

# Der Beitrag von NFDI4Health zur Nutzung von Gesundheitsdaten

Juliane Fluck

Im Auftrag des Konsortiums NFDI4Health

DHDS  
J. Fluck ZB MED  
2023-07-12 | Mainz

# Die Nationale Forschungsdateninfrastruktur (NFDI)

## ➤ Vision NFDI

Daten als gemeinsames Gut für exzellente Forschung, organisiert durch die Wissenschaft in Deutschland.

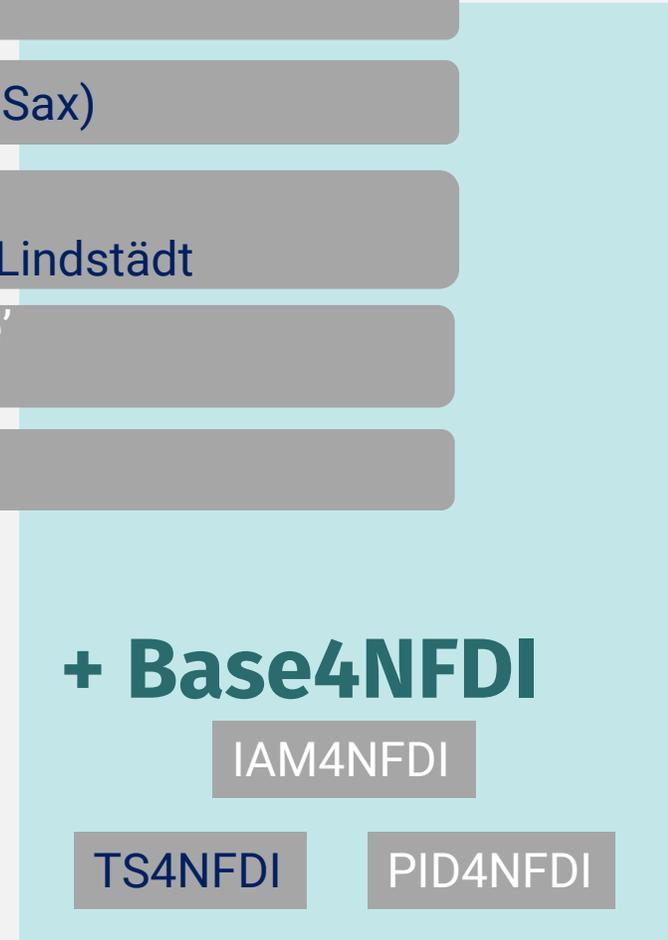
Anmerkung:

Die NFDI ist deutscher Beitrag zur ‚European Open Science Cloud‘

- in Europa sind 9 Datenräume angedacht\*, einer davon ist der
- European Health Data Space (EHDS)

\*siehe [Eine europäische Datenstrategie](#)

# NFDI - Zusammenarbeit



# NFDI Mission

- Wir verbessern die Nutzungsmöglichkeiten von Daten
- Wir schaffen rechtskonforme, interoperable und nachhaltige Dateninfrastrukturen
- Wir arbeiten national und international zusammen
- Wir bilden aus und stärken die Kompetenz im Umgang von Daten

# NFDI Konsortien mit Gesundheitsbezug

- Personenbezogene Gesundheitsdaten:
  - ❖ NFDI4Health (Start 2020),
  - ❖ GHGA (Start 2020),
- Konsortien relevant für medizinischer Forschung
  - ❖ NFDI4Microbiota (Start 2021),
  - ❖ NFDI4BiolImage (Start 2023),
  - ❖ NFDI4Immuno (Start 2023),

# NFDI4Health Mission

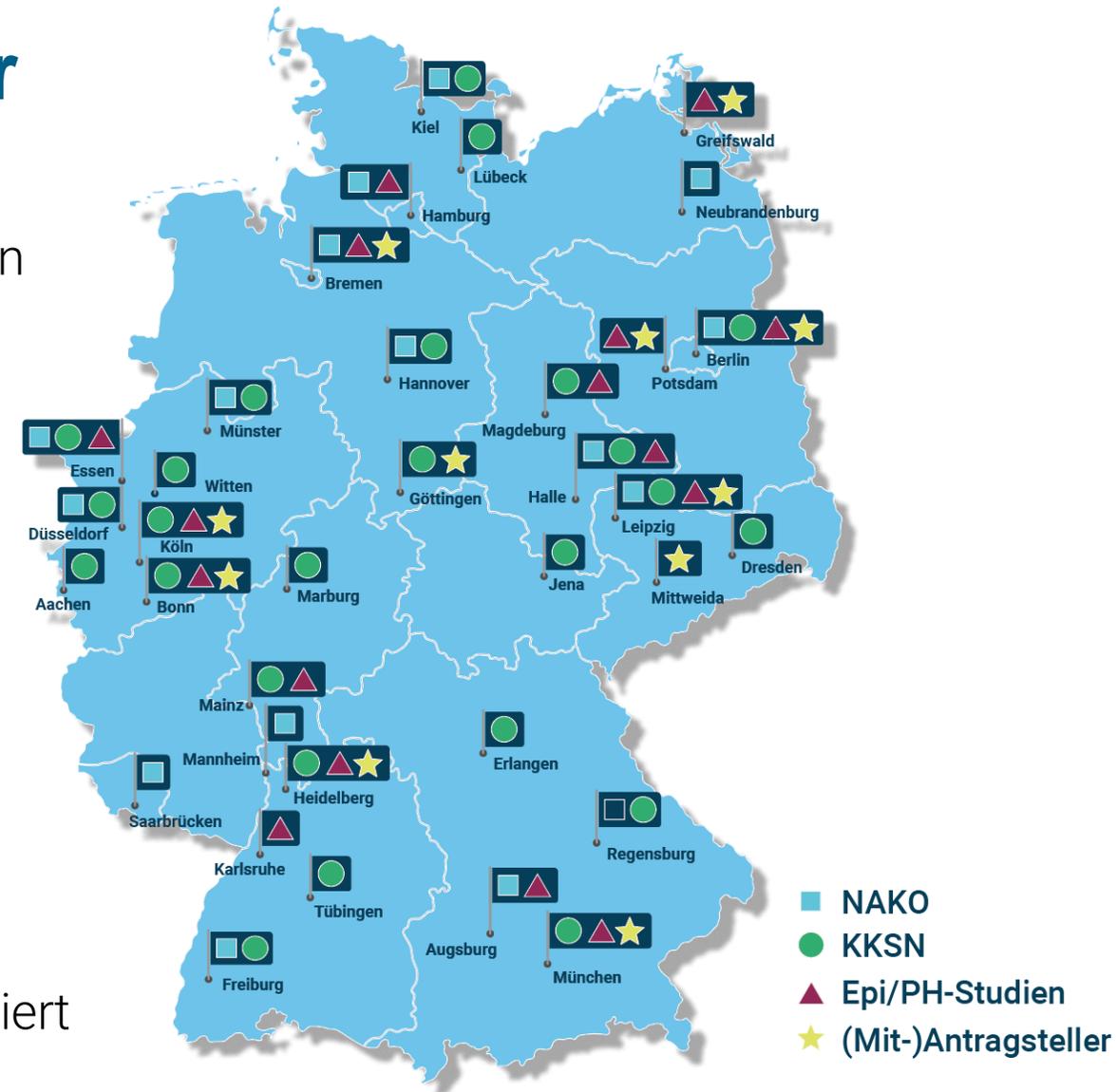
Unser Auftrag:

- Steigerung des Wertes der Forschung in den Bereichen Epidemiologie, öffentliche Gesundheit und klinische Studien basierte Medizin

indem wir qualitativ hochwertige Daten nach den **FAIR-Prinzipien** international zugänglich machen

# Forschungsdateninfrastruktur

- Viele unabhängige Forschungsdatenzentren
  - ❖ Heterogene, sehr umfangreiche komplexe Daten
  - ❖ Daten sind oft nur lokal verfügbar
  - ❖ Eigene Version der Einwilligung der Teilnehmer
  - ❖ Struktur zur Beantragung der Datennutzung lokal vorhanden
  - ❖ Record Linkage zu Registerdaten u. Krankenkassendaten oft notwendig aber extrem schwierig zu etablieren
- Klinische Studien oft ohne Data Sharing Strukturen
  - Sehr heterogene Datenkörper
  - Eigene Version der Einwilligung der Teilnehmer
  - Dateninfrastruktur oft nur zur Archivierung etabliert



# Vernetzung im deutschen Gesundheitsdatenraum

## NFDI

Gesundheits-  
Studiendaten

Menschliche  
Omics - Daten



Krankenkassen-  
Daten



Federal Institute  
for Drugs  
and Medical Devices

Gesundheitsregister-  
Daten



Zentrum für  
Krebsregisterdaten  
ROBERT KOCH INSTITUT

Versorgungsdaten



Unterschiede: Inhalt | Verantwortlichkeiten | Rechtliche Regulierungen

# Ähnliche Situation im Deutschen Gesundheitsraum

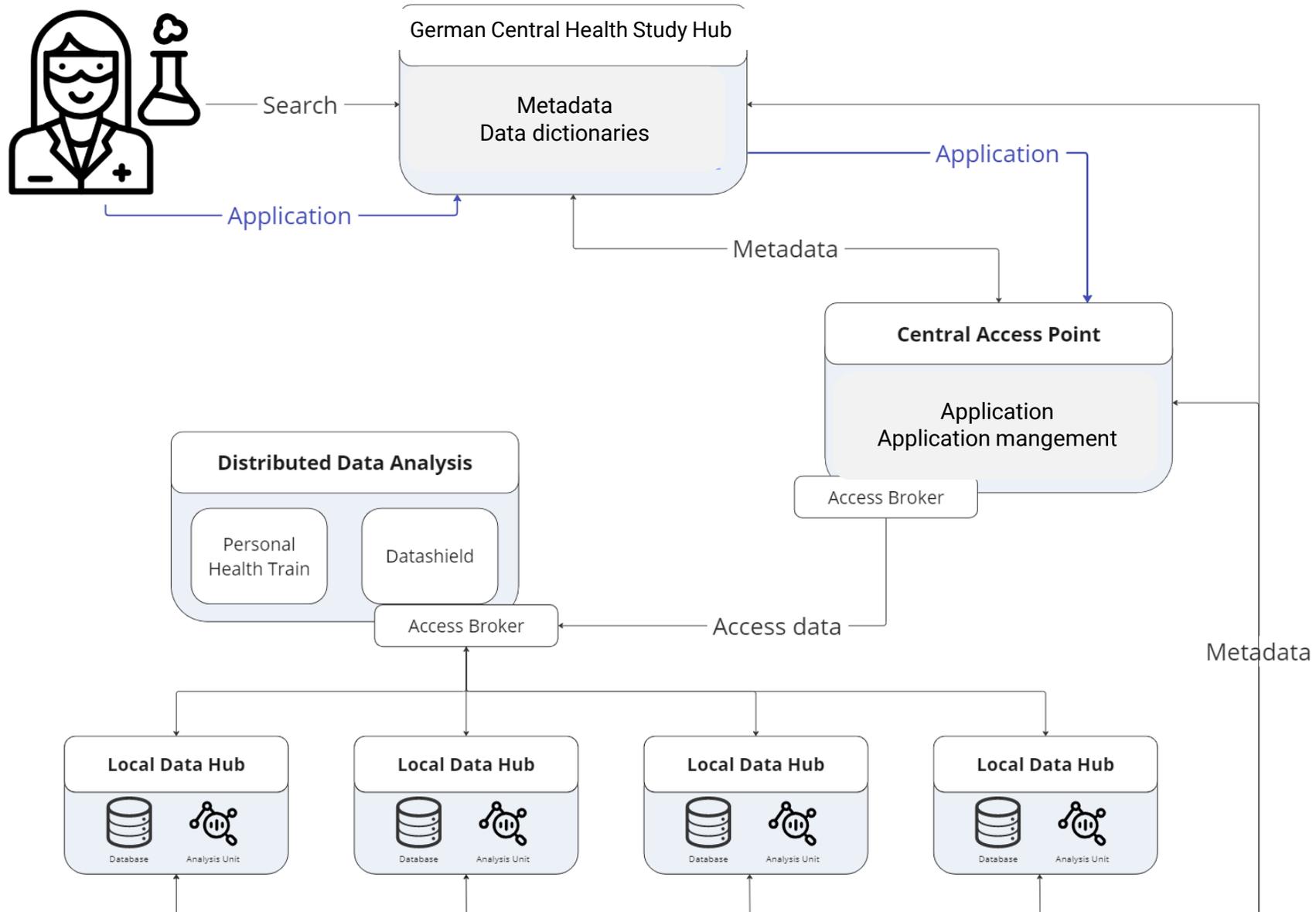
## ➤ Viele unabhängige Forschungsdatenzentren

- ❖ Heterogene Daten
- ❖ Daten sind oft nur lokal verfügbar (?)
- ❖ Eigene Version der Einwilligung der Teilnehmer oder keine Einwilligung
- ❖ Struktur zur Beantragung der Datennutzung lokal vorhanden
- ❖ Record Linkage oft notwendig aber extrem schwierig zu etablieren

NFDI4Health konzipiert, implementiert und erprobt rechts-konforme, interoperable und nachhaltige Gesundheitsdateninfrastruktur-Lösungen.

Diese können über NFDI4Health hinaus als erprobte Konzepte für zentrale Dienste im deutschen Gesundheitsdatenraum dienen.

# NFDI4Health Konzept der zentralen Dienste

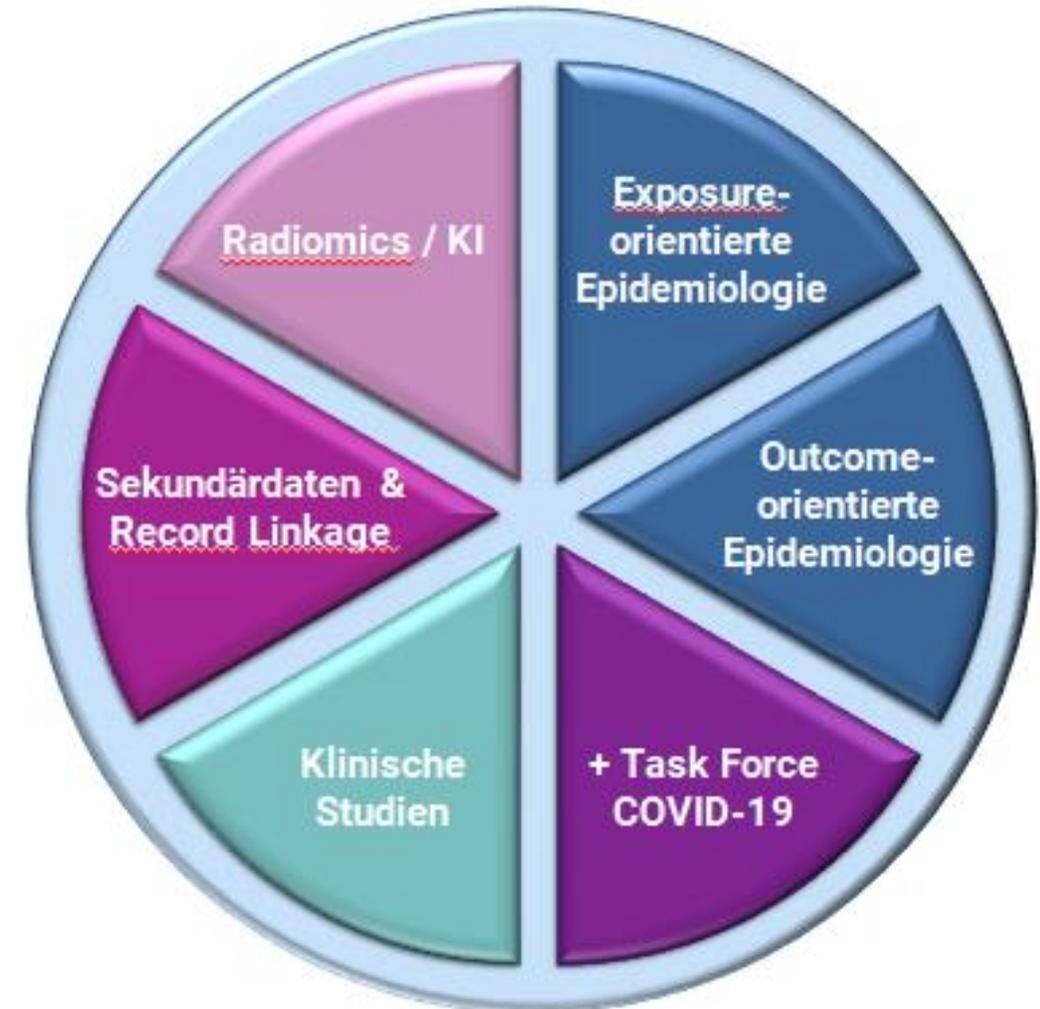


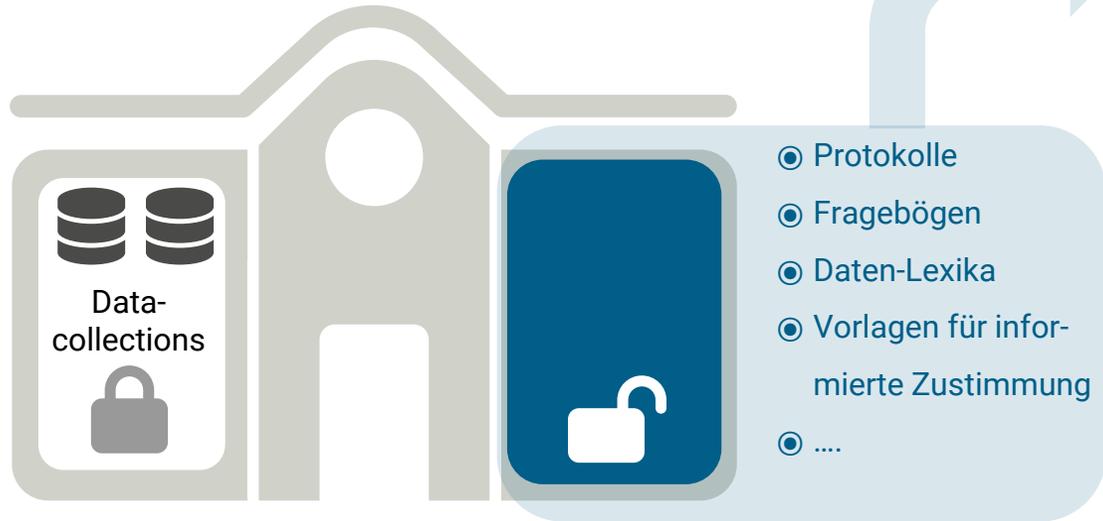
# Use Cases

## Auswahl der Use Cases, um

- die **verschiedenen User Communities** abzudecken  
**wichtigsten Aspekte** personenbezogener Gesundheitsdaten beispielhaft zu adressieren
- „**Blueprint**“ für zukünftige Studien und Datensätze
- **Services und Standards** für zentrale Datensätze in Deutschland zu implementieren
- Ansätze zur Sicherung der **Vertraulichkeit** und der Einhaltung von **Datenschutzregularien** und **ethischen Prinzipien** beispielhaft erarbeiten

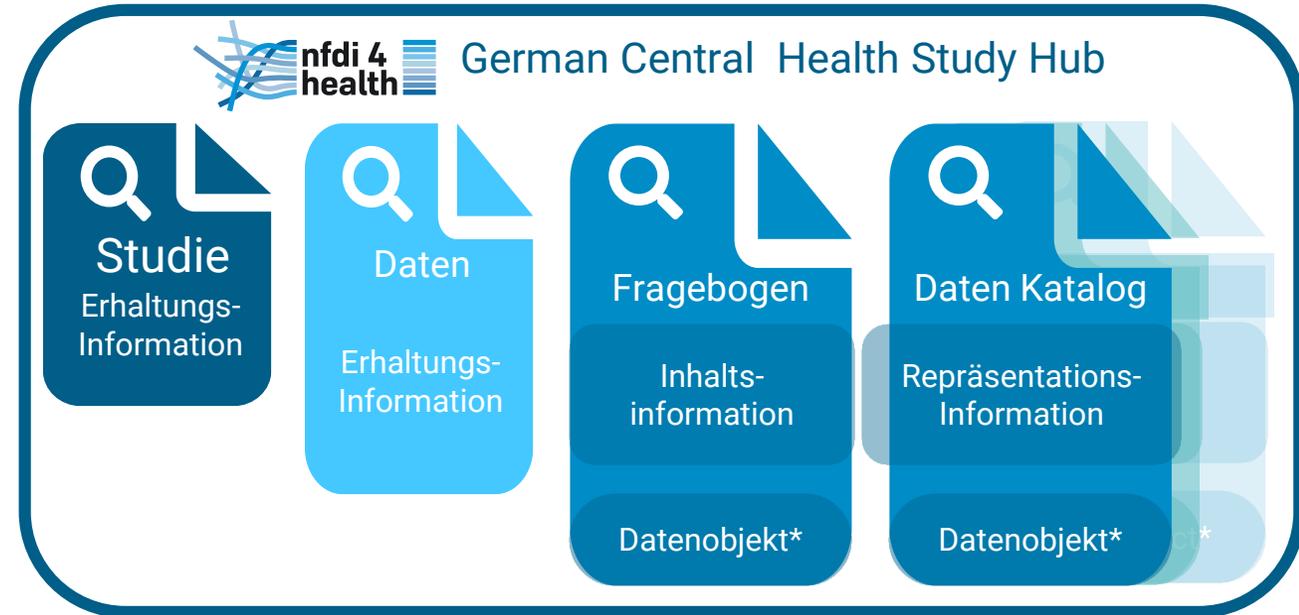
## ➤ **Pilotierung der FAIR-Prinzipien**





## Forschungseinrichtung

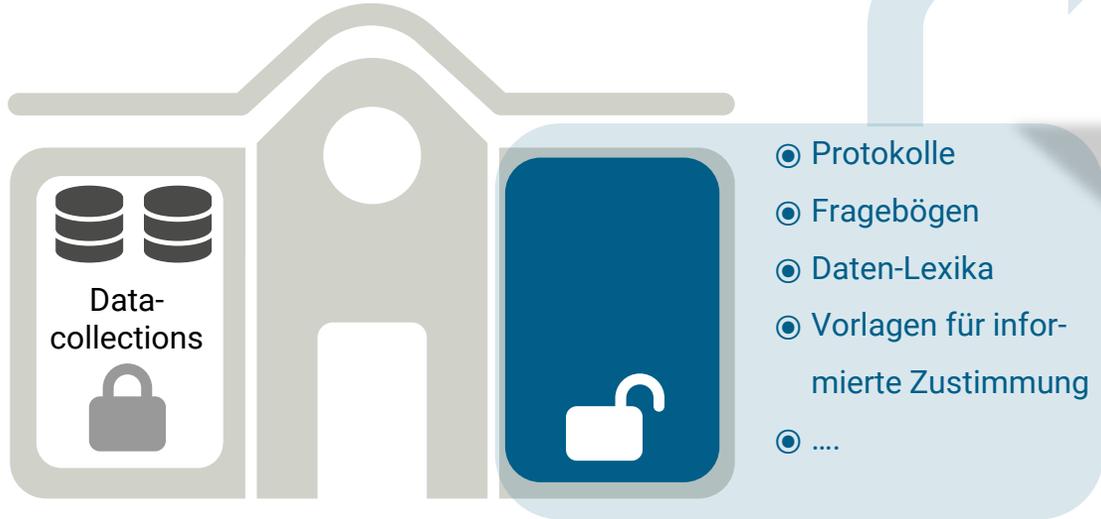
Beteiligt an der Datenerhebung und Analyse von persönlichen Gesundheitsdaten



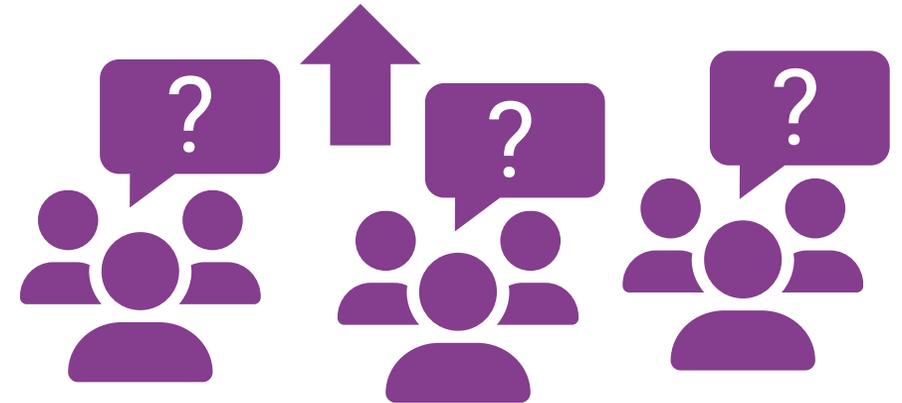
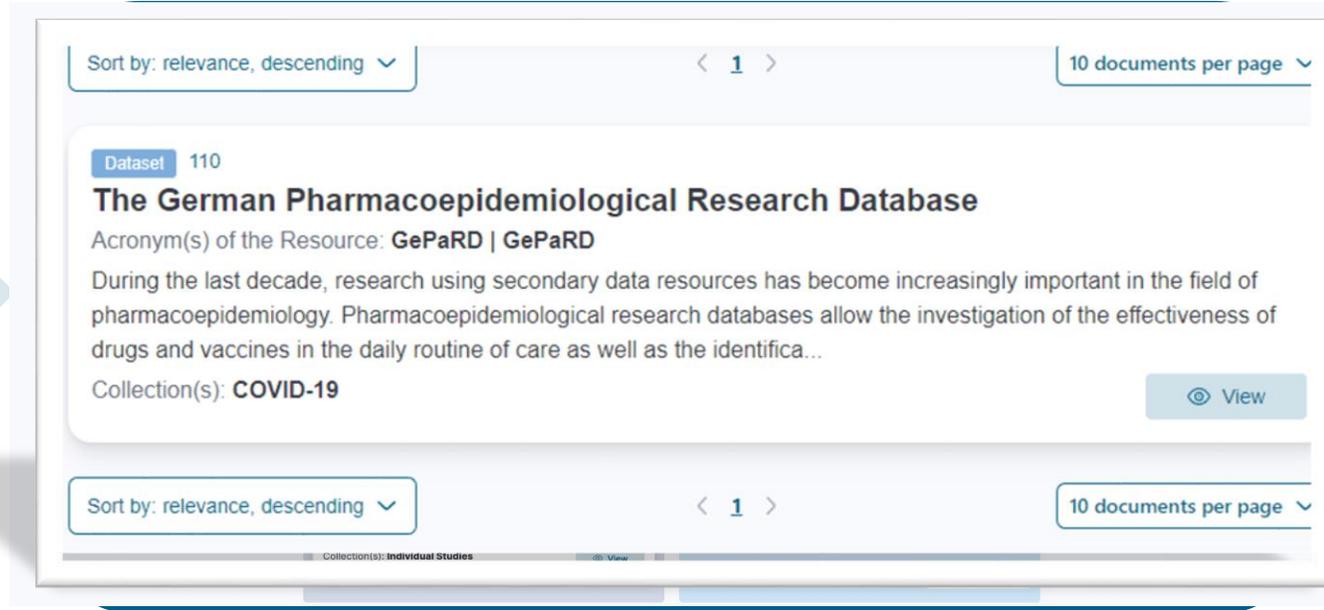
\* Optional



Wissenschaftler:innen



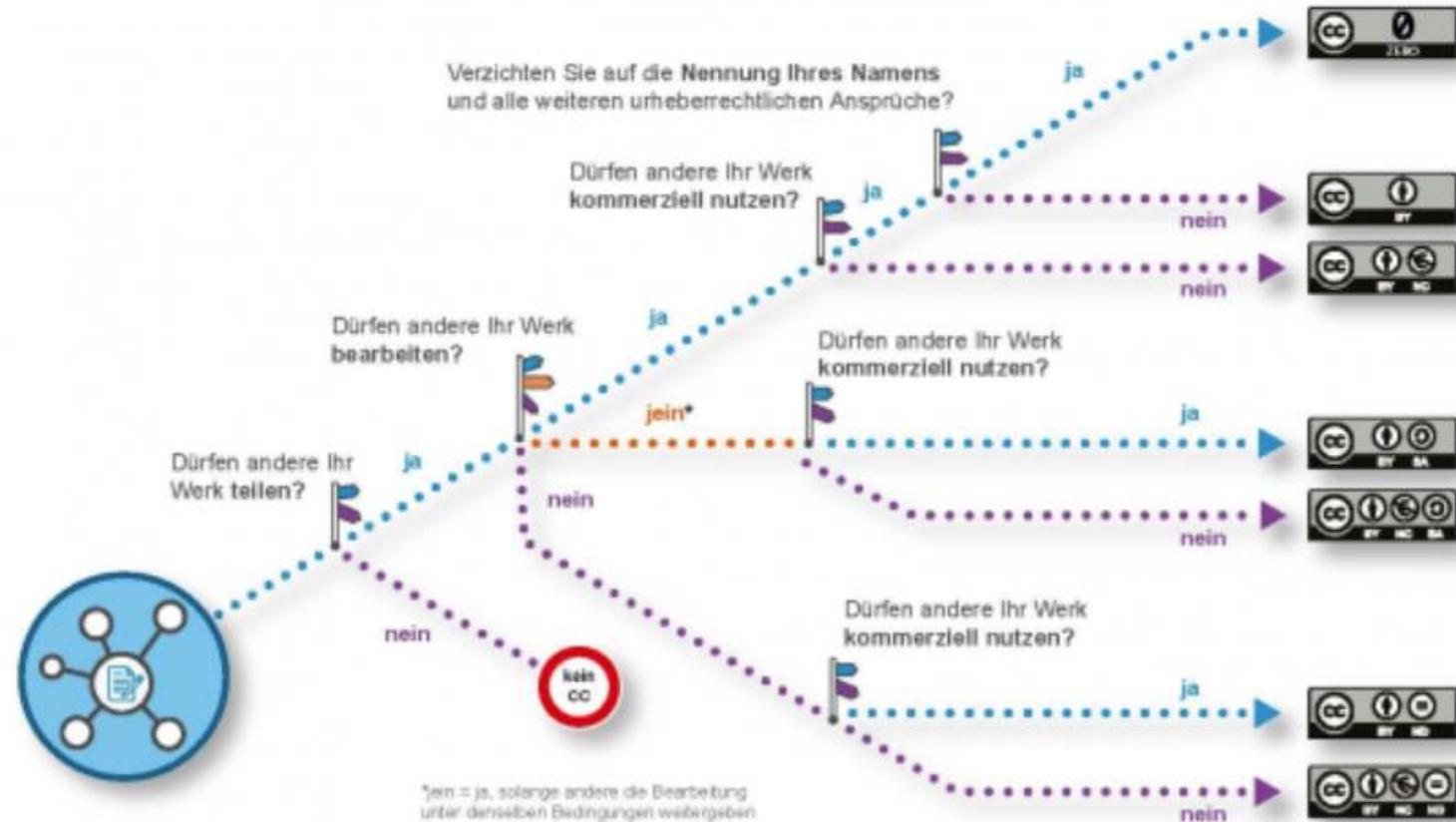
Aktuell >1600 Einträge



Wissenschaftler:innen

Erste Sekundärdatenbank eingetragen, Kollektion Krebsregisterdaten geplant

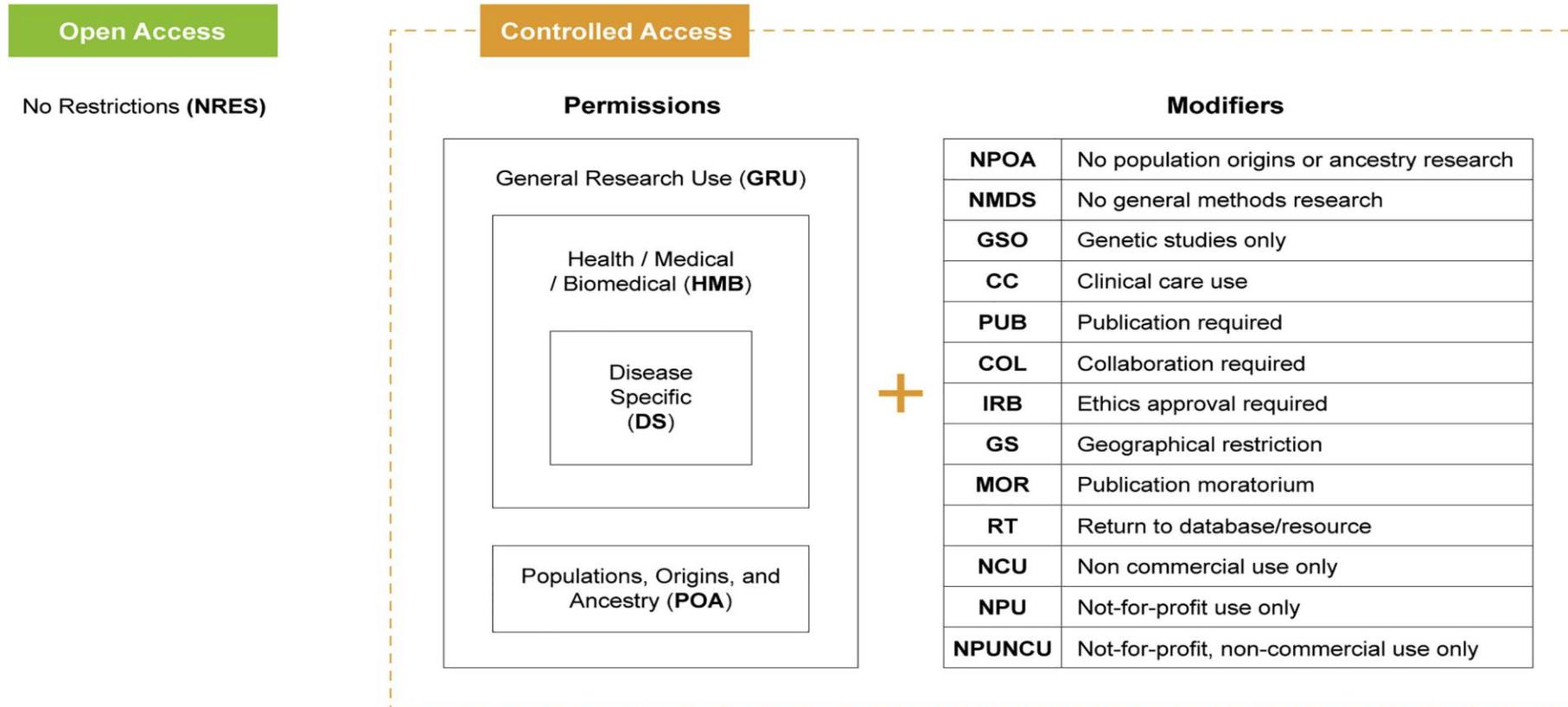
# Publikation von Studiendokumenten und Erhebungsinstrumenten – CC-Lizenzen auf einem Blick



Quelle: von Barbara Holz und Julia Paul-Mehrdorf für nfdi 4 health UMR CC BY/SA/3.0

# Nächster Schritt: Maschinen-lesbare Datennutzungsbedingungen

## Konzeption GA4GH: Data Use Ontology permissions and modifiers

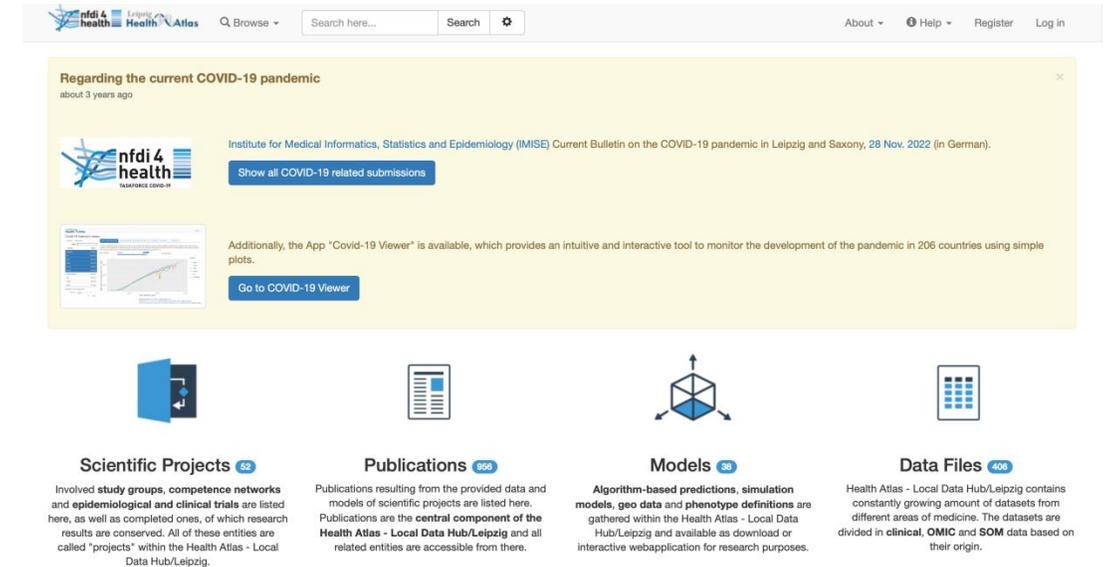


From Lawson et al. *The Data Use Ontology to streamline responsible access to human biomedical datasets*. *Cell Genomics*, Volume 1 Issue 2 (November 2021), DOI: 10.1016/j.xgen.2021.100028

# Findability

# Local Data Hub

- Forschungsdatenmanagement-Lösung für dezentralisierte/institutionelle Verwaltung der Daten
- Verschiedene Betriebsmodelle
  - On-Premise, Software als Service
- Anbindung an zentrale Dienste der NFDI4Health
  - für Auffindbarkeit und Zugänglichkeit
- Optional: dezentrale Datenanalyse-Komponente

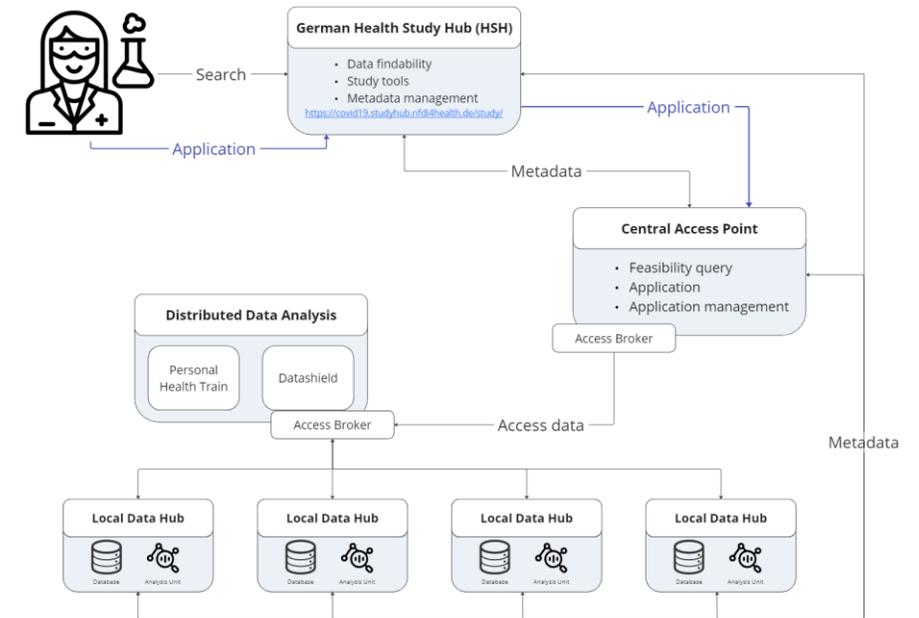


## Use Case Klinische Daten

- Unterstützung bei Teilen von **Daten klinischer Studien** durch die Einrichtung von Local Data Hubs (LDH)
- Unterstützung bei Harmonisierung von **Variablen klinischer Studien** durch Einrichtung eines Common Data Elements Katalogs



- Einrichtung eines **zentralen Zugangspunkts**, der als Einstiegspunkt für alle in NFDI4Health verfügbaren Daten dient
- Datenanalysten können ihre **Datenzugriffsanfragen über zentralen Zugangspunkt** stellen
- **Anfragen werden** an datenhaltende Organisationen weitergeleitet
- Basierend auf von Medizininformatik-Initiative (MII) entwickelten 'Forschungsdatenportal Gesundheit' (FDPG)



- Einrichtung eines **zentralen Zugangspunkts**, der als Einstiegspunkt für alle in NFDI4Health verfügbaren Daten dient
- Datenanalysten können ihre **Datenzugriffsanfragen über zentralen Zugangspunkt** stellen
- **Anfragen werden** an datenhaltende Organisationen weitergeleitet
- Basierend auf von Medizininformatik-Initiative (MII) entwickelten 'Forschungsdatenportal Gesundheit' (FDPG)
- Zwei Stakeholderworkshops mit breiter Beteiligung durchgeführt, um die Anforderungen für ein vereinfachtes einheitliches digitales Antragsmanagement



Workshop 05/2023: Modulare Antragsformulare für den Zugang zu förderiert organisierten Forschungsdaten

- Erstellung einer Infrastruktur für verteilte Analysen
  - für Forschung an sensiblen Daten ohne Übermittlung der Daten
- Bereitstellung von Tools für verteilte Durchführung von Analysen unter Wahrung der Privatsphäre
  - DataSHIELD und Personal Health Train: Algorithmen werden zu Daten gebracht und nicht umgekehrt
  - Dateneigentümer behalten Kontrolle über Daten behalten
- Implementierung für Use Cases Ernährungsepidemiologie/Epidemiologie chronischer Erkrankungen
  - Teilnahme von 17 deutsche epidemiologische Studien:
    - ❖ Eintrag im German Central Health Study Hub
    - ❖ Aufbau einer DataSHIELD Analyse Unit
    - ❖ Variablenharmonisierung für eine gemeinsame Datenanalyse

# Aktueller Stand Datenanalyse

➤ DataSHIELD Implementation und Harmonisierung aufwendiger als erwartet

Study name	Data Holding Organization	Co-applicant / Participant (C/P)	Financial compensation in 2022 (participants only)	Progress		
				Metadata collection	DataSHIELD	Data harmonisation
C				<div style="width: 75%; background-color: yellow;"></div>	<div style="width: 100%; background-color: green;"></div>	<div style="width: 0%; background-color: yellow;"></div>
C				<div style="width: 75%; background-color: yellow;"></div>		<div style="width: 0%; background-color: yellow;"></div>
L				<div style="width: 75%; background-color: yellow;"></div>		<div style="width: 0%; background-color: yellow;"></div>
K				<div style="width: 75%; background-color: yellow;"></div>		<div style="width: 0%; background-color: yellow;"></div>
E				<div style="width: 75%; background-color: yellow;"></div>		<div style="width: 10%; background-color: yellow;"></div>
C				<div style="width: 75%; background-color: yellow;"></div>	<div style="width: 5%; background-color: yellow;"></div>	<div style="width: 0%; background-color: yellow;"></div>
F				<div style="width: 100%; background-color: green;"></div>		
F				<div style="width: 75%; background-color: yellow;"></div>	<div style="width: 0%; background-color: yellow;"></div>	<div style="width: 0%; background-color: yellow;"></div>
K				<div style="width: 75%; background-color: yellow;"></div>	<div style="width: 100%; background-color: green;"></div>	<div style="width: 0%; background-color: yellow;"></div>
M				<div style="width: 10%; background-color: yellow;"></div>	<div style="width: 5%; background-color: yellow;"></div>	<div style="width: 0%; background-color: yellow;"></div>
A				<div style="width: 100%; background-color: green;"></div>	<div style="width: 100%; background-color: green;"></div>	<div style="width: 10%; background-color: yellow;"></div>
E				<div style="width: 0%; background-color: yellow;"></div>	<div style="width: 0%; background-color: yellow;"></div>	<div style="width: 0%; background-color: yellow;"></div>
E				<div style="width: 100%; background-color: green;"></div>	<div style="width: 100%; background-color: green;"></div>	<div style="width: 0%; background-color: yellow;"></div>
E				<div style="width: 100%; background-color: green;"></div>	<div style="width: 100%; background-color: green;"></div>	<div style="width: 10%; background-color: yellow;"></div>
I				<div style="width: 75%; background-color: yellow;"></div>	<div style="width: 50%; background-color: yellow;"></div>	<div style="width: 0%; background-color: yellow;"></div>
L				<div style="width: 0%; background-color: yellow;"></div>	<div style="width: 0%; background-color: yellow;"></div>	<div style="width: 0%; background-color: yellow;"></div>
S				<div style="width: 75%; background-color: yellow;"></div>	<div style="width: 100%; background-color: green;"></div>	<div style="width: 0%; background-color: yellow;"></div>

= in progress     = completed

# White Paper Verbesserung des Record Linkage für die Gesundheitsforschung in Deutschland

Breite Beteiligung der Stakeholder bei der Erarbeitung des White Papers mit mehreren Workshops - Veröffentlichung für diese Woche geplant.

- Empfehlungen zur Verbesserung des Record Linkage
  - ❖ Einführung eines Unique Identifiers und Etablierung von bereichsspezifischen Pseudonymen
  - ❖ Einführung einer federführenden Aufsichtsbehörde inkl. Genehmigungskompetenz zur Schaffung von Rechtssicherheit
  - ❖ Etablierung einer dezentral-föderierten Forschungsdateninfrastruktur mit zentralen Komponenten für das Record Linkage
  - ❖ Schaffung der Voraussetzung für Anknüpfung an europäischen Raum für Gesundheitsdaten

# Einverständniserklärung für Record Linkage – Muster, Texte, Beispiele

<Studienlogo (falls vorhanden)>

<Logo der durchführenden Einrichtung/Institutslogo (falls vorhanden)>

Hinweis: Die Mustertexte können nicht ohne weiteres übernommen werden und bedürfen der Prüfung und konkreten Anpassung an den jeweiligen Kontext durch den zuständigen Datenschutzbeauftragten!

Legende: **gelb** markierten Textstellen in Winkelklammern müssen, falls nötig, Studienspezifika eingetragen werden. Die **blau** markierten Textstellen in eckigen Klammern dienen zur Erklärung bzw. zur Angabe von Beispielen und können bei Verwendung dieses Dokuments gelöscht werden. Der **rote Text** beinhaltet rechtliche Hinweise sowie Erläuterungen und kann später bei Nutzung der Mustertexte gelöscht werden.

### Einwilligungserklärung

zur Teilnahme an der Studie <Name der Studie> (<Akronym der Studie>)

**Einwilligungserklärung Studienteilnahme**

Ich stimme der **Teilnahme** an der Studie freiwillig zu. Ich weiß, dass ich diese Zustimmung jederzeit, ohne Angabe von Gründen und ohne Nachteile widerrufen kann.

**Einwilligungserklärungen Datenverarbeitungen**

Wichtig ist, dass die Einwilligungserklärung für die Datenverarbeitung nicht allgemein ausfällt, sondern jeden Datenverarbeitungsvorgang (Erhebung, Verknüpfung etc.) und dessen jeweiligen Verarbeitungszweck konkret benennt. Über die einzelnen Datenverarbeitungsvorgänge muss in dem Datenschutzhinweis informiert werden.

Ich stimme der **Datenverarbeitung** meiner im Rahmen der Studie freiwillig angegebenen und erhobenen Daten <(Daten)> **[diese sind konkret zu benennen, z.B. Daten aus dem Kontaktformular, Befragungsdaten (Hinweis: Gesundheitsdaten müssen aufgrund des Ausdrücklichkeitserfordernisses in Art. 9 DSGVO explizit benannt werden)]** auf Grundlage meiner Einwilligung (Art. 6 Abs. 1 lit. a DSGVO und Art. 9 Abs. 2 lit. a DSGVO) zu.

Ich möchte über Studienergebnisse informiert werden.

- Per Post
- Per E-Mail

Neben der Verwendung meiner <in dieser Studie erhobenen Daten> willige ich außerdem ein,

<dass meine <Krankenversicherungsnummer> [falls zutreffend] erhoben, gespeichert und im Rahmen dieser Studie genutzt werden darf.> [falls zutreffend]

<dass meine <zu verknüpfende Daten> [z.B. Sekundärdaten „Versichertendaten“] bei der <Einrichtung benennen> [z.B. Krankenkasse] angefragt und durch diese für die

Musterdokument V1.0      NFDI4Health      1. März 2022

<Studienlogo (falls vorhanden)>

<Logo der durchführenden Einrichtung/Institutslogo (falls vorhanden)>

Hinweis: Die Mustertexte können nicht ohne weiteres übernommen werden und bedürfen der Prüfung und konkreten Anpassung an den jeweiligen Kontext durch den zuständigen Datenschutzbeauftragten!

Legende: **gelb** markierten Textstellen in Winkelklammern müssen, falls nötig, Studienspezifika eingetragen werden. Die **blau** markierten Textstellen in eckigen Klammern dienen zur Erklärung bzw. zur Angabe von Beispielen und können bei Verwendung dieses Dokuments gelöscht werden. Der **rote Text** beinhaltet rechtliche Hinweise sowie Erläuterungen und kann später bei Nutzung der Mustertexte gelöscht werden.

### Studieninformation und Datenschutzhinweise nach Art. 13 und 14 DSGVO (Direkterhebung und Dritterhebung)

<Name der Studie> <Akronym der Studie>

Wir freuen uns über Ihre Teilnahme an der Studie. Nachfolgend möchten wir Sie über den konkreten Studienverlauf sowie über den Umgang mit Ihren Daten gemäß Artikel 13 und 14 Datenschutz-Grundverordnung (DSGVO) informieren.

#### Studieninformation

Welche Ziele verfolgt diese Studie?

- <Ziel 1>
- <Ziel 2>
- <Ziel 3>

#### Studien-Teilnahmebedingungen und Möglichkeit zum Widerruf

- <Beschreibung der Teilnahmebedingungen>
- <Hinweis zur Freiwilligkeit der Teilnahme>
- Sie können Ihre **Einwilligung zur Teilnahme** jederzeit und ohne Angabe von Gründen mündlich oder schriftlich, ganz oder teilweise widerrufen, ohne dass Ihnen hierdurch Nachteile entstehen. Bei Rücktritt von der Studie haben Sie das Recht, das Löschen bereits erhobener Daten zu fordern, sofern diese noch nicht anonymisiert sind.

#### Datenschutzhinweise

Welche Daten erheben wir mit welcher Rechtsgrundlage?

<Beispieltext für die Datenverarbeitung durch Einwilligung: Alle Angaben sind freiwillig. Die Datenverarbeitung erfolgt auf Grundlage Ihrer Einwilligung gemäß Artikel 6 Absatz 1 Buchstabe a DSGVO und, sofern es sich um besondere Kategorien personenbezogener Daten handelt, gemäß Artikel 9 Absatz 2 Buchstabe a DSGVO. Ihre Einwilligung kann jederzeit mit Wirkung für die Zukunft widerrufen werden, ohne dass dies die Rechtmäßigkeit der bisher erfolgten Verarbeitung berührt.> [falls zutreffend]

Musterdokument V1.0      NFDI4Health      1. März 2022

# Guideline zur Verarbeitung personenbezogener Daten zu Forschungszwecken



## 2. Guideline: Was muss ich bei der Verarbeitung personenbezogener Daten zu Forschungszwecken beachten?

### Gliederung der Guideline:

- I. Findet eine Datenverarbeitung im Sinne der DS-GVO statt?
- II. Werden hierbei personenbezogene Daten verarbeitet?
- III. Werden die Daten pseudonymisiert oder anonymisiert verarbeitet?
- IV. Handelt es sich bei den verarbeiteten personenbezogenen Daten um einfache Daten oder besondere Kategorien personenbezogener Daten?
- V. Werden die Verarbeitungsgrundsätze für den Umgang mit personenbezogenen Daten nach der DS-GVO beachtet?
- VI. Welche Personen/Einrichtungen sind an der Datenverarbeitung beteiligt?
- VII. Erfolgt im Rahmen der Einbeziehung Dritter eine Datenübermittlung in Länder außerhalb der EU/des EWR ohne Angemessenheitsbeschluss?
- VIII. Welche Rechtsgrundlagen sind zur Legitimation der Verarbeitung der personenbezogenen Daten im Forschungskontext heranzuziehen?
- IX. Werden betroffene Personen gem. Art. 13 und 14 DS-GVO transparent über die Datenverarbeitung informiert?
- X. Werden die datenschutzrechtlichen Betroffenenrechte berücksichtigt und eingehalten?
- XI. Werden die technischen und organisatorischen Maßnahmen zur Datensicherheit angemessen umgesetzt?
- XII. Muss eine Datenschutzfolgenabschätzung gem. Art. 35 DS-GVO durchgeführt werden?
- XIII. Ist im Rahmen der Datenverarbeitung eine Datenpanne erfolgt?
- XIV. Gibt es ein Konzept zur Speicherung und Aufbewahrung bzw. Löschung von Daten nach Ablauf ihrer Lebensdauer?

### I. Findet eine Datenverarbeitung im Sinne der DS-GVO statt?

#### ➤ Definition:

Der sachliche Anwendungsbereich der DS-GVO ist gem. Art. 2 eröffnet, wenn eine ganz oder teilweise automatisierte Verarbeitung personenbezogener Daten erfolgt. Eine Verarbeitung im Sinne der DS-GVO umfasst jede Handlung im Zusammenhang mit personenbezogenen Daten, namentlich verändern, speichern, übermitteln, erheben, löschen, verknüpfen. Der Begriff der Datenverarbeitung ist somit weit gefasst und es kommt nicht auf den Umfang oder die Dauer einer Datenverarbeitung an. [Erläuterung in D6.2 Gliederungspunkt B.I.]

- Ja  
 Nein

### II. Werden hierbei personenbezogene Daten verarbeitet?

#### ➤ Definition:

Personenbezogene Daten sind gem. Art. 4 Nr. 1 DS-GVO alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person, die sog. „betroffene Person“, beziehen. Zu unterscheiden ist somit zwischen der Datenverarbeitung basierend auf personenidentifizierenden Merkmalen und der Datenverarbeitung, die keine personenidentifizierenden Merkmale verwendet. Identifizierbar ist eine Person gem. Art. 4 Nr. 1 DS-GVO dann, wenn sie direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung (Namen, Kennnummer etc.) oder sonstigen besonderen Merkmalen, wozu auch physische, physiologische, genetische, psychische und soziale Merkmale gehören können, identifiziert werden kann. Insbesondere dann, wenn miteinander kombinierte Informationen den Rückschluss auf eine Einzelperson zulassen, kann von personenbezogenen Daten gesprochen werden. [Erläuterung in D6.2 Gliederungspunkt B.II.]

- Ja  
 Nein

### III. Werden die Daten pseudonymisiert oder anonymisiert verarbeitet?



Die Prüfung, ob eine hinreichende Pseudonymisierung oder Anonymisierung vorliegt, sollte durch einen Datenschutzbeauftragten und ggf. unter Einbeziehung der Aufsichtsbehörden erfolgen.

#### ➤ Definition:

Die Verarbeitung personenbezogener Daten zu wissenschaftlichen Forschungszwecken unterliegt gem. Art. 89 Abs. 1 S. 1 DS-GVO „geeigneten Garantien für die Rechte und Freiheiten der betroffenen Person“. Der Grundsatz der Datenminimierung sowie die Wahrung der Rechte und Freiheiten betroffener Personen sollen durch technische und organisatorische Maßnahmen sichergestellt werden. Im Hinblick auf diese Maßnahmen nennt Art. 89 Abs. 1 S. 3 DS-GVO die Pseudonymisierung. Die Anonymisierung wird in dieser Vorschrift nicht explizit genannt, der Erwägungsgrund 26 der Verordnung legt dahingehend aber fest, dass die DS-GVO nicht für „anonyme Informationen“ gilt.

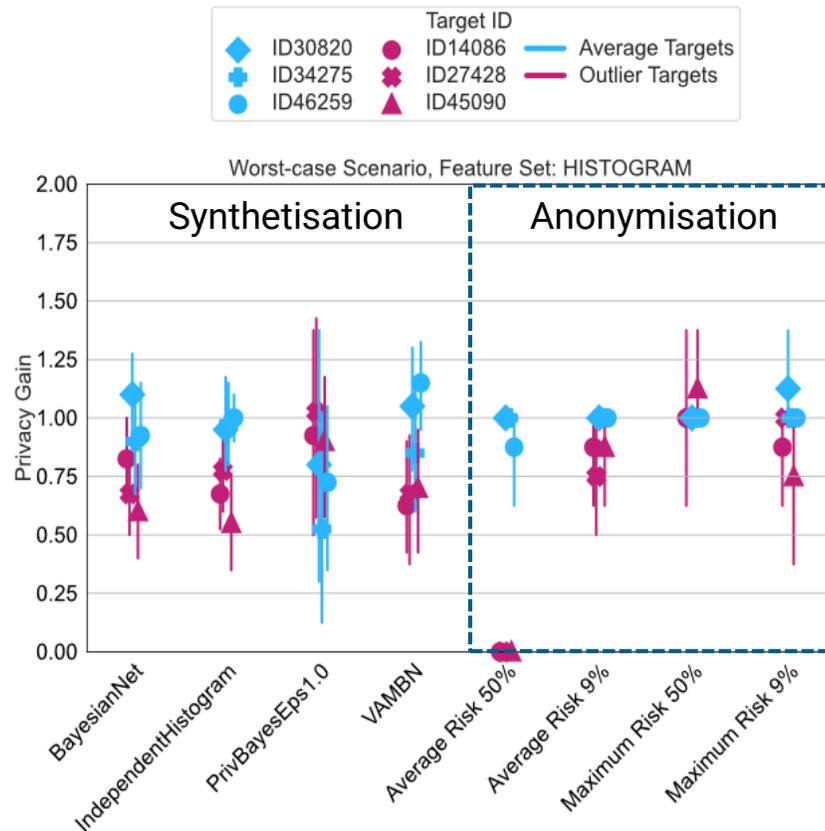
#### 1. Pseudonymisierung

#### ➤ Definition:

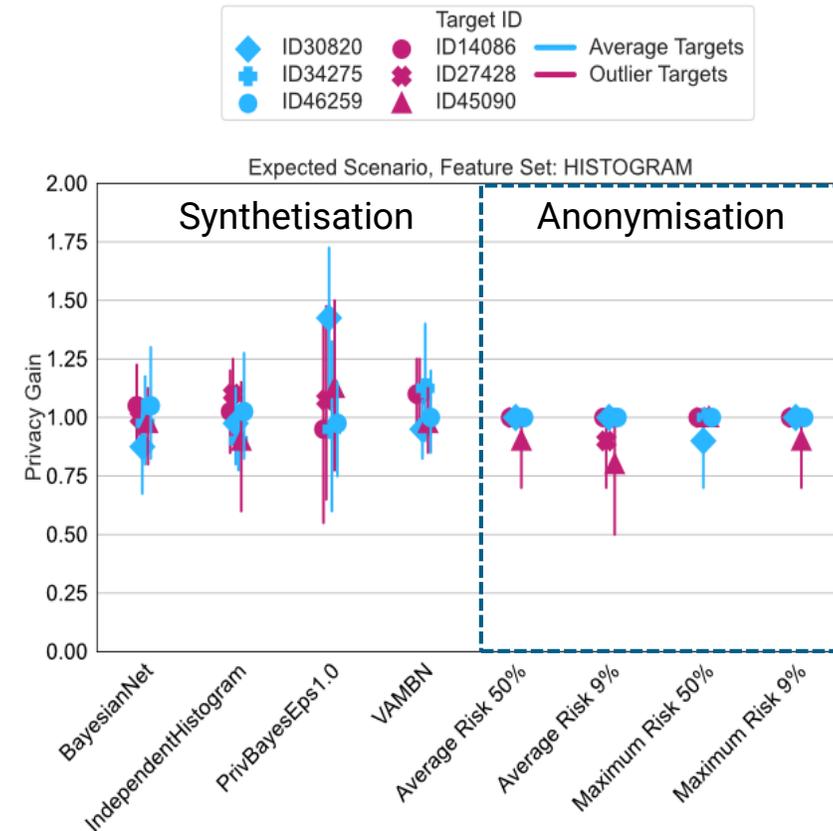
Die Pseudonymisierung von personenbezogenen Daten ist in Art. 4 Nr. 5 DS-GVO legaldefiniert als die Verarbeitung personenbezogener Daten in der Weise, dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können. Dazu müssen die

# Analyse der Re-Identifikationsrisiken synthetischer Daten

- Laufende Zusammenarbeit mit Use Cases zum Vergleich der verbleibenden Risiken bei anonymisierten und synthetischen Daten
- Erste Ergebnisse basierend auf Krebsregisterdaten



Ohne adverse Modellierung



Mit adverbser Modellierung

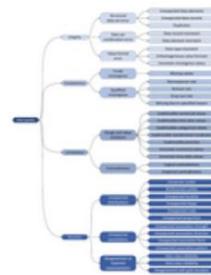
## Standards und Tools zur Beurteilung der Datenqualität in epidemiologischen Studien



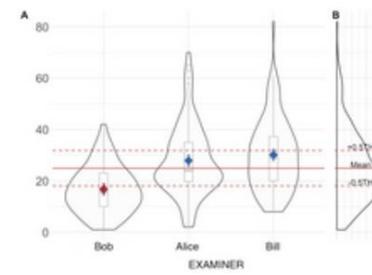
About Data Quality



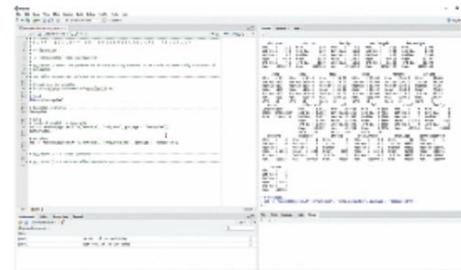
Tutorials



Data Quality Concept



Software



Publications & Activities



Literature

# Konsortium NFDI4Health

- Forschungsinstitute involviert in Datenerhebung und -analyse personenbezogener Gesundheitsdaten



- Infrastruktureinrichtungen / Expertise in Entwicklung von Standards und Methoden



- Institutionen mit Fokus auf Einbezug der Nutzer-Community



- 47 namhafte Institutionen als Beteiligte

# Questions?



Contact:

[contact@nfdi4health.de](mailto:contact@nfdi4health.de)

[www.nfdi4health.de](http://www.nfdi4health.de)

Gefördert durch  
**DFG** Deutsche  
Forschungsgemeinschaft

