

# NFDI4Health

## The National Research Data Infrastructure for Personal Health Data

Juliane Fluck

NFDI4Health Spokesperson

University of Bonn

2023-01-17

# What is the mission of NFDI?

## Bottom up approach to build NFDI based on scientific communities

“ The aim of the National Research Data Infrastructure (NFDI) is to

- **systematically manage scientific and research data,**
  - **provide long-term storage, backup and accessibility and**
  - **network the data both nationally and internationally. “\***
- 
- Final expansion stage 27 consortia (start 1.round 10/20 and 3. round 03/23)  
26 domain specific consortia + Base4NFDI for developing basic services for NFDI

Funding secured for the first 10 (5+5) years

Long-term establishment of NFDI as German contribution to the European infrastructure, the European Open Science Cloud (EOSC) envisaged.

\*Quote taken from the DFG NFDI Website: <https://www.dfg.de/foerderung/programme/nfdi/>

## Sections

Common Infrastructures

Ethical, Legal & Social Aspects (Co-Lead U. Sax, NFDI4Health)

Training & Education

Section (Meta)data, Terminologies & Provenance

↑ ↑  
**19 First and Second-  
round Consortia**

↑  
**+7 third-round  
Consortia**

**2023  
+ Base4NFDI**



# Mission of NFDI4Health

Our mission:

- to **increase the value of research** in epidemiology, public health, and clinical trial based medicine
- by making high quality data internationally accessible according to **FAIR principles**

**Vision:**

to boost scientific exploitation of personal health data

# NFDI4Health consortium

- Research institutes involved in data collection and analysis of personal health data



- Infrastructures / experts in development of standards and methods



- Institutions with focus on involvement of the user community



- 47 renowned institutions as participants

# NFDI4Health consortium

- Research institutes involved in data collection and analysis of personal health data

Prof. Nöthlings,  
Donald Study



- Infrastructures / experts in development of standards and methods

Prof. Fluck,  
Joint working group U Bonn

Prof. Fröhlich,  
Synthetic data generation



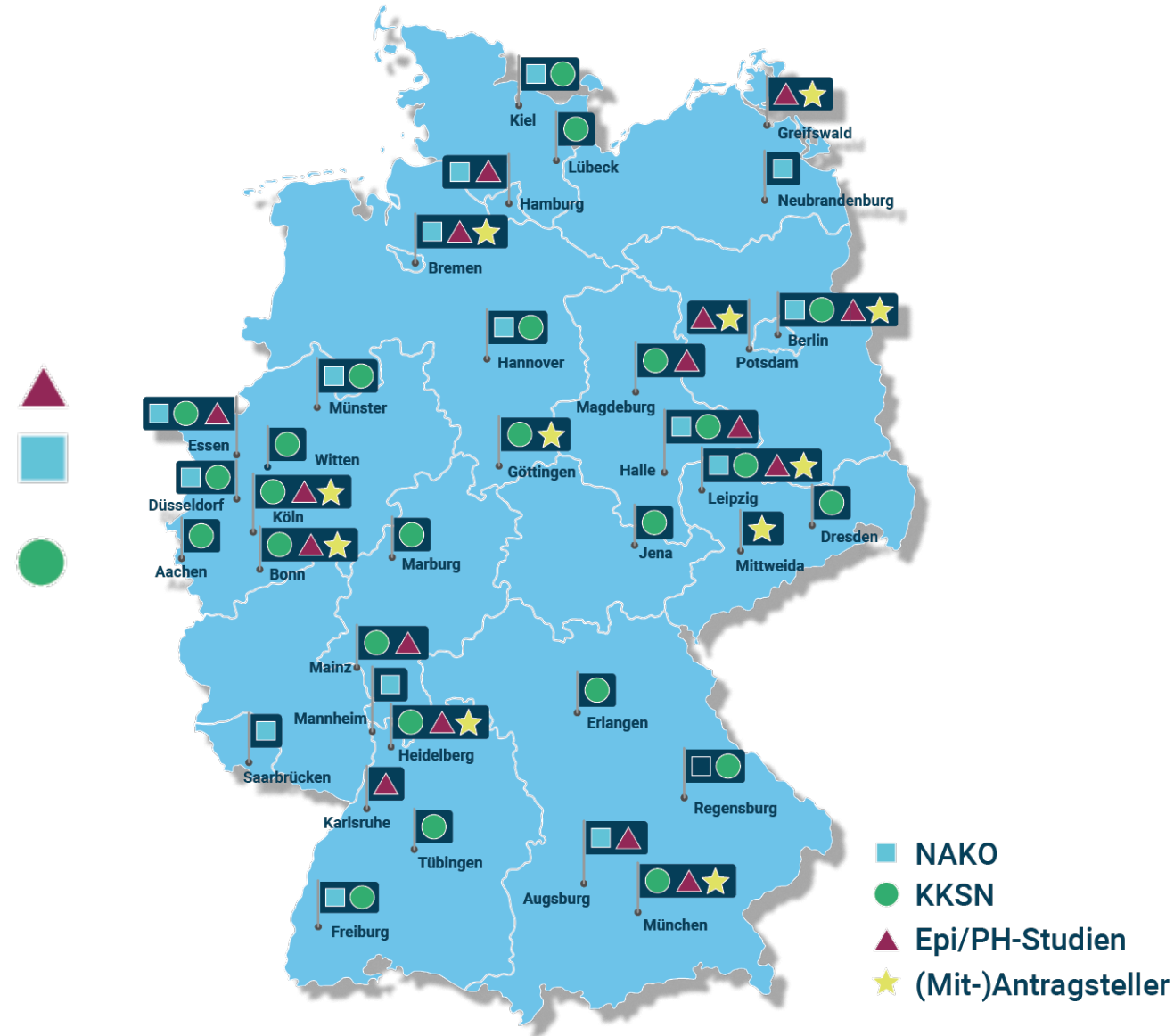
# Data body in NFDI4Health

## Primary focus

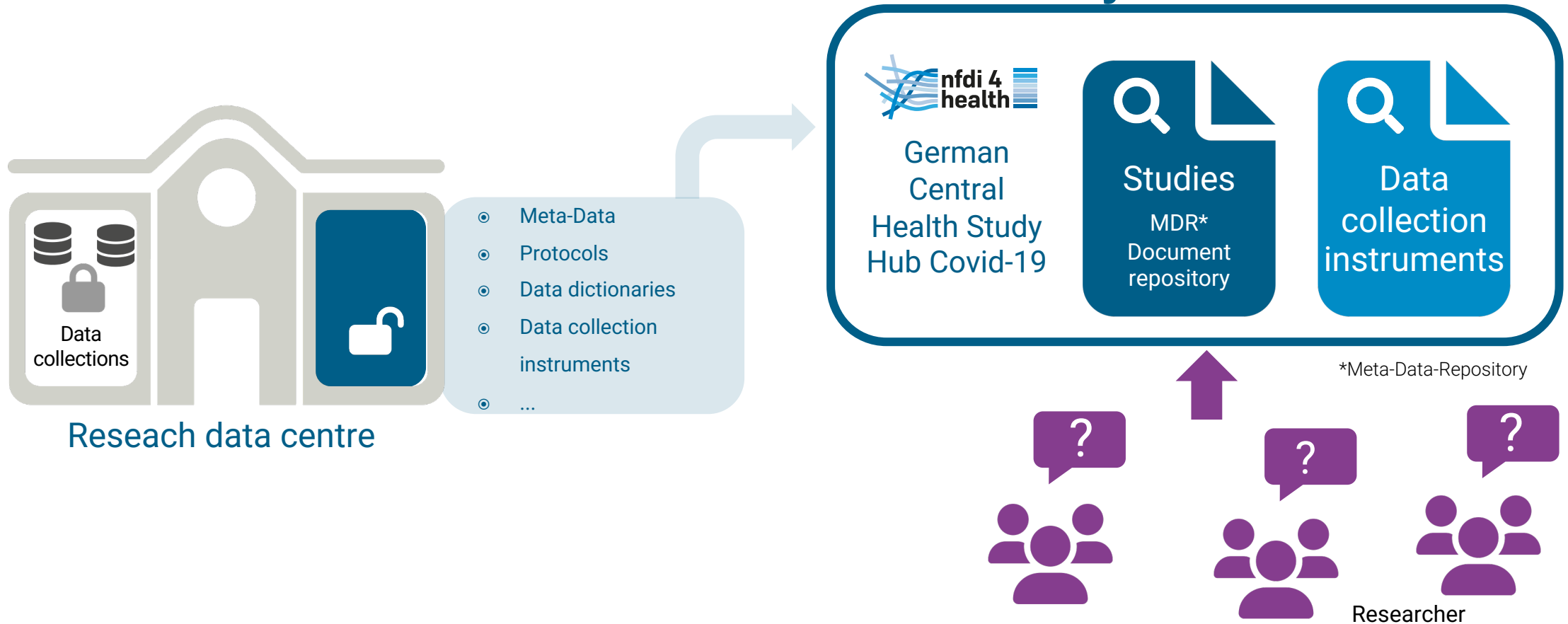
- Epidemiological and public health studies
  - 26 local studies with > 400.000 participants
    - German National Cohort
- Clinical studies
  - 27/ university study centres

## Secondary focus

- Health registries
- Administrative health databases



# NFDI4Health – Solutions: 1. Central Findability of Studies



<https://covid19.studyhub.nfdi4health.de/>



# German Central Health Study Hub Covid-19



unified simple UI



Backend



**Begin your search. 1305 studies, instruments and other documents are there to explore**

🔍 Type to search

Explore

The following syntax is supported:

- whitespace as split characters, each token is combined with a disjunction.
- + signifies AND operation
- | signifies OR operation
- - negates a single token
- " wraps a number of tokens to signify a phrase for searching

📄 Your data is not yet displayed?

Is your study not displayed yet? Please contact us via [studyregistration@nfdi4health.de](mailto:studyregistration@nfdi4health.de) and we will assist you with the publication of your research. Please note that study documents, including data collection instruments, can be published as well – as supplementary materials to your study or as independent research outputs. For further details please look at our [publication policy](#) and [flyer here](#).

**1259 Studies**

Explore Studies →

**27 Data dictionary**

Explore Data dictionary →

<https://covid19.studyhub.nfdi4health.de/>

# German Central Health Study Hub Covid-19



unified simple UI



Backend



**Begin your search. 1305 studies, instruments and other documents are there to explore**

🔍 Type to search

Explore

The following syntax is supported:

- whitespace as split characters, each token is combined with a disjunction.
- + signifies AND operation
- | signifies OR operation
- - negates a single token
- " wraps a number of tokens to signify a phrase for searching

📄 Your data is not yet displayed?

Is your study not displayed yet? Please contact us via [studyregistration@nfdi4health.de](mailto:studyregistration@nfdi4health.de) and we will assist you with the publication of your research. Please note that study documents, including data collection instruments, can be published as well – as supplementary materials to your study or as independent research outputs. For further details please look at our [publication policy](#) and [flyer here](#).

**1259 Studies**

Explore Studies →

**27 Data dictionary**

Explore Data dictionary →

- Aggregation of studies with German contributions from
  - ❖ Clinical Trial Registers:  
DRKS, WHO, ICTRP and [clinicaltrials.gov](https://clinicaltrials.gov)
  - ❖ Epidemiological studies:  
Manually collected studies from data holding organisations and data stewards
- API and UI to deposit data, search and explore

<https://covid19.studyhub.nfdi4health.de/>

# German Central Health Study Hub Covid-19

**Refine your search.**Currently 1259 studies, instruments and other documents are matched

🔍 Type to search [Explore](#) [Reset](#)

The following syntax is supported:

- whitespace as split characters, each token is combined with a disjunction.
- **+** signifies AND operation
- **|** signifies OR operation
- **-** negates a single token
- **"** wraps a number of tokens to signify a phrase for searching

Type of the resource

Study

Primary design of the study

Specific type of the study

1259

444

686

4

12

12

14

23

105

1

2

3

4

5

...

126

Sort by: relevance, ascending

10 documents per page

P-pVAC-SARS-CoV-2: Phase I Single-center Safety and Immunogenicity Trial of Multi-peptide Vaccination to Prevent COVID-19 Infection in Adults

ID of the resource used for identification within NFDI4Health: [NCT04546841](#)

Type of the resource: Study

Data Source: Automatically uploaded: ClinicalTrials.gov

Overall status for the study: Completed: Recruitment, data collection, and data quality management completed normally

Description of the resource

Safety and Effectiveness Evaluation of Seraph 100 Microbind Affinity Blood Filter (Seraph 100) in the Treatment of Patients With COVID-19

ID of the resource used for identification within NFDI4Health: [NCT04547357](#)

Type of the resource: Study

Data Source: Automatically uploaded: ClinicalTrials.gov

<https://covid19.studyhub.nfdi4health.de/>

Workshop: Strategies for Data Science and Data Management

# German Central Health Study Hub Covid-19

## COVID-GNC survey

### General Information

#### Title of the resource

**DE (German)** NAKO COVID-19-Sonderbefragung

#### Description of the resource

**EN (English)** The supplemental COVID-19 survey conducted among all NAKO participants is implemented as part of the NAKO health study (NAKO) in order to gain evidence about the spread, course, and consequences of COVID-19 in Germany. To receive the instrument please refer to the contact person of NAKO.

**DE (German)** Die COVID-19-Befragung aller TeilnehmerInnen ist eine Forschungsaktion der NAKO Gesundheitsstudie, um Erkenntnisse über Verbreitung, Verlauf, und Auswirkungen von COVID-19 in Deutschland zu gewinnen.

#### Details

**Type of the resource** ⓘ Study

**A more general term describing the resource type** ⓘ  
Other

**Acronym of the resource** ⓘ NAKO COVID-19

**Key word describing the resource** ⓘ COVID-19, Mental Health, Health Status

**Primary language of the resource** ⓘ DE (German)

**Web page of the resource** ⓘ [Visit Webpage](#) ↗

**Data Source** ⓘ Manually collected

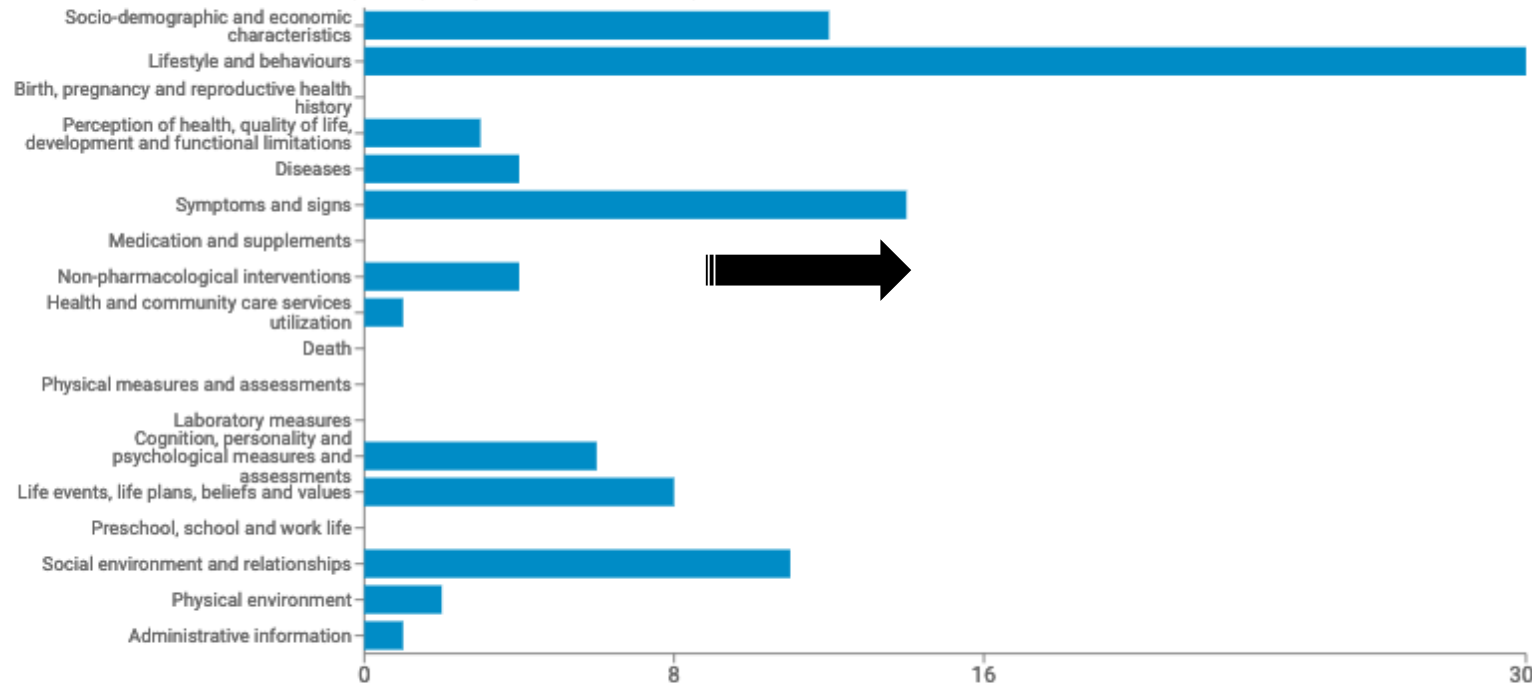
**License defining the rights to re-use the resource** ⓘ Not applicable

<https://covid19.studyhub.nfdi4health.de/>

# Machine readable data collection instruments

## Variables Classification

Classification into areas of information (developed by Maelstrom Research)

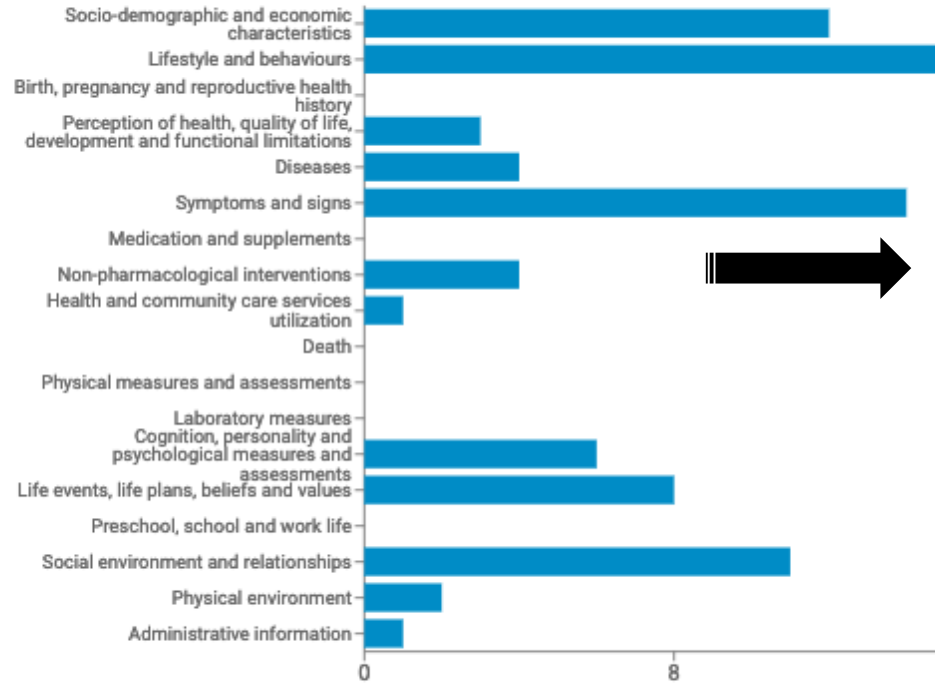


[Explore all Variables](#)

# Machine readable data collection instruments

## Variables Classification

Classification into areas of information (developed by Maelstrom Research)



[Explore all Variables](#)

nfdi 4 health Mica


Home Repository Search Q

Search Criteria



- ☒ Variables
- ☐ Areas of Information
- ☐ Source & target
- ☐ Properties
- ☒ Datasets
- ☐ Properties
- ☒ Studies
- ☐ Properties

<input type="checkbox"/>	Name	Label	Value type	Annotations	Study
<input type="checkbox"/>	q10g	Warum wurden Sie auf das Corona-Virus getestet? Ich hatte zu diesem Zeitpunkt Symptome, die auf eine Infektion mit dem Corona-Virus hinwiesen.	Integer	Symptoms and signs involving the circulatory and respiratory systems (R00-R09)	NAKO
<input type="checkbox"/>	q10ga	Wenn ja, welche Symptome? Abgeschlagenheit	Integer	General symptoms and signs (R50-R69)	NAKO
<input type="checkbox"/>	q10gb	Wenn ja, welche Symptome? Übelkeit	Integer	Symptoms and signs involving the digestive system and abdomen (R10-R19)	NAKO
<input type="checkbox"/>	q10gc	Wenn ja, welche Symptome? Gliederschmerzen	Integer	General symptoms and signs (R50-R69)	NAKO
<input type="checkbox"/>	q10gd	Wenn ja, welche Symptome? Durchfall	Integer	Symptoms and signs involving the digestive system and abdomen (R10-R19)	NAKO
<input type="checkbox"/>	q10ge	Wenn ja, welche Symptome? Atemprobleme	Integer	Symptoms and signs involving the circulatory and respiratory systems (R00-R09)	NAKO
<input type="checkbox"/>	q10gf	Wenn ja, welche Symptome? Fieber	Integer	General symptoms and signs (R50-R69)	NAKO
<input type="checkbox"/>	q10gg	Wenn ja, welche Symptome? Husten	Integer	Symptoms and signs involving the circulatory and respiratory systems (R00-R09)	NAKO
<input type="checkbox"/>	q10gh	Wenn ja, welche Symptome? Riechstörung	Integer	Symptoms and signs involving the circulatory and respiratory systems (R00-R09)	NAKO
<input type="checkbox"/>	q10gi	Wenn ja, welche Symptome? Kopfschmerzen	Integer	General symptoms and signs (R50-R69)	NAKO
<input type="checkbox"/>	q10gj	Wenn ja, welche Symptome? Schüttelfrost	Integer	General symptoms and signs (R50-R69)	NAKO

# Machine readable data collection instruments

 Mica
 

[Home](#)
[Repository](#)



[Sign in](#)
[Sign up](#)

Search Criteria

Variables

Areas of Information

Source & target

Properties

Datasets

Properties


Studies

Properties

<input type="checkbox"/>	Name	Label	Value type	Annotations	Study	Instrument
<input type="checkbox"/>	q10g	Warum wurden Sie auf das Corona-Virus getestet? Ich hatte zu diesem Zeitpunkt Symptome, die auf eine Infektion mit dem Corona-Virus hinwiesen.	Integer	Symptoms and signs involving the circulatory and respiratory systems (R00-R09)	NAKO	NAKO
<input type="checkbox"/>	q10ga	Wenn ja, welche Symptome? Abgeschlagenheit	Integer	General symptoms and signs (R50-R69)	NAKO	NAKO
<input type="checkbox"/>	q10gb	Wenn ja, welche Symptome? Übelkeit	Integer	Symptoms and signs involving the digestive system and abdomen (R10-R19)	NAKO	NAKO
<input type="checkbox"/>	q10gc	Wenn ja, welche Symptome? Gliederschmerzen	Integer	General symptoms and signs (R50-R69)	NAKO	NAKO
<input type="checkbox"/>	q10gd	Wenn ja, welche Symptome? Durchfall	Integer	Symptoms and signs involving the digestive system and abdomen (R10-R19)	NAKO	NAKO
<input type="checkbox"/>	q10ge	Wenn ja, welche Symptome? Atemprobleme	Integer	Symptoms and signs involving the circulatory and respiratory systems (R00-R09)	NAKO	NAKO
<input type="checkbox"/>	q10gf	Wenn ja, welche Symptome? Fieber	Integer	General symptoms and signs (R50-R69)	NAKO	NAKO
<input type="checkbox"/>	q10gg	Wenn ja, welche Symptome? Husten	Integer	Symptoms and signs involving the circulatory and respiratory systems (R00-R09)	NAKO	NAKO
<input type="checkbox"/>	q10gh	Wenn ja, welche Symptome? Riechstörung	Integer	Symptoms and signs involving the circulatory and respiratory systems (R00-R09)	NAKO	NAKO
<input type="checkbox"/>	q10gi	Wenn ja, welche Symptome? Kopfschmerzen	Integer	General symptoms and signs (R50-R69)	NAKO	NAKO
<input type="checkbox"/>	q10gj	Wenn ja, welche Symptome? Schüttelfrost	Integer	General symptoms and signs (R50-R69)	NAKO	NAKO



# Machine readable data collection instruments

 Mica

Home Repository Search

Sign in Sign up

Search Criteria

Variables

Areas of Information

Source & target

Properties

Datasets

Properties

Studies

Properties

<input type="checkbox"/>	Name	Label	Value type	Annotations	Study	Instrument
<input type="checkbox"/>	q10g	Warum wurden Sie auf das Corona-Virus getestet? Ich hatte zu diesem Zeitpunkt Symptome, die auf eine Infektion mit dem Corona-Virus hinwiesen.	Integer	Symptoms and signs involving the circulatory	NAKO	NAKO
<input type="checkbox"/>	q10ga	Wenn ja, welche Symptome? Abgeschlagenheit				
<input type="checkbox"/>	q10gb	Wenn ja, welche Symptome? Übelkeit				
<input type="checkbox"/>	q10gc	Wenn ja, welche Symptome? Gliederschmerzen				
<input type="checkbox"/>	q10gd	Wenn ja, welche Symptome? Durchfall				
<input type="checkbox"/>	q10ge	Wenn ja, welche Symptome? Atemprobleme				
<input type="checkbox"/>	q10gf	Wenn ja, welche Symptome? Fieber				
<input type="checkbox"/>	q10gg	Wenn ja, welche Symptome? Husten				
<input type="checkbox"/>	q10gh	Wenn ja, welche Symptome? Riechstörung				
<input type="checkbox"/>	q10gi	Wenn ja, welche Symptome? Kopfschmerzen				
<input type="checkbox"/>	q10gj	Wenn ja, welche Symptome? Schüttelfrost				

Dataschema Variable / q10gb

Wenn ja, welche Symptome? Übelkeit

Overview

Value type

Integer

Nature

Categorical

Entity type

Participant

Categories

Name	Label	Missing
0	Nein	
1	Ja	

Annotations

Source

Questionnaire

Target

Participant

Symptoms and signs

Symptoms and signs involving the digestive system and abdomen (R10-R19)

Definition

Protocol

NAKO

Initiative

NAKO

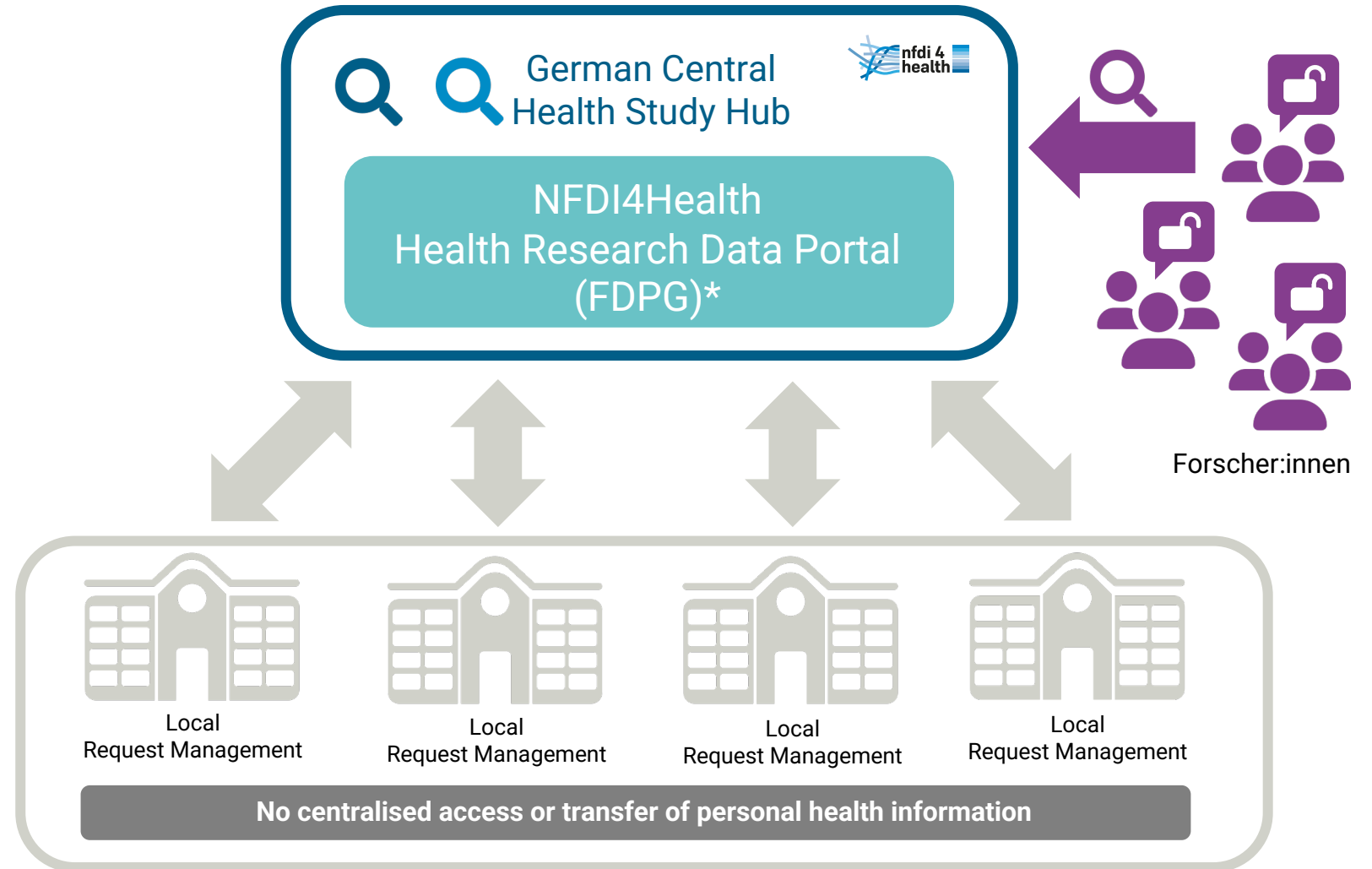
Population

Die NAKO Gesundheitsstudie-COVID

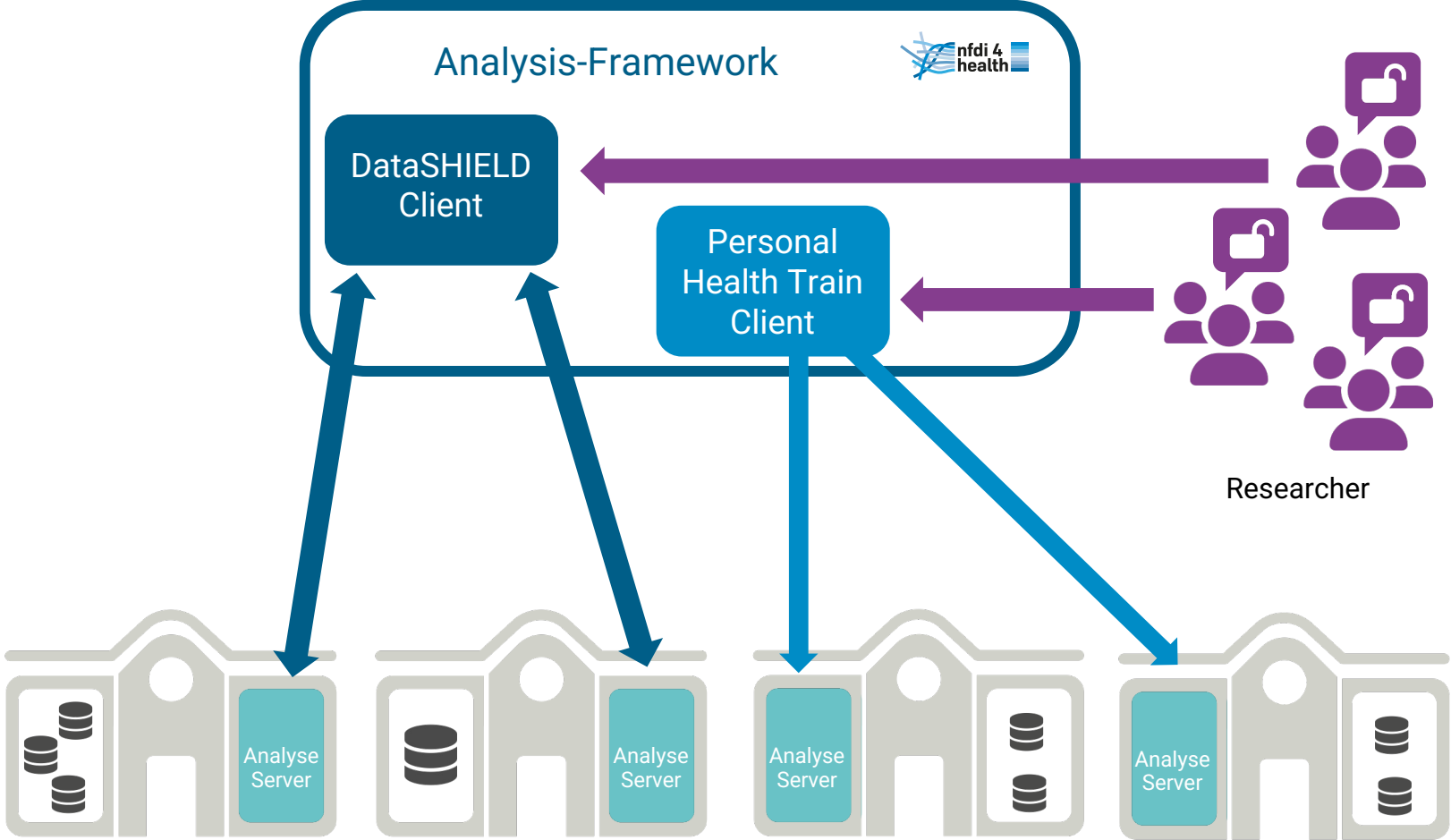


# Data Access

In the  
specification phase

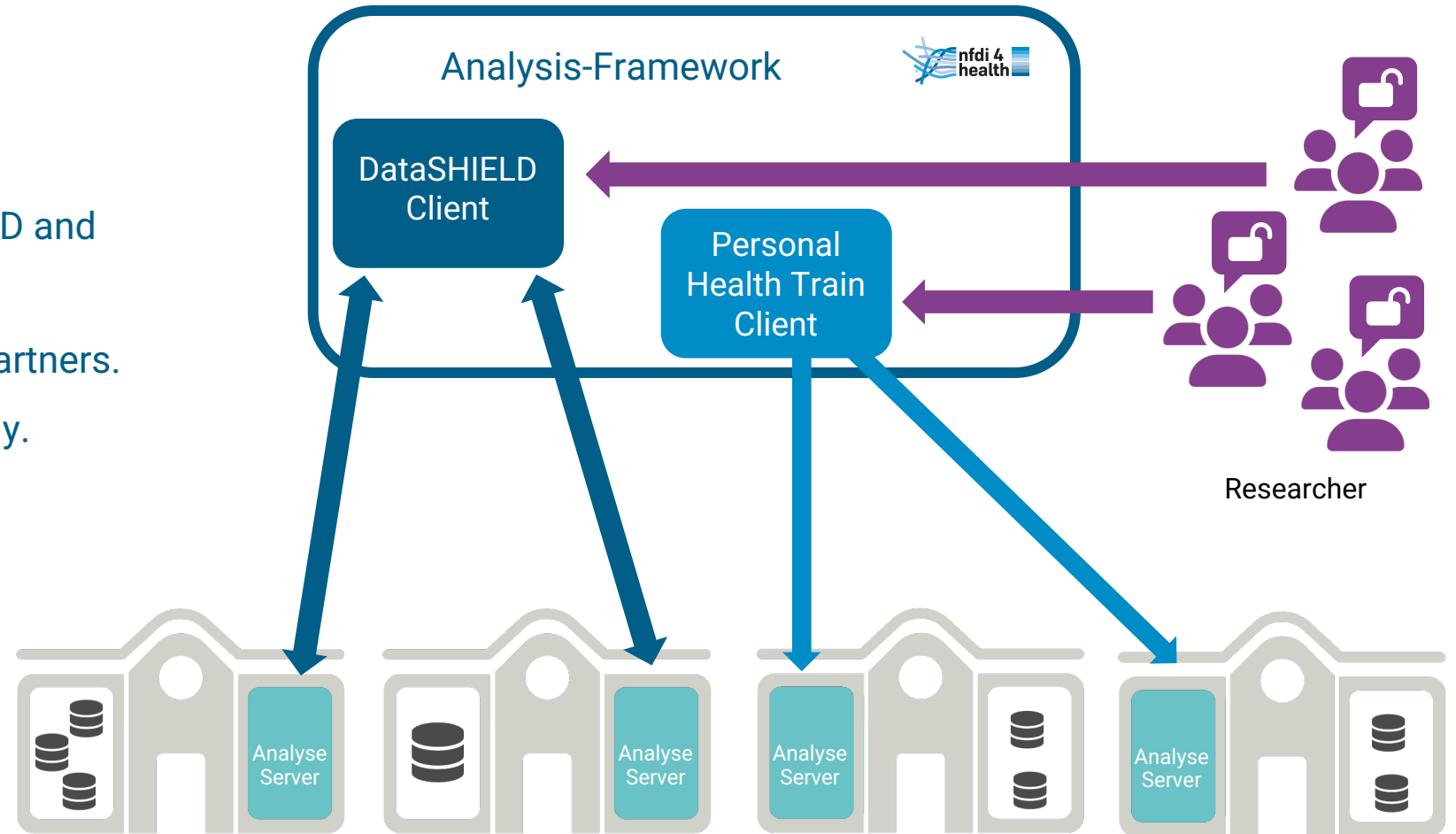


# Enabling automatic analysis in distributed analysis platform



# Enabling automatic analysis in distributed analysis platform

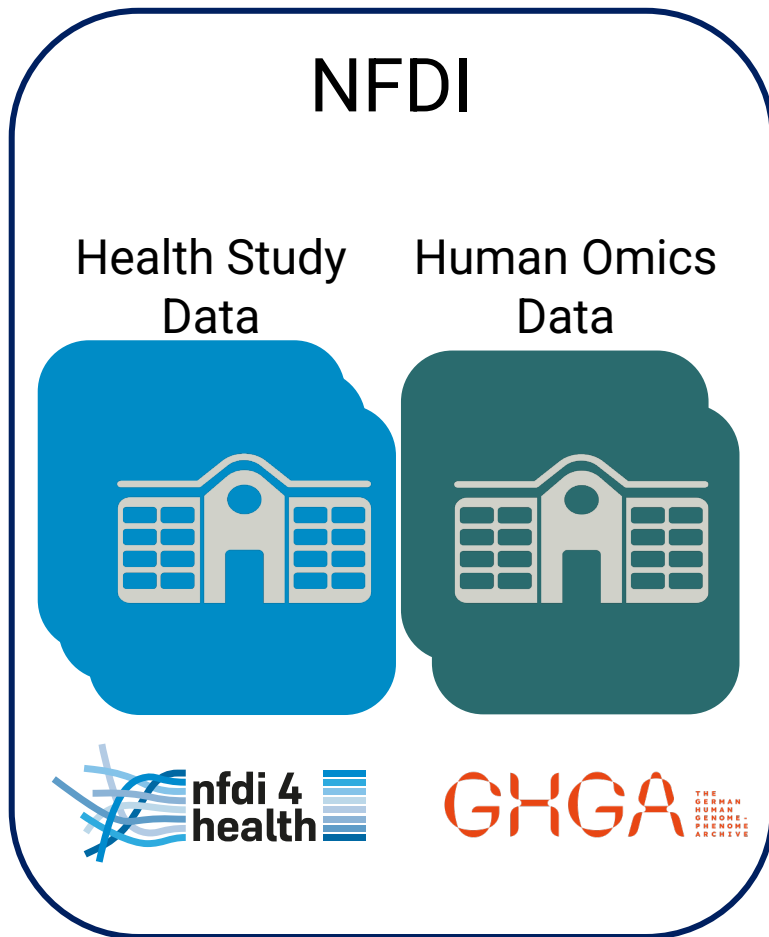
- ◉ Central Client-Server for DataSHIELD and Personal Health Train (PHT).
- ◉ Application-related integration of partners.
- ◉ Partner select and share data locally.
- ◉ Data harmonisation necessary.



# Many processes are needed to support distributed analysis

- **Harmonization of data and information models**
  - Core data sets
  - Standards: e.g. SNOMED, LOINC or FOODEX
  - Interfaces, e.g. FHIR
- **Tools supporting standards**
  - Terminology Service for data annotation, search and queries
  - Metadata Annotation Workbench for semi-automated annotation support
- **Data quality standards and tools**
- **Local data hubs**
- **Legal requirements addressing data privacy**
- **Tools for accessing data privacy**
- **Synthetic data generation**

# German Landscape Health Research Data Infrastructures



# German Landscape Health Research Data Infrastructures

## NFDI

Health Study  
Data



Human Omics  
Data



Differences: Content | Responsibilities | Legal Regulations

Health Insurance  
Data



Health Data Lab



Federal Institute  
for Drugs  
and Medical Devices

Health Registry  
Data



> 300

German Centre for  
cancer registry  
ROBERT KOCH INSTITUT

Health Care  
Data

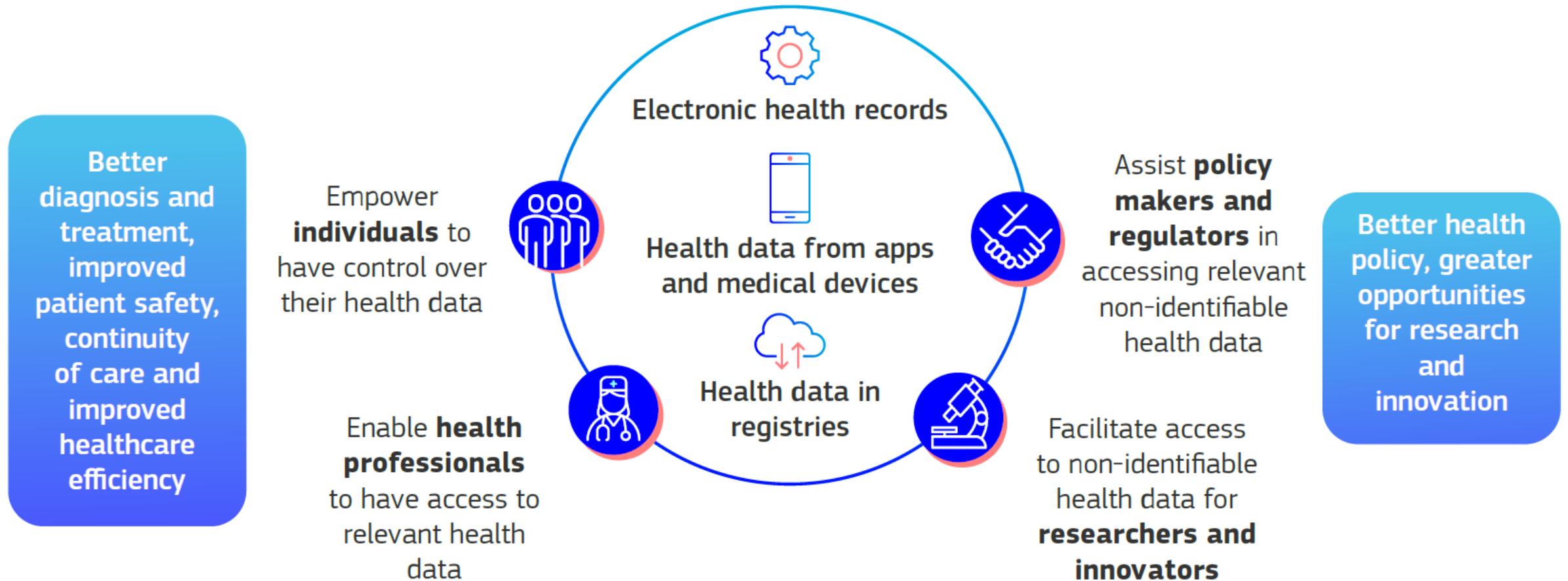


33 Data integration  
centres



+ many research related infrastructures , e.g. University Medicine Network (NUM), German Centers for Health Research (DZGs), Helmholtz Health.

# European Health Data Space



May 3, 2022: [https://ec.europa.eu/commission/presscorner/detail/en/fs\\_22\\_2713](https://ec.europa.eu/commission/presscorner/detail/en/fs_22_2713)

# Questions?



## Kontakt:

**Prof. Dr. Juliane Fluck**

fluck@zbmed.de

**ZB MED – Information centre for  
Life Sciences**

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

Gefördert durch  
**DFG** Deutsche  
Forschungsgemeinschaft



**Acknowledgments:** This work has been developed as part of the NFDI4Health and the NFDI4Health Task Force COVID-19 ([www.nfdi4health.de](http://www.nfdi4health.de)).  
We thank the German Research Foundation (DFG) for the financial support– project numbers 451265285 und 442326535.

