2 (Website, Mai 2024)</p> <p>E3.2.2: Deployment der Phase-2-Demonstratoren abgeschlossen (Software, Februar 2024, MS3)</p> <p>E3.3.3: Erstes stabiles Release mit vollständiger Integration aller Services (Software, Dezember 2023)</p> <p>E3.4.3: Erste stabile Version des Compliance-Monitoring Service (Software, November 2023)</p> <p>E3.5.7: Authentifizierungsservice, Autorisierungsservice und Organisationservice (final) (Software, Februar 2024)</p> <p>E3.5.8: Dokumentation der Schnittstellen und Implementierung (Bericht, April 2024)</p> <p>E3.6.3: Terminology Service, finale Version (Software, Dezember 2023)</p> <p>E3.6.4: Terminology-Service-Einbindungen (Bericht, April 2024)</p> <p>E3.7.4: Integration PHT-Plattform finale Version (Software, November 2023)</p> |

|  |  |
|--|--|
|  | <p>E3.7.6: Roll-out und Installation von PHT-Stationen Version 2 (Software, September 2023)</p> <p>E4.1.6: Ownership-Modell &amp; Compliance (Software, Oktober 2023)</p> <p>E4.1.7: Erweiterter Demonstrator und User Stories, Phase 2 (Software, Januar 2024)</p> <p>E4.1.8: Evaluation und Dokumentation, Phase 2 (Bericht, April 2024)</p> <p>E4.2.5: Erweiterter Demonstrator: Gruppensynchronisation (Software, November 2023)</p> <p>E4.2.6: Erweiterter Demonstrator: Analyseworkflows mit hohem Ressourcenbedarf (Software, Februar 2024)</p> <p>E4.2.7: Dokumentation und Bewirtschaftungskonzept (Bericht, April 2024)</p> <p>E4.3.8: Erweiterter Demonstrator v2 (Software, Februar 2024)</p> <p>E4.3.9: Weitere PHT Einsatzszenarien final beschrieben (Bericht, Februar 2024)</p> <p>E4.3.10: Evaluation und Dokumentation (Bericht, April 2024)</p> <p>E4.4.3: Website mit Links zu den Demonstratoren aus Phase 2 (Website/Software, Mai 2024)</p> <p>E4.4.4: Bericht Open Call Phase 2 (Bericht, Mai 2024)</p> <p>E4.5.5: ELSA-Trainingscurriculum Version 2 (Bericht Mai 2024)</p> |
|--|--|

### 3.2.2. Risiken

In dem anvisierten Projekt sehen wir vor allem die im Folgenden dargestellten Risiken zusammen mit einer graduellen Abschätzung ihrer Eintrittswahrscheinlichkeit und des Schadens sowie den im geplanten Gegenmaßnahmen, um die Eintrittswahrscheinlichkeit zu senken bzw. den Schaden zu begrenzen.

| Risiko  | Eintrittswahrscheinlichkeit | Schaden bei Eintritt | Gegenmaßnahmen   | Betroffene AP     |
|---|-----------------------------|----------------------|--|-------------------|
| Komponenten passen technisch nicht zusammen     | gering                      | mittel               | Die Entwicklung von technischen Komponenten wird entlang von Standards vorgenommen, die über Schnittstellen miteinander kommunizieren. Zugleich wird mit Workshops und Kommunikationskanälen (z.B. Slack; gemeinsamer Issue Tracker) die Zusammenarbeit zwischen verschiedenen Partnern gefördert. Im Rahmen der DevOps-Aktivitäten in AP3.2 wird die interoperable Funktionalität der Komponenten praktisch getestet. | AP3               |
| Technische Probleme beim Austausch echter Daten | gering                      | gering               | Zum Austausch von Daten werden standardisierte und wohlbekannte Datenformate verwendet; die semantische Interoperabilität mit Hilfe von anerkannten Terminologien und Ontologien umgesetzt. Die Verwendung dieser Formate und Terminologien wird im Projekt abgesprochen, dokumentiert und technisch unterstützt (vgl. AP3.6).   | AP3               |
| Rechtliche Probleme beim Austausch echter Daten | mittel                      | gering               | Rechtliche Probleme können sich z.B. bei personenbezogenen Daten ergeben oder solchen, die ein Betriebsgeheimnis darstellen. Diesem Problem begegnen wir von rechtlicher wie auch technischer Seite. AP2.2 berät die technischen Arbeiten sowohl   | AP2<br>AP3<br>AP4 |

|  |        |        |   |              |
|--|--------|--------|---|--------------|
|  |        |        | allgemein als auch mit konkretem Bezug zu den Demonstratoren. Verteilte Analysen auf Basis der über den Anwendungsbereich Gesundheit hinaus gültigen Personal-Health-Train-Infrastruktur sind eine weitere Möglichkeit, solchen Problemen zu begegnen.  |              |
| Geringes Interesse seitens der Wirtschaft  | mittel | gering | Wir werden über die Netzwerke der Partner, insbesondere auch den German Gaia-X Hub, die Interessenlage in der Wirtschaft kontinuierlich prüfen und ggf. weitere Incentives sowie verstärkte Kommunikationsmaßnahmen ergreifen.  | AP1<br>AP4.4 |
| Geringe Beteiligung am Open Call           | mittel | gering | Wir werden geeignete Kommunikationsmaßnahmen ergreifen, auch mit Unterstützung aus AP1, und ggfs mit weiteren Incentives und in Absprache mit dem BMBF eine weitere Ausschreibungsrunde anvisieren.   | AP1<br>AP4.4 |
| Open-Call-Unterauftragnehmer liefert nicht | mittel | gering | Wenn ein Unterauftragnehmer der ersten Open-Call-Runde nicht liefert, prüfen wir, ob mit den dadurch noch verbliebenen Restmitteln ein geeigneter Bewerber beauftragt werden kann, den wir in der ersten Auswahlrunde aus Budget-Gründen nicht berücksichtigen konnten. Sollte dies nicht möglich sein, stehen die Restmittel für Phase 2 zusätzlich zur Verfügung und erhöhen dort unseren Handlungsspielraum. | AP4.4        |
| Partner verlässt Konsortium                | mittel | mittel | In den am Projekt beteiligten sowie weiteren NFDI-Konsortien und mit Hilfe von NFDI e.V. sowie in der Gaia-X-Community mit Unterstützung des German Hub adäquaten Ersatz suchen, ggf. auch aus den am Open Call Interessierten  | alle         |
| Rekrutierung von Personal                  | mittel | gering | 1. An dem Projekt sind viele verschiedene Partner beteiligt. Damit wird das Risiko, geeignetes Personal für das Gesamtprojekt zu finden, gemindert.<br>2. Wir werden insbesondere in Phase 1 bereits vorhandenes Personal in das Projekt involvieren.   | alle         |

## 4. Verwertungsplan

### 4.1. Wirtschaftliche Erfolgsaussichten

Die beiden beteiligten Institute der Fraunhofer-Gesellschaft (**FhG**) planen, die aus AP3 hervorgehenden Kern-Dienste für Datenaustausch und -analyse in Gaia-X-Ökosystemen kundenspezifisch anzupassen und in Form von Integrations- und Umsetzungsprojekte mit der Industrie im Zuge eines Open Source Entwicklungs- und Geschäftsmodells zu verwerten. Dazu werden u.a. die Projektergebnisse über AP1.2 hinaus in der Gaia-X-Community und in der International Data Spaces Association verbreitet, auch in Form

der schon bestehenden Open-Source-Software-Initiative<sup>28</sup>. Die Erkenntnisse aus dem Demonstrator für Cross-Plattform FAIR Datenanalyse aus AP4.3 sowohl hinsichtlich der rechtlichen Umsetzbarkeit als auch des technischen Zusammenspiels der Komponenten sollen dazu dienen, Angebote für Dienste zur verteilten Analyse in daten-intensiven Branchen wie der Mobilitäts-, Logistik-, oder Pharma-Industrie vorzubereiten, wo schon Kontakte bestehen. Für dieselbe Zielgruppe soll durch tiefere Einbindung des vom Fraunhofer FIT geleiteten Weiterbildungszentrums<sup>29</sup> das ELSA-Schulungsprogramm aus AP4.5 als Bestandteil der Data-Scientist-Schulungen der Fraunhofer Academy<sup>30</sup> weiterentwickelt werden; weitere Erfahrungen werden der zielgruppengerechten Weiterentwicklung schon bestehender Angebote zu Cloud-nativer Softwareentwicklung (ISST), Ontologie- und Vokabular-Management (FIT), Datenmanagement<sup>31</sup> (FIT, ISST) und im speziellen Datensouveränität<sup>32</sup> (FIT, ISST) dienen. Schließlich ist geplant, geeignete Unterauftragnehmer des Open Call als zukünftige Kunden zu gewinnen.

Das **NFDI**-Direktorat plant keine direkte wirtschaftliche Verwertung.

Die **JLU** plant keine direkte wirtschaftliche Verwertung. Die entwickelten Software-Werkzeuge werden grundsätzlich unter Open-Source-Lizenzen zur Verfügung gestellt und frei zugänglich gemacht. Diese können von der Industrie dann in kommerziellen Produkten und Anwendungen genutzt werden. Auch im Rahmen zukünftiger Forschungsprojekte mit translationalen Aspekten können alle Software-Komponenten in neue Entwicklungen einfließen.

Die **TIB** plant keine direkte wirtschaftliche Verwertung, wird jedoch die anderen Projektpartner bei einer möglichen wirtschaftlichen Verwertung unterstützen. Darüber hinaus plant die TIB die im Vorhaben gewonnenen Erkenntnisse und Technologien zum Forschungsdatenmanagement auch anderen Unternehmen zur Nutzung und wirtschaftlichen Verwertung anzubieten. Evtl. ergeben sich in diesem Zusammenhang nach Projektende Anschlussprojekte und/oder Beratungsaufträge, die dann von der TIB realisiert werden.

Die **UzK** plant keine direkte wirtschaftliche Verwertung, sondern beabsichtigt, langfristig verteilte Analysedienste und den Cross-Plattform FAIR Datenanalyse-Demonstrator als Plattform für die Analyse von sensiblen und großen Datensätzen innerhalb von NFDI zu etablieren und verfügbar zu machen. Diese verteilte Analyseplattform wird die Möglichkeit zur schnellen Entwicklung von KI- und Data-Science-Produkten sowohl durch Industrie- als auch Forschungspartner bieten. Entwickelte Quellcodes und Dokumentationen werden unter einer Open-Source-Lizenz zur Verfügung gestellt. Das von der UzK entwickelte ELSA-Curriculum wird als Open Source zur Verfügung stehen und kann von Dritten weiterentwickelt werden, um Trainingsmaterial für die Ethik der Datenwissenschaft zur Verfügung zu stellen.

Die **UL** plant keine direkte wirtschaftliche Verwertung. Sie wird die von ihr entwickelten Algorithmen (Demonstratoren) Open Source unter einer Open-Source-Lizenz zur Verfügung stellen. Dies gilt auch für Infrastrukturkomponenten, die während des Projekts entstehen. Die Infrastrukturkomponenten sollen in Folgeprojekten eingesetzt werden, um darauf aufbauend ein Sharing von Daten und eine verteilte

---

<sup>28</sup> Siehe <https://www.dataspaces.fraunhofer.de/de/software.html> und <https://github.com/International-Data-Spaces-Association/>

<sup>29</sup> Siehe <https://www.learning.fraunhofer.de/>

<sup>30</sup> Siehe <https://www.bigdata-ai.fraunhofer.de/de/datascientist.html>

<sup>31</sup> Siehe <https://www.fit.fraunhofer.de/de/fb/life/data-manager-in-sciences.html>

<sup>32</sup> Siehe <https://www.dataspaces.fraunhofer.de/de/referenzen/seminare.html>

Datenanalyse zu vollziehen. Die Ergebnisse des Projekts werden in themenspezifischen Journalen publiziert und auf Konferenzen/Workshops vor Fachpublikum vorgestellt. Wir werden ausgewählte Ergebnisse, d.h. Algorithmen und Modelle über den vom BMBF mitfinanzierten Health Atlas<sup>33</sup> der medizinischen Community bereitstellen. Wir erwarten aufbauend auf den erzielten Ergebnissen die Einwerbung weiterer Drittmittelprojekte. So lassen sich medizinische Daten kombinieren und auswerten, insbesondere mit den Beteiligten der MII und der NFDI. Ferner lassen sich die gewonnenen Erkenntnisse, Software-Produkte und KI-Algorithmen in der Ausbildung einsetzen.

Die **RWTH** plant keine direkte wirtschaftliche Verwertung, will jedoch Kern-Dienste und den Demonstrator langfristig als Plattform für Qualitätssicherung innerhalb der NFDI etablieren und zur Verfügung stellen. Damit entsteht ein Marktplatz für Daten und Workflows, an dem sich Wirtschaft und Wissenschaft jeweils in gleichem Maße als Dienstleister und Dienstnehmer beteiligen können. Entwickelte Quellcodes und Dokumentationen werden unter einer Open-Source-Lizenz zur Verfügung gestellt.

Das **FIZ** plant keine direkte wirtschaftliche Verwertung der Projektergebnisse. Die Ergebnisse der datenschutzrechtlichen Analyse der Verknüpfung von NFDI und Gaia-X sollen als Vorlage für die datenschutzkonforme Entwicklung ähnlicher Cloud-basierter Datenräume für Wirtschaft und Wissenschaft dienen und sind daher wirtschaftlich verwertbar. Die datenschutzrechtliche Analyse der Demonstratoren soll auch als Vorlage für die datenschutzkonforme Entwicklung ähnlicher Dienste dienen und ist daher wirtschaftlich verwertbar.

Das **UKHD** plant keine direkte wirtschaftliche Verwertung der Ergebnisse. Die generischen Mustertexte für die Einwilligung zur Sekundärnutzung personenbezogener Daten sowie die prozeduralen und substantiellen Kriterien für eine gute Governance von gemeinschaftlichen Daten-Initiativen aus dem akademischen und privatwirtschaftlichen Sektor sind jedoch für public-private Dateninitiativen verwendbar. Die Mustertexte werden als Creative Commons angeboten.

Die **HAdW** beabsichtigt keine wirtschaftliche Verwertung der Ergebnisse. Allerdings ist eine enge Verzahnung mit den Transfervorhaben der verschiedenen Dateninfrastrukturen der EU geplant, um die Anschlussfähigkeit der Ergebnisse auf mehreren Ebenen zu gewährleisten. Das Gleiche gilt für die internationalen Infrastrukturvorhaben und den Austausch mit ihren Betreibern.

Das **EMBL** plant keine direkte wirtschaftliche Verwertung. Die erarbeiteten Anschlussmöglichkeiten im Hinblick auf Datentypen (vor allem sensible Daten) und ihre Verarbeitung durch verschiedene nationale und europäische und grenzüberschreitende Dateninfrastrukturen als Modell für die Ermöglichung grenzüberschreitender Forschung können unmittelbar angewendet werden und sind daher wirtschaftlich verwertbar.

Die **UMR** plant keine direkte wirtschaftliche Verwertung, sondern beabsichtigt durch die Förderung des Ausbaus der Open-Source-Plattform Geo Engine die entwickelten Softwarekomponenten einem wachsenden Benutzerkreis zur Verfügung zu stellen. Geo Engine soll in weiteren Projekten für den Aufbau fachlicher, Cloud-basierter Datenportale sowohl in Wirtschaft als auch in Wissenschaft genutzt werden, bei denen die Herausforderungen von Big (Spatial) Data gemeistert und die Potentiale der Daten genutzt werden sollen. Damit einhergehend soll Geo Engine auch verstärkt als Werkzeug in der Lehre verwendet werden.

---

<sup>33</sup> Siehe <https://health-atlas.de>

Die von der **WWU** entwickelten Lösungsmodelle (Musterverträge/Codes of Conduct/Gutachten) sind für die Datenwirtschaft unmittelbar anwendbar und daher wirtschaftlich verwertbar.

Die **EKUT** plant keine direkte wirtschaftliche Verwertung der Projektergebnisse. Eine Nutzung der Infrastrukturkomponenten und der im Projekt entwickelten Standards und Best-Practice-Prozesse ist im Rahmen weitergehender Projekte geplant, ebenso die wissenschaftliche Verwertung in Form von Publikationen.

Die **ZBW** plant keine direkte wirtschaftliche Verwertung. Die ZBW wird die nationale und internationale Forschungsdaten-Community, die über das Projekt erreicht wird, auch nach Ablauf des Projekts entweder unmittelbar, d.h. über eigene Community-Instrumente, oder mittelbar, d.h. über Community-Aktivitäten gemeinsam mit Dritten wie GO FAIR, weiter pflegen und für einen kontinuierlichen Informations- und Erfahrungsaustausch zusammenhalten.

**Atos** wird die aus dem Projekt entwickelten Methodologien zur Entwicklung einer Roadmap und von Architekturkonzepten nutzen um die Integration von weiteren Industriellen- und Forschungsplattformen skalierbar und effizient zu gestalten. Atos kann als Systemintegrator im Rahmen der NFDI/Gaia-X kompatiblen Forschungsdatenplattform seinen kommerziellen Kunden neue Lösungsmöglichkeiten aufzeigen

## 4.2. Wissenschaftliche und technische Erfolgsaussichten

Daten gewinnen an Relevanz für die Wissenschaft, weshalb der Austausch von personenbezogenen und sonstigen Daten zwischen Forschung und Industrie ebenso immer häufiger stattfinden wird. Vor diesem Hintergrund enthält die Arbeit von **FIZ Karlsruhe** zu datenschutzrechtlichen Fragen einer Verknüpfung von NFDI und Gaia-X und der **WWU** zu immaterialgüterrechtlichen Fragen der Datenwirtschaft gewinnbringende Informationen für Wissenschaft und Praxis. Unter anderem sollen sie als Grundlage für zukünftige wissenschaftliche Veröffentlichungen zur Entwicklung von rechtmäßigen wissenschaftlichen Informationsinfrastrukturen und Anwendungen sowie für die Entwicklung weiterer Forschungsprojekte im diesem Bereich dienen.

Die im Rahmen der Kern-Dienste und Demonstratoren gewonnenen Erkenntnisse, Systeme und Dienste sollen langfristig zur Verbesserung der Infrastruktur beitragen. An der **RWTH** werden so die Voraussetzungen für zukünftige wissenschaftliche Handlungs- und Publikationsmöglichkeiten geschaffen. Über den inhaltlichen, sowie technischen Anschluss an NFDI4Ing sind die Basistechnologien erschlossen und darauf aufbauende Systeme können Nutzenden aus Wirtschaft und Wissenschaft so zur Verfügung gestellt und nachhaltig von der Community weiterentwickelt werden. An der **UMR** (wie auch an allen anderen Hochschulen, also **JLU, Uzk, UL, RWTH, WWU** und **EKUT**) werden Open-Source Werkzeuge, wie z. B. die Softwareplattform Geo Engine, in Wissenschaft und Lehre vermehrt genutzt. Die Anbindung von Geo Engine in Cloud-Infrastrukturen wie NFDI und Gaia-X führt zur Erschließung neuer Datenquellen und damit einhergehend neue Möglichkeiten wissenschaftliche Erkenntnisse zu gewinnen. Durch die technische Integration von Geo Engine in Multi-Cloud-Infrastrukturen wird erwartet, dass zukünftig eine größere Entwickler-Community entsteht, die sich an der Open-Source Entwicklung aktiv beteiligt. NFDI strebt eine Verbreitung von Projektergebnissen auch zu neuen, in Runde 2 und 3 ausgewählten Konsortien, an.

Technische Lösungen für das Forschungsdatenmanagement versprechen leichter in zukünftigen Vorhaben auch auf europäischer und internationaler Ebene wiederverwendbar zu sein, wenn sie konform zu den an

Gewicht gewinnenden Standards von Gaia-X und seiner Grundlage International Data Spaces sind. Die in den Communities zu Gaia-X und IDS schon aktiven Partner **FhG** und **Atos** können durch das Vorhaben ihre Reputation in diesen Kreisen als Experten für das Forschungsdatenmanagement schärfen; analog kann die **ZBW** in ihren internationalen Netzwerken noch stärker als Expertin für Gaia-X auftreten.

Die Ergebnisse der **HAdW** zur Anschlussfähigkeit der verknüpften Dateninfrastruktur in Form von Berichten werden nationalen und europäischen Entscheidungsträgern zur Verfügung gestellt, um die weitere Etablierung und den Ausbau von Infrastrukturen im Mehrebenensystem zu informieren und zu fördern. Es wird ein enger Austausch mit den europäischen und internationalen Infrastrukturen geplant. Die Ergebnisse werden in interdisziplinären und rechtswissenschaftlichen Publikationsorganen veröffentlicht und dem Fachpublikum auf Konferenzen und Tagungen dargelegt. Auf dieser Grundlage besteht die Erwartung, weitere Drittmittelprojekte einzuwerben und die Forschung zu erweitern.

Die Ergebnisse des **EMBL** zur Anschlussfähigkeit der verknüpften Dateninfrastruktur sollen als Grundlage für wissenschaftliche Veröffentlichungen dienen und so auch anderen Anwendungen (und deren Weiterentwicklungen) zur Verfügung stehen.

Die Ergebnisse des **UKHD** werden mit Blick auf die generischen Mustertexte anschlussfähig sein an die Einwilligungstexte der Medizininformatik-Initiative als auch an die von GHGA für die Sekundärnutzung von Genomdaten. Das AP baut auf entsprechende Vorarbeiten auf. Wissenschaftliches Neuland mit Blick auf die Ethik ist die Ausarbeitung prozeduraler und substantieller Kriterien für eine gute Governance von gemeinschaftlichen Daten-Initiativen aus dem akademischen und privatwirtschaftlichen Sektor. Die Ergebnisse werden in interdisziplinären und bioethischen Konferenzen und Fachjournalen veröffentlicht.

### 4.3. Wissenschaftliche und wirtschaftliche Anschlussfähigkeit

Das Vorhaben erschließt Forschungsdatenmanagement als neue Anwendungsdomäne für Gaia-X und demonstriert gleichzeitig die industriekompatible Umsetzung von Lösungen wie PID-Diensten, die bisher vor allem für das Forschungsdatenmanagement realisiert worden sind. Somit kann die Fraunhofer-Gesellschaft (**FhG**) zukünftigen Kunden im Bereich der forschungsintensiven Industrie seine auf dem Gebiet des Forschungsdatenmanagements schon zuvor erworbenen und in diesem Vorhaben vertieften Kompetenzen noch zielgerichteter demonstrieren. (Zu der gemeinsam mit **UzK** und **UL** umgesetzten verteilten Analyse siehe den Abschnitt zu diesen Partnern unten.)

Das **NFDI**-Direktorat bringt sich maßgeblich ein beim Roadmapping, beim Aufbau der gemeinsamen Community sowie beim Projektmanagement. Damit wird vor allem eine koordinierende Rolle wahrgenommen und zudem die Anschlussfähigkeit von NFDI-Konsortien an Gaia-X erleichtert.

Alle Ergebnisse der **TIB** werden zur maximalen Nachnutzbarkeit offen (unter offenen Lizenzen) zugänglich gemacht. Damit wird eine Aufnahme in der ganzen Breite industrieller und wirtschaftlicher Anwendungen ermöglicht. Durch die Anschlussfähigkeit an domänenspezifische Initiativen zur Einrichtung von Dateninfrastrukturen können die Entwicklungen des Projektes einfach in diesen und weiteren Dateninfrastrukturen integriert werden. Die Anpassung des Leibniz Data Managers zeigt, wie eine etablierte Software-Lösung aus einer anderen Disziplin effizient in neuen Bereichen angewendet werden kann. Der Aufbau auf eine gemeinsame Software-Lösung erleichtert das zukünftige disziplinübergreifende Datenmanagement. Der Wert der Datenräume, in denen die Projektentwicklungen demonstriert werden, wird deutlich erhöht und die Nutzung der Daten attraktiv oder erst sinnvoll möglich gemacht. Es ergeben sich neue Nutzungsmöglichkeiten der Daten und über entsprechende Lizenzierungsmodelle können die



Besitzer der eingestellten Daten über eine Nachnutzung in der Forschung hinaus auch Erlöse durch Angebote an die Industrie generieren.

Die **UzK** und **UL** tragen gemeinsam mit dem Fraunhofer FIT (FhG) vor allem mit Infrastrukturen zur Analyse der verteilt vorliegenden Datenbestände bei. Dazu wird im Projekt eine Infrastruktur aufgebaut und adaptiert. Demonstrator-Anwendungen zeigen die Nützlich- und Nutzbarkeit des Ansatzes über den Einsatzbereich Gesundheit hinaus. Hiernach stehen die Analyseinfrastruktur den Partnern aus Wissenschaft und Wirtschaft zur Verfügung; die Verbreitung wird von beiden Partnern z.B. mit ihrer Beteiligung an NFDI-Konsortien (z.B. NFDI4Health) vorgenommen.

Die Beiträge der **RWTH** werden in den Kern-Diensten und Demonstratoren die Übertragbarkeit von den, im Kontext der NFDI etablierten, Strukturen für das FAIRe Management von Identitäten und Daten und in Cloud-Infrastrukturen sicherstellen. Dabei findet der Aspekt der Multi-Cloud Umgebungen besondere Beachtung. Die Arbeiten fügen sich daher in die Bestehenden wirtschaftlichen und wissenschaftlichen Handlungsfelder: Bewirtschaftung und Provisionierung von wissenschaftlichen IT-Ressourcen sowie Management und Validierung von (Meta-)Daten in verteilten Systemen ein. Für die langfristige Nachnutzbarkeit werden alle Ergebnisse als Whitepaper, Open Access Publikationen oder Programmcodes unter Open-Source Lizenzen zugänglich gemacht und können so auch in anderen Feldern angewendet werden. Analoges gilt für die Arbeiten der **JLU** zur Cloud-native Architektur, Storage-Infrastruktur sowie Compliance & FAIRness Monitoring.

Die Arbeit von **FIZ Karlsruhe** wird klären, wie die Verarbeitung personenbezogener Daten von den Demonstratoren in rechtmäßiger Weise erfolgen kann. Auch soll im Vorfeld geklärt werden, wie die Demonstratoren nach datenschutzrechtlichen Prinzipien entwickelt werden können (Stichwort Privacy by Design bzw. Data Protection by Design). Die Ergebnisse dieser Arbeiten sollen auch in der Zukunft als Vorlage für die datenschutzkonforme Entwicklung ähnlicher Dienste dienen. Ähnlich wird die **WWU** die immaterialgüterrechtliche Fragen der Nutzung und Verwertung von Daten (Urheberrecht, Patentrecht, Geheimnisschutz) klären und Lösungsansätze über Codes of Conduct/Musterverträge entwickeln. Die durch die **UMR** erzielten Ergebnisse bei der Weiterentwicklung der Open-Source Softwareplattform Geo Engine sollen genutzt werden, um die datengetriebenen Wissenschaftsdisziplinen insbesondere an der Universität Marburg im Bereich der Modellierung komplexer Systeme (Umwelt, Pandemie und Biodiversität) zu stärken. Lösungsansätze in diesen Bereichen können durch die Verarbeitung, Analyse und Visualisierung raumbezogener Daten mit Hilfe der Plattform Geo Engine entwickelt werden.

Die Arbeiten des **UKHD** werden via Mustertexte für die Einwilligung die Erlaubnisgrundlage für die Sekundärnutzung personenbezogener Daten in Gaia-X/ NFDI schaffen. Die Kriterien für eine gute Governance und wechselseitige Nachnutzung der Daten durch akademische und privatwirtschaftliche Forschungspartner tragen zu deren Ermöglichung und Vertrauensbildung bei. Dies soll auch dazu beitragen, ethische Standards zu entwickeln, die über den nationalen Kontext hinaus den Datenaustausch ermöglichen.

Die Arbeiten der **HAdW** werden dazu beitragen, die europäischen und internationalen Entwicklungen im Regelungskontext von Dateninfrastrukturen im Mehrebenensystem zu bewerten und bei der Entstehung der verknüpften nationalen Dateninfrastrukturen unmittelbar zu berücksichtigen. Dies wird auch dazu beitragen, dass die nationale Entwicklung stets im Einklang mit den europäischen und internationalen

Vorhaben geschieht und dass die nationalen Infrastrukturen jederzeit auch rechtlich interoperabel bleiben.

Die Arbeiten des **EMBL** sollen Möglichkeiten des Anschlusses nationaler Dateninfrastrukturen an europäische und grenzüberschreitende Infrastrukturen aufzeigen durch und durch diese für grenzüberschreitende Forschung ermöglichen.

Die Arbeiten der **EKUT** werden direkt in weitere Projekte zur Datennutzung einfließen, insbesondere in Projekte zur Analyse von Hochdurchsatzdaten im medizinischen Kontext (genomDE) sowie andere Dateninfrastrukturen (z.B. Netzwerk Universitätsmedizin, Medizininformatik Initiative).

## 5. Arbeitsteilung / Zusammenarbeit mit Dritten

### 5.1. Verbundpartner

#### **Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V. (FhG)**

Die **Fraunhofer-Gesellschaft** ist die größte Organisation für angewandte Forschung in Europa. Sie betreibt in Deutschland derzeit 72 Institute und selbstständige Forschungseinrichtungen. Das **Fraunhofer-Institut für Angewandte Informationstechnik FIT** erforscht Grundlagen und Anwendungen einer menschenzentrierten Digitalisierung. Kernthemen liegen in der Verwendung und Analyse von Daten, im menschenzentrierten Design digitaler Technologien und in deren kooperativer Verwendung sowie in der Anpassung existierender Geschäftsmodelle. Anwendungsdomänen im Fokus des FIT sind neben den Lebenswissenschaften auch Energie, Landwirtschaft/Ernährung und Mobilität. Die Abteilung Data Science und Künstliche Intelligenz bringt in dieses Vorhaben ihre Erfahrungen ein mit Forschungsdatenmanagement, semantischen Technologien, Datensouveränität, FAIR Data, verteilter Analyse, den Dateninfrastrukturen IDS, NFDI und Gaia-X sowie der Koordination von Verbundforschungsprojekten. Das vom Fraunhofer FIT geleitete Weiterbildungszentrum<sup>34</sup> konzipiert für das FIT und weitere Fraunhofer-Institute Weiterbildungen: blended, hybrid oder vollständig online, auf Wunsch mit Prüfung und Zertifikat. Das **Fraunhofer-Institut für Software- und Systemtechnik ISST** identifiziert zusammen mit Unternehmen den strategischen Wert ihrer Daten und macht sie nutzbar – von der Datenaufbereitung bis zur Entwicklung neuer Geschäftsmodelle bieten wir komplette Systemlösungen für Ihr Unternehmen. Expertinnen und Experten des ISST erforschen den Wert von und den souveränen Umgang mit Daten für die Logistik, das Gesundheitswesen und die Datenwirtschaft. Das ISST entwickelt Lösungen für das Datenmanagement und den Aufbau von Datenarchitekturen und schafft so gemeinsam mit seinen Kunden aus der Wirtschaft und als Berater der Politik mit den International Data Spaces den gesamtwirtschaftlichen Rahmen für die sichere und kontrollierbare Datennutzung über Unternehmensgrenzen hinweg. Das ISST hat sich der Digitalisierung von Unternehmen und dem Wissenstransfer von der Forschung in die Praxis verschrieben. Sein Angebot umfasst Software-Engineering für neue digitale Lösungen, Machbarkeits- und Marktstudien sowie Situations- und Potenzialanalysen.

---

<sup>34</sup> Siehe <https://www.learning.fraunhofer.de/>

## **Nationale Forschungsdateninfrastruktur (NFDI) e.V.**

Bund und alle 16 Länder fördern NFDI gemeinsam und sind die Gründungsmitglieder des Vereins Nationale Forschungsdateninfrastruktur (NFDI) e.V., der am 12.10.2020 gegründet wurde. Maßgeblich für die Ausgestaltung des Vereins ist die Satzung (siehe [www.nfdi.de/verein](http://www.nfdi.de/verein)). Der Verein hat als wesentlichen Zweck die Förderung von Wissenschaft und Forschung durch eine Nationale Forschungsdateninfrastruktur, die ein übergreifendes Forschungsdatenmanagement in Deutschland etabliert und fortentwickelt und die Effizienz des gesamten deutschen Wissenschaftssystems steigert. Der Zweck wird in erster Linie verwirklicht durch die Steuerung und Koordination beim Aufbau einer vernetzten Informationsinfrastruktur, bei der Entwicklung eines nachhaltigen interoperablen Forschungsdatenmanagements sowie bei der Schaffung eines verlässlichen Angebots von datenbasierten Diensten für Wissenschaft und Forschung.

Als „NFDI“ wird in diesem Projektantrag das Vereinsorgan Direktorat, welches aus dem Vereinsvorstand und der Geschäftsstelle des Vereins besteht, bezeichnet. Weiterhin am Projektantrag beteiligt sind Organisationen, die an NFDI-Konsortien beteiligt sind.

## **Justus-Liebig-Universität Gießen (JLU)**

Die Justus-Liebig-Universität (JLU) Gießen wurde 1607 gegründet und ist die zweitgrößte Universität in Hessen. Alexander Goesmann ist Inhaber der Professur für Systembiologie und Koordinator des BiGi Service-Zentrums für mikrobielle Bioinformatik innerhalb von de.NBI. Seine Gruppe hat eine umfassende Bioinformatik-Softwareplattform mit Schwerpunkt auf webbasierte Datenbanksysteme für die Genom- und Postgenomforschung eingerichtet. Die JLU Gießen bietet Zugriff auf die IT-Infrastruktur und Softwareanwendungen, um ausgewählte Anwendungsfälle und die Entwicklung von Research Data Commons zu unterstützen. Die JLU bietet Zugriff auf eine zentrale Registrierung standardisierter Software-Container und Bioinformatik-Workflows, die basierend auf unserer Kubernetes-Infrastruktur ausgeführt werden können. Die JLU Gießen ist einer der sechs Standorte der de.NBI-Cloud-Infrastruktur und bietet über die Bioinformatics Core Facility (BCF) eine zentralisierte IT-Infrastruktur für die Hochdurchsatz-Datenanalyse im Bereich Bioinformatik und Systembiologie. Zu den Einrichtungen gehören ein Hochleistungs-Computercluster für die Parallelverarbeitung und mehrere Anwendungsserver mit hoher Kapazität mit einer Hauptspeicherkapazität (RAM) von bis zu 3 TB und bis zu 160 CPU-Kernen auf einem einzelnen Rechner. Die auf OpenStack basierende dedizierte Cloud-Computing-Infrastruktur des deutschen Netzwerks für Bioinformatik-Infrastruktur (de.NBI Cloud) am Standort Gießen bietet umfangreiche Speicher- und Rechenressourcen inklusive GPU-basierter Systeme. Diese Hardware ist jedoch zweckgebunden für bioinformatische Analysen.

## **TIB Leibniz Informationszentrum Technik und Naturwissenschaften**

TIB Leibniz Informationszentrum Technik und Naturwissenschaften in Hannover hat eine herausragende Expertise in der Bereitstellung digitaler Dienste für ihre Fachgebiete. In einer weltweiten Zusammenarbeit mit wissenschaftlichen Bibliotheken und Informationseinrichtungen führte die TIB 2005 den Registrierungsprozess des Digital Object Identifier (DOI) für Forschungsdaten ein, die 2009 in der Gründung des internationalen DataCite-Konsortiums mündete. Mit dem AV-Portal stellt die TIB eine nachhaltige Plattform zur Bereitstellung und Nutzung von wissenschaftlichen audiovisuellen Materialien bereit. Ein zentraler Dienst ist in diesem Kontext der Open Research Knowledge Graph (ORKG) zur

semantischen Beschreibung wissenschaftlicher Erkenntnisse. Mit dem Leibniz Data Manager entwickelt die TIB eine Plattform zum Management von heterogenen Forschungsdaten mit semantischen Technologien. Die TIB ist an verschiedenen NFDI-Konsortien beteiligt und bringt den TerminologyService ein.

## **Universität zu Köln (UzK)**

Die Universität zu Köln (UzK), ursprünglich im Jahr 1388 gegründet, ist eine der ältesten und größten Universitäten in Europa mit 50.000 Studierenden, verteilt auf sechs Fakultäten. Die UzK hat eine große kritische Masse an exzellenten Forschungsgruppen am Life Science Forschungscampus Köln aufgebaut, die sich auf grundlegende Prozesse konzentrieren, die für die Medizin und die Pflanzenbiologie relevant sind. Die Universität zu Köln beherbergt das Cologne Center for Genomics, 15 DFG-geförderte Sonderforschungsbereiche und 4 Exzellenzcluster, darunter der Exzellenzcluster für Pflanzenwissenschaften (CEPLAS) und den Exzellenzcluster „Cellular Stress Responses in Aging-Associated Diseases“ (CECAD), die den Kern von zwei Key Profile Areas bilden. Die UzK ist an insgesamt zwei NFDI-Konsortien beteiligt: NFDI4Health und GHGA. Die UzK ist mit ihren neu eingerichteten Professuren und ihrem starken Datenintegrationszentrum im Bereich der medizinischen Datenwissenschaft stark positioniert. Das Team der Medizinischen Informatik unter der Leitung von Prof. Beyan ist für die Implementierung der verteilten Datenanalyse-Demonstratoren verantwortlich. Beyans Team ist bereits im Rahmen der Medizininformatik-Initiative an der Taskforce Verteilte Analysen beteiligt.

## **Universität Leipzig (UL)**

Die Universität Leipzig wurde 1409 gegründet und ist damit die zweitälteste, durchgehend betriebene Universität Deutschlands. Sie bietet eine einzigartige Fächervielfalt von Geistes- und Sozialwissenschaften, Natur- und Lebenswissenschaften. Die Medizin genießt sowohl aus Forschungssicht als auch in Hinsicht auf die Behandlung einen hervorragenden Ruf. Die Universitätsmedizin Leipzig ist im Rahmen der Medizininformatik-Initiative in das Konsortium SMITH involviert, eines von vier geförderten Konsortien; die UL ist Konsortialführer von SMITH und treibt die Entwicklungen in diesem Bereich auf unterschiedlichen Ebenen. Dazu zählen die Themen Interoperabilität, Data Sharing und Datenschutz. Darüber hinaus ist die UL an NFDI mit dem Konsortium NFDI4Health beteiligt. Die UL bringt hierzu Vorarbeiten aus dem BMBF-Projekt Leipzig Health Atlas ein, in dem eine publikationsorientierte Sharing Plattform inkl verschiedenster Demonstratoren im medizinischen Bereich und Modelle eingebracht und geteilt werden. Im Rahmen der MII (SMITH) etabliert die Universitätsmedizin ein Datenintegrationszentrum. Eng damit verbunden ist das in Gründung befindliche Institut für Medical Data Science, das ab August 2021 von Prof. Toralf Kirsten (TK) geleitet wird. Seit 2020 leitet TK das Konsortium für KI-basierte Diagnoseunterstützung der Familie Seltener Leukodystrophieerkrankungen<sup>35</sup>. Im Auftrag des SMITH Konsortiums ist TK in verschiedenen nationalen Gremien an konsortiumsübergreifenden Abstimmungen (z.B. bzgl. Interoperabilität & Data Sharing) eingebunden. Darüber hinaus ist TK im GO FAIR Implementation Network Personal Health Train (PHT) aktiv und arbeitet an einer Entwicklung des PHT. TK ist in das Konsortium NFDI4Health als Partner (Verteilte Analysen, Datenschutz) eingebunden.

---

<sup>35</sup> Siehe <https://leukoexpert.hs-mittweida.de>

## **RWTH Aachen University**

Die Exzellenzuniversität RWTH Aachen University beheimatet zwei Exzellenzcluster (EXCs) mit Ingenieurwissenschaftlichem Fokus. Das Team für das Forschungsdatenmanagement an der RWTH besteht aus Mitarbeitenden des IT Centers, der Universitätsbibliothek und Forschenden aus Verbundprojekten und bearbeitet gemeinsame Projekte und bietet Beratungsleistungen für die gesamte Hochschule. Die darauf aufbauende FDM Strategie der RWTH bindet so Data Stewards aus den drei EXCs und vier SFBs ein. Das IT Center koordiniert verschiedene Initiativen um die digitale Unterstützung des Forschungsprozesses voranzutreiben (IdM.nrw, HPC.nrw, AcademicGroupware.nrw, FDS.nrw) und kooperiert mit FDM-Experten aus TU9, CESAER, DH-NRW, DINI/nestor und RDA. In den überregionalen Verbänden NFDI4Ing, NFDI4Chem und NHR4CES stellt das IT Center Anwendungen sowie Rechen- und Datendienste für ein breites Spektrum der wissenschaftlichen Nutzergemeinschaft in Deutschland zur Verfügung. Dr. Marius Politze ist Teil des FDM-Teams und von NFDI4Ing; am IT Center leitet er die Gruppe „Anwendungsentwicklung Forschung“. In der Gruppe verantwortet er die Entwicklung prozessunterstützender Services für die Bewirtschaftung und Nutzung wissenschaftlicher IT-Ressourcen. Seine Forschungsinteressen liegen im Bereich verteilter, serviceorientierter Softwarearchitekturen und Semantic Web.

## **FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur GmbH (FIZ)**

FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur GmbH ist eine der großen Infrastruktureinrichtungen in Deutschland und Mitglied der Leibniz-Gemeinschaft. FIZ Karlsruhe hat den öffentlichen Auftrag, Wissenschaft und Forschung weltweit mit wissenschaftlicher Information zu versorgen, dafür entsprechende Produkte und Dienstleistungen zu entwickeln und angewandte Forschung zu betreiben. Damit verbunden ist das Ziel, den nationalen und internationalen Wissenstransfer zu stärken, Innovationen zu fördern und die Zusammenarbeit in Wissenschaft und Forschung zu unterstützen. Die von Prof. Franziska Boehm geleitete Forschungsabteilung Immaterialgüterrechte in verteilten Informationsinfrastrukturen (IGR) analysiert urheberrechtliche Vorgaben und Reformentwicklungen und leitet daraus Empfehlungen für digitale Wissenschaft einschließlich deren Infrastruktureinrichtungen ab. Neben dem fachjuristischen Forschungsschwerpunkt zu (Forschungs-)Daten werden urheber-, lizenz- und datenschutzrechtliche Lösungen für Bereiche wie z.B. Informationsinfrastruktur, Open Access, Software, Big-Data-Anwendungen, kollaborative Forschungswerkzeuge, vernetzte Lehrformate und digitale Distribution von Fachmaterialien erstellt.

## **Universitätsklinikum Heidelberg (UKHD), Sektion für Translationale Medizinische Ethik**

Die Exzellenzuniversität Heidelberg ist an zwei Konsortien von NFDI beteiligt: GHGA und NFDI4Culture. Ziel des NFDI-Projektes GHGA ist der Aufbau eines Genomarchivs als nationaler, förderierter Knoten des Europäischen Genom-Phänom-Archivs (EGA). Als Teil der Nationalen Forschungsdateninfrastruktur integriert GHGA dabei breite Teile der Genomforschungsgemeinde in Deutschland inklusive der großen Genomsequenzierzentren. GHGA wird die sichere Speicherung, den Zugriff und die Analyse menschlicher Omics-Daten in einem einheitlichen ethisch-rechtlichen Rahmen ermöglichen. Für die Entwicklung dieses ethischen Rahmens ist die Sektion Translationale Medizinische Ethik (TME) unter der Leitung von Frau Prof. Winkler verantwortlich. Bereits involviert ist die TME in die Ausgestaltung der Datensouveränität im

Rahmen der Medizininformatik-Initiative (AG Consent) und die Konzeption der Zentren Personalisierter Medizin. Ergebnisse des Ethischen APs beinhalten generische Mustertexte für die Einwilligung in die Sekundärnutzung personenbezogener Daten und Governance-Empfehlungen im Spannungsfeld zwischen akademischer und privatwirtschaftlicher Forschung.

## **Heidelberger Akademie der Wissenschaften (HAdW)**

GHGA möchte Forschende in Deutschland in die Lage versetzen, humane Genomdaten rechtskonform entsprechend der FAIR-Richtlinien auszutauschen und für die Forschung zu nutzen. Zudem kann GHGA dabei internationale Standards zum Datenaustausch stärker mitgestalten. GHGA ist dabei eingebunden in flankierende internationale Forschungsnetzwerke wie etwa die europäische 1+ Million Genomes und die Beyond 1+ Million Genomes (B1MG)-Initiative. Die Entwicklung des rechtlichen Rahmens für GHGA und des Anschlusses an europäische Infrastrukturen erfolgt durch die Heidelberger Akademie der Wissenschaften (HAdW) unter der Leitung von Frau Dr. Fruzsina Molnar-Gabor. Die HAdW ist eine von acht Landesakademien in Deutschland. Seit ihrer Gründung im Jahr 1909 fördert die Akademie die interdisziplinäre Forschung durch fakultätsübergreifende und internationale Hochschulkooperationen. Sie bringt juristische und politische Expertise in internationale Forschungsverbünde ein, die an Lösungen sowie unabhängigen Empfehlungen für Politik und Gesellschaft im Bereich der gesundheitsbezogenen Datenverarbeitung und -weitergabe arbeiten, wie früher z. B. europäisch im European Open Science Cloud Pilot der EU-Kommission (EOSCPilot), und zurzeit neben nationalen Initiativen auch international zum Europäischen Humangenom-Phänom-Archiv.

## **Europäisches Laboratorium für Molekularbiologie (EMBL)**

Die internationale Einbettung der Infrastruktur wird maßgeblich von Dr. Jan Korbel, Head of Data Science am Europäischen Laboratorium für Molekularbiologie (EMBL), vorangetrieben. Das EMBL (<http://www.embl.org/>) wurde 1974 als internationale Forschungsorganisation gegründet und wird von über zwanzig Ländern unterstützt. Das EMBL ist ein Exzellenzzentrum für molekularbiologische Grundlagenforschung. Es ist eine verteilte internationale Forschungsorganisation (27 Mitgliedsstaaten) mit dem Hauptlabor in Heidelberg. Die Forschung am EMBL betont die experimentelle Analyse auf mehreren Ebenen der biologischen Organisation, vom Molekül bis zum Organismus. Eine der fünf Kernaufgaben des EMBL ist es, neue Technologien für die Lebenswissenschaften zu entwickeln und verfügbar zu machen. Dementsprechend hat das EMBL eine umfangreiche Erfolgsbilanz bei der Entwicklung neuer Biologie-getriebener Technologien und der Bereitstellung serviceorientierter Core Facilities, einschließlich der EMBL-EBI-Daten-Ressourcen. Eine weitere Kernaufgabe des EMBL ist die Weiterbildung von Forschern, die vom EMBL International Centre for Advanced Training (EICAT) organisiert wird und das EMBL International PhD Programme (EIPP), das EMBL Postdoctoral Programme und das Visitors Programme sowie eine große Anzahl von Workshops, Kursen und Konferenzen umfasst. Das EMBL unterhält starke interaktive Beziehungen zu den wichtigsten biowissenschaftlichen Forschungseinrichtungen in Europa und verfügt über mehr als 40 Jahre Erfahrung in der Integration der europäischen Forschung.

## **Philipps-Universität Marburg (UMR)**

Die Philipps-Universität Marburg ist die traditionsreichste Hochschule Hessens. 1527 gegründet, bietet sie heute ihren rund 23.500 Studierenden Lehre in einem breiten Fächerspektrum an insgesamt 16 Fachbereichen. Sie ist an verschiedenen NFDI-Projekten beteiligt und koordiniert für das Land Hessen die

hessischen Forschungsdateninfrastrukturen (HEFDI). Prof. Bernhard Seeger vertritt das Gebiet Datenbanksysteme am Fachbereich Mathematik und Informatik. Seine Gruppe arbeitet an der Entwicklung von Zeit- und Geo-Datenbanksystemen sowie im Bereich Ereignissysteme. Seit 2010 widmet er sich Problemstellungen im Anwendungsbereich der Biodiversität und war 2013 Mitinitiator des DFG-geförderten Projekts GFBio, in dem seine Gruppe mit dem VAT-System eine Plattform für die Transformation, Analyse und Visualisierung von Biodiversitätsdaten entwickelt hat, das inzwischen über GFBio hinaus zum Aufbau von Forschungsdatenportalen verwendet wird. Er war federführend an der Antragstellung von NFDI4Biodiversity beteiligt und leitet dort die Task Area für die Entwicklung des Research Data Commons. Neben seinen Arbeiten im Bereich Datenbanksysteme hat er erfolgreiche Ausgründungen aus seiner Arbeitsgruppe begleitet und ist derzeit Mentor eines durch das BMWi geförderten Forschungstransferprojekts.

### **Westfälische Wilhelms-Universität Münster (WWU)**

Das ITM (Institut für Informations-, Telekommunikations- und Medienrecht) ist eine seit Jahrzehnten in der Forschung und Lehre tätige Einrichtung der Westfälischen Wilhelms-Universität (WWU). Mehr als 20 ForscherInnen beschäftigen sich dort seit 1997 mit Rechtsfragen der Digitalwirtschaft und der Informationsgesellschaft, voran mit Problemen des Immaterialgüterrechts (Urheberrecht, Patentrecht, Geheimnisschutz) und deren Auswirkungen im Vertragsrecht. Das ITM entwickelt zusammen mit der Industrie und der Informatik Lösungsmodelle für Rechtsprobleme neuer digitaler Technologien (Blockchain, KI u.a.); dieses Wissen fließt dann auch in Forschungsprojekte der Europäischen Kommission und des BMBF ein. Das ITM kooperiert dabei in der Gestalt des Institutsleiters und FIT Adjunct Professors Dr. Thomas Hoeren mit der Fraunhofer-Gesellschaft. und verfügt über gute Beziehungen zum FIZ.

### **Eberhard Karls Universität Tübingen (EKUT)**

Die Eberhard Karls Universität Tübingen ist eine 1477 gegründete Universität des Landes Baden-Württemberg in Tübingen. Im Juni 2012 erhielt die Universität im Rahmen der dritten deutschen Hochschul-Exzellenzinitiative den Exzellenzstatus, welcher 2019 im Zuge der Exzellenzstrategie verlängert wurde. Sie gehört damit zu den elf Exzellenzuniversitäten Deutschlands. Oliver Kohlbacher (OK) ist Professor für Angewandte Bioinformatik an der Universität Tübingen, Direktor des Instituts für Translationale Bioinformatik am Universitätsklinikum Tübingen und Fellow am Max-Planck-Institut für Entwicklungsbiologie. Ein Schwerpunkt seiner Forschung der letzten Jahre liegt im Bereich der translationalen Bioinformatik, insbesondere bei der Integration von klinischen und molekularen Daten. OK ist stellvertretender Leiter des Konsortiums DIFUTURE der MII und Direktor des Medizinischen DIZ am Department für IT und Angewandte Medizininformatik am UK Tübingen. Die Ideen der GO FAIR-Initiative und des PHT implementiert er seit mehreren Jahren in verschiedenen Projekten und im „PHT Implementation Network“. OK hat über 250 begutachtete wissenschaftliche Arbeiten als Co-Autor veröffentlicht. Er ist Sprecher der Fachgruppe Bioinformatik (FaBI, <http://www.bioinformatik.de>) und Mitglied in zahlreichen internationalen Beiräten und Gutachtergremien.

### **ZBW Leibniz Informationszentrum Wirtschaft**

Die ZBW – Leibniz-Informationszentrum Wirtschaft ist die weltweit größte Informationsinfrastruktur für Wirtschaftswissenschaften mit Standorten in Kiel und Hamburg. Als nationale Informationsinfrastruktureinrichtung arbeitet die ZBW seit vielen Jahren auf dem Gebiet des

elektronischen Publizierens. So betreibt sie das fachliche Repository *EconStor*, das über 200.000 Publikationen wirtschaftswissenschaftlicher Fakultäten, Institute und Fachgesellschaften frei verfügbar als Volltext bereitstellt. Zudem betreibt die ZBW das wirtschaftswissenschaftliche Fachportal *EconBiz*, in dem ca. 12 Millionen Publikationen aus unterschiedlichsten Quellen nachgewiesen werden; im Jahr 2020 wurden die digitalen Volltexte der ZBW (lizenzierter und open access) knapp 12 Millionen Mal aus aller Welt heruntergeladen. Mit fünf Professuren, von denen derzeit vier besetzt sind, betreibt die ZBW Forschung mit Fokus auf Open Science im Bereich Forschungsdatenmanagement. Die ZBW ist an verschiedenen NFDI-Konsortien beteiligt und übernimmt dort stets die Rolle der Infrastrukturentwicklung.

## **Atos Information Technology GmbH**

Atos ist ein weltweit führender Anbieter für die digitale Transformation mit 110.000 Mitarbeitern in 73 Ländern und bietet ganzheitliche IT-Lösungen für zahlreiche Branchen. In seinen deutschen Gesellschaften hat Atos ca. 10.000 Mitarbeiter. Die größte ist die Atos Information Technology GmbH (AIT) mit der am Projekt teilnehmenden Innovationseinheit C-LAB, über die ebenfalls die Technologieexperten aus weiteren Abteilungen der AIT eingebunden werden. Atos ist Gründungsmitglied der Gaia-X Association und Mitglied der International Data Space Association und in weiteren Organisationen mit Fokus auf Datenaustausch und Interoperabilität aktiv, darunter oneM2M, ETSI und FIWARE. Atos besitzt lange Erfahrung in Entwurf und Aufbau sicherer, interoperabler Datenplattformen, sowohl zentraler als auch dezentraler Natur.

### **5.2. Zusammenarbeit mit Dritten**

Die rechtliche und ethische Arbeit in AP2 wird durch Zusammenarbeit mit dem verwandten Projekt „**Anreizsystem und Ökonomie des Data Sharings** (Incentives and Economics of Data Sharing – IEDS)“ ergänzt, an dem als gemeinsamer Partner das Fraunhofer ISST beteiligt ist. Während sich die Arbeit im IEDS-Projekt auf die Ökonomie des Rechts und Möglichkeiten zur wirtschaftlichen Verwertung von Daten fokussiert, wird sich die juristische Arbeit in diesem Projekt auf die materiell-rechtliche und die ethische Sachlage fokussieren und konkrete Vorschläge für die zu entwickelnden Demonstratoren erarbeiten.

Der Demonstrator aus AP4.1 wird auf Basis der Open-Source-Plattform **Geo Engine** entwickelt, die aus dem Projekt GFBio hervorgegangen ist und in NFDI4Biodiversity als webbasiertes Tool für die zeitliche Analyse und Visualisierung von Geo-Daten in diesem Jahr zur Verfügung gestellt wird. Diese einzigartige Plattform wird vom Start-up Geo Engine weiterentwickelt und für die Entwicklung kundenspezifischer Datenportale bereits genutzt. Die Personen im Start-up sind derzeit die einzigen Experten, die sich mit dem Programmcode des Softwareprodukts Geo Engine auskennen. Um insbesondere kurzfristig einen erfolgreichen Demonstrator Ende 2021 in dem Projekt zu erstellen und über die gesamte Projektlaufzeit beratend zur Seite zu stehen, ist eine Zusammenarbeit mit dem Start-up Geo Engine zwingend erforderlich. Zudem erwarten wir durch eine solche Zusammenarbeit mit dem Start-up die Verbindung zwischen Wirtschaft und Wissenschaft zu stärken.

Das Vorhaben baut auf den Arbeiten des Projekts **Gaia-X Federation Services** auf, in dem der eco Verband der deutschen Internetwirtschaft mit Förderung des BMWi Unteraufträge für die Spezifikations-Phase 1 u.a. an die Fraunhofer-Gesellschaft vergeben hat. Die Entscheidung ist noch offen, ob weitere Fraunhofer-Institute (über das hier nicht vertretene Fraunhofer IOSB hinaus) an der Implementation der Spezifikation in der zweiten Hälfte des Jahres 2021 und an der im Jahr 2022 liegenden Phase 2 der Spezifikation und Implementation beteiligt sein werden.

Weitere externe Organisationen haben uns ihre **Unterstützung** mit den beigefügten Schreiben zugesagt:



- Die Siemens AG als führender Technologiekonzern im Bereich intelligenter Infrastruktur möchte bei einem unserer Workshops die Herausforderungen industrieller Ökosysteme präsentieren und den Open Call hinsichtlich der Definition der Kriterien und Auswahl begleiten.
- acatech, die Deutsche Akademie der Technikwissenschaften, koordiniert den Gaia-X Hub Germany. Über diesen unterstützt sie das Roadmapping und die Vernetzung der Community.
- Die Open Telekom Cloud, deren Dienste wir für wirtschaftsnahe Demonstrationen zu beauftragen planen, und die das Projekt darüber hinaus bewerben wird.

## 6. Notwendigkeit der Zuwendung

Die an diesem Projektantrag beteiligten Institutionen verfügen über grund- und drittmittelfinanziertes Personal, das allerdings für dieses Vorhaben nicht ausreichend zur Verfügung steht. Weitere öffentliche Ausschreibungen auf nationaler (BMBF, BMG, BMWi usw.) und internationaler Ebene (z.B. EU Horizon Europe) sind derzeit nicht passend verfügbar. Der aktuelle Förderwettbewerb „Innovative und praxisnahe Anwendungen und Datenräume im digitalen Ökosystem GAIA-X“ des BMWi umfasst Data Spaces als einen Schwerpunkt, adressiert jedoch Leuchtturmvorhaben ungefähr im doppelten Umfang des für FAIR-DS veranschlagten Budgets und fokussiert in größerem Maße eine Marktorientierung der Lösungen, während für den Datenaustausch zwischen Wissenschaft und Wirtschaft erst noch, wie in diesem Antrag dargelegt, grundlegende organisatorische, rechtliche und technische Herausforderungen anzugehen sind. Aus demselben Grund ist aktuell auch noch nicht mit umfangreichen F&E-Aufträgen durch die forschungsintensive Industrie zu rechnen. Auf EU-Ebene sind Ausschreibungen rund um „Data sharing in the common European data spaces“ (Work Programme „Digital, Industry and Space“) und „Supporting EOSC-Core: Enabling access to the Web of FAIR data and services“ (Work Programme „Research Infrastructures“) angekündigt, die jedoch erst gegen Ende 2021 öffnen. Entsprechende Projekte können jedoch helfen, die in der ersten Hälfte von FAIR-DS erreichten Ergebnisse auf die europäische Ebene zu heben.

Daher ist eine finanzielle Zuwendung für die Finanzierung des Personals notwendig. In den Institutionen werden nur in Einzelfällen Neueinstellungen vorgenommen; in der Regel wird auf bereits eingearbeitetes Personal zurückgegriffen.

Neben der Finanzierung des Personals ist der Aufbau von Infrastrukturkomponenten abzusichern. Hierzu kommen zwar Eigenleistungen der beteiligten Institutionen zum Tragen, die jedoch nicht ausreichen, alle Aspekte finanziell abzugelten. Bevor die Infrastruktur und das für deren Betrieb notwendige Personal über eingeworbene F&E-Aufträge finanziert werden kann, bedarf es zum Aufbau weiterer Grundlagenarbeit.

Darüber hinaus sind Abstimmungen innerhalb des Verbundes sowie mit den Aktivitäten in Deutschland (z.B. NFDI) und auf europäischer Ebene (Gaia-X) notwendig. Dafür sind in AP1 Workshop-Serien auf nationaler und internationaler Ebene geplant, die mit diesem Projekt finanziert werden.

Die von einzelnen Verbundpartnern zum Teil beantragten Kosten für hochqualifizierte Mitarbeitende höherer Entgeltgruppen z.B. für interdisziplinäre Arbeit, Koordination mehrerer Verbundpartner und komplexe Software-Architektur, sowie für projektspezifische Infrastruktur, sind in deren jeweiligen Anträgen begründet; hier nur eine kurze Zusammenfassung:

- Insbesondere die Arbeit an der Schnittstelle von AP2 zu den AP 3 und 4 erfordert interdisziplinäre Erfahrung.
- Für den Betrieb und die Entwicklung der Demonstratoren und der Software-Services wird die **JLU** eine dedizierte Infrastruktur basierend auf den bereits bestehenden technischen Grundlagen der de.NBI-Cloud bereitstellen und als Gaia-X compatible Infrastruktur betreiben.

- Für den Betrieb und die Entwicklung der Kern-Dienste und Demonstratoren plant die **RWTH** die Bereitstellung von Infrastruktur im realistischen, industrienahen Umfeld durch den Gaia-X Partner Open Telekom Cloud. Dabei wird insbesondere die Multi-Cloud-Kompatibilität der entwickelten Kern-Dienste und Demonstratoren aus AP3 und AP4 sowie die Übertragbarkeit zwischen wissenschaftlichen und kommerziellen Cloud-Infrastrukturen sichergestellt.

## 7. Literaturverzeichnis

### 7.1. Veröffentlichungen der Partner

Authmann C, Beilschmidt C, Dröner J, Mattig M, Seeger B (2015). VAT: a system for visualizing, analyzing and transforming spatial data in science. *Datenbank-Spektrum*, 15(3), 175-184.

Bader S, ..., Lange C (2020). [The International Data Spaces Information Model - An Ontology for Sovereign Exchange of Digital Content](#). International Semantic Web Conference, 176–192.

Bartram C, Eils R, Kalle C, Kirchhof P, Kulozik A, Schirmacher P, Tanner K, Winkler EC (2013), Eckpunkte für eine Heidelberger Praxis der Ganzgenomsequenzierung: Stellungnahme / Projektgruppe „Ethische und Rechtliche Aspekte der Totalsequenzierung des menschlichen Genoms“ (EURAT). Heidelberg: Marsilius-Kolleg.

Beyan O, ... Kohlbacher O et al. (2020). [Distributed Analytics on Sensitive Medical Data: The Personal Health Train](#). *Data Intelligence* 2(1-2): 96–107

BMWi (2020). [GAIA-X: Policy Rules and Architecture of Standards](#).

BMWi (2020a). [GAIA-X: Technical Architecture](#).

Boehm F (2016). Herausforderungen von Cloud Computing-Verträgen: Vertragstypologische Einordnung, Haftung und Eigentum an Daten. *ZEuP: Zeitschrift für europäisches Privatrecht*, (2), 358-387.

Grangel-González I, Baptista P, Halilaj L, Lohmann S, Vidal ME, Mader C, Auer S (2017). The Industry 4.0 standards landscape from a semantic integration perspective. *Emerging Technologies and Factory Automation*, 1-8.

Halilaj L, Petersen N, Grangel-González I, Lange C, Auer S, Coskun G, Lohmann S (2016). VoCol: An Integrated Environment to Support Version-Controlled Vocabulary Development. *Knowledge Engineering and Knowledge Management*, 303–319.

Hallinan D (2021). *Protecting Genetic Privacy in Biobanking through Data Protection Law*. Oxford (Oxford University Press).

Hallinan D (2021a). Biobank Oversight and Sanctions Under the General Data Protection Regulation. In *GDPR and Biobanking Individual Rights, Public Interest and Research Regulation across Europe*. Herausgeber: S. Slokenberga et. al. Heidelberg (Springer).

Hallinan D (2020). Broad consent under the GDPR: an optimistic perspective on a bright future. *Life Sciences Society and Policy* 16.1: (2020) 16:1. <https://doi.org/10.1186/s40504-019-0096-3>.

Hoeren T et al. (2021). *Geschäftsgeheimnisgesetz. Kommentar*, Berlin (de Gruyter).

Hoeren T, Pinelli, S (2020). [A NOVA LEI BRASILEIRA DE PROTEÇÃO DE DADOS – UMA VISÃO CRÍTICA](#). *Proteção de Dados*, 25–32.

Hoeren T (2020). [Möglichkeiten und Grenzen eines selbstständigen Beweisverfahrens in Softwaremängelprozessen](#). Festschrift für Jürgen Taeger, Frankfurt (dfv Mediengruppe), 739-747.

Hoeren T, Pinelli S (2020a). [Daten im Rechtsverkehr – Überlegungen für ein allgemeines Datenvertragsrecht](#). *Juristen Zeitung*, Mohr Siebeck, Heft 18, 879–884.

Hoeren T, Pinelli S (2020b). [Das neue brasilianische Datenschutzrecht](#), in ZD, C.H. Beck, 351–354.

Hoeren T (2020a). [Dateneigentum und Datenbesitz](#), in: Tereza Pertot (Hrsg.), *Rechte an Daten*, Tübingen, 37–48.

Hoeren T, Rabovskaja E, Bauer A (2020). [Blockchain und die Datenschutzgrundverordnung](#), Sankt Augustin.

Hoeren T, Pinelli S (2020c). [International Interest: Software Assessment for Security-Relevant Errors](#), in: *Communications Law* 25:1, 42–49.

Molnár-Gábor F, Lück R, Yakhneen S, und Korbel JO (2017). Computing Patient Data in the Cloud: Practical and Legal Considerations for Genetics and Genomics Research in Europe and Internationally. In: *Genome Medicine*, Genomic Technologies Special Issue 9:58.

Molnár-Gábor F (2018). Germany: a fair balance between scientific freedom and data subjects' rights? In: *Human Genetics* 137(8): 619-626.

Molnár-Gábor F (2019a). Verarbeitung von Patientendaten in der Cloud unter den Bedingungen der DS-GVO mit besonderer Rücksicht auf transnationale (Forschungs-)Aspekte. In: Spiecker gen. Döhmman I, Wallrabenstein A (Hrsg.) *Gesundheitsversorgung in Zeitalter der DS-GVO*, Peter Lang 121-141 (2019).

Molnár-Gábor F, Korbel JO (2020). Genomic data sharing in Europe is stumbling – Can a code of conduct prevent its fall? In: *EMBO Molecular Medicine* e11421.

Molnár-Gábor F (2021) Der Schutz der Gesundheitsdaten in Deutschland. In: Kugelmann/Lukanko (Hrsg.) *Nationale Spielräume im Datenschutzrecht, Nomos, im Erscheinen*.

Molnár-Gábor F (2021a). Medizindatenschutz. In: *Nomos Kommentar zum Medizinstrafrecht, Nomos, im Erscheinen*.

Otto B, Jarke M (2019). [Designing a multi-sided data platform: findings from the International Data Spaces case](#). *Electronic Markets* 29, 561–580.

Otto B, ..., Lange C et al. (2021). [GAIA-X and IDS](#). Position Paper v1.0. International Data Spaces Association.

Otto B, ..., Auer S, ..., Lange C et al. (2019). [IDS Reference Architecture Model](#). International Data Spaces Association.

Philips M, Molnár-Gábor F, Korbel JO, ..., Townend D, Knoppers BM (2020a), The PCAWG Consortium. Of Clouds and Genomic Data Protection. In: *Nature* 578, 31-33, doi: 10.1038/d41586-020-00082-9.

Politze M, Claus F, Brenger B, Yazdi MA, Heinrichs B, Schwarz A (2020): [How to Manage IT Resources in Research Projects? Towards a Collaborative Scientific Integration Environment](#). European Journal of Higher Education IT: 2020-1.

Politze M, Schwarz A, Kirchmeyer S, Claus F, Müller MS (2020): [Kollaborative Forschungsunterstützung: Ein Integriertes Probenmanagement](#). E-Science-Tage 2019. Data To Knowledge.

Politze M, Bensberg S, Müller MS (2019): [Managing Discipline-Specific Metadata Within an Integrated Research Data Management System](#). 21st International Conference on Enterprise Information Systems.

Politze M, Decker B (2016): [Ontology Based Semantic Data Management for Pandisciplinary Research Projects](#). 2nd Data Management Workshop, Kölner Geographische Arbeiten, 96.

Pictor M, Lewis MA, Newson AJ, ..., Molnár-Gábor F, Yamamoto B, Kaye J, Teare HJA (2019). Dynamic Consent: An Evaluation and Reporting Framework. In: *Journal of Empirical Research on Human Research Ethics* 14(5):1-12.

Schmitz D, Politze M (2018): [Forschungsdaten managen – Bausteine für eine dezentrale, forschungsnaher Unterstützung](#). In o-bib. Das offene Bibliotheksjournal 5 (3).

Spiekermann M, Tebernum D, Wenzel S, Otto B (2018). [A metadata model for data goods](#). Multikonferenz Wirtschaftsinformatik, 326–337.

Winkler EC, Knoppers BM (2020). [Ethical challenges of precision cancer medicine](#). Seminars in Cancer Biology.

## 7.2. Weitere Quellen

Bundesregierung (2018). Strategie Künstliche Intelligenz der Bundesregierung. <http://www.ki-strategie-deutschland.de/>

Europäische Kommission (2020). Eine europäische Datenstrategie. Mitteilung der Kommission. [https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy\\_de](https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_de)

Richter G et al. (2020). Secondary research use of personal medical data: attitudes from patient and population surveys in The Netherlands and Germany. *European Journal of Human Genetics* 2020: 1-8.

Voigt TH et al. (2020). Willingness to donate genomic and other medical data: results from Germany. *European Journal of Human Genetics* 28.8: 1000–1009.

Wilkinson, MD et al. (2016). [The FAIR Guiding Principles for scientific data management and stewardship](#). *Scientific Data* 3.