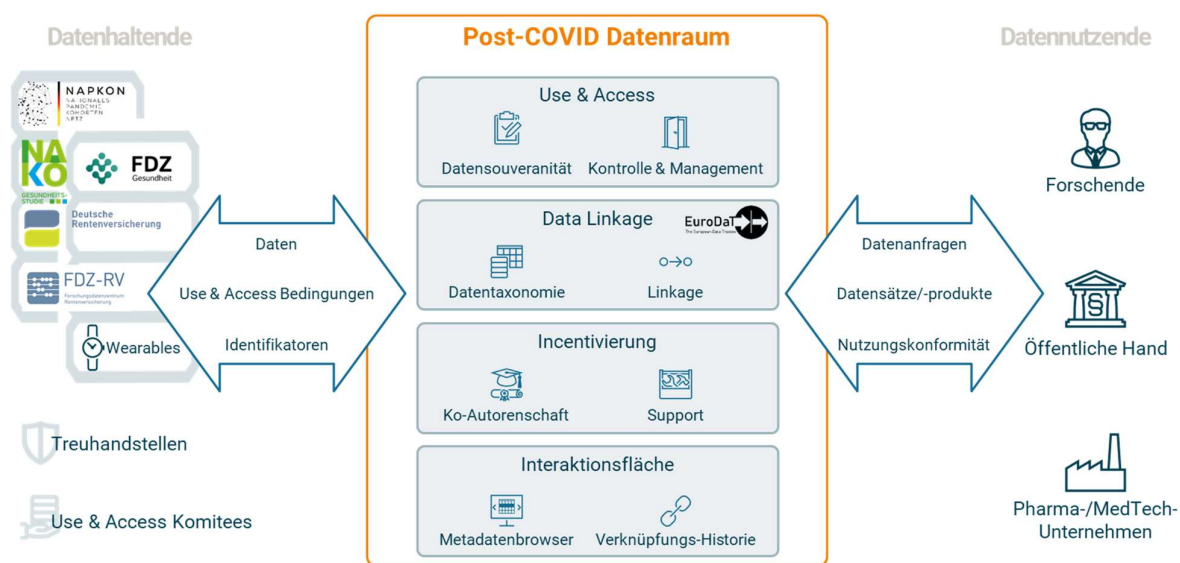


Challenge „Post-COVID-Datenmodell“



Bericht über die abgeschlossene Stufe 2

13. Dezember 2024

MANAGEMENT SUMMARY

Dieser Bericht dokumentiert die Arbeiten der zweiten Stufe im Rahmen der „Post-COVID-Datenmodell“-Challenge mit Laufzeit vom 23. September 2024 bis zum 27. Dezember 2024. In dieser Phase haben wir wie geplant die Implementierung unseres Datenökosystems vorangetrieben und dafür die notwendigen Stakeholder-Prozesse ausgearbeitet, das Architekturdesign der technischen Infrastruktur entworfen und aufgebaut sowie relevante Datensätze eingeworben und ihre Anbindung vorbereitet. Hierbei haben wir von Anfang an rechtliche Rahmenbedingungen wie Datenschutz und aufkommende nationale und europäische Regularien berücksichtigt.

HAUPTPUNKTE DES BERICHTS

In diesem Bericht beschreiben wir unseren Implementierungsansatz. Wir stellen hierfür formal strukturierte Geschäftsprozesse vor, die die fachliche Logik unserer Lösung technisch sicher umsetzen. Darüber hinaus beschreiben wir die Komponenten, die das Teilen und das Verknüpfen von Daten gemäß unserem technisch, prozedural und rechtlich integrierten Ansatz ermöglichen. Diese Ergebnisse stellen entscheidende Fortschritte dar, da sie vom Einsammeln der Daten, über deren Verarbeitung und Verknüpfung bis hin zu ihrer Auslieferung alle Ebenen des Datenmodells abbilden.

NÄCHSTE SCHRITTE

Die nächsten Schritte unserer Arbeit werden den Betrieb der entwickelten Lösung vorbereiten. Dafür werden wir die technische Applikation in ein finales MVP weiterentwickeln, die fortlaufende Datenaufnahme und Modellaktualisierung vorbereiten und unserer Prozessbeschreibungen veröffentlichen. Des Weiteren werden wir ausnutzen, dass wir unsere Ergebnisse stets allgemein formuliert haben, um unsere Arbeit in ein Konzept für die sektorunabhängige Entwicklung offener Datenmodelle zu übertragen. Schließlich möchten wir unterschiedliche Betriebsmodelle, einschließlich kommerzieller, gemeinnütziger und staatlich-hoheitlicher Ansätze, am Markt testen.

UNSER UNIQUE SELLING POINT (USP)

Der USP unseres Ansatzes liegt in einem systemisch umfassenden Konzept, das eine integrierte Sicht auf alle technischen, prozeduralen und rechtlichen Aspekte eines offenen, dezentralen Datenökosystems bietet. Insbesondere legen wir großen Wert auf Datensouveränität, Rechtssicherheit und einfache Zugänglichkeit des Datenökosystems. Zudem fördern wir die Verknüpfung von Medizin, Rechts- und Sozialwissenschaften und lösen für diese das langjährige Problem des Record Linkage. Hierdurch eröffnen wir einerseits umfassende neue Möglichkeiten für die Post-COVID Forschung und entwickeln andererseits neue konzeptionelle Kompetenzen für das Dateninstitut als Enabler einer datengetriebenen und interdisziplinären Wertschöpfung.

INHALTSVERZEICHNIS

Management Summary	i
01. Einleitung.....	1
02. Gesamtbeschreibung der geleisteten Arbeiten.....	2
02.01 Beschreibung der geleisteten Arbeiten	2
02.02 Entwickeltes geistiges Eigentum und Veröffentlichungen	6
03. Konzepte	6
03.01 Data Governance Konzept.....	6
03.02 Konzept der Datenmodellpflege	14
03.03 Schulungskonzept.....	14
03.04 Grundlegende Fachprozesse	18
04. Forschungsobjekt.....	25
04.01 Ausrichtung des Datenmodells an den Anforderungen der Forschung.....	25
04.02 Prozesse zur Einbindung von Stakeholdern.....	27
04.03 Konzept zur Veröffentlichung des Datenmodells.....	30
04.04 Innovationsgrad des Ansatzes.....	36
05. Datenmodell	39
05.01 Vorliegende Datensätze.....	39
05.02 Orientierung an Branchen-Standards	44
05.03 Verknüpfung verschiedener Datenquellen	45
05.04 Strukturierung des Datenmodells (Typ der Datenbank).....	50
06. Prozesse und Architektur.....	52
06.01 Planung der IT-Infrastruktur	52
06.02 Aufbau der Infrastruktur	56
06.03 Prozesse zur Datenintegration.....	57
06.04 Prozesse zur Datenaktualisierung	59
07. Betrieb und Nachnutzung des Datenmodells	64

01. EINLEITUNG

Die COVID-19-Pandemie hat tiefgreifende Auswirkungen auf die Gesellschaft und das Gesundheitssystem hinterlassen. Und noch immer bleiben die komplexen Wirkungsweisen und Symptome, die sowohl durch die virale Infektion als auch durch veränderte Lebensumstände während der Pandemie ausgelöst wurden, ein zentrales Forschungsfeld. Für eine effektive Forschungsarbeit ist dabei eine umfassende und praktisch nutzbare Datengrundlage wichtig. In Stufe 1 der Post-COVID-Challenge haben wir daher ein Datenökosystem konzipiert, in Stufe 2 einen detaillierten Implementierungsplan entworfen und mit dessen Umsetzung begonnen. Hierfür haben wir die notwendigen Stakeholder-Prozesse aufgebaut und die technische Infrastruktur vorbereitet, mit der wir relevante Datensätze effektiv an das Ökosystem anbinden und nutzbar machen werden.

Technisches und konzeptionelles Kernstück des Datenökosystems soll dabei ein neuartiger transaktionsbasierter Treuhänder werden, der Data Governance, Datenschutz, Datenaktualität und Datensouveränität der Datengebenden automatisiert absichern kann und über den Use & Access- und Incentivierungs-Verfahren sicher abgebildet werden können. Grundlage unseres Implementierungsansatzes sind dabei die von uns in Stufe 1 konzipierten Prozesse zur Stakeholder-Einbindung und die technische Datengabe über das Ökosystem. Diese Prozesse haben wir in Stufe 2 auf konkret ausdetaillierte Geschäftsprozesse abgebildet, die nahtlos in den logischen Workflow der von uns entwickelten Datenapplikation integriert werden können. So stellen wir sicher, dass unsere Lösung von Anfang an auf eine einfache Nutzung und einen adäquaten Funktionsumfang optimiert ist.

Unser Minimum Viable Product (MVP) schafft auf dieser Basis eine benutzerfreundliche Plattform, die rechtssichere Datentransaktionen ermöglicht. Des Weiteren legen wir einen besonderen Schwerpunkt auf die Implementierung von Data Linkage, insbesondere Record Linkage, um die Verknüpfung von Daten zu ermöglichen und so tiefere Einblicke und Analysen zu erlauben. Aufbauend auf den Konzepten und Infrastrukturen eines transaktionsbasierten Datentreuhänders, wie z.B. [EuroDaT](#), gestalten wir das MVP so, dass sowohl die Integrität als auch die Zugänglichkeit der Daten gewahrt bleibt. Hierbei ist es für uns entscheidend, den Fokus auf mehrwertstiftende Elemente des MVPs zu legen, um nicht nur einen funktionalen Prototyp zu schaffen, sondern auch um die Motivation der Akteure zu stärken, weiterhin aktiv am Vorhaben teilzunehmen.

Darüber hinaus haben wir die Stakeholder-Prozesse gezielt erweitert und verfeinert und unsere Gespräche mit vier zentralen datenhaltenden Stellen vertieft. Hierdurch konnten

wir bereits beispielhafte Daten für drei besonders hochwertige Datensätze einsammeln und haben für einen vierten wertvollen Datensatz eine Lieferzusage erwirken können.

Die Interaktionsfläche für die Nutzenden des Datenökosystems haben wir durch die Entwicklung einer benutzerfreundlichen [Website](#) und die Einrichtung eines [GitLab-Repositories](#) (aktuell noch zugangsbeschränkt) optimiert. Die sich im Aufbau befindende [Website](#) fungiert als zentrales Portal, das den Zugang zu Metadaten, zu Informationsinhalten über das Ökosystem und zum Datenmodell vereinfacht und eine intuitive Nutzeroberfläche bietet, die die Sichtbarkeit der verfügbaren Daten verbessert und die Teilhabe der Datenhaltenden sowie -nutzenden an der Plattform fördert. Parallel dazu dient das [GitLab-Repository](#) als Code-Repository und Zugangspunkt für externe Nutzende sowie auch als interaktive Plattform, die den aktiven Austausch und eine Beteiligung an der Modellentwicklung ermöglicht, einschließlich der Möglichkeit, Änderungen vorzuschlagen und Feedback zu geben. Durch diese beiden Angebote schaffen wir eine umfassende und effektive Kommunikationsstruktur, die die Kollaboration innerhalb des Datenökosystems fördert.

In Summe haben wir für jede der in Stufe 1 identifizierten Herausforderungen Lösungen gefunden. Mit der Treuhänder-Plattform und dem dazugehörigen Datenkatalog stellen wir sicher, dass eine rechtssichere Grundlage für das Teilen von Daten geschaffen und die Datenauffindbarkeit verbessert wird. Durch die Einbindung relevanter Stakeholder können wir eine umfassende Qualitätssicherung der Daten erreichen und gleichzeitig die Antragsprozesse für den Zugang und die Nutzung harmonisieren. Die Kommunikation mit externen Stellen wie [NFDI4Health](#), [HealthDCAT-AP](#) oder [THS Greifswald](#) hat dazu beigetragen, existierende Standards des medizinischen Datenmanagements einzubinden und Anreize für das Teilen von Forschungsdaten zu schaffen. Schließlich haben wir durch die Entwicklung einer klaren Datentaxonomie die Vereinfachung der Vertragsgestaltung für Datenteilungsvereinbarungen erreicht und gleichzeitig die Einordnung in bestehende sowie entstehende Regulatorik vorbereitet.

02. GESAMTBESCHREIBUNG DER GELEISTETEN ARBEITEN

02.01 BESCHREIBUNG DER GELEISTETEN ARBEITEN

Der Hauptfokus der Stufe 2 lag darin, die praktische Implementierung des von uns in Stufe 1 konzipierten Datenmodells voranzutreiben und dafür insbesondere die notwendigen Stakeholder-Prozesse auszubauen, die Infrastruktur vorzubereiten und die Anbindung der relevantesten Datensätze zu planen. Dieses Kapitel beschreibt unsere dafür geleisteten Arbeiten und fasst die Erkenntnisse und Ergebnisse zusammen. Tiefergehende Details sind im weiteren Dokument verfügbar und verlinkt.






 Stakeholder	<ul style="list-style-type: none"> • Erweiterung und Verfeinerung der Stakeholder-Liste mit Akteuren aus den Gruppen Datenhaltende, -nutzende und Intermediäre • Vertiefende Gespräche mit vier datengebenden Stellen: NAPKON, NAKO, Rentenversicherung, Garmin Health • Austausch und Abstimmung mit Initiativen der Gesundheitsdaten-ökonomie: NFDI4Health, MII, NUM, HealthDCAT-AP, THS Greifswald
 Datensätze	<ul style="list-style-type: none"> • Anfragen und einsammeln von Metadaten und Dummy-Daten • Analyse gewonnener Metadaten und Dummy-Daten • Konzept der Datenspende ausgearbeitet, insb. Übersicht Datenschutz • Konzeptionierung der Datenverknüpfung mittels EuroDaT
 Datenökosystem	<ul style="list-style-type: none"> • Design und Ausgestaltung der wichtigsten Business-Prozesse, z.B. Datenabruf, -verknüpfung, -suche, -integration und Informationssuche • Einrichtung der Website als Portal zur Interaktionsfläche • Erstellen von Wireframes und Inhalten als Zielbild der Website • Exemplarische Integration des Datentreuhänders EuroDaT
 Technik	<ul style="list-style-type: none"> • Anforderungsaufnahme und Implementierung des grundlegenden Technologie-Stacks mit Vue JavaScript Frontend, Kotlin Spring-Boot-Backend, PostgreSQL DB, Authentisierungs- und Autorisierungskomponente Keycloak • Design der technischen IT-Architektur inkl. <ul style="list-style-type: none"> • Konzipierung der Client-Applikation für Datengebende • Konzipierung der angestrebten Zielarchitekturen
 Dokumentation	<ul style="list-style-type: none"> • Dokumentierte Ergebnisse in dem vorliegenden Bericht • Einrichtung GitLab Repository • Website für integrierte Informationsbereitstellung und Zugang zum Datenökosystem

Tabelle 1: Zusammenfassende Darstellung der Ergebnisse der Stufe 2.

Zur Strukturierung der folgenden Ausführungen fassen wir unsere zentralen Ergebnisse in Tabelle 1 zusammen.

Im Folgenden stellen wir unsere Ergebnisse detailliert dar und verweisen auf die entsprechenden Kapitel dieses Berichts mit weiterführender Dokumentation.

STAKEHOLDER

Ein wichtiger Bestandteil der Stufe 2 war die Fortführung und Konkretisierung der Kommunikation mit den Stakeholdern. Hierbei spielten insbesondere die Datenhaltenden eine elementare Rolle. Die Ausrichtung unseres Modelles an den Anforderungen der Forschung sowie die aktiven Prozesse zur Einbindung von Stakeholdern und Abfrage ihrer Datensätze beschreiben wir in Kapiteln 04.01 und 04.02.

DATENSÄTZE

Wir haben eine Priorisierung der Datensätze vorgenommen und von den vier wichtigsten datengebenden Stellen sowohl Metadatenbeschreibungen als auch explorative Datensätze entweder bereits einwerben können oder Zusagen für eine

Datengabe erhalten. Anschließend haben wir die uns zur Verfügung gestellten Metadaten analysiert und auf Basis geteilter Testdaten erste Integrationstests unternommen. Hierdurch haben wir ein eingehendes Verständnis der Daten aufgebaut und einen exemplarischen Use-Case entwickelt. Informationen hierzu sowie zu den Datensätzen, dem weiter ausgearbeiteten Bereich der Datenspende, sowie den ausgearbeiteten Konzepten zur Verknüpfung von Daten mittels [EuroDaT](#) können im Kapitel 05.01 und 05.03 gefunden werden.

DATENÖKOSYSTEM

Das Datenökosystem wurde von uns weiter ausgestaltet. Insbesondere haben wir hierfür die in Stufe 1 konzipierten Prozessabläufe des Datenökosystems in konkrete Geschäftsprozesse übersetzt und weiter ausdetailliert. Exemplarisch stellen wir hierzu in Kapitel 03.04 die Business-Prozesse zu Datenabruf, Datenverknüpfung und Datensuche vor. Außerdem haben wir eine [Website](#) als Anlaufstelle für das Ökosystem designt und das technische Hosting aufgebaut, s. Kapitel 04.03. Als Zielbild für die [Website](#) haben wir Wireframes und wesentliche Inhalte gestaltet, die kontinuierlich in die öffentlich zugängliche Instanz überführt werden. Um Datenverknüpfungen zu ermöglichen, wurde die Integration des Datentreuhänders [EuroDaT](#) weiter ausgearbeitet, s. Kapitel 03.01, 04.04 und 06.01. Unser Ansatz ist dabei grundsätzlich technologisch agnostisch und kann jeden transaktionsbasierten Treuhänder zur Umsetzung nutzen. Aktuell ist uns außer EuroDaT allerdings kein vergleichbares Angebot bekannt, weswegen wir diesen Treuhänder für unser anvisiertes Datenökosystem exemplarisch anbinden. Durch unsere modulare Architektur gewährleisten wir jedoch einen einfachen Austausch von Komponenten – so auch die Anbindung an weitere transaktionsbasierte Treuhänder.

TECHNIK

Ein bedeutender Fokus unserer Arbeiten in dieser Stufe lag auf der technischen Umsetzung des MVP. Hierzu wurde die Implementierung des Front- und Backends gestartet, der entsprechende Tech-Stack aufgebaut, die notwendige Infrastruktur vorbereitet, Clients für Datengebende konzipiert und begonnen, die oben beschriebenen Business-Prozesse in dieser Infrastruktur abzubilden. Unsere IT-Architekturdokumentation orientiert sich am [arc42](#)-Standard. Details hierzu finden sich in Kapitel 05.04, 06.01 und 06.02.

DOKUMENTATION

Der vorliegende Bericht bildet die Grundlage für die weitere Ausarbeitung des MVP und die nächsten Schritte. Als Erweiterung zur Dokumentation wurde ein [GitLab-Repository](#) eingerichtet und der Aufbau der oben beschriebenen [Website](#) begonnen, über die unsere

Arbeiten für die Öffentlichkeit dokumentiert werden. Genauere Informationen hierzu sind in Kapitel 04.03 beschrieben.

FACHLICHE FRAGESTELLUNGEN

Neben den eben beschriebenen konkreten Implementierungsarbeiten des Datenökosystems, haben wir uns zur Ausrichtung unserer Arbeiten kontinuierlich weiter mit den fachlichen Fragestellungen der Post-COVID-Forschung befasst. Hierbei sind nach wie vor viele Fragen offen, insbesondere zu den komplexen Wirkungsweisen und Symptomen, die sowohl durch die virale Infektion selbst als auch durch veränderte Lebensumstände während der Pandemie verursacht werden können. Wichtige Forschungsfragen betreffen vor allem die negativen Langzeitwirkungen von COVID, die Clusterung von Symptomen, qualitative und quantitative Aspekte der Symptomatik sowie Unterschiede zu anderen Viruserkrankungen wie Influenza.

Um diese Fragen bearbeiten zu können, ist es von besonderem Wert, Datenquellen aus verschiedenen Sektoren zu verknüpfen, da somit tiefergehende Fragestellungen erschlossen werden können, basierend auf bereits aufwändig erhobenen Daten und gewonnen Erkenntnissen. Unter diese Datenquellen fallen etwa Krankenversicherungsdaten, Studiendaten zu COVID und Post-COVID, Rentenversicherungsdaten sowie Informationen zu Lebensstil und Gesundheit. Diese Daten helfen, ein tieferes Verständnis für die Zusammenhänge, Auswirkungen und gesellschaftliche Relevanz von Post-COVID zu entwickeln.

Zur inhaltlich sinnvollen Erprobung unseres Vorhabens zur Datenverknüpfung konzentrieren wir uns auf die folgende, durch die Medizinsäule des Projektteams definierte exemplarische Fragestellung: **Wie wirkt sich Post-COVID auf die Erwerbsfähigkeit aus?** In diesem Zusammenhang planen wir, die Daten der [NAPKON](#)- und [NAKO](#)-Studien zusammenzuführen und diese mit Informationen der [Rentenversicherung](#) zu ergänzen. Aus Gesprächen mit unseren medizinischen Konsortialpartnern hat sich ergeben, dass diese Fragestellung von großer Relevanz ist, bislang jedoch nur schwer zu erforschen war. Ein positives Ethikvotum für diese Forschungsfrage ist in Arbeit.

Aktuell befinden wir uns im Antragsverfahren, um Zugang zu den Daten von sowohl [NAPKON](#) als auch [NAKO](#) zu erhalten. Im nächsten Schritt werden wir zusätzlich das erforderliche Verfahren für den Datenzugriff auf die Informationen der [Rentenversicherung](#) in Angriff nehmen. Unser Ziel ist es, sowohl die rechtlichen Hürden für den Datenzugriff wie beispielsweise die Genehmigung von Nutzungsanträgen, agil zu überwinden, als auch die notwendige technische Infrastruktur bereitzustellen und zu

integrieren. Mit diesem parallelen Ansatz gewährleisten wir schnellstmögliche Fortschritte.

Die Ansätze und Methoden, die wir zur Erprobung der Datenverknüpfung im Kontext der Post-COVID-Forschung verfolgen, sind nicht nur auf diese spezifische Fragestellung begrenzt, sondern bieten auch wertvolle Erkenntnisse für eine Vielzahl anderer Forschungsfragen. Durch die Möglichkeit der Kombination von Daten aus verschiedenen Studien und Institutionen schaffen wir eine flexible Infrastruktur, die sich leicht an unterschiedliche Forschungsbedarfe anpassen lässt.

Diese Methodik der interdisziplinären Datenintegration ermöglicht es, umfassendere Analysen durchzuführen und komplexe Zusammenhänge zwischen verschiedenen Gesundheitsaspekten und sozialen Faktoren zu erkennen. Darüber hinaus können die gewonnenen Erkenntnisse als Modell und Wissensbasis für zukünftige Projekte in anderen Bereichen der Gesundheitsforschung dienen, sei es bei der Untersuchung chronischer Erkrankungen, übertragbarer Krankheiten oder sozialer Determinanten von Gesundheit.

02.02 ENTWICKELTES GEISTIGES EIGENTUM UND VERÖFFENTLICHUNGEN

Die im bisherigen Projektverlauf und insbesondere in dieser Stufe entwickelten Ergebnisse haben wir in Kapitel 02.01 zusammengefasst. An diesen Ergebnissen haben wir keine geistigen Eigentumsrechte angemeldet, um die Nachnutzbarkeit des Ansatzes zu wahren und im Interesse maximaler Transparenz eine frühzeitige Veröffentlichung unserer Ergebnisse schon in der Entwicklungsphase zu ermöglichen. Ein solch transparenter Ansatz wäre nicht möglich, wenn wir ein vollständig entwickeltes Produkt als Schutzmarke anmelden würden, was typischerweise erst am Ende der Projektarbeit erfolgen kann.

Für eine detaillierte Zusammenfassung der entwickelten Veröffentlichungskanäle verweisen wir auf Kapitel 04.03.

03. KONZEPTE

03.01 DATA GOVERNANCE KONZEPT

Unter Data Governance verstehen wir ein Regelwerk, das festlegt, wie man mit Daten umgehen darf, und das zusammen mit einem formalen Prozessgerüst eine effektive Kontrolle ihrer Einhaltung ermöglicht. Indem wir klare Prozesse, Rollen und Aufgaben definieren und geeignete Kontrollmechanismen etablieren, stellen wir die Vertrauenswürdigkeit und Neutralität des angebotenen Datenservices sicher.

GRUNDLEGENDE ASPEKTE UNSERES ANSATZES

Ein Kernaspekt unseres Ansatzes ist es, die Einhaltung vorgegebener Regeln und Rollenkonzepte mit technischen Mitteln automatisch zu garantieren. Im Folgenden wollen wir daher beschreiben, wie unser Ansatz eine weitgehend automatisierte Data Governance ermöglicht und so Integrität, Vertraulichkeit und Verfügbarkeit der geteilten Daten gewährleistet.

Als zentrales Merkmal der Data Governance übertragen wir daher ein Prinzip des Datentreuhänders [EuroDaT](#) auf den vorliegenden Anwendungsfall: Grundsätzlich ist der Treuhänder so aufgebaut, dass er keinerlei wirtschaftliches Interesse an der Nutzung oder Vermarktung von über ihn geteilten Daten haben und entwickeln kann. In diesem Sinne soll die Governance des vorliegenden Datenmodells ebenfalls so ausgestaltet sein, dass die betreibende Stelle ausschließlich für ihre Dienstleistungen bezahlt wird, hingegen keinerlei Nutzen aus den Daten selbst ziehen darf und kann. Darüber hinaus muss die Governance sicherstellen, dass Datennutzende und Datengebende nicht direkt miteinander in Kontakt kommen müssen, sondern ein neutraler Intermediär als eigenständiger Markt- und Vertragspartner zwischengeschaltet ist. Dieser Intermediär verwaltet dabei auf Grundlage der Metadatenbeschreibungen, der Einwilligungen der Datenhaltenden sowie der gesetzlichen Vorgaben lediglich die Zugriffsrechte aller teilnehmenden Organisationen auf geteilte Daten und deren Analysen. Die Primärdaten und die Kontrolle über ihre Nutzung verbleiben dauerhaft bei den Datengebenden, die die Nutzungsrechte nach eigenem Ermessen jederzeit gewähren und entziehen können.

Eine zentrale Eigenschaft des hier vorgestellten Konzepts ist die Vereinigung von Data Governance mit technologischer Infrastruktur: Durch den Einsatz eines transaktionsbasierten Datentreuhänders mit ausgefeiltem Rechte- und Rollenkonzept (s. Abb. 1 aus dem Abschlussbericht der Stufe 1) wird die Einhaltung der vorgegebenen Data Governance technisch abgesichert, automatisiert und Missbrauch ausgeschlossen. Die technische Infrastruktur des Datentreuhänders zielt dabei darauf ab, die Datennutzung in einer sicheren, rechtskonformen und vertrauenswürdigen Umgebung zu ermöglichen, welche die Data Governance-Anforderungen des Post-COVID-Datenmodells technisch umsetzt. Beispielsweise schließt der Datentreuhänder Daten für jede Transaktion neu aus der Infrastruktur der Datengebenden an den Datenraum an und speichert sie nicht, was die Datensouveränität der Datengebenden sichert. Innerhalb der gesicherten Umgebung des Treuhänders werden die Daten nach Bedarf pseudonymisiert, anonymisiert oder aggregiert, um Datenschutz und Datensicherheit auch für die ausgegebenen Daten zu gewährleisten. Das Modell stellt sicher, dass nur pseudonymisierte, anonymisierte oder aggregierte Daten an die Datennutzenden übermittelt werden, wobei der Zugriff auf die Daten streng reguliert ist und nur

legitimierten Nutzenden gewährt wird. Verarbeitungsalgorithmen werden für jede Transaktion separat und ohne menschliche Einflussmöglichkeit in einer jeweils neu provisionierten gekapselten Umgebung des Datentreuhänders ausgeführt. Erst hierdurch ermöglicht unser Data Governance Ansatz so zentrale Lösungen wie Record Linkage auf Einzelpatientenebene.

Neben dieser grundsätzlichen Erwägung, umfasst unser Data Governance Konzept drei weitere Bereiche: Das Datenmanagement, die Datenqualitätssicherung sowie die kontinuierliche Anpassung und Verbesserung der Governance.

DATENMANAGEMENT

Um Daten effektiv verknüpfen und analysieren zu können, müssen Prozesse und Richtlinien festgelegt werden, die eine angemessene und vertrauenswürdige Nutzung des angebotenen Post-COVID Datenraums sicherstellen. Die Richtlinien legen dabei Rechte und Pflichten sowohl der Datengebenden als auch -nehmenden fest, deren Einhaltung eine faire Gleichbehandlung aller Akteure im Post-COVID Datenökosystem sicherstellen soll. Um eine allgemeine Akzeptanz zu sichern, sehen wir vor, dass Nutzende die vereinbarten Richtlinien als Teilnahmebedingungen des Datenökosystems bereits beim Onboarding akzeptieren müssen. Konkret sehen wir Richtlinien zu den Aspekten Zugriffsmanagement, Datenkategorisierung und Datenverarbeitung als wichtig an.

Zugriffsmanagement

Das Zugriffsmanagement legt fest, welche Nutzende auf welche Daten und Bereiche des Datenökosystems zugreifen dürfen. Die Freigabe von Daten bleibt dabei in unserem Ansatz grundsätzlich den Datengebenden vorbehalten. Der Zugriff auf das Ökosystem wird hingegen über ein klares Rollenkonzept gesteuert.

Technisch wird das vereinbarte Zugriffsmanagement wie oben beschrieben über den Datentreuhänder abgebildet. Grundlage dieser technischen Umsetzung soll ein angepasstes Rollenkonzept sein, das Rechte und Verantwortlichkeiten festlegt. Die genaue Ausarbeitung des Rollenkonzepts kann dabei erst erfolgen, sobald das Betriebsmodell festgelegt ist, für das wir aktuell verschiedene Optionen wie öffentliche oder privatwirtschaftliche Trägerschaft prüfen. In jedem Fall soll der Betrieb ohne wirtschaftliche Gewinnerzielungsabsicht erfolgen, sondern lediglich kostendeckend arbeiten. Wir skizzieren daher die Grundzüge der geplanten Rollen, s. Abbildung 1.

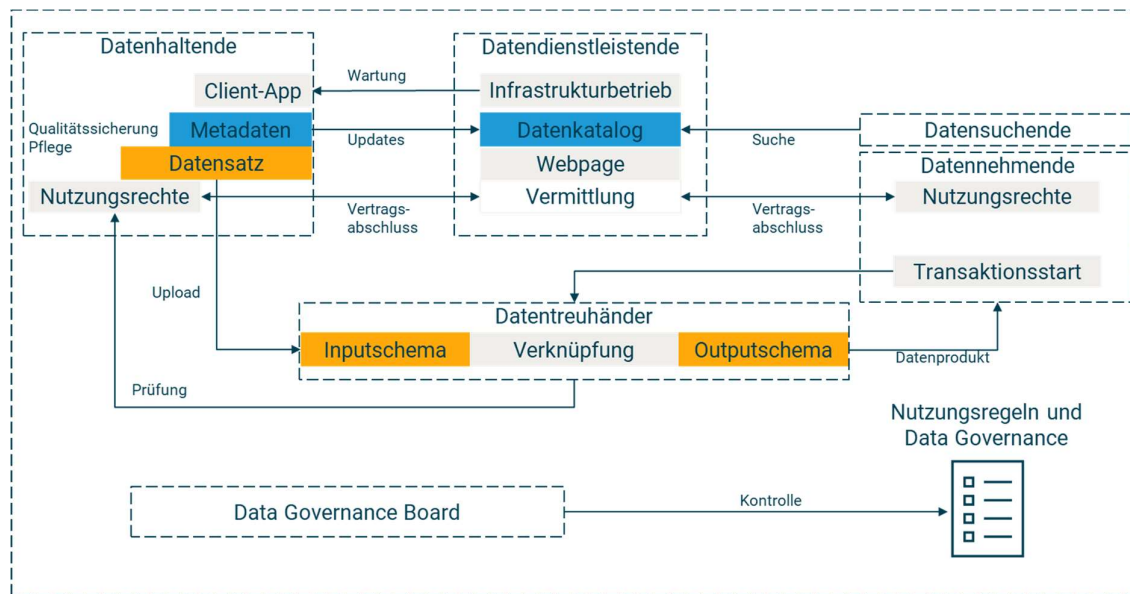


Abbildung 1: Rollenkonzept des Datenökosystems.

EuroDaT Rollen

Datengebende haben die Verantwortung hochwertige Daten bereitzustellen, eine Qualitätssicherung der Daten auszuführen und eine Metadatenbeschreibung gemäß vereinbartem Standard im Datenmodell zu pflegen. Bei Änderungen der Primärdaten oder ihrer Metadatenbeschreibung müssen sie eine Meldung stellen und für Rückfragen der Datennehmenden erreichbar sein. Technisch dürfen Datengebende ausschließlich Daten in das Inputschema des Treuhänders schreiben, s. Kapitel 03.04.

Datennehmende haben die Verantwortung, die ihnen zur Verfügung gestellten Daten nur im Rahmen der vertraglich vereinbarten Nutzungserlaubnis zu verwenden und das Datenökosystems nur im Rahmen der vereinbarten AGB zu nutzen. Technisch dürfen Datennehmende Datentransaktionen auslösen, sofern sie die Berechtigung haben, und anschließend Daten aus dem spezifischen Datenbank-Schema des Treuhänders lesen, s. Kapitel 03.04.

Datendienstleistender wird im Post-COVID Datenökosystem die betreibende Stelle. Diese hat die Verantwortung, einen reibungslosen Ablauf jeder angefragten Datentransaktionen zu sichern und dafür die Einsatzfähigkeit sowohl der technischen Infrastruktur der Interaktionsflächen wie des Webfrontends und der Client-App als auch der dahinterstehenden Infrastruktur sicherzustellen. Darüber hinaus verantworten die Datendienstleistenden die Aktualisierung des Datenmodells gemäß der von den Datengebenden abgefragten Primär- und Metadatenbeschreibungen. Technisch dürfen sie im Treuhänder Daten weder lesen noch schreiben.

Vertragsbeziehungen

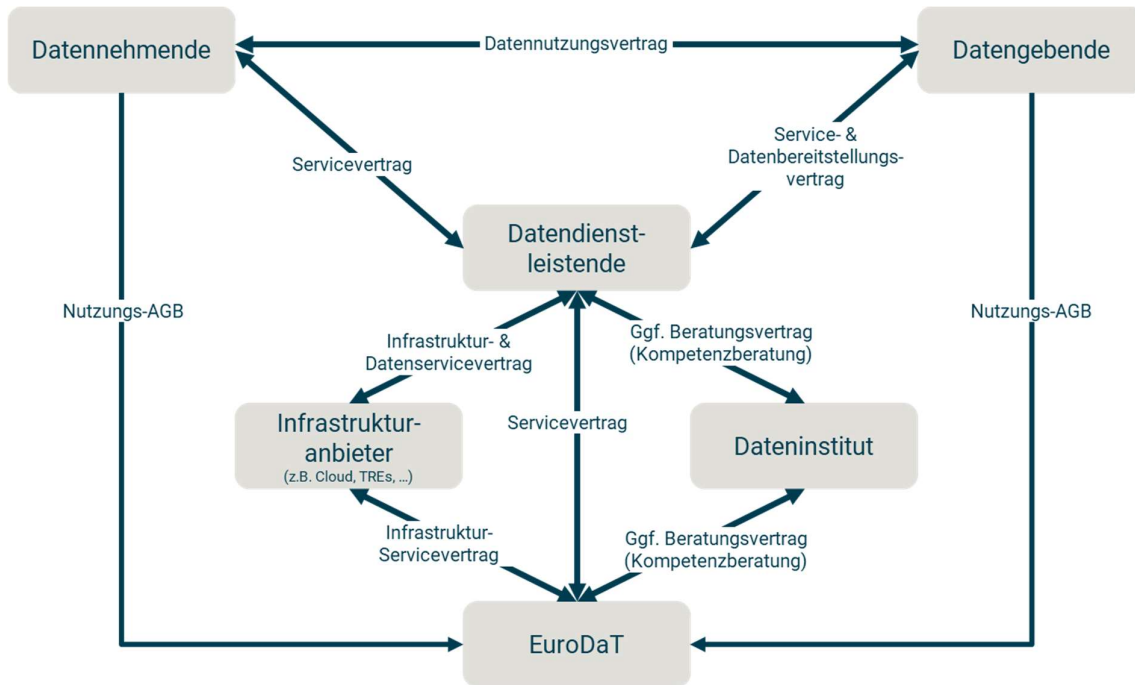


Abbildung 2: Schematische Darstellung relevanter Vertragsbeziehungen.

Die Umsetzung der skizzierten Datentreuhand verlangt nicht nur eine komplexe technische Infrastruktur, sondern es sind auch die Beziehungen zwischen den unterschiedlichen beteiligten Akteuren rechtlich abzusichern. Hierzu bedarf es eingangs einer genauen Analyse der einschlägigen Leistung-Gegenleistung-Strukturen und einer hierauf abgestimmten, der jeweiligen Austauschbeziehung entsprechenden Vertragsgestaltung. Überblicksmäßig haben wir die zu berücksichtigenden Vertragsbeziehungen in der beigefügten Übersicht dargestellt, s. Abbildung 2. Die damit skizzierte Vertragsarchitektur beinhaltet aber natürlich noch keine spezifischen oder umfassend generalisierbaren Aussagen über die konkret erforderliche Vertragssituation; das bleibt einer späteren Analyse und namentlich der Anwendungspraxis vorbehalten. Zu berücksichtigen und bereits jetzt für die weiteren Überlegungen heranzuziehen sind indes die Vorarbeiten, die mit Blick auf die standardmäßig zu erwartenden Klauseln einem sog. Vertragsbaukasten zu entnehmen sein können. Ein solcher Vertragsbaukasten wurde prototypisch für EuroDaT entwickelt und im [Rechtshandbuch Datentreuhand](#) vorgestellt. Das Baukastenprinzip ermöglicht eine arbeitserleichternde und geschwindigkeitserhöhende Typisierung, lässt jedoch hinreichend Raum, um einzelfallbezogene Elemente aufnehmen zu können, und stellt damit gerade keine unzulässig simplifizierende one-size-fits-all-Lösung dar.

Weitergehende Rollen im Datenökosystem

Um die Angemessenheit der Datennutzungsverträge sowie ihre Einhaltung zu kontrollieren, richten wir ein Data Governance Board ein. Initial planen wir jährliche Sitzungen, in denen die strategische Ausrichtung und die Einhaltung der allgemeinen Datennutzungsrichtlinien überwacht werden, sowie asynchrone Kommunikationsformate zur Behandlung von ad-hoc Fragestellungen. Dieses Gremium setzt sich aus Vertretern der betreibenden Stelle, wie beispielsweise des Dateninstituts, zusammen und wird zudem durch Vertreter aus Forschung und Wirtschaft ergänzt. Diese zusätzlichen Mitglieder arbeiten ehrenamtlich und bringen wertvolle Perspektiven und Erfahrungen ein. Das Data Governance Board stellt somit das höchste Entscheidungsgremium innerhalb der Data Governance Hierarchie dar und spielt eine entscheidende Rolle bei der Gewährleistung von Compliance und strategischer Kohärenz im Umgang mit Daten.

Und als letzte, vertraglich nicht gebundene Rolle im Datenökosystem informieren sich Datensuchende unverbindlich über die verfügbaren Daten, z.B. über das öffentliche Informationsangebot und den Datenkatalog. Ihr Ziel ist es, zu verstehen, welche Daten verfügbar sind und welche Verknüpfungen bereits erfolgreich durchgeführt wurden. Sie nutzen die öffentlich zugängliche Website, um den Datenkatalog zu durchsuchen und Informationen zu den Daten zu finden, die für ihre Bedürfnisse relevant sind.

Datenkategorisierung

Um die Arbeit mit den Daten maximal zugänglich und erwartbare Hürden frühzeitig transparent zu gestalten, bietet unsere in Stufe 1 konzipierte und in Stufe 2 verfeinerte Datentaxonomie eine übersichtliche Kategorisierung der Daten. In Rücksprache mit Forschenden bietet eine Unterteilung in die folgenden zwei Kategorien einen guten Kompromiss zwischen Übersichtlichkeit und Nutzen für die Datennehmenden:

- **Pseudonyme oder anonyme Daten**, die ohne Identifikationsmerkmale vorliegen und für Forschungszwecke problemlos nutzbar sind, da die DSGVO nicht einschlägig ist
- **Personenbezogene Daten**, deren Verarbeitung unter strengen Datenschutzrichtlinien und Einholung der notwendigen Einwilligungen erfolgen muss.

Für personenbezogene Daten nehmen wir anhand der von uns erstellten Datentaxonomie eine Unterkategorisierung vor, s. Kapitel 04.04. Der Nutzen dieser Datentaxonomie für die hier vorgestellte Data Governance ist entscheidend, da die Taxonomie klare Zuständigkeiten und Nutzungsrechte darstellt, an deren Kategorisierung sich Nutzende des Datenmodells frühzeitig und übersichtlich zu fundamentalen Aspekten der Data Governance informieren können.

Datenverarbeitung

Unter diesem Begriff verstehen wir eine Richtlinie, die beschreibt, welche Daten mit welchen Verfahren bearbeitet werden dürfen, wie z.B. Datenbereinigung, -verknüpfung, -aggregation und -anreicherung. Insbesondere dieser Teil der Governance ist wichtig, da in [EuroDaT](#) stets eine vollautomatisierte Verarbeitung ohne menschliche Eingriffsmöglichkeit stattfindet, sodass die erlaubten Prozesse und Algorithmen klar vorgegeben sein müssen.

DATENQUALITÄTSSICHERUNG

Datenqualitätssicherung beschreibt einen Prozess, der sicherstellen soll, dass die Daten korrekt formatiert, mit passenden Metadaten beschrieben und inhaltlich fehlerfrei sind.

Da die in unserem Ökosystem bereitgestellten Daten nach aktueller Planung mehrheitlich von großen Forschungsverbünden bereitgestellt werden, müssen die Mechanismen der Datenqualitätssicherung, die die datengebenden Stellen bereits anwenden, sinnvoll und adaptiv mit in unser Konzept eingebunden werden.

Datenmonitoring

Als zentralen Prozess der Datenqualitätssicherung sehen wir ein Datenmonitoring vor. Damit bezeichnen wir einen Mechanismus, der die Daten bei ihrer ersten Bereitstellung und anschließend innerhalb des Datenökosystems jährlich prüft, s. Abbildung 3.

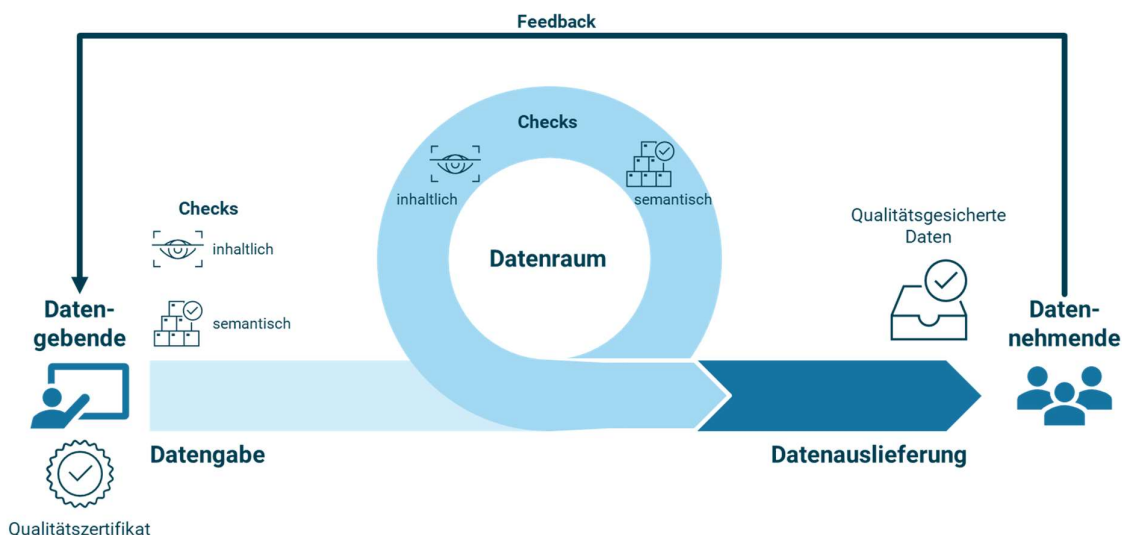


Abbildung 3: Schematische Darstellung des geplanten Datenmonitorings.

Hierbei ist zu beachten, dass der Datendienstleister aufgrund der Treuhänder-Konstruktion keinen Zugriff auf die Primärdaten haben kann. Daher sehen wir automatisierte Checks vor, für die wir zwei Arten unterscheiden

- **Semantische Checks** prüfen ob bereitgestellte Daten und Metadaten den vorgegebenen Datenmodellen und -standards entsprechen.
- **Inhaltliche Checks** prüfen ob die Daten innerhalb fachlich definierter, sinnvoll erwartbarer Grenzen liegen, z.B. würde ein Patientenalter von mehr als 200 Jahren auf einen Eingabefehler hindeuten.

Darüber hinaus können Daten, die bereits bei der gebenden Stelle ein internes Qualitätssicherungsprogramm durchlaufen haben, bei der Übergabe als qualitätsgesichert gekennzeichnet und im Datenraum so ausgeschrieben werden. Die Überprüfung durch den Betreibenden des Post-COVID Datenmodells lässt sich dann entsprechend auf das Qualitätssicherungsprogramm der datengebenden Stelle beschränken. Das jährliche interne Monitoring überprüft dann lediglich, ob die von den Datengebenden angemeldeten Qualitätsstandards weiterhin eingehalten werden.

Da wir keine menschliche Prüfinstanz vorsehen, kann es vorkommen, dass eine mangelnde Datenqualität nicht entdeckt wird. Für diesen Fall sehen wir Feedback-Loops vor, über die Datennehmende Rückmeldungen zur Datenqualität geben und so zur Verbesserung der Datenaufnahme und -verknüpfung beitragen können.

KONTINUIERLICHE ANPASSUNG UND VERBESSERUNG

Zuletzt sind die kontinuierliche Weiterentwicklung und Überprüfung der Prozesse entscheidend für eine gute Data Governance, da sie sicherstellen, dass Richtlinien an sich ändernde gesetzliche Anforderungen, technologische Entwicklungen und geschäftliche Bedürfnisse angepasst werden. Nur so kann eine ordnungsgemäße Funktion des Datenökosystems dauerhaft gewährleistet werden.

Konkret sehen wir folgende Mechanismen hierfür vor:

Governance-Reviews

Im Rahmen der Sitzungen des oben vorgestellten Data Governance Boards finden Reviews der vorliegenden Data Governance Konzepte und Prozesse statt, mit denen überprüft wird, ob die Abläufe den aktuellen Anforderungen an die Data Governance noch genügen. Bei Bedarf kann das Data Governance Board auch Arbeitsgruppen für detailliertere Untersuchungen oder Konzipierung neuer Prozesse einrichten.

Informationen und Schulungen

Wir erstellen Informationsmaterialien zu den geltenden Governance-Strukturen, die für alle Datennehmenden und -gebenden Teil des Onboardings werden. Die Kenntnisnahme dieser Schulungsmaterialien muss dabei von den Nutzenden bestätigt und protokolliert werden. Darüber hinaus werden alle an Betrieb und Weiterentwicklung des

Datenökosystems beteiligten Mitarbeitenden bei ihrem Onboarding zu Datenschutzbestimmungen und ethischen Standards der Data Governance geschult, s. auch Kapitel 03.03.

Regelmäßige Audits

Wir planen die zweijährliche Durchführung von Audits, um die Einhaltung der Data Governance Richtlinien systematisch zu überprüfen, wobei verschiedene Optionen für die Durchführung in Betracht gezogen werden, wie interne Prüfungsteams, externe Auditoren oder die Nutzung von automatisierten Prüfwerkzeugen, sowie verschiedene Methoden wie Interviews, Dokumentenanalysen und Datenanalysen zur Bewertung der Compliance und Effektivität der Prozesse.

Berichtswesen

Wir können bei Bedarf ein Berichtswesen zur Erstellung von jährlichen Berichten über Datenmanagement und -nutzung für Stakeholder einrichten. Diese Berichte können durch ihre Transparenz das Vertrauen in unsere Data Governance erhöhen und somit die gesellschaftliche Akzeptanz fördern. Mögliche Adressaten sind öffentliche Institutionen, wissenschaftliche Einrichtungen und die allgemeine Öffentlichkeit, während Veröffentlichungskanäle wie Online-Plattformen, Pressemitteilungen und Fachkonferenzen sowie Berichtformate in Form von schriftlichen Dokumenten, Infografiken und interaktiven Dashboards genutzt werden können.

03.02 KONZEPT DER DATENMODELLPFLEGE

Die Pflege des vorliegenden Datenmodells ist entscheidend, um die Qualität und Aktualität der bereitgestellten Daten für die Forschung sicherzustellen. Um Redundanzen zu vermeiden, verweisen wir hier auf Kapitel 06.04. Dort beschreiben wir die Prozesse, mit denen wir automatisiert sicherstellen, dass die geteilten Primärdaten in ihrem aktuellen Stand abgefragt werden. Zusätzlich stellen wir einen Maßnahmenkatalog zur Pflege der Metadaten sowie allgemein relevante Aspekte der Datenmodellpflege vor.

03.03 SCHULUNGSKONZEPT

Um das hier beschriebene Datenökosystem effektiv nutzen zu können, müssen sowohl die Nutzenden als auch die Betreibenden eine Reihe von Informationen erhalten. Der Informationsbedarf und geeignete Formate unterscheiden sich dabei je nach ihren Rollen und Interessen.

Im Folgenden stellen wir die notwendigen Schulungsbedarfe rund um das Datenökosystem sowohl für Nutzende als auch für Mitarbeitende der betreibenden Stelle des Datenökosystems vor. Mit dieser Aufteilung berücksichtigen wir die spezifischen Bedürfnisse der verschiedenen Zielgruppen und gewährleisten so eine effektive und praxisnahe Wissensvermittlung. Auf der einen Seite stehen die Nutzenden des Datenökosystems, wie Datengebende und -nehmende, die Kenntnisse zur effektiven Nutzung des Datenökosystems benötigen. Mitarbeitende der betreibenden Stelle auf der anderen Seite, benötigen ein tieferes Verständnis für die Systemadministration und Datenmanagementprozesse, um einen reibungslosen Betrieb aufrechterhalten zu können. Für beide Gruppen sehen wir eine Reihe verschiedener Informationsartefakte vor, s. Abbildung 4.

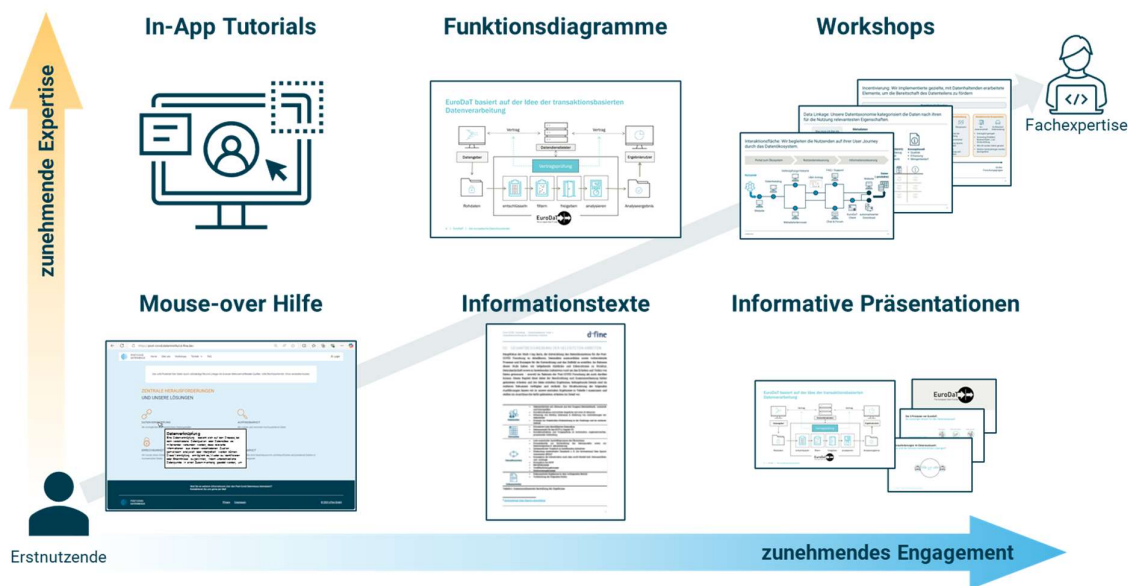


Abbildung 4: Vorgesehene Informationsartefakte je nach Engagement und Expertise der Nutzenden. Erstnutzende folgen ihrem bevorzugten Lernpfad zur Fachexpertise.

EINFÜHRUNG IN DAS DATENÖKOSYSTEM FÜR NUTZENDE

Notwendige Schulungs-Inhalte

Um das Datenökosystem optimal zu nutzen, müssen die Nutzenden zentrale Inhalte vermittelt bekommen, die die effektive Interaktion mit dem System fördern. Zunächst stellen wir hierfür grundlegende Informationen über das Datenökosystem, zu dessen Strukturen, Zielen und den relevanten Akteuren bereit. So können alle Nutzenden ein gemeinsames Verständnis der Rahmenbedingungen und Zielsetzungen entwickeln.

Für die Datennehmenden liegt ein Schulungsschwerpunkt auf dem Verständnis gängiger Zugriffsmodi auf Daten und des Datenmanagements. Hierbei müssen die

Datennehmenden insbesondere Verfahren zur Datenanfrage, die Nutzung des Metadatenkatalogs, ein Verständnis der Metadatenstruktur und Exploration bereits erfolgreicher Verknüpfungen erlernen. Dabei müssen sie unter anderem lernen Einschränkungen und Wiederverwendbarkeit bereits genutzter Konfigurationen einzuschätzen. Diese Kenntnisse sind entscheidend, um die benötigten Daten effizient zu finden und zu nutzen. Im Gegensatz dazu sind für die Datengebenden die Möglichkeiten der Datengabe von zentraler Bedeutung. Hier lernen sie die Verfahren zum Einpflegen von Daten in den Datenraum und Anforderungen an Metadatenformate kennen, was die Qualität und Nachvollziehbarkeit ihrer Beiträge sichert.

Ein weiterer wichtiger Bestandteil des Schulungskonzepts sind die Informationen zu Datenschutz- und ethischen Richtlinien, die allen Nutzenden grundlegende Kenntnisse über den verantwortungsvollen Umgang mit personenbezogenen Daten und die Einhaltung relevanter Vorschriften vermitteln. Diese Vorschriften können auch im Zuge des Onboardings der Nutzenden im Rahmen der zu akzeptierenden Teilnahmebedingungen geteilt werden, s. Kapitel 03.01.

Geeignete Schulungs-Formate

Es bieten sich konzeptionell vier Formate an, um den Nutzenden des Datenökosystems die notwendigen Schulungsinhalte effektiv zu vermitteln, s. auch Abbildung 4:

- **Statische Infomaterialien**
Diese bieten den Nutzenden die Möglichkeit sich entsprechend ihres angestrebten Grads an Engagement und vorhandener Expertise eigenständig zu informieren. Hierzu zählen wir z.B. die Mouse-Over Hilfstexte auf der Website und Informationstexte wie technischer Dokumentationen, Richtlinien und Allgemeine Geschäftsbedingungen.
- **In-App-Tutorials**
Diese Anleitungen und Hilfestellungen in der Frontend-App können insbesondere für erfahrene Nutzende beim Onboarding und Vorstellung der Funktionalitäten des Datenökosystems eingesetzt werden.
- **Präsentationen**
Lineare Formate, in denen nach einem konkreten Plan feste Inhalte vorgestellt werden, eignen sich zur breiten gut dafür, Informationen breit zu streuen und spezifische Inhalte zu vertiefen.
- **Interaktive Workshops**
Das höchste Maß an Engagement und Expertise setzen interaktive Formate voraus, in denen konkrete Fachfragen mit Experten aus der Anwendung sowie der Entwicklung und Betrieb erarbeitet werden können. Hierbei können konkrete Anwendungsfälle und Szenarien besprochen und die Nutzung des Datenraums

eingübt werden. Die Planung eines einführenden Workshops wurde bereits aufgenommen, mehr Details hierzu in Kapitel 04.02.

EINFÜHRUNG IN DAS DATENÖKOSYSTEM FÜR MITARBEITENDE DER BETREIBENDEN STELLE

Notwendige Schulungs-Inhalte

Zunächst ist für alle Mitarbeitenden eine detaillierte Schulung zu den Datenmanagementprozessen erforderlich. Dies umfasst die gesamten Abläufe der Datenakquise, -bereinigung, -validierung und -speicherung, sodass die Mitarbeitenden präzise wissen, wie Daten effizient und qualitativ hochwertig über das Datenökosystem geteilt werden können, s. auch Kapitel 03.01.

Darüber hinaus müssen die Mitarbeitenden mit den Anforderungen an Compliance und Qualitätsmanagement vertraut gemacht werden. Hierbei sollten gesetzliche Anforderungen, interne Richtlinien sowie Qualitätssicherungsmaßstäbe abgedeckt werden, um die Konformität und Integrität der Daten im Ökosystem zu gewährleisten. Durch diese umfassenden Schulungsinhalte erhalten die Mitarbeitenden das notwendige Know-how, um das Datenökosystem kompetent und effizient zu betreiben.

Mitarbeitenden im technischen Bereich muss darüber hinaus Fachwissen zu Systemkomponenten und -schnittstellen des Datenökosystems vermittelt werden. Dieses Wissen ist für die Nutzung der technischen Architektur maßgeblich und bildet die Grundlage für ein tiefgehendes Verständnis der Funktionsweise und Interoperabilität der einzelnen Teile des Systems.

Geeignete Schulungs-Formate

Die Schulungsformate für die Mitarbeitenden der betreibenden Stelle des Datenökosystems zeichnen sich durch einen höheren Betreuungs- und Begleitungsbedarf aus als die Formate für die Nutzenden des Ökosystems. Da die Mitarbeitenden eine entscheidende Rolle bei der Gewährleistung der Geschäftskontinuität und der effektiven Betriebsführung des Datenökosystems spielen, ist es wichtig, dass sie umfassend geschult werden, bevor die Betriebsphase beginnt. Dies erfordert ein intensives Training und die Unterstützung durch erfahrene Kollegen. Daher setzen wir auf Workshops, ähnlich des von uns geplanten Stakeholder Workshop zum Aufbau des Datenökosystems, s. Kapitel 04.02, um mit der Betriebsstelle des Datenökosystems interaktiv praktische Lernerfahrungen zu erarbeiten.

Zusätzlich zu den oben beschriebenen Schulungsformaten bietet sich für die Mitarbeitenden der Betriebsstelle außerdem das Format eines Mentoring-Programms an. Ein solches Programm setzt erfahrene Mitarbeitende als Mentoren für neue

Mitarbeitende ein. So kann über individuelle Unterstützung Verständnis für die spezifischen Analysen und den verantwortungsvollen Umgang mit den bereitgestellten Daten gefördert werden.

03.04 GRUNDLEGENDE FACHPROZESSE

In den vorangehenden Kapiteln haben wir dargestellt, wie wir den Teilnehmenden des Datenökosystems dessen Nutzung näherbringen und die Prozesse im Betrieb überwachen und absichern. Daran stellt sich die Frage, welche Prozesse für den Betrieb maßgeblich sind. In diesem Kapitel beschreiben wir die grundlegenden Fachprozesse methodisches Interesse, Datensuche und Datenabruf aus Sicht der Stakeholdergruppe der Datennutzenden und der Datenhaltenden.

FACHPROZESS METHODISCHES INTERESSE

Der Prozess (s. Abbildung 5) startet mit dem methodischen Interesse potenzieller Datennutzenden. Den zentralen Anlaufpunkt hierfür bildet unsere öffentlich zugängliche [Website](#), wo vorhandene Informationsangebote geprüft werden können. Bei weiter bestehendem Interesse sind auf der [Website](#) allgemeine Informationen, der Link zu dem dann öffentlichen Git-Repository, sowie Kontaktmöglichkeiten zu den Entwicklern hinterlegt. Dieser Prozess kann so lange durchlaufen werden, bis das Methodik-Interesse entweder befriedigt wurde oder erloschen ist.

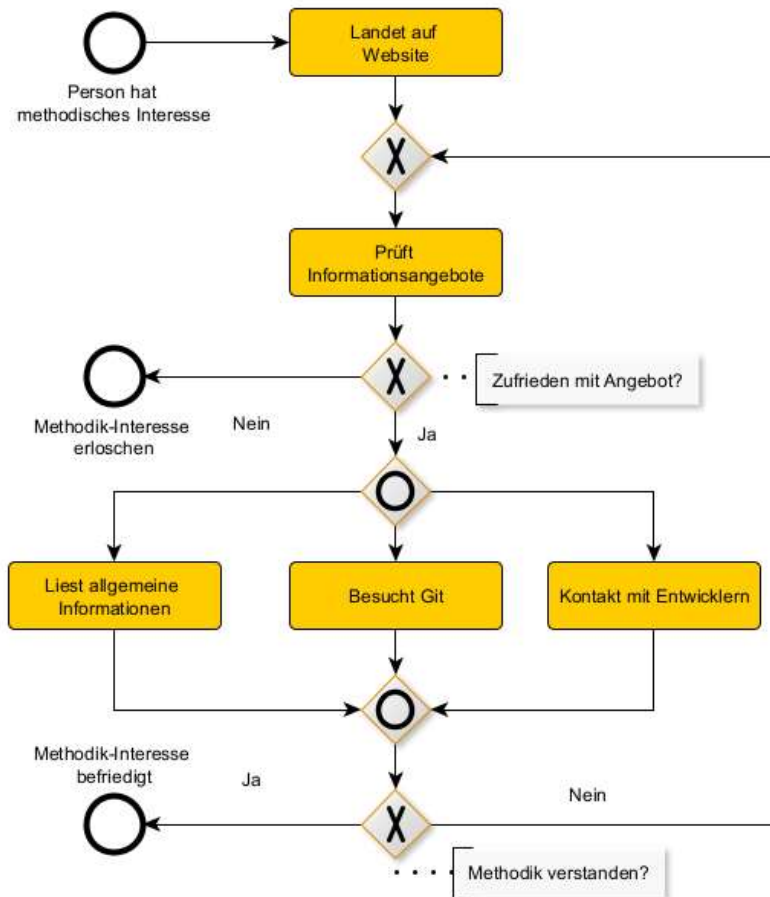


Abbildung 5: Fachprozess zur Befriedigung von methodischem Interesse.

FACHPROZESS DATENSUCHE

Der Prozess für die Suche nach geeigneten Datensätzen (s. Abbildung 6) startet in dem öffentlich zugänglichen Bereich der [Website](#) des Datenraums. Datensuchende können hier zum Datenkatalog navigieren und diesen durchsuchen, wahlweise explorativ oder durch Nutzung der Suchfunktion mit spezifischen Suchparametern. Wurde ein interessanter Datensatz identifiziert, besteht die Möglichkeit bei Bedarf nach tiefergehenden Informationen die Verknüpfungshistorie oder den entsprechenden Eintrag des Metadatenkatalogs einzusehen. Diese Informationen sind nur für registrierte Nutzende zugänglich. Für nicht registrierte Nutzende muss daher vor dem Login ein neues Konto erstellt werden, während bereits registrierte Nutzende sich direkt einloggen können. Nach dem Login stehen die Informationen aus der Verknüpfungshistorie und dem Metadatenkatalog bereit. Wurde kein geeigneter Datensatz gefunden, kann entweder eine erneute Suche gestartet oder die Suche abgebrochen werden. Zudem wird die Suche auch beendet, wenn Nutzende keine

Registrierung wünschen. Wurde ein geeigneter Datensatz (mit oder ohne Zuhilfenahme tiefergehender Informationen) gefunden, so kann entweder nach weiteren Datensätzen gesucht oder die Suche beendet werden.

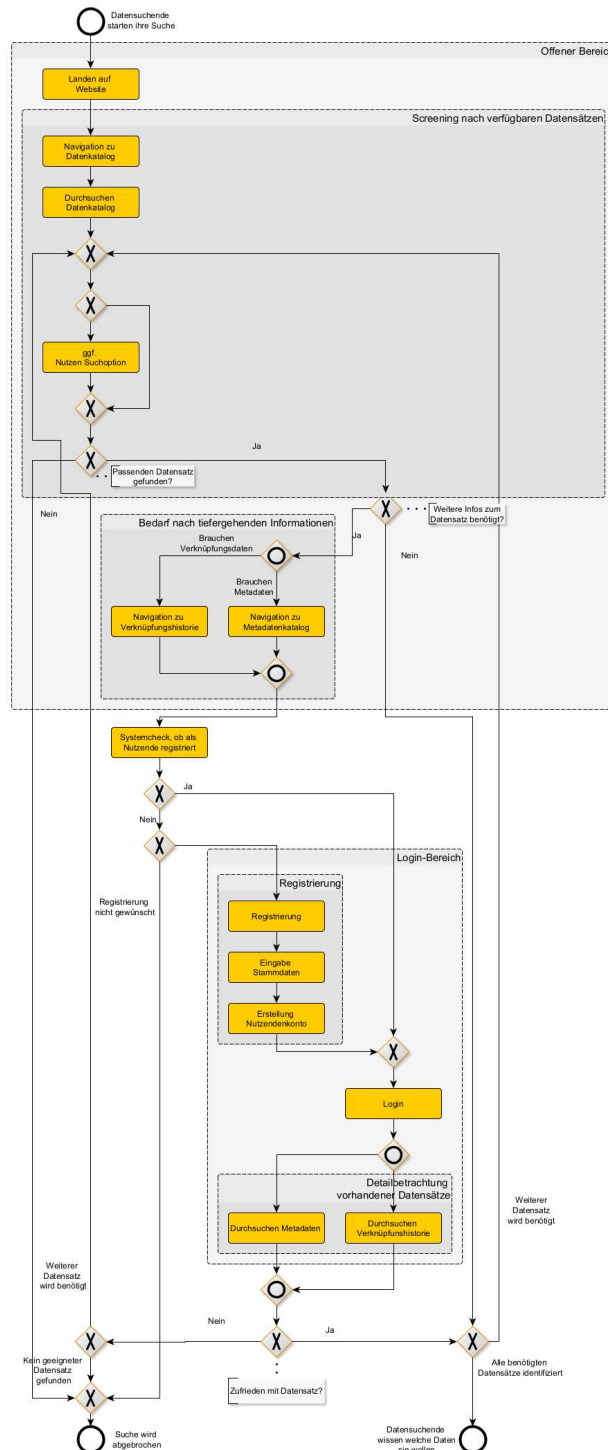


Abbildung 6: Fachprozess für die Suche nach geeigneten Datensätzen.

FACHPROZESS DATENABRUF

Der Prozess für den Datenabruf durch Datennehmende startet, nachdem die relevanten Daten ausgewählt wurden. Eine Top-Level-Ansicht des Prozesses ist in Abbildung 7 zu sehen. Zunächst müssen Datennehmende das Use & Access Verfahren durchlaufen. Nachdem dies erfolgreich abgeschlossen wurde, muss für einen Datenabruf ein Nutzungsvertrag ausgehandelt werden. Einigen sich die Parteien und unterzeichnen den Vertrag, müssen alle Datengebenden ihre Freigabe erteilen. Nachdem alle Freigaben erteilt wurden, können die Datennehmenden die Transaktion auslösen, welche in [EuroDaT](#) dann die angeforderten Daten verknüpft. Nach erfolgreicher Verknüpfung können die Datennehmenden auf den resultierenden Datensatz zugreifen. Folgend werden die einzelnen Teilprozesse erläutert.

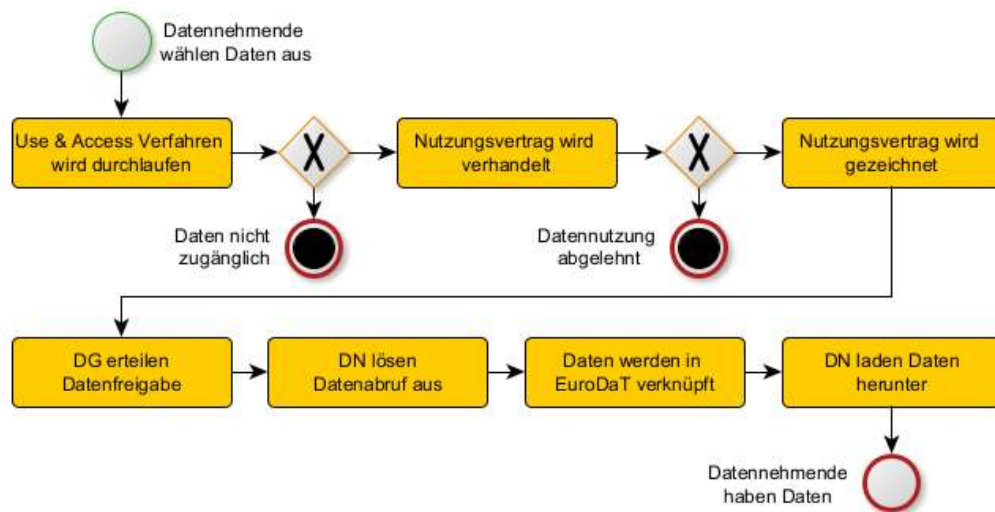


Abbildung 7: Top-Level-Ansicht des Prozesses zum Datenabruf aus Sicht der Datennehmenden. DG steht für Datengebende, DN steht für Datennehmende.

Subprozess Use & Access Management

Abbildung 8 zeigt eine Detailansicht des Subprozesses für das Use & Access Verfahren. Der Subprozess beginnt zunächst mit der Umleitung auf die Use & Access Unterseite der [Website](#) inklusive der zugehörigen Verfahren der jeweiligen Datengebenden, für welche alle relevanten Formulare ausgefüllt und eingereicht werden müssen. Nach Einreichung der Daten wird der Antragsstatus getrackt und kann stets von den Datennehmenden eingesehen werden. Die Datengebenden werden über den Antrag informiert und treffen eine Entscheidung, wobei bei fehlenden Angaben noch Informationen nachgereicht werden können. Nach der Entscheidung werden die Datennehmenden über die Entscheidung informiert und bei erfolgter Genehmigung folgt der nächste Prozessschritt.

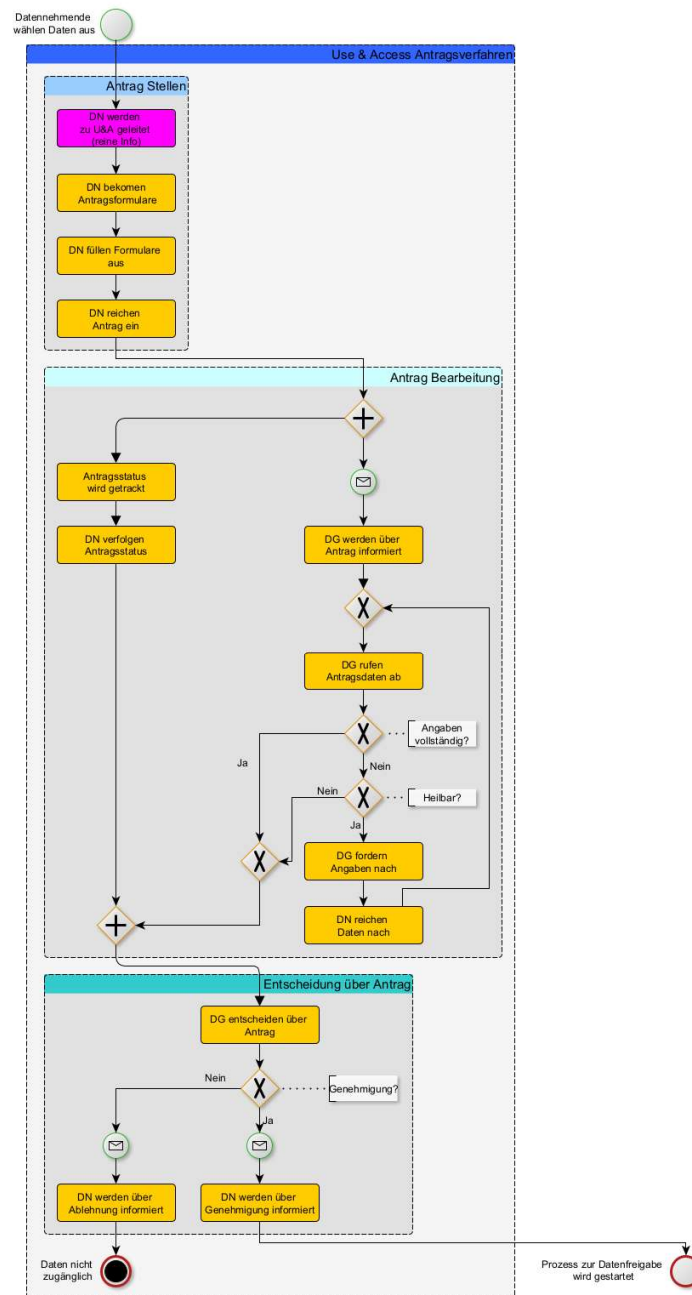


Abbildung 8: Detailansicht des Subprozesses für Use & Access Management zum Datenabruf aus Sicht der Datennehmenden. Pinke Kästchen stellen im MVP umgesetzte Teile dar. DG steht für Datengebende, DN steht für Datennehmende.

Subprozess Datenfreigabe

Abbildung 9 zeigt den Subprozess für die Freigabe von Daten. Nach dem erfolgreichen Durchlaufen des Use & Access Verfahrens muss zunächst ein Nutzungsvertrag geschlossen werden. Dieser wird von den Datengebenden aufgesetzt und den

Datennehmenden zur Verfügung gestellt. Lehnen diese den Vertrag ab, ist der gesamte Prozess beendet. Zusätzlich besteht aber auch die Möglichkeit, in Nachverhandlungen mit den Datengebenden zu treten. Wird der Vertrag nun akzeptiert, wird er archiviert und die Datengebenden müssen ihre Freigabe im Interface eintragen. Der Freigabestatus wird dabei getrackt und ist jederzeit durch die Datennehmenden einzusehen. Das zentrale Backend schickt eine Nachricht an das Backend des lokalen Clients der Datengebenden, was eine Zuordnung zu der gestellten Anfrage möglich macht. Nun können die Datengebenden die angeforderten Daten in den Client laden. Nachdem alle Datengebenden ihre Freigabe erteilt und die Daten in den Client geladen haben, wird der Datennehmende informiert und der nächste Prozessschritt wird eingeleitet.

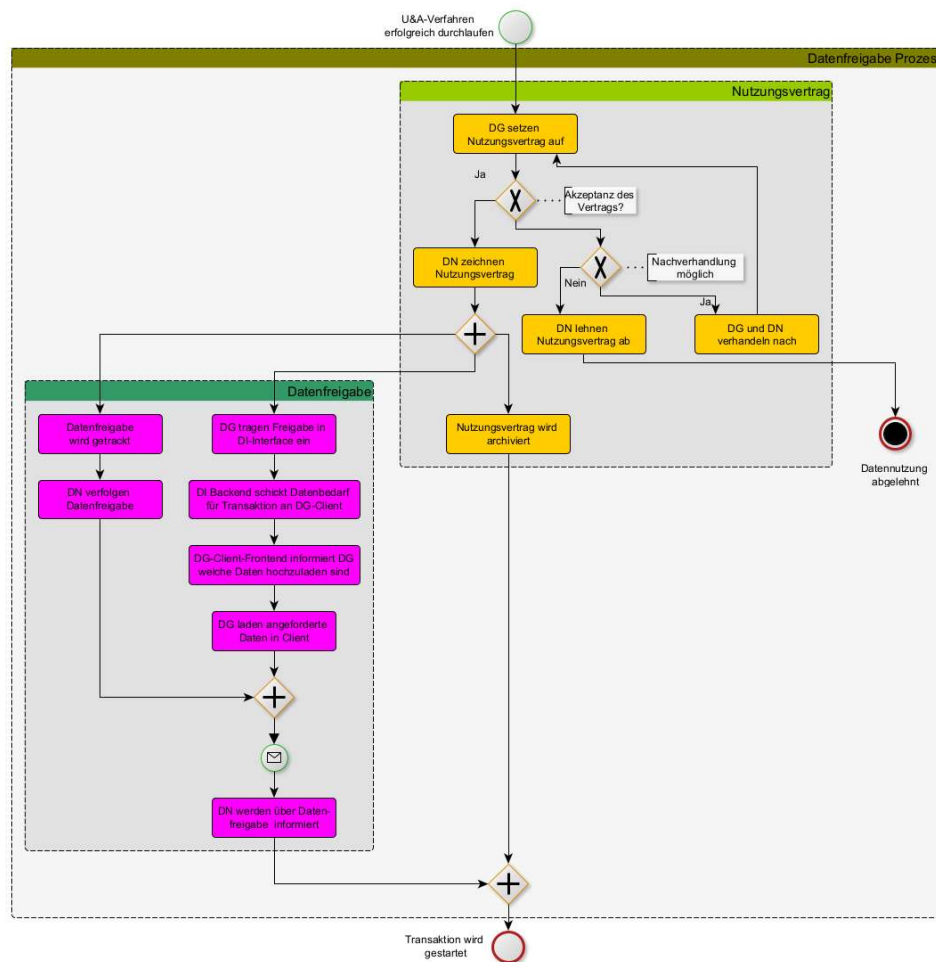


Abbildung 9: Detailansicht des Subprozesses für die Datenfreigabe zum Datenabruf aus Sicht der Datennehmenden. Pinke Kästchen stellen im MVP umgesetzte Teile dar. DG steht für Datengebende, DN steht für Datennehmende.

Subprozess EuroDaT-Transaktion

Abbildung 10 zeigt den Subprozess der nun beginnenden **EuroDaT**-Transaktion.

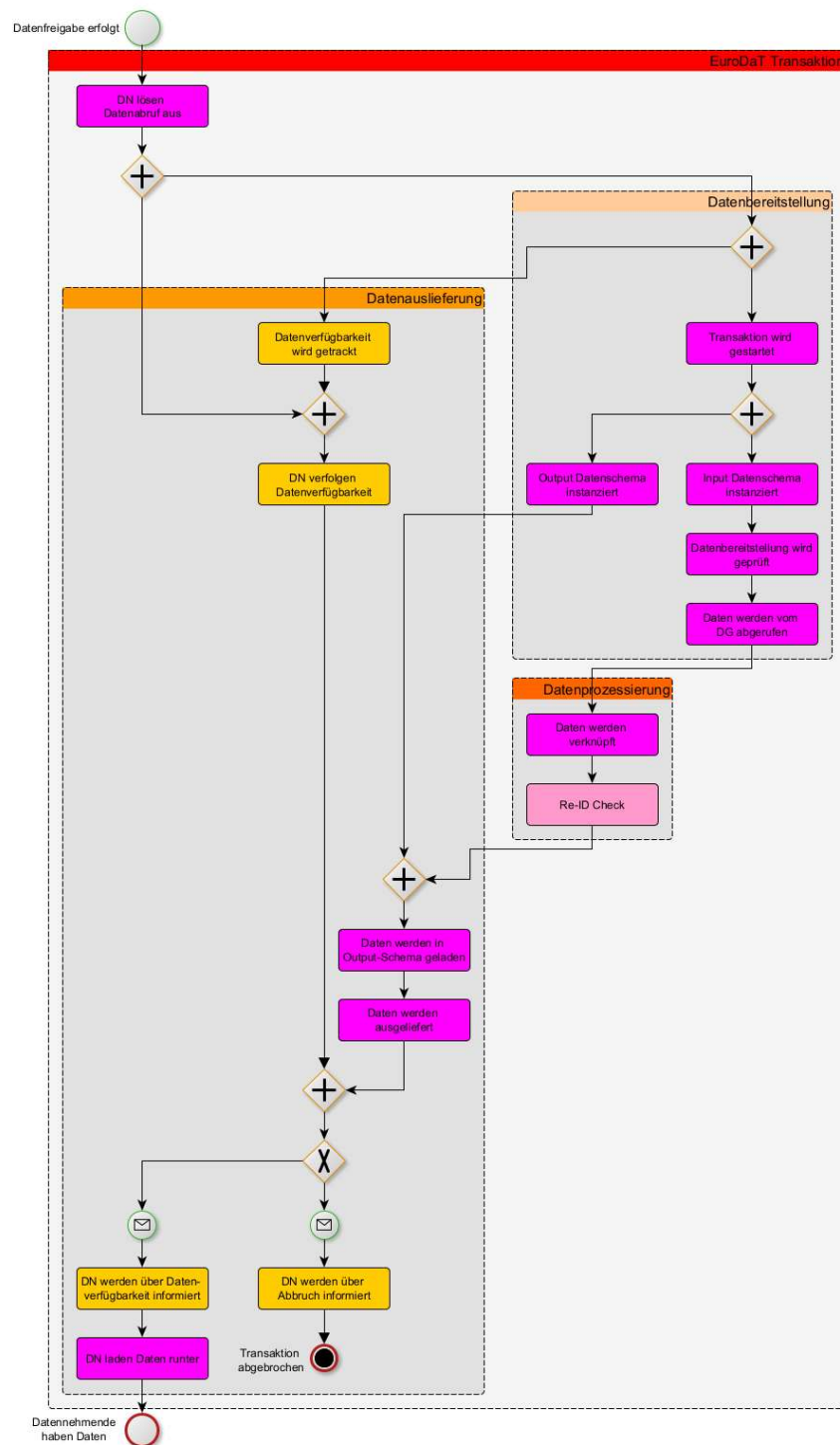


Abbildung 10: Detailansicht des Subprozesses für die EuroDaT-Transaktion zum Datenabruf aus Sicht der Datennehmenden. Pinke Kästchen: im MVP umgesetzte Teile. DN: Datennehmende.

Nach erfolgter Freigabe lösen die Datennehmenden den Datenabruf aus. Dabei wird die Datenverfügbarkeit konstant getrackt und von den Datennehmenden verfolgt. Sollte sich während des Prozesses etwas an der Freigabe oder der Verfügbarkeit von Daten ändern, wird der Prozess abgebrochen und die Datennehmenden werden informiert. Im Regelfall wird jedoch eine [EuroDaT](#)-Transaktion gestartet. Diese provisioniert eine temporäre Datenbank mit passenden Schemata zum Daten-Input für Datengebende und zum Daten-Output für Datennehmende, s. Kapitel 05.04. Nachdem alle benötigten Daten der Datengebenden geladen wurden, werden diese in [EuroDaT](#) verknüpft und einem Re-ID Check unterworfen (s. Kapitel 05.03). Nach erfolgreicher Verknüpfung und Check werden die Daten in das betreffende Schema für das Daten-Output geschrieben und stehen den Datennehmern daraufhin zur Verfügung.

04. FORSCHUNGSOBJEKT

04.01 AUSRICHTUNG DES DATENMODELLS AN DEN ANFORDERUNGEN DER FORSCHUNG

Für die Schaffung eines aktiven Datenökosystems in der Post-COVID-Forschung ist es entscheidend, dass den Akteuren ein unmittelbarer Nutzen geboten wird. Dabei müssen individuelle sowie strukturelle Interessen und Hürden berücksichtigt werden. In der ersten Stufe der Challenge haben wir hierzu Perspektiven und Herausforderungen der verschiedenen Akteure identifiziert und zentrale Anforderungen an das Datenökosystem formuliert. Inhaltlich leitet uns dabei die hochaktuelle interdisziplinäre Forschungsfrage zur Erwerbsfähigkeit, s. Kapitel 02.01.

ANFORDERUNGEN DER FORSCHUNG

Ein zentraler Aspekt der Stufe 2 war die Fortführung der Kommunikation mit verschiedenen Stakeholdern der Forschung, um Anforderungen weiterhin aktiv zu identifizieren. Dabei legten wir besonderen Wert darauf, unser Zielbild umfassend vorzustellen und einer kritischen Prüfung zu unterziehen. Aus unseren Gesprächen können wir durchweg positives Feedback berichten: Das Zielbild wurde in mehreren Diskussionsrunden intensiv hinterfragt, und es stellte sich stets heraus, dass die Forschung den Datenraum sowie die von uns vorgesehenen Funktionalitäten als dringend notwendig erachtet und ein großes Interesse daran hat. Zwei zentrale Aspekte kamen hierbei in unseren Gesprächen vermehrt auf, Datenschutz und Datensouveränität.

Datenschutz

Das Thema Datenschutz spielt eine zentrale Rolle in der medizinischen Forschung und viele Forschende berichten von regelmäßigen Einschränkungen durch die Datenschutz-

Grundverordnung (DSGVO). Häufig können Ideen nicht wie geplant umgesetzt werden, da datenschutzrechtliche Bedenken im Wege stehen. Die Aussicht, Teile dieser Problematik mithilfe technischer Lösungen, wie dem Einsatz eines neu entwickelten Modells des Datentreuhänders (s. Kapitel 06), zu entschärfen (s. Kapitel 04.04), wird durchweg positiv aufgenommen. Der Konsens unter unseren Ansprechpartnern war eindeutig: Datenschutz muss oberste Priorität haben, und es ist von großem Interesse, dass Daten ausschließlich sicher und gemäß der DSGVO verarbeitet werden. Gleichzeitig besteht der Wunsch, die Forschung bestmöglich zu unterstützen, weshalb der neue Ansatz einer technischen Lösung auf großes Interesse stößt.

Datensouveränität

Während der laufenden Entwicklung des MVP und der entsprechenden Integration der Datenhaltenden verstärkte sich unser Eindruck, dass die Datensouveränität stets bei den datenhaltenden Stellen bleiben muss. Eine klar definierte Gatekeeper-Funktion sorgt dafür, dass Daten nicht leichtfertig geteilt werden. Ein Modell, in dem Daten zentral gehalten oder verwaltet werden, haben wir deshalb ausgeschlossen. Die transaktionsbasierte Architektur der [EuroDaT](#)-App stellt sicher, dass die Daten jederzeit im Besitz der Datenhaltenden bleiben und nur auf transaktionsbasierter Basis abgerufen werden. Zudem gilt: Wenn die Auswertung der Daten innerhalb von [EuroDaT](#) erfolgt, verlassen die Rohdaten die Infrastrukturen der Datenhaltenden bis auf die typischerweise kurze Transaktionszeit nicht, sodass Dritte keinen Einblick oder Zugriff auf diese Daten erlangen können.

Indem wir so sicherstellen, dass die Kontrolle über die Daten bei den Datenhaltenden bleibt, erhöhen wir nicht nur das Vertrauen in das Datenökosystem, sondern fördern auch eine aktive Teilnahme der Nutzenden. Die transparente Handhabung von Daten, die durch den Einsatz innovativer Technologien und klar definierter Zugriffskontrollen ermöglicht wird, stärkt das Gefühl der Sicherheit und des Schutzes.

ANFORDERUNGEN AN DEN DATENRAUM

Aus den Anforderungen der Forschung resultieren eine Vielzahl an Anforderungen an den Datenraum, welche wir ausführlich im Bericht der Stufe 1 erörtert haben. Im Rahmen der Stufe 2 haben wir uns auf die Umsetzung der wichtigsten dieser Anforderungen fokussiert. Dazu zählt in erster Linie die Flexibilität des Datenraums, die es ermöglicht, unterschiedliche Forschungsansätze wirkungsvoll zu unterstützen. Eine einfache Zugänglichkeit und Sichtbarkeit der Daten spielt ebenfalls eine zentrale Rolle, da sie die Datennutzung fördert und sowohl Forschende als auch Datenhaltende zur Teilnahme anregt. Ein weiterer wesentlicher Aspekt ist die nahtlose Integration in bestehende Prozesse bei den Datengebenden und -nehmenden, sowie die Sicherstellung

der Kompatibilität mit bestehenden Standards, um Doppelstrukturen zu vermeiden und die Effizienz der Entwicklung zu steigern. Folgend beschreiben wir kurz die Umsetzungen der verschiedenen Anforderungen an den Datenraum.

Die gewählte Struktur, in der der Datenraum eine Vermittlerrolle einnimmt (s. Kapitel 03.01), ermöglicht eine flexible Erweiterung des Modells z.B. mit Daten aus nicht-medizinischen Bereichen, wodurch eine Offenheit für vielseitige Forschungsansätze sichergestellt wird.

Die Gestaltung der [Website](#) wurde mit dem Ziel optimiert, eine intuitive Nutzeroberfläche zu bieten, die die Sichtbarkeit der verfügbaren Daten verbessert. Dadurch werden sowohl Datennutzende als auch Datenhaltende animiert, aktiv an der Plattform teilzunehmen. Eine einfache Auffindbarkeit der Daten sowie die geplante Integration eines zentralen, übersichtlichen Use & Access Managements senken die Hemmschwelle für Nutzende Daten zu beantragen erheblich und fördern somit eine verstärkte Nutzung der Datenressourcen.

Angesichts der hohen Komplexität und des umfangreichen Charakters des Gesamtprojekts ist es von größter Bedeutung, wo immer möglich, auf bereits bestehende Strukturen zurückzugreifen. Nur so können wir effektiv Doppelstrukturen vermeiden und Ressourcen optimal nutzen. Ein Beispiel dafür ist unser Ziel, einen bereits bestehenden Metadatenkatalog in die Metadatenbereitstellung zu integrieren. Aktuell stehen wir hierfür mit der [NFDI4Health](#) in Kontakt. Diese Vorgehensweise verspricht nicht nur eine effizientere Nutzung bestehender Ressourcen, sondern auch eine schnellere und einfachere Zugänglichkeit zu relevanten Daten für die Forschung.

04.02 PROZESSE ZUR EINBINDUNG VON STAKEHOLDERN

Langfristig angelegte Prozesse zur Einbindung von Datenhaltenden und -nutzenden stellen die Zukunftsfähigkeit des Datenökosystems sicher und sollen beliebigen Akteuren aus der (Post-COVID) Forschung den Zugriff ermöglichen. Die hierfür nötigen Prozessschritte wurden im Kapitel 03.02 des ersten Berichtes ausführlich beschrieben. Primäres Ziel ist es, die Anforderungen und Interessen der verschiedenen Stakeholdergruppen (Datenhaltenden, Datennutzenden und Intermediäre) auch während der voranschreitenden Implementierung angemessen zu berücksichtigen. Folgend werden die in der Stufe 2 unternommenen, sowie für die nächsten Stufe vorgesehenen, Bestrebungen zur Einbindung von Stakeholdern beschrieben.

NUTZERZENTRIERTE FORMATE DER EINBINDUNG

Als primäre Stakeholder sehen wir die drei Gruppen Datenhaltende, Datennutzende und Intermediäre. Eine ausführliche Übersicht inklusive beispielhafter Auswahl an

Akteuren findet sich im Bericht zum Abschluss der ersten Stufe. Die dort beschriebene Auswahl ist weiterhin aktuell. Im Rahmen der fortschreitenden Entwicklung haben wir uns auf mehrere Arten der aktiven Einbindung der Beteiligten fokussiert, welche folgend beschrieben werden.

Einzelgespräche

Insbesondere während der Entwicklungsphase des Datenmodells messen wir dem direkten Austausch mit den verschiedenen Stakeholdern hohen Wert und Priorität bei. Zur Annäherung und zur Gewinnung als Partner für das Projekt eignen sich Einzelgespräche besonders gut. Diese haben wir aufbauend auf unsere Gespräche aus der ersten Stufe mit den verschiedenen Parteien (bspw. [NAKO](#), [Rentenversicherung](#), [NFDI4Health](#)) weitergeführt. Hierbei war ein wichtiger Punkt das Aufbauen von Vertrauen in unsere Lösung, u. a. durch das detaillierte Vorstellen des Zielbildes und der Möglichkeit der Akteure sowohl Fragen zur konkreten Implementierung des MVPs zu stellen als auch der Möglichkeit der Diskussion spezifischer Anforderungen und zukünftiger Funktionalitäten des Datenraums.

Im Rahmen des persönlichen Austausches haben wir durchweg positives Feedback erhalten; alle aus der Stufe 1 bekannten Stakeholder unterstützen unser Konzept und möchten zum Erfolg des Projektes beitragen. Neu hinzugekommen ist die [Treuhandstelle Greifswald](#), wobei hierauf in Kapitel 05.01 genauer eingegangen wird.

Review/Fokus-Gruppe

Als neues Instrument zur Einbindung von Stakeholdern haben wir die Planung einer Review- bzw. Fokusgruppe aufgenommen. Ziel dieser Gruppe ist es, relevante Stakeholder frühzeitig in die Entwicklung des Datenraums für die Post-COVID Forschung zu integrieren. Durch die gezielte Einbindung können Bedürfnisse und Erwartungen bereits in den frühen Phasen des Projekts themenspezifisch identifiziert und berücksichtigt werden. Zudem bietet der Austausch im Rahmen der Fokusgruppe die Möglichkeit, wertvolles Feedback zeitnah einzuholen, wodurch Anpassungen in der Entwicklung schnell und effizient vorgenommen werden können. Die Ausrichtung dieses Feedback-Prozesses an den spezifischen Interessen der Stakeholder trägt maßgeblich dazu bei, dass der Datenraum den Anforderungen der Nutzer gerecht wird und somit eine hohe Akzeptanz findet.

Wir sehen im Rahmen der Fokus-Gruppen regelmäßige Meetings vor, in der Regel einmal monatlich. Entscheidend ist die sorgfältige Vorstrukturierung der Termine, um den Stakeholdern die personell passende Auswahl für den Termin zu erleichtern. Hierdurch ist es möglich, während der Termine detaillierte Fragestellungen zu besprechen, gleichzeitig aber die Ressourcen der Stakeholder zu schonen und zielgenau

einzusetzen. Unser Vorgehen stößt schon jetzt auf große Resonanz; so liegen uns bereits von mehreren Akteuren positive Rückmeldung zur Teilnahme an Fokusgruppen vor.

Workshops

Als konkreten ersten Schritt zur Erhöhung der Sichtbarkeit des Datenökosystems und damit auch der Ansprache von neuen Stakeholdern, haben wir bereits die Planungen für einen Workshop zur Vorstellung und Ausgestaltung des Datenökosystems und der angeschlossenen [EuroDaT](#)-Applikation aufgenommen. Ziel dieses Workshops ist es, das Datenökosystem, das innovative Konzept des Datentreuhänders [EuroDaT](#) und die wegweisende Automatisierung der DSGVO-Erfüllung vorzustellen. Dieser Workshop ist für das erste Quartal 2025 geplant. In diesem Rahmen wollen wir Stakeholder einladen, um sie nicht nur über das Datenökosystem zu informieren, sondern auch um weitere Kooperationen und Synergien zu erkennen und zu fördern. Perspektivisch sollen regelmäßig Workshops dieser Art abgehalten werden.

THEMENSCHWERPUNKTE DER EINBINDUNG

Die Einbindung von Stakeholdern soll nicht nur integraler Bestandteil der Entwicklungsphase sein, sondern darüber hinaus auch während des Betriebs aktiv gefordert und gefördert werden. Im Folgenden gehen wir auf diejenigen Themengebiete ein, bei denen die enge Abstimmung mit den Stakeholdern den größten Mehrwert für den Projekterfolg leistet.

Use & Access

Wie bereits in Kapitel 04.01 beschrieben, gilt es stets, die Datensouveränität der Datenhaltenden zu wahren. Hiermit einher geht unter anderem, dass die Entscheidungen über Nutzungsanträge unverändert von den Datenhaltenden durchgeführt werden. Je nach Institution bestehen hierzu bereits ausführliche Begutachtungs- und Herausgabeprozesse.

Im Rahmen des Zielbildes ist vorgesehen, das Use & Access Verfahren für Datennutzende zu vereinfachen. Hierbei geht es weniger darum, die dahinterliegenden Prozesse zu verändern als darum, eine bessere Übersichtlichkeit mit einer zentralen Anlaufstelle für Datennutzende zu generieren. Hierzu stehen wir in engem Austausch mit der [NFDI4Health](#), welche dieses Vorhaben aktuell bereits für eine Institution exemplarisch umsetzt. Um Synergien bestmöglich zu nutzen, möchten wir auf diesem Wissen aufbauen und idealerweise eine Integration in unser System anstreben.

Die datenhaltenden Stakeholder sind zentral für das Gelingen des Projektes: Sie müssen weiterhin ihr Use & Access Verfahren durchführen und Bereitschaft für die Zusammenarbeit und den eventuell vorübergehend gegebenen Mehraufwand durch

beispielsweise das zusätzliche Eintragen der Freigabe in das System des Datenmodells mitgehen. Daher ist es ausgesprochen wichtig, die Datenhaltenden bereits während der Entwicklung für das Projekt zu begeistern, sodass eine intrinsische Motivation für diese Arbeiten besteht. Hierzu nutzen wir wie beschrieben insbesondere den persönlichen Austausch mit den Stakeholdern.

Vertragsbaukasten

Eine Datennutzung geht immer einher mit der Erstellung eines Datennutzungsvertrages. Diese Ausarbeitung ist teils mühsam und bedarf mehrere Iterationen der verschiedenen Parteien und Rechtsabteilungen. Im Zielbild soll dieses Verfahren vereinfacht werden, indem u. a. auf einen Vertragsbaukasten zurückgegriffen werden kann, s. Kapitel 03.01.

Analog zum Use & Access Prozess behalten die datenhaltenden Stakeholder im MVP jedoch eine wichtige Rolle bei der Vertragsschließung und der Nutzung des Vertragsbaukastens. Es müssen weiterhin Nutzerverträge geschlossen werden und eine Integration der bestehenden Mechanismen in das MVP stattfinden, weshalb die beteiligten Stakeholder bereits in die Entwicklung integriert sind und alle nötigen Prozesse für das MVP in enger und kontinuierlicher Abstimmung ausgearbeitet werden.

Technische Anbindung

Im Rahmen des MVP ist eine erste technische Anbindung von Stakeholdern an den Datenraum vorgesehen. Insbesondere bei der Entwicklung benötigen wir hier die aktive Mitarbeit der Datenhaltenden: die Client-Applikation des Datenraums muss installiert und APIs eingerichtet werden. Die Clients bieten den Stakeholdern hierbei eine benutzerfreundliche Oberfläche, um auf die benötigten Informationen zuzugreifen und diese zu verwalten. Um die Bereitschaft der Stakeholder zur Mitarbeit zu fördern, setzen wir auf den persönlichen Austausch und die aktive Einbindung in die Entwicklung.

04.03 KONZEPT ZUR VERÖFFENTLICHUNG DES DATENMODELLS

Im Folgenden stellen wir vor, wie wir das in Stufe 1 entworfene Veröffentlichungskonzept unseres Post-COVID Datenmodells in Stufe 2 ausdefiniert und mit konkreten Entwicklungsschritten untermauert haben. Hierbei haben wir die folgenden drei grundlegenden Prinzipien, aus denen sich die Anforderungen an die Veröffentlichung ableiten lassen, stets im Blick gehabt:

- einen offenen Zugang zu den Daten zu gewährleisten,
- einen offenen Zugang zu der Modellierung des Datenmodells zu gewährleisten und
- eine vergemeinschaftete kontinuierliche Weiterentwicklung zu gewährleisten.

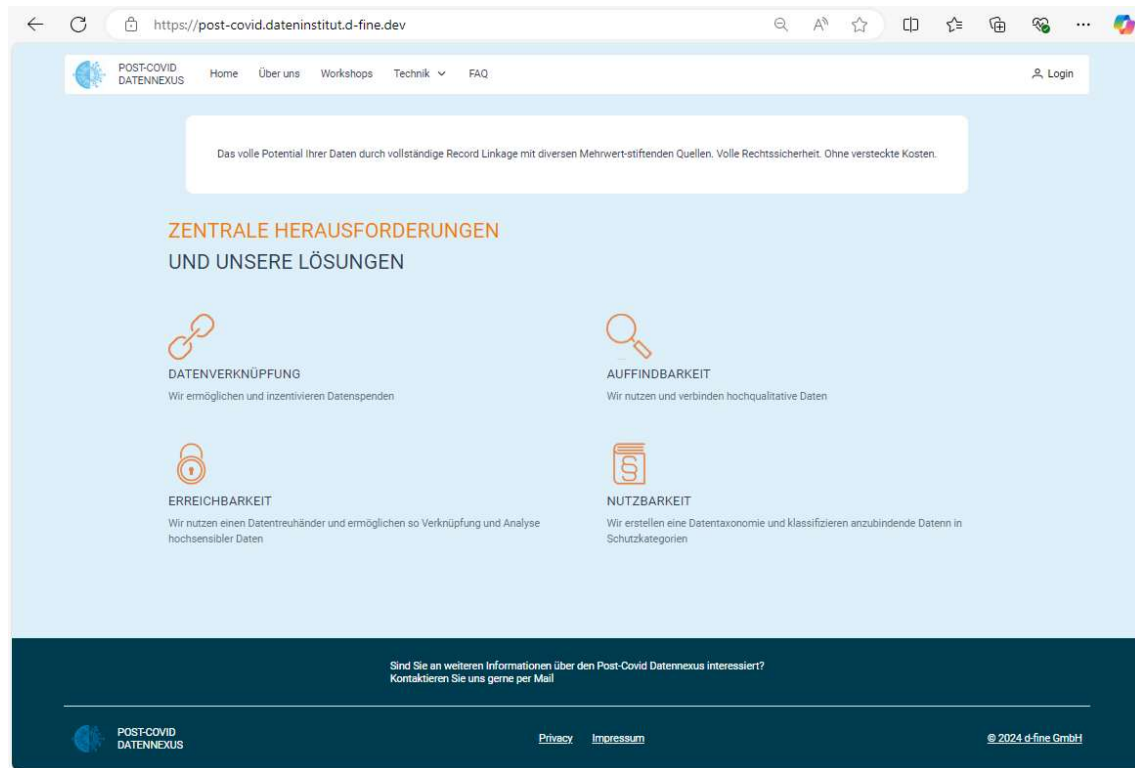


Abbildung 11: [Entwicklungsinstanz](#) des Webfrontends.

Da ein transaktionsbasierter Datentreuhänder wie z.B. [EuroDaT](#), das technische Kernstück unseres Datenökosystems bilden wird, ist die Veröffentlichung der für seine Nutzung notwendigen Datenapplikation von zentraler Bedeutung. Eine solche Datenapplikation entwickeln wir aktuell und werden sie über verschiedene Kanäle veröffentlichen, einschließlich prominenter Plätze wie der Homepages des hier vorgestellten Datenökosystems, EuroDaTs und ggf. des Dateninstituts, sowie durch Vorstellung in geeigneten Fachkreisen über Fachvorträge und andere Formate. Über die Applikation werden dann der Zugang zu den Primärdaten ermöglicht und Datentransaktionen durchgeführt, s. Kapitel 03.01, 04.04 und 06.01.

Um die oben genannten zentralen Zwecke der Öffentlichkeits-einbindung zu erreichen, ist es unserer Meinung nach notwendig, die Codebasis unserer Entwicklung öffentlich zugänglich zu machen, Informationen über die verfügbaren Datensätze und ihre Verknüpfungsmöglichkeiten niederschwellig bereitzustellen und auch Feedback-Kanäle für die Nutzenden und Entwickelnden des Datenmodells einzurichten. Für diese Zwecke entwickeln wir als zentralen Anlaufpunkt ein Webfrontend, das den Zugang zu den Metadaten, zur Codebasis und zum Datenmodell für die Nutzenden unkompliziert und übersichtlich ermöglicht. Es dient als Helikopterperspektive, die den Nutzenden eine einfache und offene Möglichkeit bietet, die verfügbaren Metadaten und die logische Struktur des Datenmodells schnell zu überblicken und sich über die Nutzung und das Onboarding in das Datenökosystem zu informieren.

Darüber hinaus ermöglicht es eine direkte Kommunikation und Beteiligung der Nutzenden an der Weiterentwicklung des

Datenmodells, indem es beispielsweise eine Diskussionsplattform stellt und Feedback ermöglicht. Daher fügt sich ein Webfrontend als zentrale Schnittstelle nahtlos in das Veröffentlichungskonzept ein und ermöglicht es, die drei herausgestellten Aspekte des Konzepts – offener Zugang, Transparenz und partizipative Weiterentwicklung – effektiv und effizient umzusetzen.

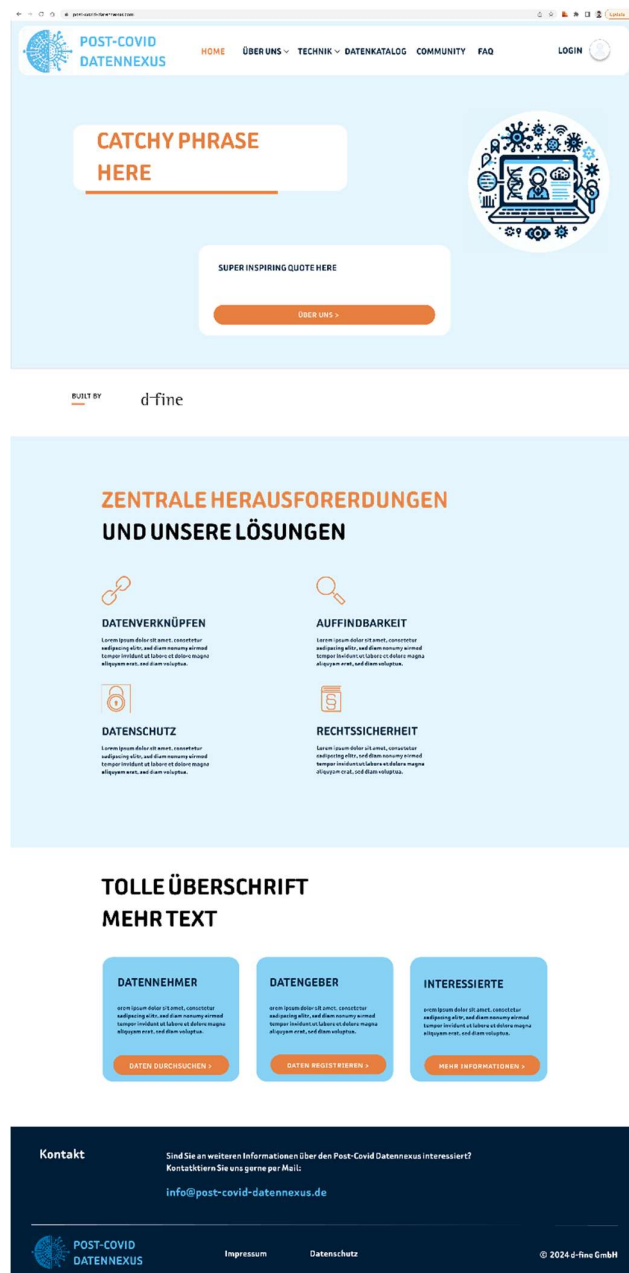


Abbildung 12: Entwurf der Landingpage des von uns entwickelten Webfrontends.

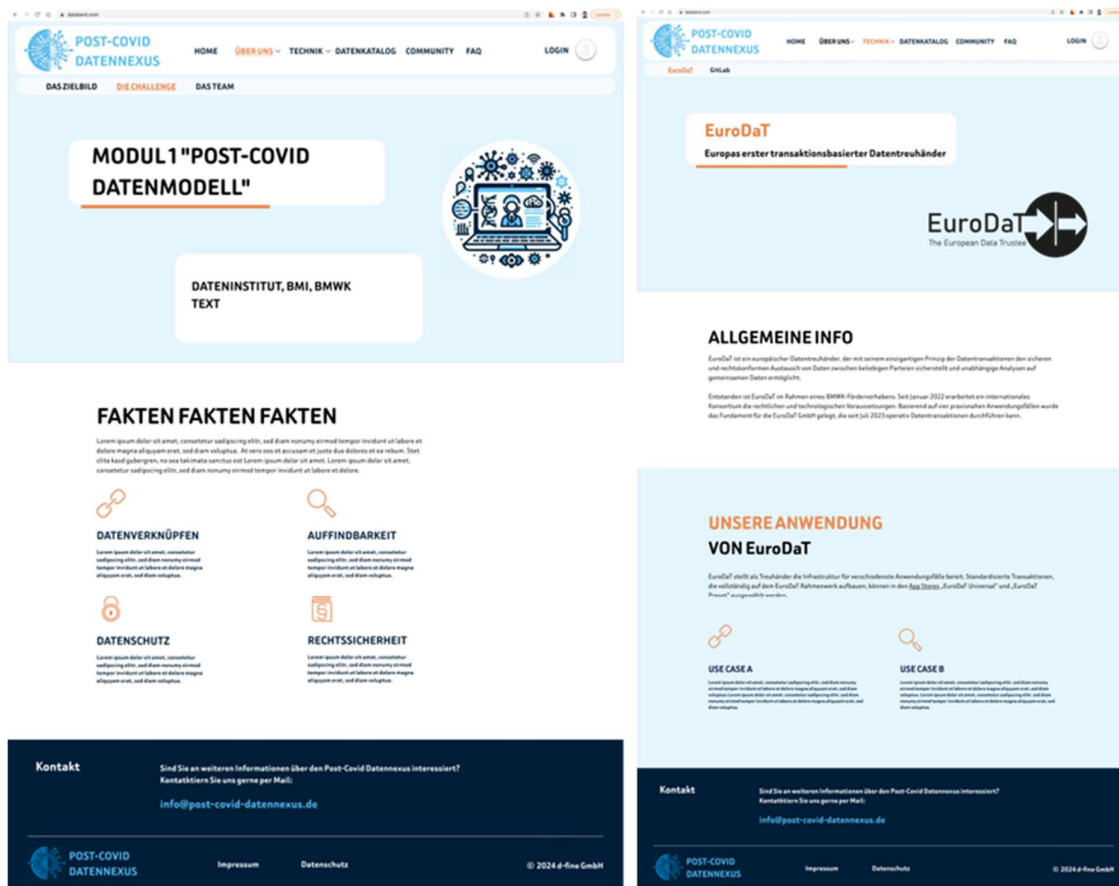


Abbildung 13: Entwürfe der Informationsseiten "Über uns" (links) und „Technik“ (rechts) des von uns entwickelten Webfrontends.

Für die Implementierung des Webfrontends haben wir eine technische Entwicklungsinstanz aufgesetzt, die bereits live unter der Adresse <https://post-covid.dateninstitut.d-fine.dev/> gehostet wird, s. auch Abbildung 11. Das unter dieser Adresse gehostete Webfrontend entwickeln wir gemäß den Anforderungen des Datenökosystems und im Rahmen des von uns gewählten agilen Projektansatzes kontinuierlich weiter. Um hierfür ein klares Zielbild vor Augen zu haben, haben wir bereits Wireframes der anvisierten finalen Entwicklungsstufe entworfen, die als strukturierte visuelle Orientierung dienen und uns dabei unterstützen, das Endprodukt effektiv zu gestalten und zu optimieren.

Das Design unseres Zielbilds folgt dabei einem bewährten Muster moderner Webapplikationen mit einer horizontalen Steuerleiste am oberen Rand der [Website](#). Diese beinhaltet Rubriken wie "Home", die von jeder Unterseite auf die Landing Page zurückführt, s. Abbildung 12. Unter der Überschrift "Über uns" stellen wir Informationen über das Konsortium und die Hintergründe der Challenge bereit. Und unter der Rubrik "Technik" erklären wir die genutzte IT-Infrastruktur und insbesondere die Anbindung

an den Datentreuhänder [EuroDaT](#). Des Weiteren gibt es den "Datenkatalog", der Nutzenden einen Überblick über die im Datenraum abrufbaren Primärdaten bietet, und die "Community", die eine Plattform für Engagement und Interaktion bietet. Zudem haben wir eine "FAQ"-Seite erstellt, um häufige Fragen zu beantworten (s. Abbildung 12). Die Steuerleiste samt spezifischen Rubriken ermöglicht es Nutzenden schnell und effizient zu navigieren und auf die relevanten Informationen und Funktionen zuzugreifen, wodurch ein intuitives und nutzerfreundliches Erlebnis gewährleistet wird.

Das Webfrontend bietet außerdem einen nahtlosen Übergang zum vorgesehenen Login-Bereich. In diesem können Nutzende nach einer Registrierung im Datenökosystem ihre Freigaben für angeforderte Daten nachverfolgen, mit Datengebenden in Kontakt treten und Datentransaktionen über den Treuhänder auslösen.

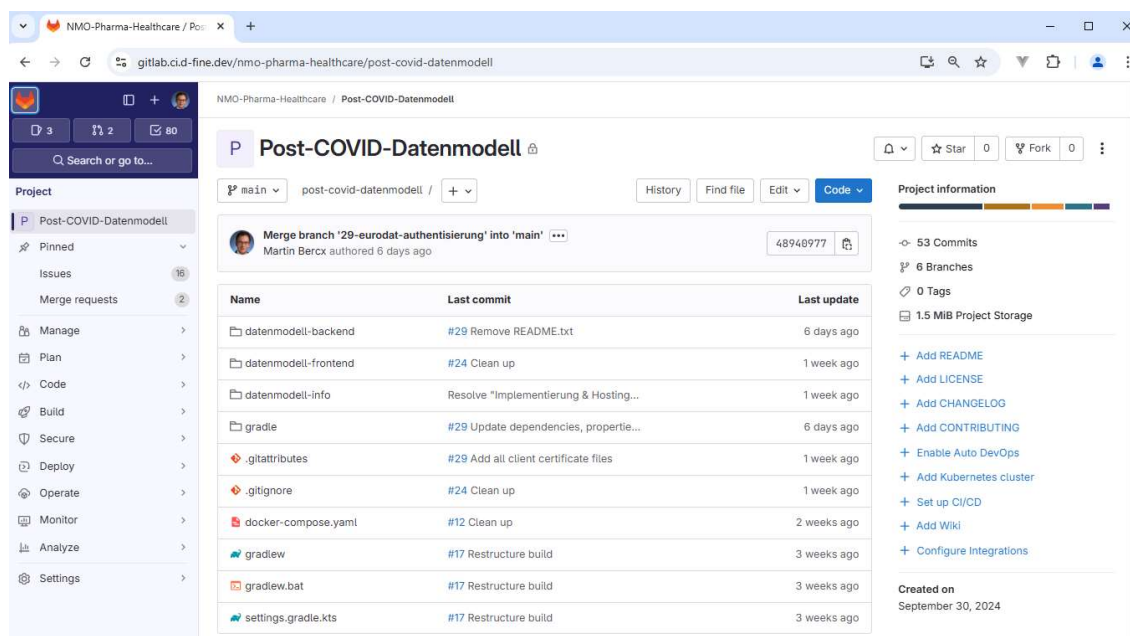


Abbildung 14: Startseite des von uns eingerichteten [GitLab-Repositories](#).

Zusätzlich zum oben beschriebenen Webfrontend, haben wir unter <https://gitlab.ci.d-fine.dev/nmo-pharma-healthcare/post-covid-datenmodell> auch ein GitLab-Repository eingerichtet, s. Abbildung 14. Dieses Repository ist in der aktuellen Entwicklungsphase noch intern, wird aber von uns nach dem Git-Industriestandard gepflegt damit es nahtlos auf das öffentliche [GitLab](#) umziehen und dort veröffentlicht werden kann, sobald es die entsprechende Reife erreicht hat. So wird das Repository als interaktive Plattform für Entwicklungen und Kollaborationen rund um das Datenökosystem dienen. Das [GitLab-Repository](#) dient als Plattform für die Modellentwicklung, indem es als Code-Repository genutzt wird, in dem die Codebasis und das Datenmodell gespeichert und versioniert werden können. So werden auch die zur Anbindung an das Datenökosystem

notwendigen [EuroDaT](#)-Tennants als Open Source-Code über das [EuroDaT-GitLab Repository](#) frei zur Verfügung gestellt. Dadurch ermöglicht das Repository eine transparente und nachvollziehbare Modellentwicklung für alle Beitragenden und steigert so die Qualität des Datenmodells.

Das [GitLab-Projekt](#) dient aber nicht nur als Code-Repository, sondern stellt auch einen effektiven Partizipationskanal für die Nutzenden und Entwickelnden dar. Durch die Integration von GitLab kann jeder Beitragende Änderungen vorschlagen und am Entwicklungsprozess des Datenökosystems partizipieren. Zusätzlich ermöglicht es einen strukturierten Feedback-Kanal und fördert Diskussionen, wodurch eine Gemeinschaft von beteiligten Akteuren entsteht, die zur Verbesserung und Weiterentwicklung des Datenökosystems beiträgt.

Zusätzlich zu diesen Angeboten ist allerdings die Erzeugung von zusätzlicher Sichtbarkeit für das Datenökosystem entscheidend, um eine breite Akzeptanz und Nutzung seiner Fähigkeiten zu gewährleisten, neue Partner zu erreichen und die Idee eines gemeinsamen, offenen Datenraums in der Post-COVID-Forschung voranzutreiben. Um eine größere Bekanntheit und Präsenz in den relevanten Communities zu erreichen, wollen wir über die eben vorgestellten statischen Veröffentlichungskanäle hinaus noch weitere, informelle Informationsformate etablieren. Hierbei denken wir z. B. daran institutionelle Nachrichtenkanäle, wie den regelmäßigen Newsletter des [NAPKON](#) Verbunds, zu nutzen, um über Fortschritte, Entwicklungen und Möglichkeiten zur Beteiligung zu informieren. Darüber hinaus bieten Veranstaltungen wie die des [BDI](#) oder Netzwerke wie [DaTNet](#) ausgezeichnete Plattformen, um das Datenökosystem zu präsentieren, Diskussionen zu fördern und potenzielle Stakeholder zu erreichen.

Als konkreten ersten Schritt zur Erhöhung der Sichtbarkeit des Datenökosystems haben wir bereits die Planungen für einen Workshop zur Vorstellung und Ausgestaltung des Datenökosystems und der angeschlossenen [EuroDaT](#)-Applikation aufgenommen, s. Kapitel 04.02.

Abschließend möchten wir noch unsere Überlegungen zur Benennung des Datenökosystems erwähnen. Wir sind uns bewusst, dass ein im Markt platzierter Datenservice einen eingängigen Namen braucht. Da wir uns aktuell in der Entwicklungsphase befinden, haben wir noch keinen finalen Namen ausgewählt, sondern lediglich mögliche Namensvarianten entwickelt, die die Kernaspekte des Datenökosystems beschreiben. Neben dem aktuell verwendeten Arbeitstitel **Datennexus** sind ebenso sprechende Akronyme in der Überlegung, mit denen eine emotionale Botschaft transportiert werden kann wie z.B. **SHARE** („Secure data Handling, Analysis and Record LinkagE), **MERLIN** (Meta data and Record LINKage) oder **LINC**

(Linkage, Integration and Collaboration). Aus diesen und weiteren Optionen können wir schnell einen Namen auswählen, sobald es notwendig wird.

04.04 INNOVATIONSGRAD DES ANSATZES

Wie im letzten Stufenbericht ausführlich dargelegt, besteht der zentrale Innovationspunkt unseres Ansatzes in der rechtssicheren Zusammenführung und Analyse von Daten, insbesondere sensibler personenbezogener medizinischer Informationen. In der aktuellen Stufe haben wir diesen innovativen Ansatz ausgebaut und erweitert, so dass die integrierten technischen und rechtlichen Schutzmechanismen Risiken für Datengebende und Datennutzende minimieren und somit ein vertrauensvolles Umfeld für den Austausch sensibler Informationen schaffen.

TRANSAKTIONSBASIERTER DATENTREUHÄNDER

Die zentrale technische und operative Innovation des so entstandenen Lösungsansatzes liegt vor allem in der Verwendung eines transaktionsbasierten Treuhänders, was eine bedeutende Neuerung in der Art und Weise darstellt, wie sensible Daten gehandhabt und analysiert werden können. Unser innovativer Ansatz bietet eine neue Dimension an Datensicherheit, indem er die Verarbeitung in einer abgesicherten Umgebung automatisiert und ohne menschliche Eingriffsmöglichkeit ausführt.

Aus einer rechtlichen Perspektive basiert das Modell der transaktionsbasierten Datentreuhand – anders als herkömmliche Vorstellungen von Datentreuhand – nicht auf einer langfristigen und umfangreichen Vorratshaltung von großen Datenbeständen („Silo-Modell“). Stattdessen werden über einen zwischengeschalteten besonderen Akteur und mit Hilfe einer spezifisch dafür entworfenen Verarbeitungsinfrastruktur im Einzelfall auf konkrete Anfragen hin – daher transaktionsbasiert – Einmal-Datensätze für bestimmte Datenanalysen generiert und anschließend umgehend wieder gelöscht. Dezentral gespeicherte Datenbestände unterschiedlicher datengebender Stellen werden damit ausschließlich zum Zweck einer einmaligen und vorab klar definierten, algorithmischen Auswertung virtuell zusammengeführt und verarbeitet. Die Datentreuhandstruktur dient insoweit dazu, die jeweils analyserelevanten Datenbestände in einer hierfür geschaffenen digitalen Umgebung gesichert zusammenzuführen, für die jeweils anvisierte Datenanalyse bereitzustellen und anschließend zu löschen. Aus diesen Kerncharakteristika lassen sich in datenschutzrechtlicher Perspektive zwei wesentliche Konsequenzen ableiten:

Erstens kann man die besondere Transaktionsgebundenheit als eine technisch-institutionelle Absicherung des Zweckbindungsgrundsatzes rekonstruieren (der ansonsten häufig praktisch leerzulaufen droht). Denn weil die durch die Treuhand

hergestellte Verbindung von Datengebenden, -nutzenden und -dienstleistenden über anlassbezogen zusammengeführte Einmaldatensätze erfolgt, ermöglicht sie eine zielgenaue und zweckorientierte Datenbereitstellung. Das widerstreitet besonders missbrauchsanfälligen explorativen Datennutzungen und erschwert damit einerseits bestimmte Big-Data-Anwendungen, die gerade auf der Unvorhersehbarkeit zukünftiger Mustererkennung aufbauen. Es reduziert andererseits aber die aus dem offenen Informationspotential großer Datenbestände resultierenden Risiken und unterbindet ein Datensammeln als verselbständigten Zweck.

Zweitens spricht einiges dafür, die besondere technische Ausgestaltung der Datenanalyseprozesse als eine im Vergleich zu alternativen Mechanismen vorteilhafte Schutzgewähr einzuordnen. Denn der Datentreuhänder verwendet eine „gekapselte“, also gegen jede Form der Zweckentfremdung abgeschirmte und uneinsehbare Auswertung des Einmaldatensatzes. Dabei wird die Datenanalyse von Anfang an auf bestimmte Anfragen hin ausgerichtet und erfolgt (nur) nach diesen Vorgaben. Deshalb ist die Zweckbindung nicht nur in zeitlicher Hinsicht (durch die transaktionsbasierte Bildung des ephemeren Datensatzes), sondern durch die Abschirmung auch in sachlich-gegenständlicher Hinsicht abgesichert. Denn die Datenauswertung kann nur diejenige Art der Information generieren, für die das Analyseinstrument technisch konfiguriert und von den Beteiligten konkret beauftragt und freigegeben wurde. Anstatt also am Datenmaterial anzusetzen und zunächst umfassend generierte Daten mit Personenbezug erst in einem zweiten Schritt einer Anonymisierung zuzuführen (etwa durch nachträgliche Verschlüsselung, Schwärzung o.ä.), erfolgt durch diese besondere technische Ausgestaltung eine früher ansetzende Prozessanonymisierung. Das wirkt Reidentifizierungen entgegen, weil und soweit von Anfang an auf Datensätze verzichtet wird, die spezifische Identitätsmerkmale beinhalten. Gleichzeitig wird durch den prozessbezogenen Ausschluss entsprechend sensibler Informationsgehalte eine weiterreichende Analytik sonstiger, unter zahlreichen Gesichtspunkten relevanter und unverdächtigter Informationen ermöglicht. Ohne Effektivitätsverluste mit Blick auf den Datenschutz in Kauf nehmen zu müssen, werden so gleichermaßen die Effizienz der Datenanalyse erhöht wie die diesbezüglichen Anwendungsoptionen erweitert.

In der Fachliteratur wird daher auch vom Übergang von der [Verarbeitung personenbezogener Daten zur personenbezogenen Datenverarbeitung](#) gesprochen. Ein transaktionsbasierter Treuhänder wie [EuroDaT](#) ermöglicht somit eine risikominimierte, effiziente und rechtlich konforme Datenverknüpfung und -analyse. Es handelt sich hierbei nicht nur um eine konzeptionelle Innovation, sondern um eine technische Lösung, die bereits im Modell integriert ist. Dieser Ansatz ist in seiner Form einzigartig und es existiert derzeit kein vergleichbares Angebot auf dem Markt.

Unser Ansatz berücksichtigt durch die aktive Einbindung von Rechtsexpertise außerdem bereits in der Aufbauphase aktuelle und aufkommende regulatorische Standards wie den Europäischen Gesundheitsdatenraum ([EHDS](#)), das Gesundheitsdatennutzungsgesetz ([GDNG](#)), den [AI Act](#) und den [Data Act](#).

DATENTAXONOMIE

Um den bis hierher beschriebenen fundamentalen Innovationsansatz zu unterfüttern, haben wir die von uns konzipierte Datentaxonomie weiterentwickelt und ausgebaut.

Die Datentaxonomie leistet weiterhin einen besonderen Mehrwert für die datenschutzrechtliche Klassifizierungen verfügbarer Datensätze, um sicherzustellen, dass die Verarbeitung der Daten den strengen Datenschutzbestimmungen entspricht. Als Neuentwicklung haben wir in der aktuellen Stufe damit begonnen, die Datentaxonomie für konkrete Datentypen und -gebende auszufüllen, um den Mehrwert für Forschende schnell zu heben. Darüber hinaus haben wir ein schnell zugängliches Ampelschema implementiert, um die erwartbaren Herausforderungen beim Zugriff auf Daten verschiedener Kategorien von verschiedenen Datengebern zu klassifizieren.

Den neuesten Stand der Datentaxonomie werden wir auf dem Abschlussevent der Stufe 2 am 14. Januar 2025 präsentieren.

VERKNÜPFUNG

Ein weiterer entscheidender Innovationsfaktor ist die Schaffung einer nachnutzbaren Verknüpfung bisher getrennter wissenschaftlicher Disziplinen. Durch unsere Arbeit in der aktuellen Challenge-Stufe haben wir Dummy-Datensätze erhalten und Perspektiven auf Zugang zu den dahinterstehenden Realdaten aufgebaut. Dies wird es uns ermöglichen Daten, aus den Sozialwissenschaften, wie jenen der [Rentenversicherung](#), mit medizinischen Forschungsdaten des [NAPKON](#) und der [NAKO](#) sowie persönlichen Gesundheitsdaten von Wearable-Anbietern wie [Garmin](#) zu integrieren und gemeinsam nutzbar zu machen. Diese interdisziplinäre Vernetzung ermöglicht es, umfassendere und vielschichtige Analysen durchzuführen, die in der Vergangenheit aufgrund der Trennung dieser Disziplinen nicht möglich waren. Insbesondere die von einzelnen Nutzenden gehaltenen Wearable-Daten wollen wir nahtlos über das neuartige Konzept der Datenspende integrieren, s. auch Kapitel 05.01, gemäß dessen Nutzende ihre Daten für gesellschaftlich relevante Zwecke freiwillig zur Verfügung stellen.

Ein weiterer Baustein der Zukunftssicherheit dieses Modells ist die kontinuierliche technologische Weiterentwicklung. Das Datenökosystem bleibt flexibel und reagiert proaktiv auf neue Technologien und Trends in der Post-COVID-Forschung. Die Bereitstellung des Datenraums und -modells unter einer Open Source-Lizenz ermöglicht

eine dynamische Anpassung an künftige Entwicklungen und fördert die Zusammenarbeit innerhalb der Forschungscommunity. Diese Agilität und vorausschauende Orientierung sind essenziell für die Stabilität und Kontinuität des Modells und machen es zu einer zuverlässigen und zukunftsicheren Plattform für die Post-COVID-Forschung.

In Summe manifestiert sich der hohe Innovationsgrad unseres Ansatzes in der Verbindung von datenschutzgerechter Datenintegration, interdisziplinärer Zusammenarbeit und zukunftsweisender Technologie, was ihn als wegweisend für die wissenschaftliche Analyse und Erkenntnisgewinnung im Gesundheitssektor positioniert.

05. DATENMODELL

05.01 VORLIEGENDE DATENSÄTZE

Im Rahmen dieser Stufe der Challenge haben wir uns auf eine Auswahl an Datensätzen fokussiert. Hierzu greifen wir auf unsere Vorarbeiten aus der ersten Stufe zurück, in der es galt, zunächst relevante Datensätze zu identifizieren. Hierbei haben wir sowohl eine Long-List an potenziellen Datensätzen erstellt sowie eine konkrete Auswahl an fünf verschiedenen Datensätzen, auf welche wir uns in dieser Stufe fokussieren wollten. Unsere Auswahl sowie unser Ergebnisstand sind in folgender Tabelle beschrieben.

Datensatz	Verfügbarkeit	Dummy-Daten	Metadaten
NAPKON	Use & Access	Liegen vor	Liegen vor
NAKO	Use & Access	Werden für uns erstellt	Liegen vor
Abrechnungsdaten	Use & Access	Public Use File wird erstellt	Liegen vor
Rentenversicherungsdaten	Datenabhängig	Liegen vor	Liegen teils vor
Wearables	Datenabhängig	Liegen vor	Liegen teils vor

Tabelle 2: Übersicht der für die MVP-Umsetzung priorisierten Datensätze.

Die Auswahl der von uns fokussierten Datensätze blieb im Vergleich zur Stufe 1 unverändert. Mit unserer Auswahl decken wir ein möglichst breites Feld ab, streben einen großen Mehrwert für die Forschung an, und zielen darauf ab, möglichst viele Hürden bei der Verknüpfung und Integration der Daten aufzudecken. Durch den Aufbau dieser breiten Datenbasis wird es künftig einfach möglich sein, beliebige Datensätze in den Datenraum zu integrieren. Aufgrund der Tatsache, dass sich das neue [FDZ-](#)

[Gesundheit](#) noch im Aufbau befindet bzw. aktuell die Bereitstellung der Abrechnungsdaten vorbereitet, war eine aktive Zusammenarbeit in dieser Stufe aus Gründen der Ressourcenschonung nicht vorgesehen. Sobald die Daten verfügbar sind, stehen wir auch hier im Kontakt. Unsere Energien haben wir in dieser Stufe auf die weiteren vier Datensätze und deren Bereitsteller gebündelt und können folgende positiven Entwicklungen berichten.

Ziel für diese Stufe war es, vor dem Hintergrund unserer interdisziplinären Forschungsfrage zur Erwerbsfähigkeit (s. Kapitel 02.01) einen möglichst guten Überblick über die Daten und insbesondere deren Metadaten zu erhalten. Durch das Sammeln von Dummy-Daten wollen wir ermöglichen, die Machbarkeit der Integration und Verknüpfung der Daten exemplarisch aufzuzeigen, sodass das Datenmodell später nahtlos auf die Verknüpfung von Realdaten umgestellt werden kann. Auch die Verfügbarkeit der jeweiligen Daten wurde von uns mitbetrachtet. Hierzu wurde insbesondere der direkte Austausch mit den Datenhaltenden gesucht. Die Resonanz war hierbei durchgängig positiv. Es besteht offenkundig ein großes Interesse an der Teilnahme an dem Projekt sowie ein hohes Maß an Offenheit und Kooperation bezüglich des Datenteilens. Im Folgenden gehen wir detailliert auf die von uns zunächst ausgewählten Datenquellen ein:

NAPKON

Durch die Säule Medizin unseres Teams besteht hier ein ausgesprochen guter Austausch. Im Rahmen der Stufe 2 wurden sowohl Dummy-Daten als auch umfangreiche Metadaten bereitgestellt. Wir bereiten außerdem gerade einen Antrag auf die Nutzung der Realdaten vor, s. Kapitel 04.01, um mittels eines exemplarischen Use-Cases die Machbarkeit unseres Ansatzes zu demonstrieren. Auch hierbei werden wir tatkräftig unterstützt. Ebenso erhoffen wir für die prototypische Integration eines [EuroDaT](#)-Clients bei Datengebenden die volle Unterstützung des [NAPKON](#)-Teams.

NAKO

Mit der [NAKO](#) gab es mehrere Termine, bei denen wir sowohl unser Zielbild genauer vorgestellt haben als auch detaillierte Fragen zu den Daten, den technischen Rahmenbedingungen sowie dem Use & Access Verfahren beantworten konnten. Das Team steht einer Zusammenarbeit sehr offen gegenüber und unterstützt uns aktuell in der Antragsstellung zur Nutzung der Daten, der on-premise Integration eines [EuroDaT](#)-Clients, sowie der Bereitstellung von Dummy-Daten. Metadaten liegen uns bereits umfangreich vor.

DATEN DER RENTENVERSICHERUNG

Um den Fokus nicht nur auf rein medizinische Daten zu legen, sondern das Spektrum möglichst breit und bereichsübergreifend aufzustellen, stehen wir weiterhin im Austausch mit der deutschen [Rentenversicherung](#). Auch hier besteht eine große Bereitschaft zur Teilnahme am Projekt. Es wurde uns beispielsweise zugesagt, dass auch die Integration eines Hashalgorithmus in die Datenbasis vorstellbar ist. Aufgrund der Datenstruktur und der sehr hohen Datenschutzbedarfe seitens der [Rentenversicherung](#), ist hier eine Integration in das System und das Erstellen von Hashes nötig, um Record Linkage möglich zu machen. Teil unserer Arbeiten war des Weiteren die Exploration der Metadaten, wobei wir hier noch Potential sehen, den umfangreichen Datenschatz der [Rentenversicherung](#) zu erfassen. Wir planen deshalb die Erstellung eines Metadatensatzes mit den wichtigsten Kennzahlen. Dummy-Daten zu exemplarischen Themenbereichen liegen uns vor. Die Beantragung von Echt-Daten sehen wir für die Stufe 3 vor.

DATEN VON WEARABLES

Exemplarisch für die Bereitstellung von Wearable Daten befinden wir uns im Austausch mit einem der größten Smart- und Multisportuhren Herstellern weltweit, [Garmin](#). Garmin unterstützt mit seinen Wearables einerseits verschiedenste Studien, andererseits werden insbesondere die Smart- und Sport-Watches des Herstellers von zahlreichen Privatpersonen dauerhaft genutzt. Die hierbei erfassten Daten eignen sich ausgesprochen gut für die exemplarische Ausarbeitung eines Datenspende-Konzeptes.

Ziel ist es, Nutzenden von Garmin Uhren zu ermöglichen ihre Daten für die Forschung zur Verfügung zu stellen. Hierbei sollen Daten direkt geteilt werden, wie des beispielsweise mit anderen Apps wie [Strava](#) möglich ist. Dort werden Daten zu Aktivitäten automatisch via API übertragen und den Aktivitäten des Profils hinzugefügt. Um die Forschung bestmöglich zu unterstützen, wird sich das Zielbild nicht nur auf Daten zu Aktivitäten beschränken, sondern zusätzlich allgemeine Daten mit abrufen.

Zur Umsetzung dieses Vorhabens gibt es zwei wichtige Schritte, welche eine nähere Betrachtung benötigen. Der erste Schritt ist die technische Umsetzung. Hierzu stehen wir im Austausch mit Garmin und wollen in Stufe 3 mittels des [Garmin Connect Developer Program](#) eine API erstellen, durch die Nutzende ihre Daten direkt spenden können. Hierzu befinden wir uns bereits im Antragsverfahren. Da die Daten in speziellen Formaten übermittelt werden, stellen wir zusätzlich sicher, dass die nötigen Schritte zur Datenverarbeitung bekannt und umsetzbar sind. Hierfür liegen uns Beispieldaten vor.

DATENSPENDE

Ein weiterer elementarer Schritt zur Nutzung von Wearable Daten ist die Ausarbeitung der Datenspende und die nähere Betrachtung der Datenschutzbestimmungen hierzu. Wie bereits aus unserem ersten Bericht hervorgeht, fehlt für die Datenspende eine klare rechtliche Basis. Ziel dieser Stufe war es, aus rechtlicher Perspektive eine Übersicht der aktuellen Lage zu erarbeiten.

Der zuvor eher nicht gängige Begriff der Datenspende wird erstmals in der Stellungnahme des Deutschen Ethikrates zu [Big Data und Gesundheit](#) (S. 44, 184 f.) prominent verwendet. Allerdings wird schon hier deutlich, dass es sich nicht um einen bereits klar umgrenzten Rechtsbegriff handelt, sondern um einen heuristischen Begriff, der rechtspolitischer Absicherung bedarf. An dieser Ausgangslage hat sich ungeachtet nachfolgender Annäherungs- und Konkretisierungsbemühungen, s. z. B. [„Datenspende“ – Bedarf für die Forschung, ethische Bewertung, rechtliche, informationstechnologische und organisatorische Rahmenbedingungen](#), wenig geändert. Auch das aktuelle Gesundheitsdatennutzungsgesetz enthält keine konkreten Rechtsgrundlagen für eine (gesundheitsdatenbezogene) Datenspende. Allerdings gibt es Anzeichen dafür, dass zumindest im Gesundheitsbereich selbst eine – in sich indes widersprüchliche – nicht einwilligungsbasierte Datenspende auf breiten Rückhalt in der Bevölkerung gestützt werden könnte (vgl. [Link](#)).

Versucht man sich vor diesem Hintergrund an einer terminologisch-konzeptionellen Arbeitsdefinition, ist charakteristisch für Datenspenden das Fehlen unmittelbarer Anreize für die Datenhingabe/Zugriffsgewährung, s. [Sovereignty and Data Sharing](#). Jedoch schließen (ggf. indirekte oder immaterielle) Gegenleistungen (etwa im Kontext gruppennütziger Forschung) den Spendencharakter nicht pauschal aus. Bedenken bestehen hingegen bereits auf begrifflicher Ebene gegenüber Ansätzen, auch nicht-einwilligungsbasierte Datenzugangseröffnungen als „Spende“ zu qualifizieren. Eine unfreiwillige „Spende“ dürfte den Rahmen sinnvoller Begriffsbildung überdehnen.

Eine solche Datenspende im Sinne einer auf freiverantwortlich getroffenen und informierten Entscheidungen der Betroffenen beruhenden, zielgerichteten Abgabe eigener Daten wird in der DSGVO nicht explizit erwähnt. Sie kann aber unter bestimmten Voraussetzungen als hiernach zulässige Option angesehen werden. Voraussetzung hierfür ist, dass die Grundkonzeption der Datenspende nicht den Grundanliegen der DSGVO (harmonisierte Regulierung von Vertraulichkeitsschutz und freiem Datenverkehr) widerspricht. In der Tat finden sich in der DSGVO und ihr folgenden Vorschriften Erleichterungen zugunsten privilegierter Verarbeitungszwecke, die eine praktische Umsetzung der Datenspende stützen können:

- umfassendere Zweckbestimmung bei Einwilligung (EG 33) und eigener Rechtsgrund bei sensiblen Daten (Art. 9 II lit. j) DSGVO i.V.m. § 27 I BDSG);
- Einschränkung von Betroffenenrechten aus Kap. III der DSGVO & ausgedehnte Speicherbegrenzung (Art. 5 I lit. e) HS 2 DSGVO;
- Privilegierung sekundärer „Weiterverarbeitung“ (Art. 5 I lit. b) HS 2 DSGVO; EG 50 S. 2).

Deutlich wird jedoch auch, dass das Fehlen einer expliziten Regelung der Datenspende gewisse rechtliche Unsicherheiten mit sich bringt und die praktische Umsetzung erschwert. Beispielsweise dürfte den Anforderungen der DSGVO umso einfacher entsprochen werden, je zielgerichteter eine Spende erfolgt. Gerade eine strenge Zweckverpflichtung ist indes bei Datenspendemodellen schwierig, bei denen die Initiative vollständig von den betroffenen Personen ausgeht und bewusst offen erfolgen soll. Der (Daten-)Spende (etwa im Gegensatz zum „Geschenk“) liegt zudem schon konzeptionell eine gewisse Distanz gegenüber nachfolgenden Verwendungsoptionen inne. Darüber hinaus sind indes die besonderen Garantien für die Rechte und Freiheiten der Betroffenen bis hin zu einer vorrangigen Anonymisierung zu berücksichtigen; dies dürfte der pauschalen Bevorzugung von Gemeinwohlbelangen entgegenstehen. Schließlich ist – namentlich im Kontext der besonders sensiblen Sozialdaten (§ 67b I 1, 4 SGB X) näher zu eruieren, wie sich das Modell einer möglichen (ggf. rechtspolitisch abgesicherten) Datenspende zu dem vorhandenen Privileg der erleichterten sekundären „Weiterverarbeitung“ verhält.

In der nächsten Stufe der Challenge sehen wir vor, das Datenspende-Modell weiter auszuarbeiten und ein konkretes Konzept für Wearabledaten zu erarbeiten.

TREUHANDSTELLEN

Oftmals ist beim Umgang mit sensiblen Daten, wie z.B. in der klinischen Forschung, das Hinzunehmen einer Treuhandstelle zur Überwachung und Koordination der Einverständniserklärungen sowie der personenidentifizierenden Daten der Teilnehmenden unumgänglich. Da wir für das Verknüpfen von Daten explizit die personenidentifizierenden Daten benötigen, um ein Record Linkage durchführen zu können, war die Kontaktaufnahme zu Treuhandstellen ein Teil dieser Stufe. Die [Treuhandstelle der Universität Greifswald](#) ist hierbei für viele deutsche medizinische Forschungsstudien zuständig, so auch für [NAPKON](#) und [NAKO](#), weshalb wir im Austausch mit dieser Stelle stehen. Es war uns im persönlichen Austausch möglich, die Treuhandstelle für unser Projekt zu begeistern.

Erwähnenswert ist hier, dass eine Treuhandstelle wie die THS Greifswald ein anderes Betriebsmodell hat als der Datentreuhänder [EuroDaT](#) und somit beide in dem von uns

anvisierten Datenökosystem eine relevante Rolle spielen. Während die Treuhandstelle sensible Daten geliefert bekommt und statisch verwaltet, bietet [EuroDaT](#) eine transaktionsbasierte Datenverarbeitung an. Dadurch kann [EuroDaT](#) nahtlos Drittdaten anbinden, die Datensouveränität der Datengebenden automatisch wahren und standardisierte Analysen auf den verknüpften Daten durchführen. All dies sind wichtige Funktionen unseres Zielbilds, die eine klassische Treuhandstelle nicht abbilden kann, s. auch Kapitel 04.04.

Aktuell stehen wir in Gesprächen zu den nächsten Schritten, sowie die Möglichkeit des Verwendens eines Hashfilters, der von der [Treuhandstelle Greifswald](#) in einem früheren Projekt entwickelt wurde (Anwendung s. Kapitel 05.03), und sich ggf. gut für unseren Anwendungsfall wiederverwenden lässt. Informationen zu den personenidentifizierenden Merkmalen, welche die Treuhandstelle überwacht, liegen uns vor und werden als Basis für zu entwickelnde Verknüpfungsalgorithmen genutzt.

05.02 ORIENTIERUNG AN BRANCHEN-STANDARDS

Zusammenfassend konzentrieren wir uns in unserem Datenraum vor allem darauf, dass die von den Datenhaltenden verwendeten Standards nahtlos in unser System integriert werden können. Dadurch möchten wir den Aufwand für die Datenhaltenden minimieren, während gleichzeitig für die Dateninteressierten eine größtmögliche Menge an Informationen bereitgestellt wird. Ziel ist es, eine benutzerfreundliche Handhabung für beide Parteien zu schaffen.

Im Gesundheitssektor gibt es zahlreiche Standards, von denen eine Auswahl der relevantesten im Kapitel 04.02 des ersten Berichts dargestellt wurde. Bereits während der Implementierung legen wir großen Wert auf die optimale Integration und Nutzung der branchenüblichen Standards.

In unserem Zielbild, in dem der Datenraum als Vermittler zwischen verschiedenen Einheiten fungiert und Primärdaten von den Datenhaltenden direkt übernommen werden, werden die von den Datenhaltenden implementierten semantischen Standards automatisiert weitergegeben. Dies liegt im Wesentlichen daran, dass nicht das Ziel verfolgt wird, sämtliche Daten in eine festgelegte Struktur zu überführen, sondern vielmehr sollen Daten transaktionsbasiert miteinander verknüpft werden und so den Datennutzenden zur Verfügung gestellt werden. Dafür genügen grundlegende Voraussetzungen an die Datenstruktur.

Besonders bedeutend ist die Orientierung an Standards jedoch für den von uns geplanten Metadatenkatalog. Hierbei beobachten wir aktiv die neuesten Entwicklungen, beispielsweise die des [EHDS](#), der [MII](#), der [NUM](#) und der [NFDI4Health](#). Angesichts der

bereits großen Vielfalt an existierenden Metadatenkatalogen und -standards ist es unser Ziel, auf bestehenden Strukturen aufzubauen. Hierbei planen wir aktuell den Daten- und Metadatenkatalog der [NFDI4Health](#) nachzunutzen. Dieser wurde gemäß der [FAIR-Prinzipien](#) entwickelt, ist frei zugänglich und lässt sich über APIs anbinden. Bei der Entwicklung wurde viel Wert auf die Integration von Standards gelegt; so ist beispielsweise das Metadatenschema als Profil in [FHIR](#) implementiert.

Wir werden im Austausch bezüglich weiterer Standardisierungen wie beispielsweise [HealthDCAT-AP](#) (europäischer Standard für Metadatenkataloge und Ersteller von Metadaten-Standards) bleiben und Entwicklungen in diesem Umfeld aktiv weiterverfolgen.

05.03 VERKNÜPFUNG VERSCHIEDENER DATENQUELLEN

Im noch sehr jungen und dynamischen Forschungsbereich der Post-COVID Forschung wurden in den letzten nahezu fünf Jahren zahlreiche Studien unter hohem Druck durchgeführt. Ziel dieser Anstrengungen war und ist es, das neuartige Krankheitsbild zu verstehen und die medizinische Versorgung optimal zu gestalten. In unserem Austausch mit forschenden Stakeholdern sowie der Säule Medizin hat sich zunehmend der Eindruck herauskristallisiert, dass die meisten „Quick-Wins“, die durch klinische Studien erforscht werden können, bereits untersucht wurden. Die verbleibenden offenen Fragestellungen zeigen eine hohe Komplexität und lassen sich oft nicht durch einen einzelnen Datensatz beantworten.

Durch die Bereitstellung von Möglichkeiten zur Datenverknüpfung sollen neue Perspektiven für die Forschung sowie neue Forschungsfelder eröffnet werden. Mögliche Forschungsfragen sind hier beispielsweise die Untersuchung der Auswirkungen von Post-COVID auf die Berufstätigkeit einzelner und der Gesellschaft (individuelle Arbeitskraft, Renteneintritt) sowie Vergleichsstudien zur allgemeinen Gesundheit mit historischen Kontrollgruppen.

In der ersten Stufe wurden verschiedene Ansätze zur Datenverknüpfung aufgezeigt und erläutert. In dieser Stufe haben wir den Schwerpunkt bewusst auf das Record Linkage gelegt, das in der deutschen Medizinlandschaft derzeit nur begrenzt Anwendung findet, insbesondere aufgrund von datenschutzrechtlichen Herausforderungen. Durch die Entwicklung einer sicheren und benutzerfreundlichen Lösung, die es ermöglicht, Daten aus unterschiedlichen Quellen auf Personenebene miteinander zu verknüpfen, eröffnen sich wie oben beschrieben neue Forschungsfelder. Wir erkennen hierin sowohl den größten Mehrwert für die Forschung als auch den dringendsten Bedarf seitens der Forschenden. Record Linkage wird hierbei durch den Einsatz des Datentreuhänders

[EuroDaT](#) möglich. Einige der Besonderheiten und Chancen, die sich aus der Nutzung von [EuroDaT](#) ergeben, wurden bereits in den Kapiteln 03.01 und 04.04 erörtert. Zusammenfassend zeichnet sich [EuroDaT](#) durch eine hohe Rechtssicherheit und den damit verbundenen Mehrwert für Sicherheit und Vertrauen aus, sowohl seitens der Datengebenden als auch Datennehmenden. Die Datengebenden behalten hierbei jederzeit die Souveränität über ihre Daten.

Insgesamt sehen wir für das Zielbild drei verschiedene Varianten, wie Daten mit Hilfe von [EuroDaT](#) miteinander verknüpft werden können: **Record Linkage über PII**, **Record Linkage über Hashes (PPRL)** und **Record Linkage über Matching Tabelle**. Bevor wir die verschiedenen Möglichkeiten einzeln detailliert betrachten, beschreiben wir die grundlegenden Voraussetzungen für eine Datenverknüpfung. Wir gehen davon aus, dass in den betrachteten Beispielen zwei Datenquellen miteinander verknüpft werden, wobei das Beispiel um beliebig viele Datenquellen erweitert werden kann. Beide Datenhaltende halten ihre personenidentifizierenden Informationen (PII) getrennt von den (Studien-) Datensätzen, entweder bei sich oder bei einer Treuhandstelle. Die Datensätze sind pseudonymisiert. Es existiert pro Patienten jeweils eine Studien-ID bzw. ein Pseudonym, welches im Datensatz und den PII hinterlegt ist. Die später beschriebene Pseudonym-Matching Tabelle kann man sich als Tabelle mit z.B. dem Eintrag, dass Patient 248 in Datensatz A dem Patient 73 in Datensatz B entspricht, vorstellen.

Ausgehend von diesem Szenario beschreiben wir im Folgenden die drei von uns konzipierten Verknüpfungs-Alternativen:

RECORD LINKAGE ÜBER PII

1. Die PII mit den Pseudonymen, sowie die Datensätze inklusive der Pseudonyme werden von den Treuhandstellen/Datenhaltenden an die [EuroDaT](#)-App übermittelt.
2. Diese PII-Pseudonym-Tabellen werden mittels eines Record Linkage Algorithmus in eine Pseudonym-Matching-Tabelle überführt.
3. Die PII werden nach der Erstellung der Pseudonym-Matching-Tabelle gelöscht.
4. Die Datensätze werden basierend auf der Pseudonym-Matching-Tabelle miteinander verknüpft.
5. Zur Minimierung des Reidentifikationsrisikos wird eine Reidentifikationsprüfung durchgeführt.
6. Basierend auf dem Ergebnis der Re-ID-Prüfung und den Datennutzungsbedingungen werden entweder die verknüpften Daten oder Analyseergebnisse herausgegeben.

Vorteil dieses Verfahrens ist die hohe Genauigkeit beim Linkage, da aufgrund der Bereitstellung von PII spezielle Algorithmen zur Verknüpfung genutzt werden können, welche eine höhere Präzision vorweisen als die in der nächsten Methode verwendeten

Privacy-Preserving Record Linkage (PPRL)-Verfahren. Nachteil ist jedoch, dass sich die meisten Datenhaltenden trotz der hohen Sicherheit von [EuroDaT](#) sehr zurückhaltend zeigen, was das Teilen ihrer PII angeht. Hier gibt es eine sehr hohe Hemmschwelle und es muss zunächst durch Stakeholdermanagement ein gewisses Vertrauen in die Technik aufgebaut werden. Hierzu eignet sich folgende Methodik.

RECORD LINKAGE ÜBER HASHES (PPRL)

1. Basierend auf einem von uns bereitgestellten Hash-Algorithmus werden die personen-identifizierenden Daten gehasht.
2. Die Hash-Pseudonym-Tabellen sowie die Datensätze inklusive der Pseudonyme werden von den Treuhandstellen/Datenhaltenden an die [EuroDaT](#) App übermittelt.
3. Diese Hash-Pseudonym-Tabellen werden mittels eines PPRL-Algorithmus in eine Pseudonym-Matching-Tabelle überführt.
4. Die Hashes werden nach der Erstellung der Pseudonym-Matching-Tabelle gelöscht.
5. Die Datensätze werden basierend auf der Pseudonym-Matching-Tabelle miteinander verknüpft.
6. Zur Minimierung des Reidentifikationsrisikos wird eine Reidentifikationsprüfung durchgeführt.
7. Basierend auf dem Ergebnis der Re-ID-Prüfung und den Datennutzungsbedingungen werden entweder die verknüpften Daten oder Analyseergebnisse herausgegeben.

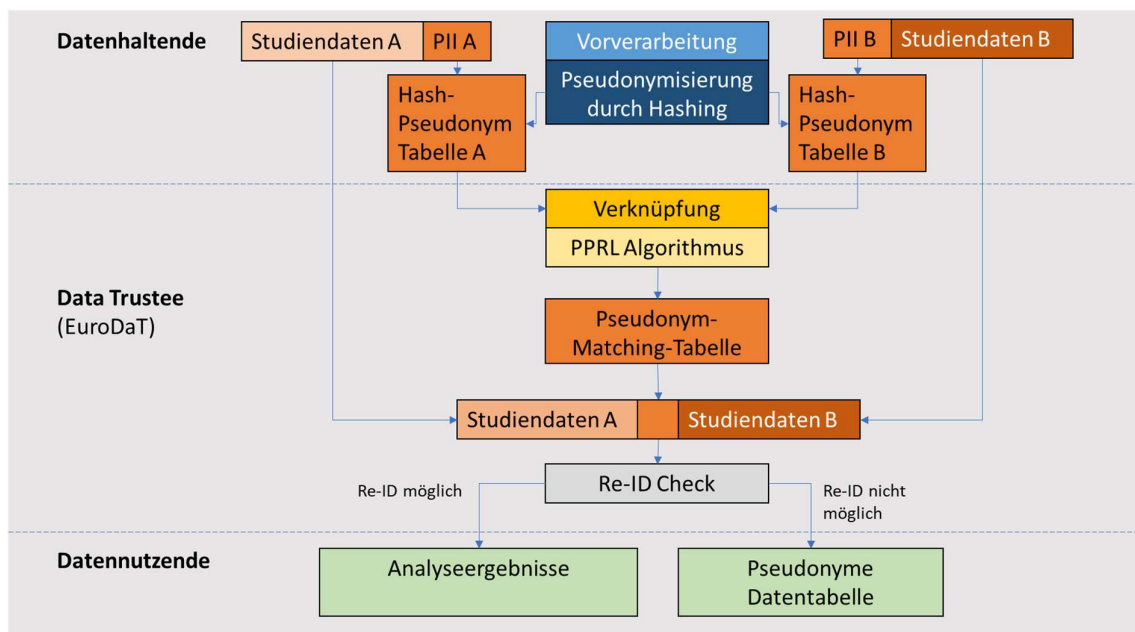


Abbildung 15: Schaubild zum Prozess der Datenverknüpfung mittels Hashes (PPRL).

Vorteil dieser Variante ist der hohe Datenschutz. Zusätzlich zu der bereits abgesicherten Verarbeitung mittels [EuroDaT](#) werden die personenidentifizierenden Merkmale mittels

Hash-Algorithmus anonymisiert. Die Bereitschaft der Datenhaltenden zum Datenteilen ist auf diesen Weg höher und wir haben von mehreren Stellen positives Feedback hierzu erhalten. Insbesondere zur Erprobung und Einführung der neuen Methodik ist dieser Ansatz deshalb gut geeignet. Nachteil des Verfahrens ist jedoch die höhere Fehlerquote beim Verknüpfen der Daten mittels PPRL-Algorithmen.

RECORD LINKAGE ÜBER MATCHING TABELLE

1. Die PII mit den Pseudonymen oder Hash-Pseudonym-Tabellen inklusive der Pseudonyme werden von den Treuhandstellen/Datenhaltenden an die [EuroDaT](#) App übermittelt.
2. Diese PII-Pseudonym-Tabellen oder Hash-Pseudonym-Tabellen werden mittels eines (PP)RL-Algorithmus in eine Pseudonym-Matching-Tabelle überführt.
3. Die PII oder Hashes werden nach der Erstellung der Pseudonym-Matching-Tabelle gelöscht.
4. Die Pseudonym-Matching-Tabelle wird von [EuroDaT](#) an die Datennutzenden übermittelt.
5. Die Datensätze werden von den Datenhaltenden direkt an den Datennutzenden übermittelt.
6. Der Datennutzende verknüpft auf Basis der Pseudonym-Matching-Tabelle die Datensätze.

Der Vorteil dieser Methode ist die maximale Flexibilität der Datennutzenden in der Verarbeitung der Daten. Durch das Bereitstellen der einzelnen Datensätze sowie der Verknüpfungstabelle kann individuell über die Verknüpfung und Auswertungsmethodiken entschieden werden. Nachteil ist jedoch das Einbüßen in der Datensicherheit, da eine Re-ID-Prüfung in dieser Variante nicht möglich ist.

Im Rahmen des MVP setzen wir unseren Fokus zunächst auf das Record Linkage über Hashes. Dies liegt primär an der höheren Bereitschaft der Datenhaltenden ihre PII gehasht und somit mit höherem Datenschutz zur Verfügung zu stellen. Perspektivisch werden auch die beiden weiteren Methoden mit aufgenommen und den Nutzenden zur Verfügung gestellt.

PROZESS ZUM CHECK DER DATEN-REIDENTIFIZIERBARKEIT

Die Anonymisierung von sensiblen und im Allgemeinen hochdimensionalen Daten auf Datensubjektebene ist eine zentrale Herausforderung bei der Verwendung solcher Daten. Das Problem ist die Abwägung zwischen einer hohen Nutzbarkeit der Daten und der Wahrung der Privatsphäre der Datensubjekte. Hochdimensionale Daten bieten zwar reichhaltige Informationen, erhöhen jedoch das Reidentifizierungsrisiko einzelner

Datensubjekte ([Gadotti](#)). Aufgrund dessen ist in unserem anvisierten Zielbild eine Reidentifikationsprüfung vorgesehen.

Ausgangslage und Kontext des Reidentifizierungsrisikos

Neben direkten Identifikatoren wie Name oder Adresse ermöglichen auch Quasi-Identifikatoren wie Geburtsdatum oder Staatsangehörigkeit eine Verknüpfung herausgegebener Daten mit schon existierenden Datensätzen. [Morehouse et al.](#) argumentieren, dass alle Informationen, die bereits in anderen Datensätzen veröffentlicht wurden, als Quasi-Identifikatoren betrachtet werden sollten. Da immer mehr Datensätze veröffentlicht werden steigt somit die Anzahl an Quasi-Identifikatoren stetig an. Zum Problem wird nun, dass das Reidentifizierungsrisiko multiplikativ in der Anzahl an Quasi-Identifikatoren ist. Deshalb steigt das Risiko mit jedem weiteren Quasi-Identifikator drastisch an. Dies wurde veranschaulicht durch [Sweeney](#), die nur mittels Geschlecht, Geburtsdatum und Postleitzahl etwa 87% aller US-Bürger reidentifizieren konnte.

Ansätze zur Minimierung des Reidentifikationsrisikos

Es gibt verschiedene Ansätze, um dieses hohe Reidentifikationsrisiko zu verringern. Die „Safe Harbor“ Methode, dargestellt im US-amerikanischen Datenschutzgesetz für Gesundheitsdaten „Health Insurance Portability and Accountability Act“ ([HIPAA](#)), setzt vor allem auf die Entfernung von allen direkten Identifikatoren sowie die Generalisierung von geographischen Daten und Datumsangaben. Die Deutsche [Bundesbank](#) hingegen rundet alle Euro-Werte zufällig und generalisiert Berufsbezeichnungen sowie das Alter.

Einen weiteren Schritt in Richtung Privatsphäre könnten Algorithmen wie „MinBlur“ ([Morehouse et al.](#)) oder „Quasi-Identifizierung Algorithm“ ([Mansour](#)) darstellen. Diese identifizieren Quasi-Identifikatoren und schlagen eine einfache Modifizierung des Datensatzes vor, um das Reidentifizierungsrisiko zu reduzieren.

Abgeleitete Maßnahmen im Rahmen des Prototyps

Aufgrund der immer steigenden Anzahl an veröffentlichten Daten und der somit steigenden Anzahl an Quasi-Identifikatoren ([Morehouse et al.](#)) erachten viele Forscher Anonymisierung von hochdimensionalen Daten als schlechten Kompromiss zwischen Nutzbarkeit der Daten und Privatsphäre der Datensubjekte ([Gadotti](#)). Allerdings sind die oben genannten Ansätze zur Reduzierung des Reidentifizierungsrisikos teilweise sehr zeitaufwendig in der Umsetzung. Deshalb setzen wir im Rahmen des Prototyps darauf die maximale Spaltenanzahl bei der Ausgabe von Daten zu limitieren. Dadurch wird sichergestellt, dass keine hochdimensionalen Daten auf Datenschubjektebene

herausgegeben werden. Dies verringert das Reidentifikationsrisiko und geht Hand-in-Hand mit dem Prinzip der Datensparsamkeit.

05.04 STRUKTURIERUNG DES DATENMODELLS (TYP DER DATENBANK)

In diesem Kapitel beschreiben wir das logische Datenmodell des Datenraums und legen den Fokus vorrangig auf die beteiligten relationalen Datenbanken und auf das Datenbank-Design, s. Tabelle 3 für einen Überblick über die Datenbanken.

DB-Name	Zugang & Berechtigungen	Beschreibung
Audit- & Management-DB des Datenraums	Nur technische Benutzer	Verwaltung der Metadaten des Datenkatalogs und der Transaktionen, sowie Monitoring
Transaktions-DB des Daten-treuhänders	Nur die an der Transaktion teilnehmenden und autorisierten Datennutzenden und Datenhaltenden, nur während der Laufzeit der Transaktion, granulare Berechtigungen durch Row-Level-Security-Policies, s. Kapitel 06.01	Temporär existierende DB, um die Eingangsdaten aufzunehmen, zu analysieren und den Datennutzenden bereitzustellen
Lokale Client-DB der Daten-haltenden	Nur Datenhaltende	Lokale (on-premise) Ablagestelle für das Datenprodukt des Datenhaltenden

Tabelle 3: Überblick über die Datenbanken des Datenraums.

AUDIT- UND MANAGEMENT-DB DES DATENRAUMS

Wir haben uns beim Datenbank-Design für den [Data-Vault](#)-Ansatz entschieden, um flexibel genug auf Änderungen und Erweiterungen gemäß den Anforderungen der Stakeholder während der Projektphase reagieren zu können. Das Entity-Relationship-Diagramm für die Hub-, Satelliten- und Link-Tabellen ist in Abbildung 16 zu sehen. Die wesentlichen Aufgaben dieser zentralen Datenbank sind:

- Verwaltung des Datenkatalogs: Attribute (Spaltenname, Typ) und weitere Metadaten (bspw. Kurzbeschreibung, Herkunft).
- Verwaltung der Datennutzenden und Datenhaltenden, in Verbindung mit der Authentisierungs- und Autorisierungslösung Keycloak.
- Versionierung der Datenmodelle und der Verknüpfungen, v.a. um die Nachnutzbarkeit zu befördern.

- Tracking und Tracing der Transaktionen, um Statusänderungen abzubilden und Nachvollziehbarkeit zu garantieren.
- Allgemeines Monitoring, um die Fehlerbehebung zu erleichtern und so die Robustheit der Architektur zu unterstützen.

Das Tracking und Tracing der Transaktionen schafft Mehrwert für die Stakeholder:

- Datenhaltende: Dokumentation wer die eigenen Daten abgerufen und wie verknüpft hat,
- Datennutzende: Wiederverwendung von erprobten Verknüpfungen und Information über die Güte des Matchings.

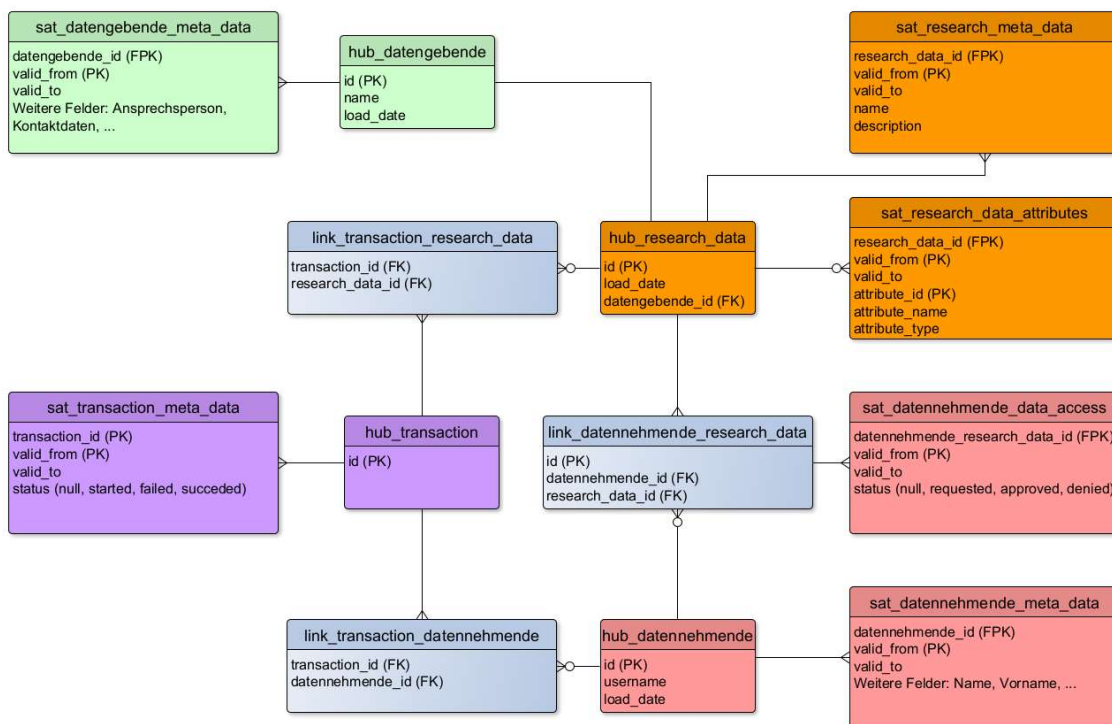


Abbildung 16: Entity-Relationship-Diagramm des Data Vaults der relationalen Datenbank des Datenraums für Audit und Metadatenmanagement. Farbliche Gruppierung in Datenebene (grün), Hub (blau), Forschungsmetadaten (orange) & Transaktionsmanagement (rot).

TRANSAKTIONS-DB DES DATENTREUHÄNDERS

Da der Datentreuhänder [EuroDaT](#) transaktionsbasiert arbeitet, wird nur für die typischerweise kurze Laufzeit einer Transaktion eine temporäre Datenbank provisioniert, die nach Ende der Transaktion gelöscht wird. Die Ergebnisse der Analyse werden in ein Output-Datenschema geschrieben, von wo die Datennehmenden es abrufen können. Datengebende hingegen schreiben ihre Daten in ein separates Input-Datenschema. Die Kommunikation mit dieser Datenschicht ist durch eine Daten-

Management-API abgebildet. Die Transaktions-DB hat drei festgelegte Schemata für das Einlesen von Primärdaten (Input-Schema), die Analyse (Intermediate-Schema) und das Bereitstellen der Ergebnisdaten (Output-Schema). Über Row-Level-Security-Policies für das Output-Schema kann jedem Teilnehmer einer Transaktion exklusiver Zugriff auf individuelle Ergebnisdaten ermöglicht werden, s. Kapitel 06.01.

06. PROZESSE UND ARCHITEKTUR

06.01 PLANUNG DER IT-INFRASTRUKTUR

EINFÜHRUNG UND ZIELE

Bei der Planung der IT-Infrastruktur leiten uns die aus Stufe 1 bekannten ausgewählten Qualitätsziele¹, basierend auf dem [ISO/IE 25010 Standard](#):

- Benutzerfreundlichkeit / Usability,
- Kompatibilität / Interoperabilität, und
- IT-Sicherheit.

Unsere öffentliche Projektseite <https://post-covid.dateninstitut.d-fine.dev/> dient als Schaufenster für unseren Lösungsansatz und wird im weiteren Projektfortschritt kontinuierlich ausgebaut.

LÖSUNGSSTRATEGIE UND ARCHITEKTURENTSCHEIDUNGEN

Der Quellcode des Datenraums soll als Open-Source-Software veröffentlicht werden, was die Auswahl der Komponenten maßgeblich beeinflusst. Diese Wahl und die weiteren wichtigen Architekturentscheidungen mit zentraler Bedeutung für das Gesamtsystem sind in Tabelle 4 aufgeführt und begründet.

Entscheidung	Kontext	Beschreibung und Auswirkungen
Open-Source-Software Stack	Unsere Lösung soll kostengünstig zu betreiben sein und großen Mehrwert für Forschende und die Öffentlichkeit bieten.	Der Open-Source-Software-Ansatz bietet Mehrwert in den Dimensionen <ul style="list-style-type: none">• Sicherheit: Veröffentlichung ermöglicht unabhängige Experten-Reviews• Transparenz: Funktionsweise der Datenraum-Plattform ist einsehbar

¹ Für eine detaillierte Diskussion der Qualitätsziele, der Randbedingungen und ihrer Rationale sei auf Kap. 05.01 des Abschlussberichts der Phase 1 verwiesen.

		<ul style="list-style-type: none"> Nutzerbeteiligung: Stakeholder können sich aktiv an der Weiterentwicklung beteiligen
Transaktionsprinzip	Die rechtlichen Vorgaben zur Speicherung von Forschungsdaten, dem Schutz von personenbezogenen Daten und zur Datenverknüpfung setzen enge Grenzen für den Datenaustausch.	Das Transaktionsprinzip gewährleistet einen sicheren & rechtskonformen Datenaustausch: Da die Transaktions-DB nur zur Laufzeit der Transaktion existiert, werden Daten nie dauerhaft gespeichert.
EuroDaT als Datentreuhänder	Der Datenraum benötigt eine Zwischenschicht, die den sicheren und automatisierten Datenaustausch unter den beteiligten Entitäten leisten kann.	EuroDaT arbeitet transaktionsbasiert, d.h. die nötigen Forschungsdaten werden nur während der Transaktionszeit abgerufen und in einer In-Memory-DB gespeichert. Analyseskripte aggregieren Forschungsdaten zu Ergebnisdaten. Nach Abschluss der Transaktion wird die gesamte virtualisierte Infrastruktur wieder abgerissen. Gleichzeitig sei betont, dass sich unsere Gesamtarchitektur agnostisch zur Auswahl eines speziellen Datentreuhänders verhält, solange dieser den sicheren transaktionsbasierten Datenaustausch inkludiert.
Client-Applikation für Datenhaltende	Zur Interaktion der Datenhaltenden mit dem Datenraum wird ein sicherer Einstiegspunkt benötigt.	Die Client-Applikation wird on-premise bei den Datenhaltenden installiert, bspw. auf einer virtuellen Maschine. Der Integrationsaufwand in den bestehenden Freigabeprozess ist dabei minimal, insbesondere verbleibt die Hoheit über die Datenweitergabe bei den Datenhaltenden.
Lokale Datenbank für Audit und Metadatenmanagement.	Um einen Datenkatalog abzubilden und die Anforderungen an Nachnutzbarkeit vorheriger Verknüpfungen,	Die Datenbank folgt dem Data-Vault -Schema und speichert lediglich die Metadaten der Datenprodukte und der Transaktionen. Insbesondere hat sie keine Kenntnis über tatsächliche

	Monitoring und Zugriffskontrolle zu erfüllen, benötigt der Datenraum eine Persistenzschicht.	Forschungsdaten, s. Kapitel 05.04 für Details.
Datenschließfach	Perspektivisch soll die Architektur in der Lage sein, Datenspenden zu persistieren und Forschungsinstitutionen ohne eigene Infrastruktur anzubinden	Die Safe-Deposit-Box des Datentreuhänder EuroDaT stellt ein Datenschließfach dar, in dem Daten sicher und in Verbindung mit einem Rechtenkonzept verwahrt werden können.

Tabelle 4: Überblick über zentrale Architekturentscheidungen für den Datenraum.

Die technische Implementierung der [EuroDaT](#)-Plattform ist als Open-Source-Lösung über das [EuroDaT-Code-Repository](#) öffentlich zugänglich. Die grundlegenden Bausteine und Rollen innerhalb EuroDaT werden in Abbildung 17 gezeigt. Für eine detaillierte Diskussion des Rollenkonzepts von EuroDaT im Kontext unserer Data Governance sei auf Kapitel 03.01 verwiesen, ebenso auf Kapitel 04.04 für die DSGVO-Implikationen, die wir für den Betrieb von EuroDaT abgeleitet haben.

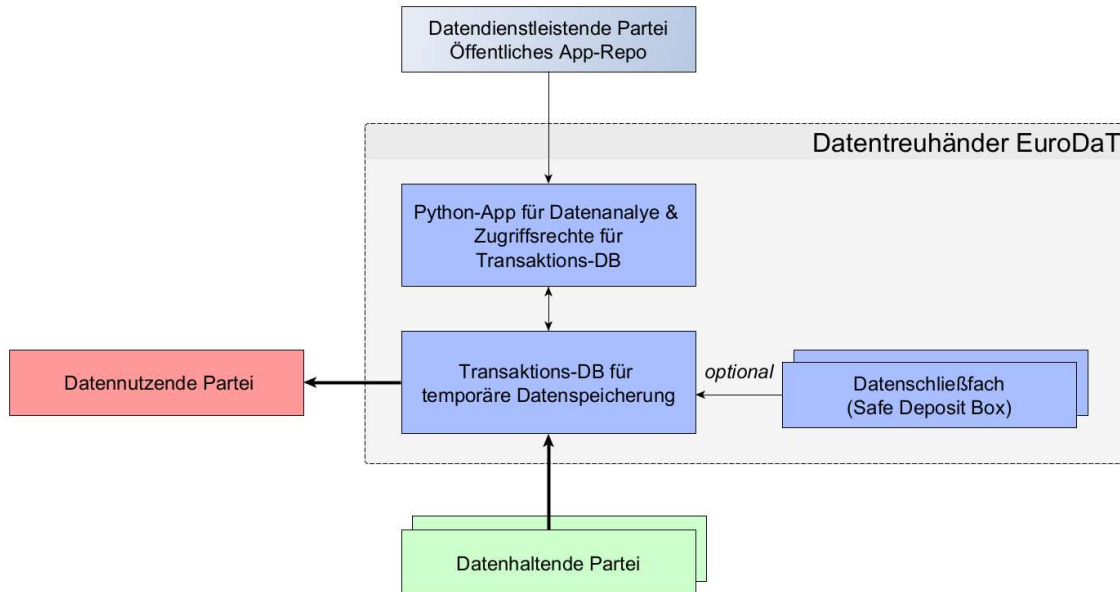


Abbildung 17: Überblick über die zentralen Bausteine des Datentreuhänders EuroDaT mit den Rollen der Datennutzenden, Datenhaltenden und den Datendienstleistende.

FACHLICHER & TECHNISCHER KONTEXT

Der fachliche Kontext der IT-Architektur des Datenraum spezifiziert die externen Schnittstellen, d.h. er beschreibt alle Kommunikationspartner und die übermittelten Informationsobjekte. Um Dopplungen zu vermeiden, verweisen wir für den fachlichen Kontext auf die grundlegenden Prozessdiagramme in Kapitel 03.04, welche zusätzlich die internen Kommunikationsprozesse beinhalten.

Für den technischen Kontext soll der Fokus auf dem technisch anspruchsvollsten Subsystem liegen: der Integration eines Datentreuhänders mitsamt der Anbindung der einzelnen Datengebenen, s. Abbildung 17 und Abbildung 18.

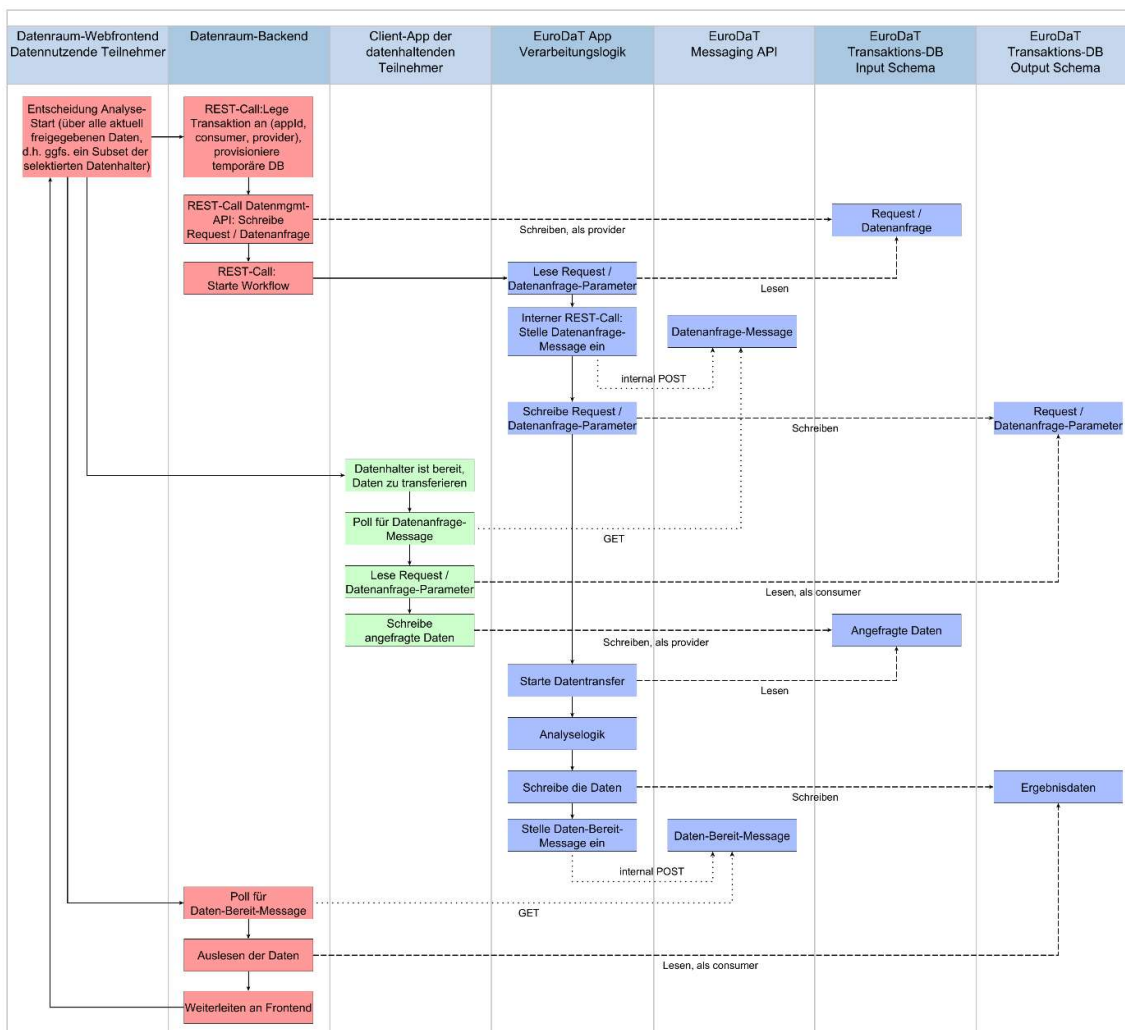


Abbildung 18: Darstellung der Datenflüsse zwischen Datenkatalog-Applikation für Datennutzende, Datentreuhänder [EuroDaT](#) und der Client-Applikation der datenhaltenden Partei.

Die wesentlichen technischen Kennzeichen dieser Datenschnittstelle sind:

- Die Datenflüsse werden als REST-Anfragen abgebildet.
- Die Kommunikation von Datenanfragen und bereitstehenden Ergebnisdaten erfolgt über eine Messaging-API.
- Die Zugangsberechtigungen zu den Ergebnistabellen wird über Row-Level-Security-Policies gesteuert: Diese Regeln legen pro Tabelle granular auf DB-User-Ebene fest, welcher DB-User welche Tabellenzeilen sehen und bearbeiten darf und verhindern so unbefugten Datenzugriff.
- Das Auslösen einer Transaktion unterliegt ebenfalls einem Berechtigungsmanagement.
- Die Authentisierung und Autorisierung der datennutzenden und datenhaltenden Entitäten erfolgen durch TLS-Client-Zertifikate.

06.02 AUFBAU DER INFRASTRUKTUR

BAUSTEINSICHT

Die Bausteinsicht der Datenrauminfrastruktur beschreibt die statische Zerlegung des Gesamtsystems in einzelne Funktionsbausteine, s. die Überblicksdarstellung in Abbildung 19 und die detaillierte Beschreibung in Tabelle 5.

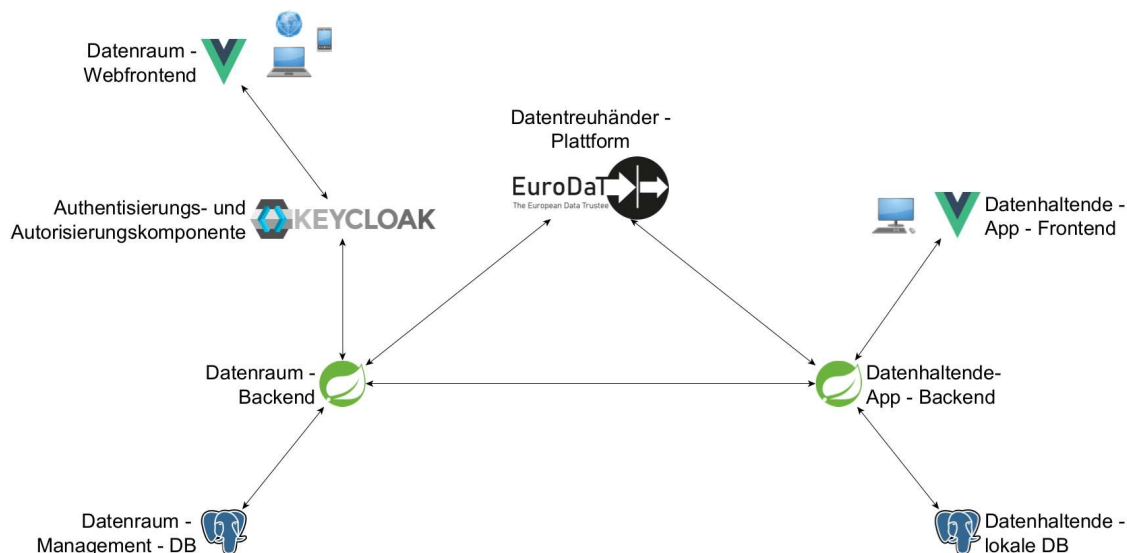


Abbildung 19: Whitebox-Darstellung der Datenrauminfrastruktur.

Bausteinname	Beschreibung	Open-Source-Technologien
Datenraum-Webfrontend	Öffentliche Website des Datenraums, mit den Funktionalitäten Datenkatalog, Datenverknüpfung, Login-Bereich für	JavaScript & Vue-Framework, Docker Container

	Datennutzende und Datenhaltenden, sowie perspektivisch U&A-Verfahren	
Datenraum-Backend	Steuerung der fachlichen Prozesslogik, der Kommunikation mit dem Datentreuhänder und mit den Datenhaltende-Client-Apps und der Datenbankzugriffe	Kotlin Spring Boot, Docker Container
Datenraum-Management-DB	Audit und Management der Metadaten der Datenprodukte und der Transaktionen	PostgreSQL DB, Docker Container
Keycloak	Authentisierungs - & Autorisierungskomponente	Docker Container
Datentreuhänder EuroDaT	Transaktionsbasierter Datenaustausch und Datenanalyse	Externer Service
Datenhaltende-Frontend	Lokale nicht-öffentliche Applikation auf Seite der Datenhaltenden, zur Aufnahme der Datenprodukte für die Datennutzenden	JavaScript & Vue-Framework, Docker Container
Datenhaltende-Backend	Kommunikation mit dem Datenraum und dem Datentreuhänder	Kotlin Spring Boot, Docker Container
Datenhaltende-Lokale-DB	Temporäre lokale Speicherung der für den Weitertransport bestimmten Datenprodukte	PostgreSQL DB, Docker Container

Tabelle 5: Überblick über die wesentlichen Bausteine der IT-Architektur des Datenraums, ihrer Zuständigkeiten und Schnittstellen.

Insbesondere senkt die Containerisierung der Client-Anwendung mit Docker die Schwelle zur Anbindung für die Datenhaltenden erheblich, indem sie die on-premise Integration in die lokale Infrastruktur vereinfacht.

06.03 PROZESSE ZUR DATENINTEGRATION

In diesem Kapitel stellen wir unseren Prozess für eine möglichst niederschwellige Integration von Forschungsdaten in den Datenraum vor: Von der initialen Informationssuche durch Datenhaltende auf unserer Web-Seite, dem Anmelde-Vorgang, dem Einpflegen der Datensatzbeschreibung sowie, optional, von Zusatzinformation bis hin zum abschließenden Teilprozess für die Prüfung und Freigabe durch die Datenraum-Betreibenden. Dieser fachliche Gesamtprozess ist in Abbildung 20 dargestellt und sieht die folgenden Abläufe vor:

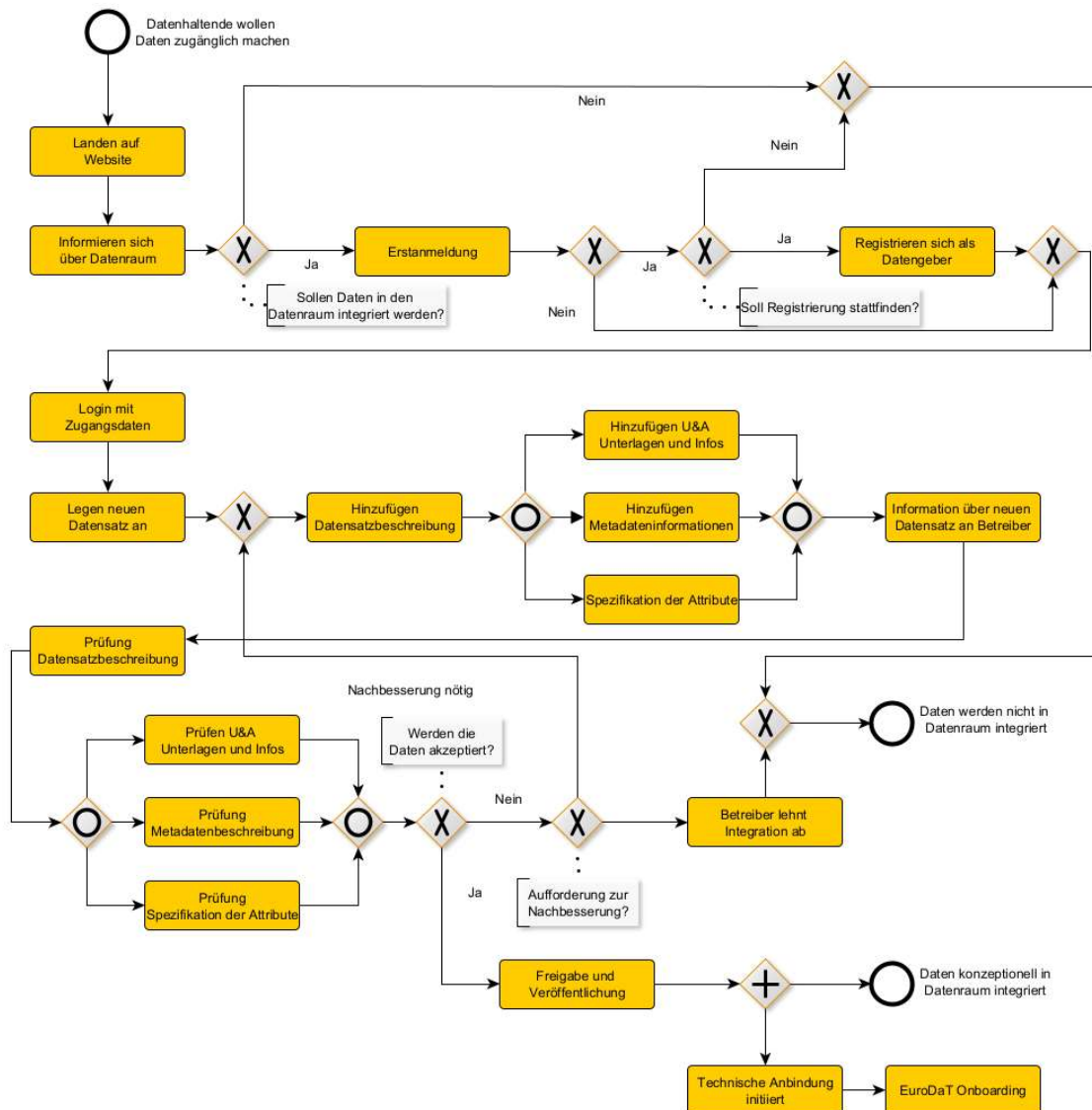


Abbildung 20: Fachprozess für die Integration von Daten, aus Sicht der Datenhaltenden.

Um eigene Daten in den Datenraum zu integrieren, müssen sich Datenhaltende über die Website mit ihren Zugangsdaten anmelden, bzw. bei Erstanmeldung zunächst registrieren. Um das Integrationsangebot für Datenhaltende so niederschwellig wie möglich zu gestalten, ist bei Erstaufnahme lediglich die Datensatzbeschreibung verpflichtend: U&A Unterlagen, Metadateninformationen und die Spezifikation der Spaltenattribute sind zwar essenziell für die spätere Verwendung, aber erstmal optional und können zu einem späteren Zeitpunkt nachgepflegt werden. Alle eingereichten Dateninformationen werden dann in geeigneter Weise automatisiert bzw. im MVP noch tlw. manuell geprüft und mit den Qualitätsstandards des Datenraums abgeglichen. Sollten die Daten nicht den Qualitätsstandards genügen wird ggfs. eine

Nachbesserungsschleife angestoßen oder der Datensatz vom Betreibenden abgelehnt. Bei erfolgreicher Prüfung wird der neue Datensatz in den Datenraum integriert und die technische Anbindung an EuroDaT samt Onboarding angestoßen.

06.04 PROZESSE ZUR DATENAKTUALISIERUNG

In diesem Kapitel soll der Fokus auf den Prozessen zur Pflege der Daten liegen, mit denen das Datenmodell aktuell gehalten wird. Dazu zählen Maßnahmen zur Aktualisierung der Primär- und Transaktionsdaten, zur Pflege der Metadaten, sowie zur Datenmodellpflege.

PRIMÄRDATENAKTUALISIERUNG

Die zentrale Innovation des Post-COVID Datenökosystems spiegelt sich auch in der Datenmodellpflege wider: Durch den Einsatz eines transaktionsbasierten Datentreuhänders können wir die Pflege der Primärdaten automatisieren und so garantieren, dass bei jeder Datentransaktion der zum Zeitpunkt des Abrufs jeweils aktuelle Datenstand von den Datengebenden abgerufen wird. Der Datentreuhänder gewährleistet dies, da er konsequent die Datensouveränität der Datenhaltenden sichert, indem die Daten innerhalb der technischen Infrastruktur der Datengebenden verbleiben und – bis auf die typischerweise kurze Zeitspanne der Transaktion – nicht physisch herausgegeben werden. Stattdessen vermittelt der Treuhänder Datentransaktionen, bei denen die Zugriffsrechte auf die Daten geprüft und die Daten anschließend gemäß vorliegenden Rechten im jeweils aktuellen Stand direkt von den Datengebenden abgerufen werden. Im Treuhänder erfolgt lediglich die Verknüpfung dieser Daten, die anschließend an die Berechtigten ausgeliefert werden.

TRANSAKTIONS DATENAKTUALISIERUNG

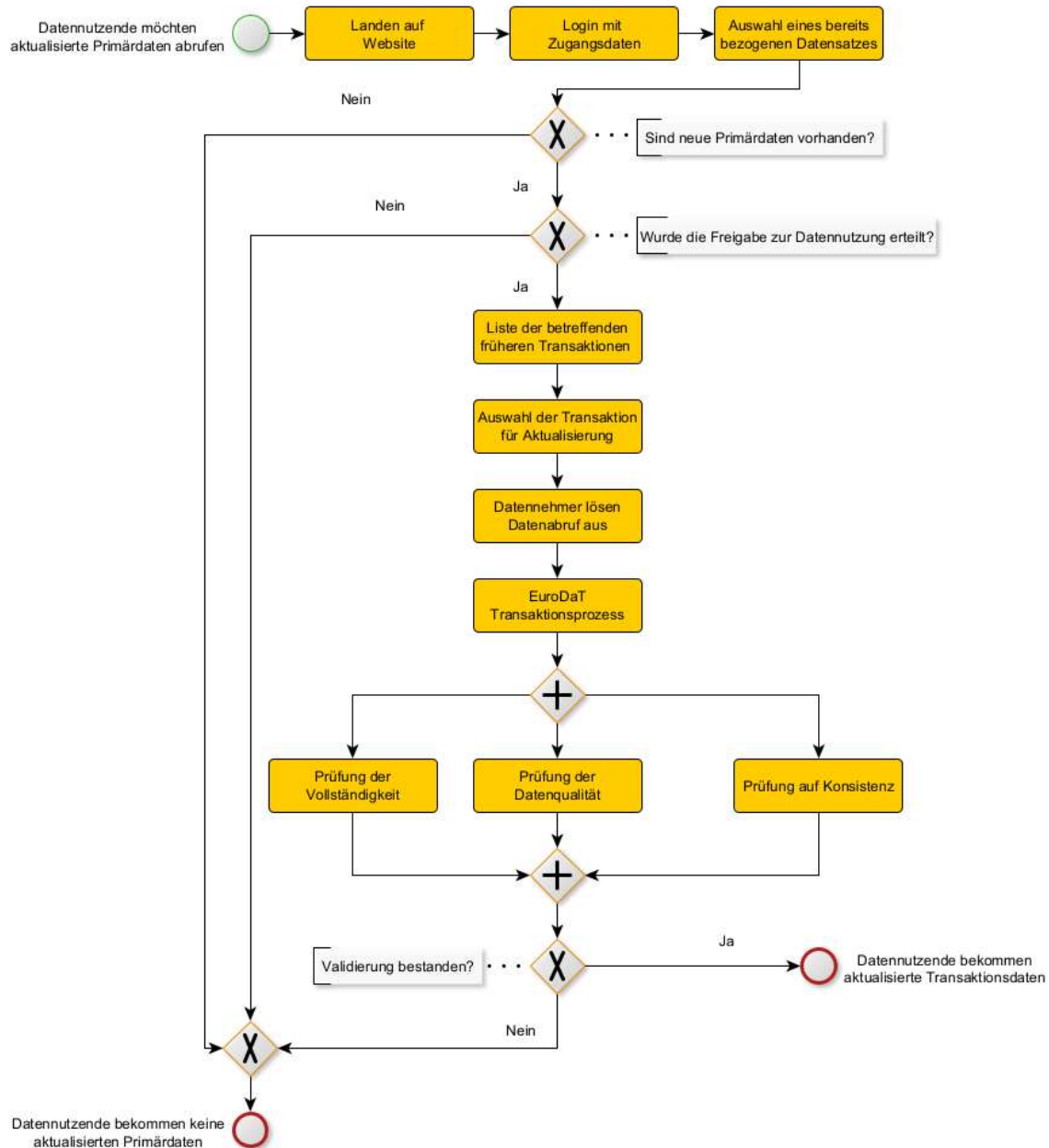


Abbildung 21: Fachprozess zur Aktualisierung der Transaktionsdaten nach Änderung der Primärdaten, in einer vereinfachten Ausprägung für den aktuellen MVP.

Da es auch nach erfolgreich durchgeführten Datentransaktionen zu Änderungen der Primärdaten kommen kann, muss zusätzlich ein Konzept zur Aktualisierung der Transaktionsdaten implementiert werden. Dieser Fachprozess ist in vereinfachter Ausprägung für den MVP in Abbildung 21 dargestellt.

Kommt es nach Durchführung einer Datentransaktion zu Änderungen an den Primärdaten, muss diese Information von Seite der Datenhaltenden an den Datenraum weitergegeben werden. In einer späteren Ausbaustufe lässt sich über die Datenapplikation ein Mechanismus installieren, über den alle Datennutzenden, die in vorherigen Transaktionen diese Daten genutzt haben, über die Aktualisierung informiert werden. Bei Freigabe und Bereitstellung der neuen Daten durch die Datenhaltenden, können die Datennutzenden eine neue Datentransaktion starten und erhalten im Anschluss die aktualisierten Transaktionsdaten.

METADATENAKTUALISIERUNG

Im Gegensatz zu den Primärdaten und Transaktionsdaten, sollen die Metadatenbeschreibungen im Datenraum vorgehalten werden, um den Nutzenden Informationen über die verfügbaren Daten einfach zugänglich zu machen und so einen niederschweligen Zugang zum Datenökosystem zu ermöglichen. Die Metadaten müssen also aktiv gepflegt werden. In Abbildung 22 ist unser Fachprozess für die Aktualisierung der allgemeinen Informationen zu einem Datensatz, inklusive der Aktualisierung der Metadaten durch Datenhaltende im Datenraum abgebildet.

Primär setzen wir hierbei auf die Eigenverantwortung und regelmäßige Aktualisierung durch die Datenhaltenden. Hierzu werden ergänzend halbjährlich Erinnerungen an die Datenhaltenden geschickt, in denen sie aufgefordert werden, ihre Metadaten auf deren Aktualität zu überprüfen.

Ein wichtiger Aspekt sind zusätzlich regelmäßige Schulung und die Einrichtung von Feedback-Runden. Hierdurch können das Bewusstsein für korrekte und aktuelle Metadaten geschärft und deren Bedeutung für die Forschung hervorgehoben werden. Ergänzend dazu werden Feedback-Mechanismen etabliert, durch die Forschende und Nutzende aktiv Rückmeldungen zu den Metadaten geben können. Dies fördert die Identifikation von etwaigen Lücken oder Fehlern und ermöglicht eine prompte Korrektur, s. auch Kapitel 04.03.

Schließlich sehen wir vor, einen Katalog zu erstellen, also ein Tool zur visuellen und systematischen Überwachung der Metadaten. Das hierfür entwickelte Dashboard stellt perspektivisch den aktuellen Status der Metadaten, etwaige Lücken und die Verknüpfungen von Daten übersichtlich dar. Es befähigt die Nutzenden somit mögliche Verknüpfungen von Forschungsdaten einfach zu identifizieren und zu aktualisieren.

Zusammen ermöglichen diese Mechanismen eine effektive Pflege der Metadaten im Datenökosystem, gewährleisten deren Aktualität und fördern eine transparente und reibungslose Datenverarbeitung für die Forschung.

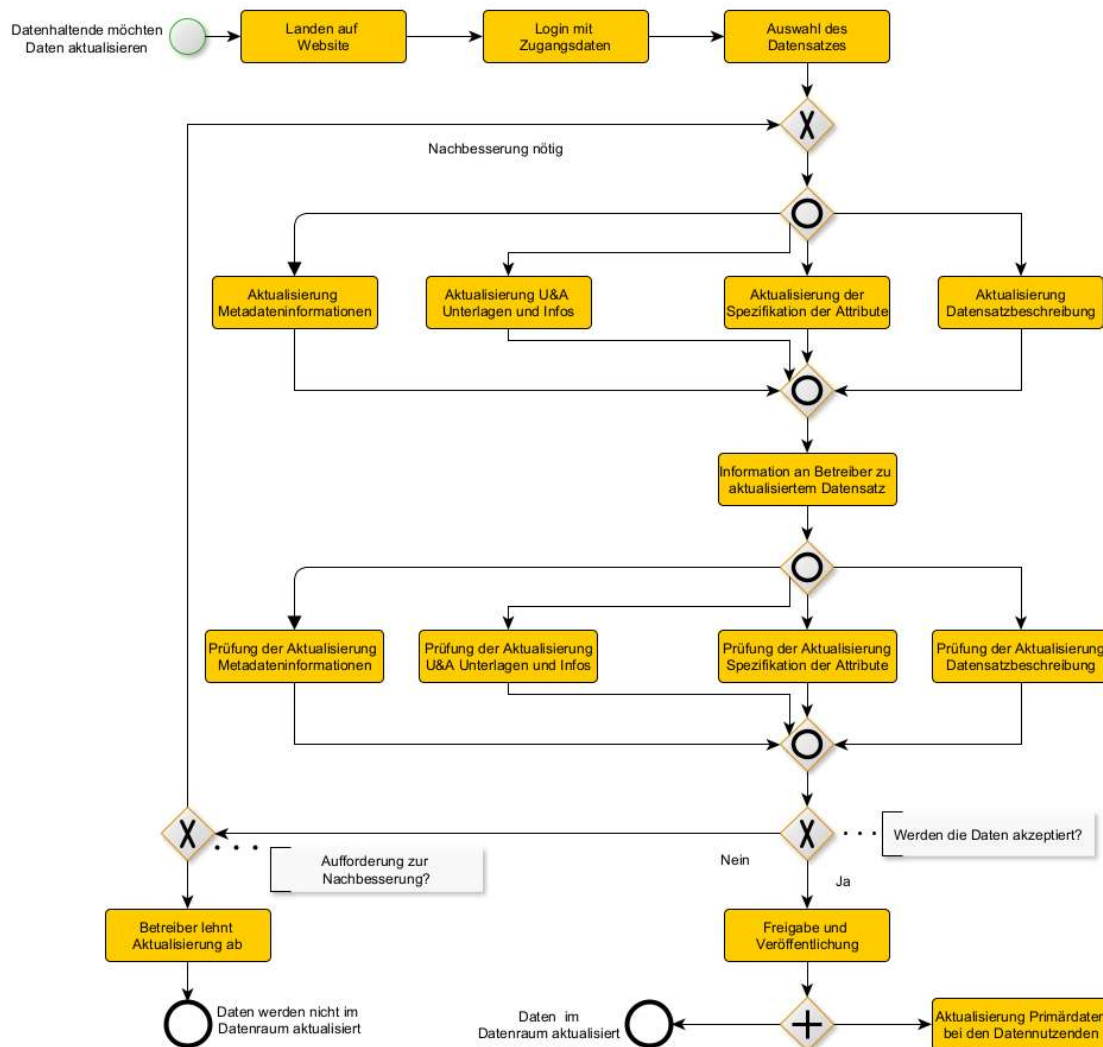


Abbildung 22: Fachprozess für die Aktualisierung der allgemeinen Informationen zu einem Datensatz, inklusive der Metadaten durch Datenhaltende im Datenraum.

DATENMODELLPFLEGE

Um die Daten des Datenraumes aktuell halten zu können, muss auch das Datenmodell selbst gepflegt werden. Nur so kann die Integrität und Relevanz der Daten gesichert werden.

Hierzu ist die Compliance und der Datenschutz von zentraler Bedeutung, da eine ständige Überprüfung der Datenhaltung sicherstellen muss, dass alle Aktivitäten den geltenden Datenschutzrichtlinien, wie der DSGVO, und ethischen Standards entsprechen, s. auch Kapitel 04.01. Zusätzlich ist es unserer Meinung nach sinnvoll, langfristig eine Versionierung des Datenmodells einführen, um Änderungen systematisch nachzuvollziehen und frühere Versionen bei Bedarf abrufen zu können,

was entscheidend für die Rückverfolgbarkeit ist. Zudem ist eine kontinuierliche Überwachung unerlässlich; durch die Implementierung von Monitoring- und Reporting-Systemen können potenzielle Probleme in der Datenmodellpflege frühzeitig erkannt und adressiert werden, vgl. Kapitel 03.01. Nicht zuletzt spielt die Stakeholder-Einbindung eine wichtige Rolle, da der regelmäßige Austausch mit allen Beteiligten (z.B. Forschungsverbünde, Krankenkassen) dazu beiträgt, die Datenakzeptanz und -relevanz zu sichern und den gemeinsamen Nutzen zu maximieren, s. auch Kapitel 04.02.

07. BETRIEB UND NACHNUTZUNG DES DATENMODELLS

Das Betriebs- und Nachnutzungskonzept des Datenmodells haben wir bereits im Abschlussbericht der Stufe 1 ausführlich vorgestellt. Das dort vorgestellte Konzept haben wir in Stufe 2 weiter konkretisiert und spezifische Prozesse und Mechanismen entwickelt, um den Betrieb, die Pflege und die Nachnutzung des Datenökosystems zu gewährleisten. Unser Ansatz hat dabei eine sehr positive Resonanz von vielen Stakeholdern erhalten, welche die Notwendigkeit des Record Linkage hervorheben und dies als einen bedeutenden Faktor für einen erfolgreichen Betrieb und eine effektive Nachnutzung ansehen. Die Schwierigkeiten des Record Linkage, insbesondere aufgrund des Fehlens eines Unique Identifiers in der medizinischen Forschung, stellen in der Tat ein erhebliches Problem dar. So ist die Verknüpfung relevanter Informationen unter Einhaltung der relevanten DSGVO-Bestimmungen nur unter unproportional hohem Aufwand und häufig auch gar nicht möglich. Die Möglichkeit, nun eine DSGVO-konforme Umsetzung eines Record Linkage zu realisieren, eröffnet nicht nur neue Perspektiven in der medizinischen Forschung, sondern ist auch für finanzstarke Wirtschaftsunternehmen bspw. aus der Pharmaforschung von höchstem Interesse. Diese Einschätzung teilt auch das ZEG Berlin, das ein Testimonial zu unserer Lösung abgegeben hat, s. unten.

Darüber hinaus kann die Nachnutzung unseres Datenmodells als Public Code einen signifikanten Mehrwert für zukünftige Entwicklungen und Implementierungen bieten. Da wir unsere Arbeitsergebnisse sowohl zentral als Teil des hier konzipierten Datenökosystems als auch dezentral für eigenständige Implementierungen zur Verfügung stellen, haben Teilhabende leichteren Zugang zu den Daten und den damit einhergehenden Verknüpfungsmöglichkeiten. Des Weiteren ermöglicht unser Ansatz den Austausch und die Zusammenarbeit mit einer Vielzahl von Stakeholdern. Durch verschiedene Kommunikationsformate und partizipative Kanäle wie ein öffentliches Git-Repository oder ein strukturierter Feedback-Kanal haben wir eine Umgebung geschaffen, die die Beteiligung und Zusammenarbeit aller Interessierten fördert. Dies ebnet den Weg für eine Vertiefung und Verbesserung der Zusammenarbeit, die das volle Potenzial des vernetzten Datenraums ausschöpft, der somit für die Post-COVID-Forschung gewinnbringend nachgenutzt werden kann.

Das ZEG Berlin – Zentrum für Epidemiologie und Gesundheitsforschung (<https://zeg-berlin.de/>) ist ein Auftragsforschungsunternehmen, das sich maßgeblich auf das Themengebiet der Arzneimittel- und Patientensicherheit fokussiert und hier innovative Konzepte der Pharmakoepidemiologie zur Untersuchung des Risikoprofils von zugelassenen pharmazeutischen Produkten etabliert hat. Im Pharma und Healthcare Sektor erstellt das ZEG Berlin mit Fachexpertise und hochqualitativen Daten aussagekräftige Analysen, optimiert das Design prospektiver nicht-interventioneller Studien, sichert die Patientensicherheit und weist für medizinische Innovationen den Mehrwert nach. Hierzu bedient sich das ZEG Berlin auch sog. Sekundärdatenquellen z.B. zur Erforschung des Risikoprofils bestimmter Arzneimittel bzw. zum Aufzeigen des Zusatznutzens eines pharmazeutischen Präparates.

Die zugrundeliegenden Sekundärdatensätze enthalten aber häufig nicht alle benötigten Informationen, unter anderem da es aktuell sehr aufwendig bis unmöglich ist, sensible Patientendaten aus verschiedenen Quellen miteinander zu verknüpfen. Eine vielversprechende Möglichkeit solche benötigten Datenverknüpfungen herzustellen, bietet nun der Ansatz, den d-fines Konsortium bei der Post-Covid Challenge verfolgt. Der erwartbare Mehrwert für die Datennutzung in Deutschland, wenn diese Lösung umgesetzt wird, ist substantiell und wäre auch für uns beim ZEG Berlin spürbar. Durch die Möglichkeit Record Linkage auf Einzelpatientenebene durchführen zu können, eröffnen sich völlig neue Perspektiven in der datengetriebenen medizinischen Forschung, die die Bereitschaft zu bedeutenden Investitionen in Projekte, die diese Lösung nutzen, antreiben werden. Wir sehen daher einen klaren Business Case für die von d-fine mitentwickelte Lösung, da die erwartbaren Effizienz- und Reichweitengewinne die prognostizierten Kosten der Nutzung übersteigen sollten. Hierdurch würde unserer Meinung nach die medizinische Forschung und die Wettbewerbsfähigkeit Deutschlands in einem zunehmend datengetriebenen Markt nachhaltig gestärkt.

Dr. Marc Pignot, PhD, MSc
Chief Strategy Officer, ZEG Berlin

Dieses und weitere uns zurückgespielten positiven Rückmeldungen verstärken unsere Überzeugung, dass unser Datenmodell das Potenzial hat, die Post-COVID-Forschung maßgeblich voranzubringen und auch als Blaupause für die Entwicklung weiterer Datenräume in anderen Sektoren dienen kann. In Kombination mit der positiven Resonanz der Stakeholder und der überzeugenden Leistungsfähigkeit unseres Datenmodells, bestärkt uns dies in der Überzeugung, dass wir durch unser Datenmodell ein großes Marktpotential erschließen könnten.



Ihr Kontakt

Dr. Robert Görke

Partner

+49 069 907370

Healthcare@d-fine.com

d-fine GmbH
An der Hauptwache 7
60313 Frankfurt
Deutschland

d-fine

analytical. quantitative. tech.