How the HDL is enabling health data research with AI
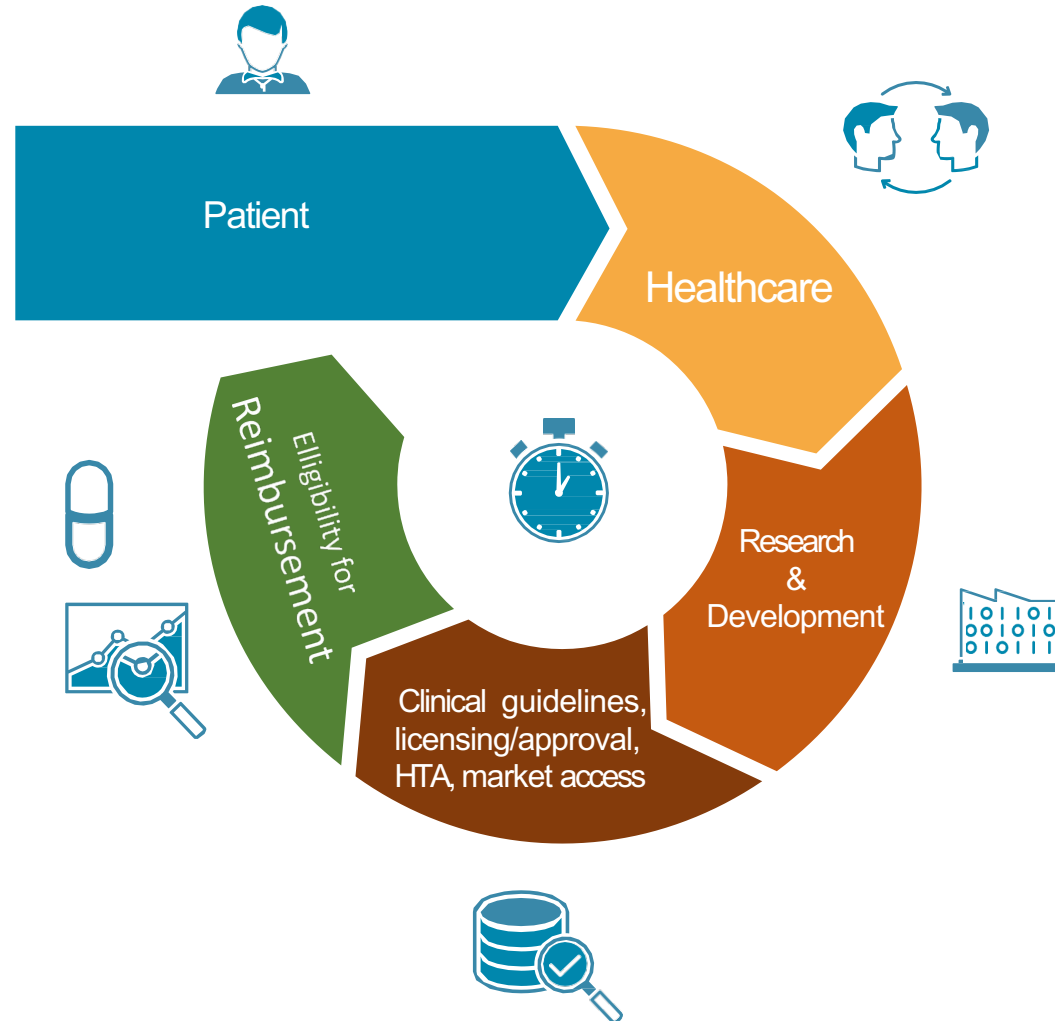
Steffen Heß

30th Nov. 2023, London

Synthetic Data Summit 2023

How can we increase the impact on individual health?

Personal communication N. Schneider

Health Data Lab at BfArM

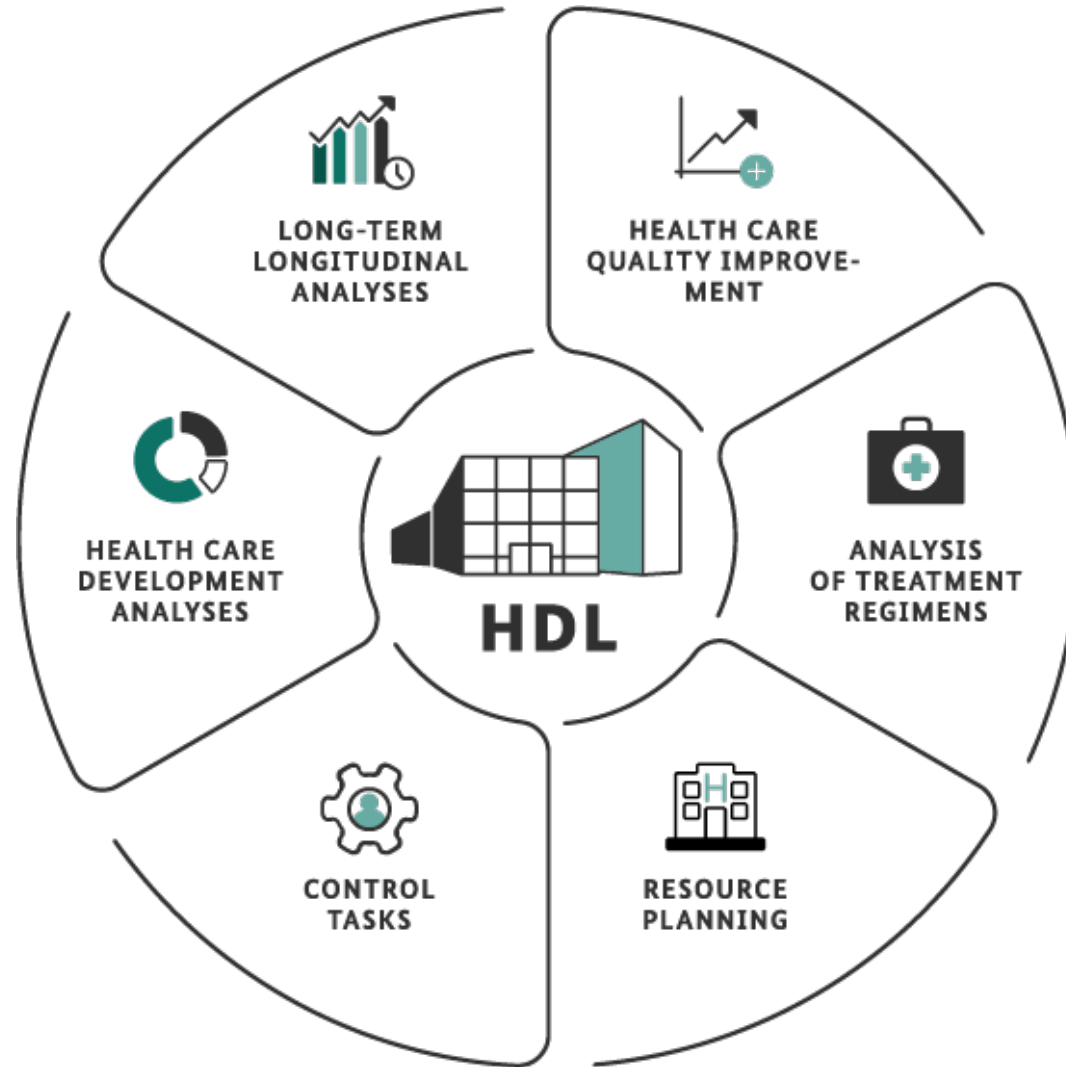The Health Data Lab (HDL) enables secure health data research by balancing:

Research
- **Facilitate access to health data**
- Close collaboration with data users

Security
- Consideration of data sensitivity
- Close collaboration with information security and data protection authorities (BSI, BfDI)

Federal Institute for Drugs and Medical Devices

HDL

Scope of the HDL

The central hub is labeled **HDL**, surrounded by six segments:

- **LONG-TERM LONGITUDINAL ANALYSES**
- **HEALTH CARE QUALITY IMPROVE-MENT**
- **ANALYSIS OF TREATMENT REGIMENS**
- **RESOURCE PLANNING**
- **CONTROL TASKS**
- **HEALTH CARE DEVELOPMENT ANALYSES**

According § 303e SGB V

Data Characteristics

- 73 million people with statutory health insurance in Germany

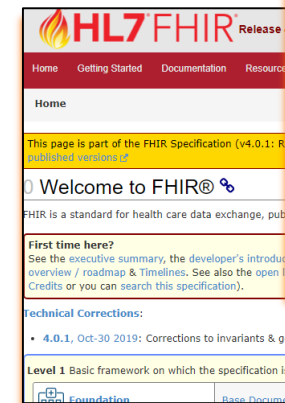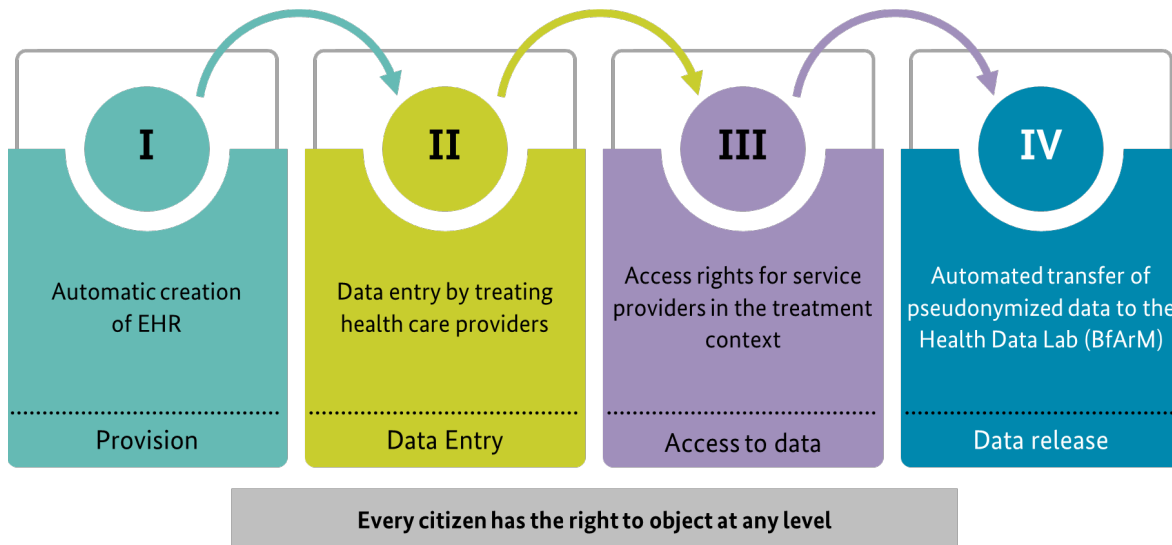- Information from all health care sectors linked on the individual level

- Longitudinal data starting from 2009

- Interoperability with established code systems (e.g. ICD10, ATC)

Prescriptions

- PZN (Central pharmaceutical number)
- ATC
- Date of prescription
- ID of doctors / institution
- Quantity / fraction
- Date of prescription
- Number of product prescriptions
- …

Outpatient Data

- ID of doctors / institution
- Quarter
- Doctor ID
- Diagnoses (ICD-10)
- Localization
- Date of service
- Number and type of Procedures: OPS codes
- Dental procedures
- …

Inpatient Data

- Dates of admission and discharge
- Hospital ID
- Department
- Admitting doctor
- Primary and secondary diagnoses (ICD-10)
- Localization
- Type of treatment
- Procedures
- DRGs
- …

Personal information

- Year of birth
- Sex
- insurance coverage times
- Insurance status
- Place of residence
- Reason for leaving insurance
- Death
- …

Others

- Aids and remedies
- Midwife care
- Digital Health Applications (DiGA)
- Volunteered electronic health records
- …

Federal Institute for Drugs and Medical Devices

HDL

Electronic Health Record (eHR)

Voluntarily shared electronic Health Records:

➢ Structured medical information objects (MIO) in HL7/FHIR®, e.g.:

    ➢ Digital maternity record

    ➢ Digital child examination booklet

I	Automatic creation of EHR	Provision
II	Data entry by treating health care providers	Data Entry
III	Access rights for service providers in the treatment context	Access to data
IV	Automated transfer of pseudonymized data to the Health Data Lab (BfArM)	Data release

Every citizen has the right to object at any level

https://mio.kbv.de
https://www.hl7.org/fhir/

Typical Data Provisioning Process

User envisions analysis

runs analysis on some original data somewhere

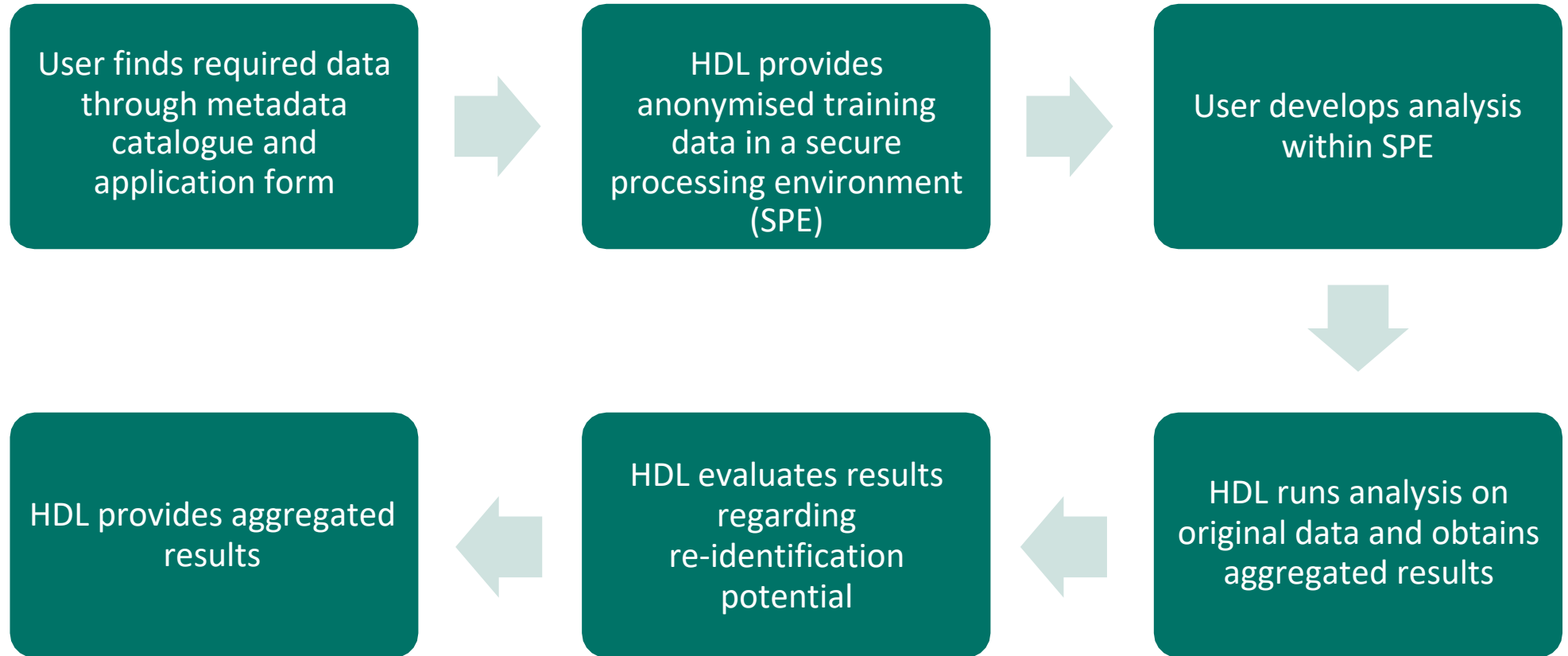# Typically unFAIR Data Provisioning Process

➤ Process to **find** data is unclear

➤ Process to **access** data is unclear,
especially when several data sources need to be linked

➤ Usability and safety of **processing** the data is unclear,
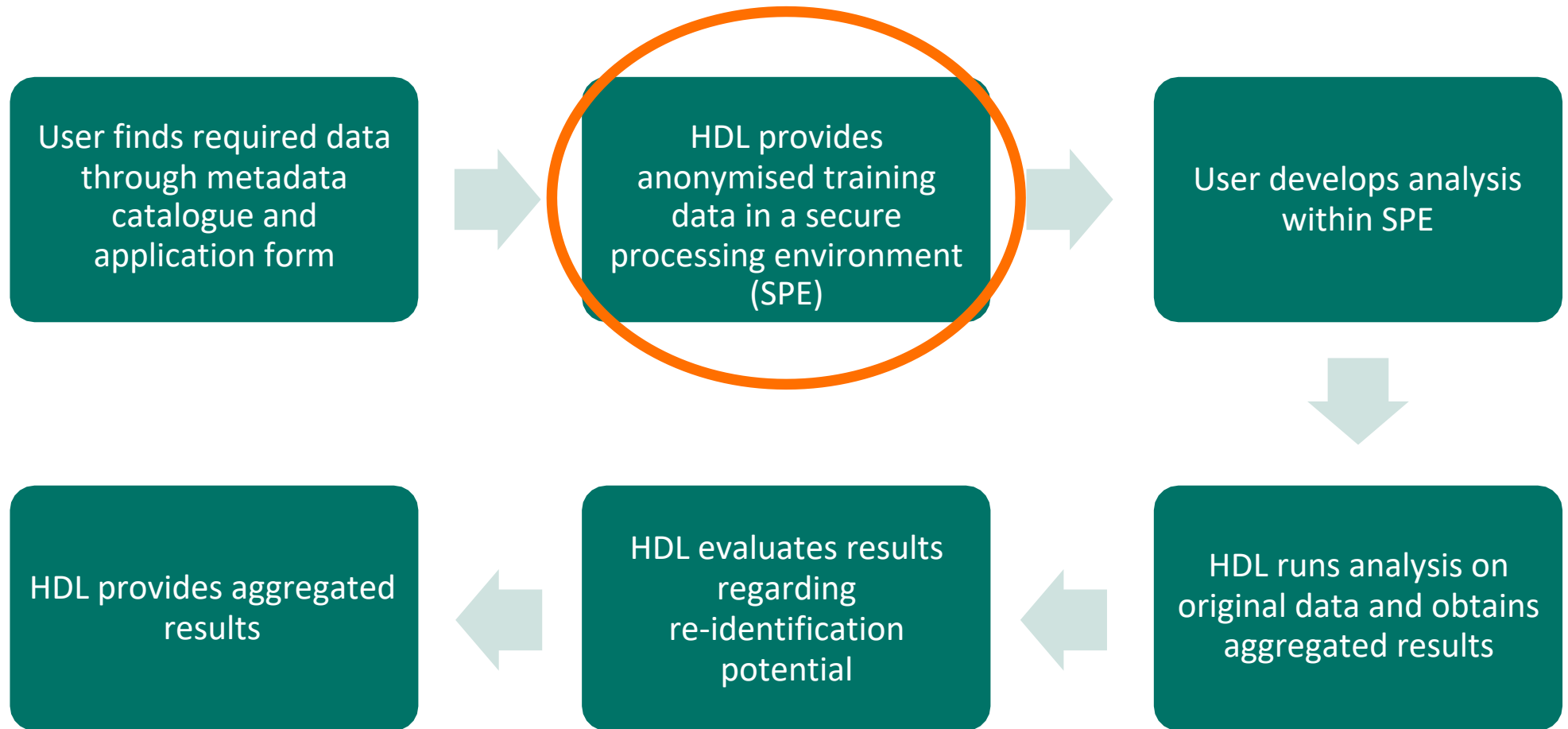especially concerning processing environment

User envisions analysis
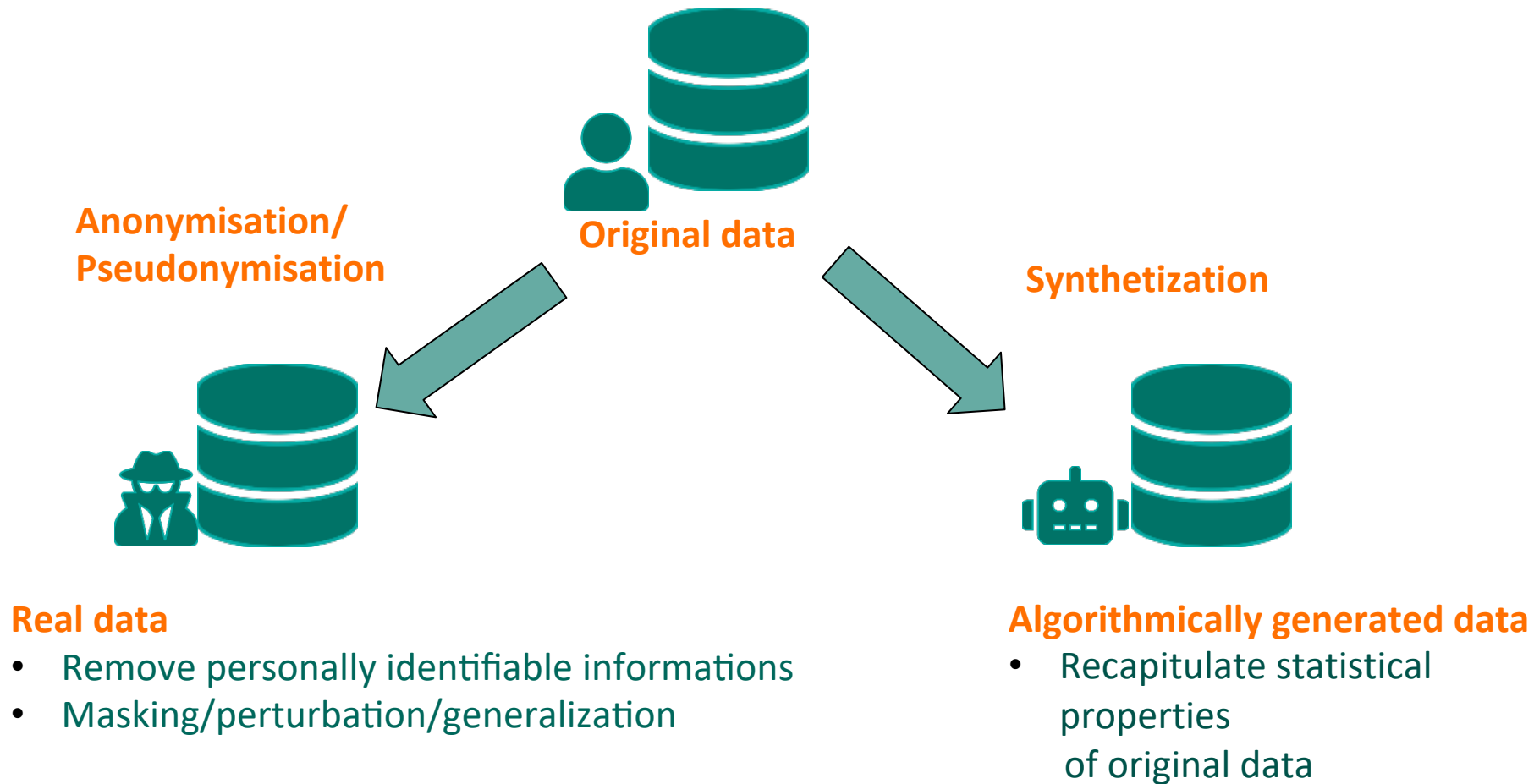
?

runs analysis on
some original data
somewhere

Steffen Heß|
HDL BfArM |
24.11.2023
59

Federal Institute
for Drugs
and Medical Devices

HDL

Data Provisioning Process @HDL

User finds required data through metadata catalogue and application form

→

HDL provides anonymised training data in a secure processing environment (SPE)

→

User develops analysis within SPE

↓

HDL provides aggregated results

←

HDL evaluates results regarding re-identification potential

←

HDL runs analysis on original data and obtains aggregated results

Data Provisioning Process @HDL

User finds required data through metadata catalogue and application form

HDL provides anonymised training data in a secure processing environment (SPE)

User develops analysis within SPE

HDL provides aggregated results

HDL evaluates results regarding re-identification potential

HDL runs analysis on original data and obtains aggregated results

Federal Institute for Drugs and Medical Devices

HDL

Anonymised Training Data - Privacy Protection

**Anonymisation/
Pseudonymisation**

Original data

Synthetization

Real data
- Remove personally identifiable informations
- Masking/perturbation/generalization

Algorithmically generated data
- Recapitulate statistical
  properties
  of original data

Federal Institute
for Drugs
and Medical Devices

HDL

Synthetic Data and AI-Readiness

Goals:

1. Comparison of classic anonymisation and synthetic data in terms of privacy and utility

2. Evaluation of AI readiness
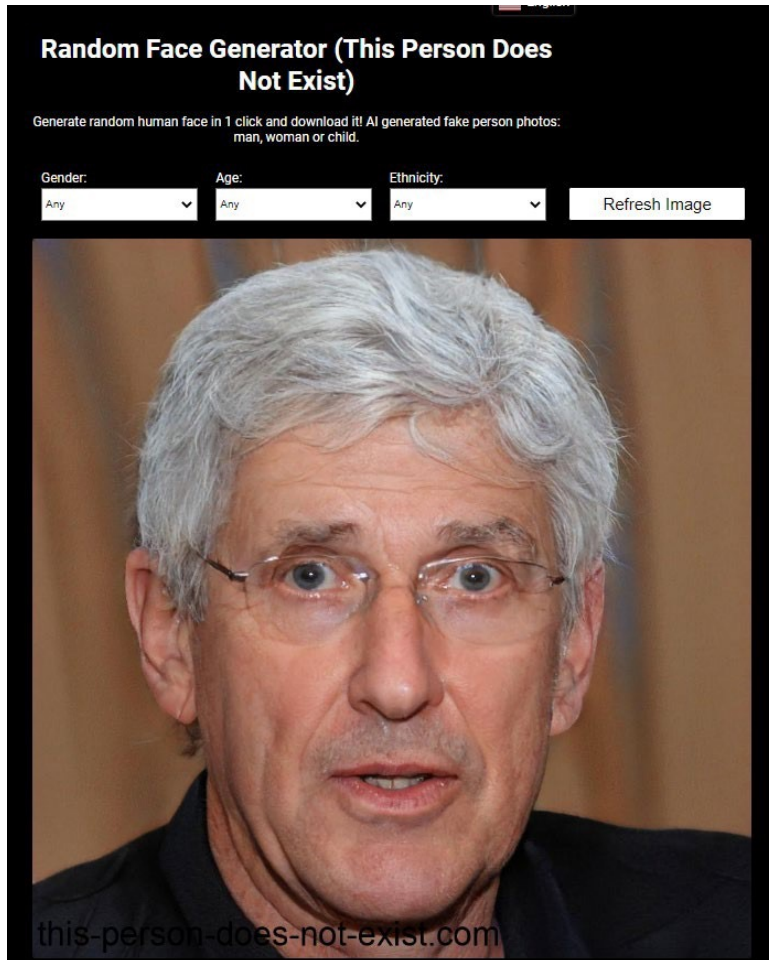
3. International applicability

high utility,
no privacy

high privacy,
no utility

Source: https://medium.com/birds-view/solving-the-data-innovation-versus-privacy-trade-off-the-rationale-behind-our-investment-in-c135b71d8d18
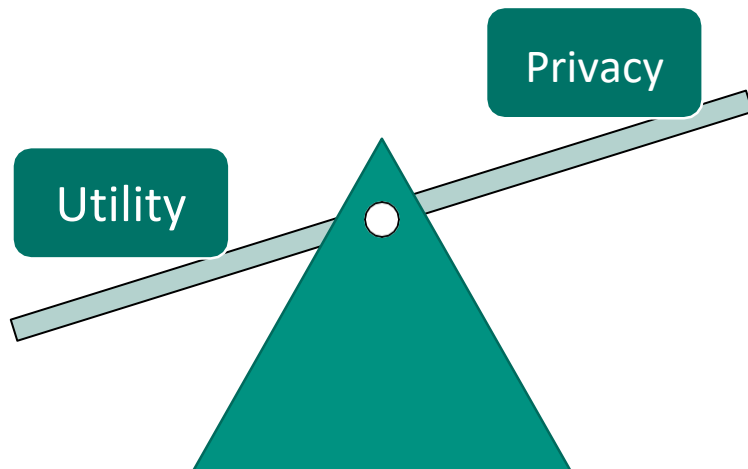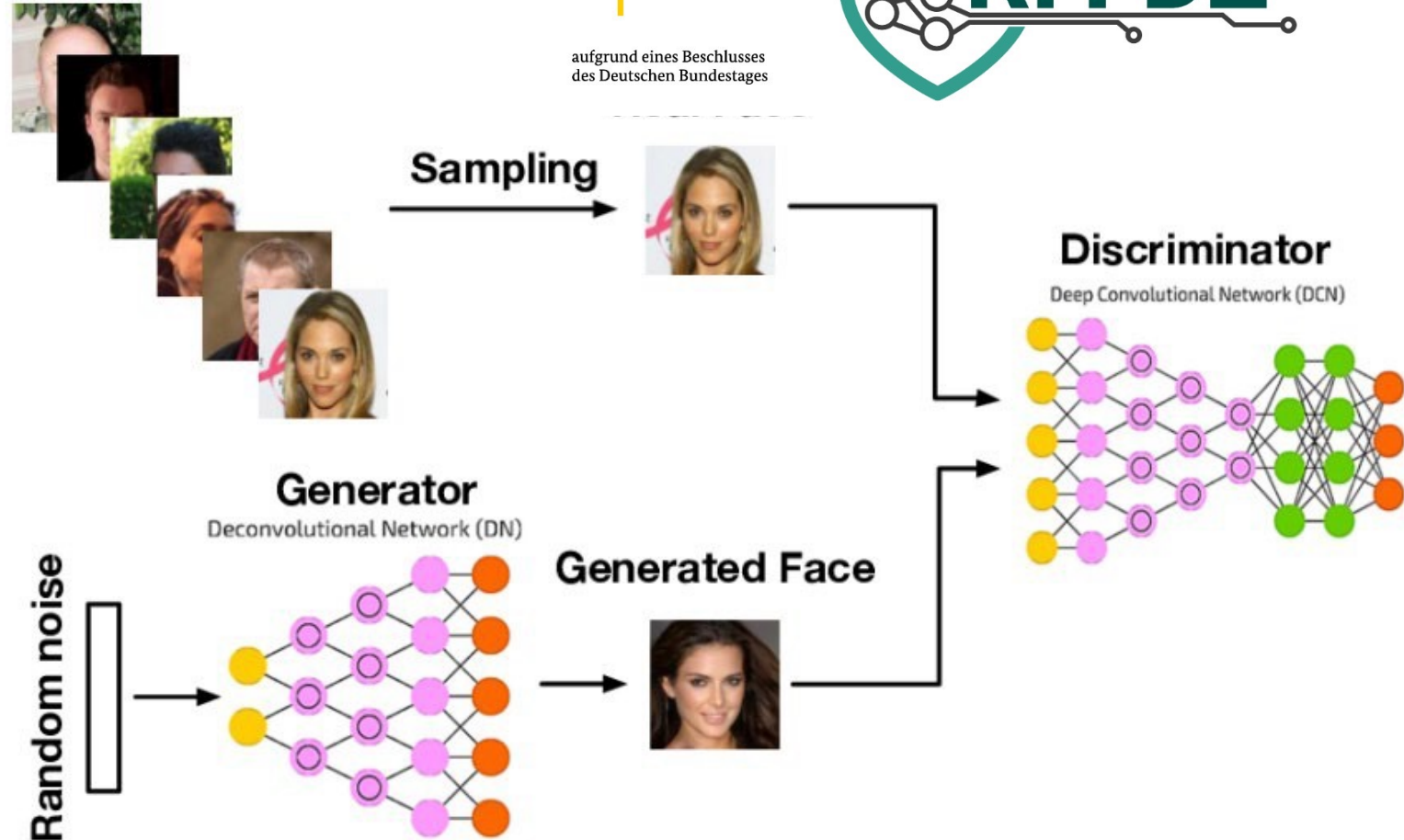
Example of Synthetic Data based on GANs

KI-FDZ

Random Face Generator (This Person Does Not Exist)

Generate random human face in 1 click and download it! AI generated fake person photos: man, woman or child.

Gender: Any
Age: Any
Ethnicity: Any
Refresh Image

this-person-does-not-exist.com

https://this-person-does-not-exist.com/

Sampling

Discriminator
Deep Convolutional Network (DCN)

Generator
Deconvolutional Network (DN)

Random noise

Generated Face

Federal Institute for Drugs and Medical Devices

HDL

KI-FDZ – Project

AI for Synthetic Data

- Best privacy protection
- Most flexible applications
- Creating and evaluating synthetic data with AI-methods

AI - Readiness

- Evaluation of AI-readiness
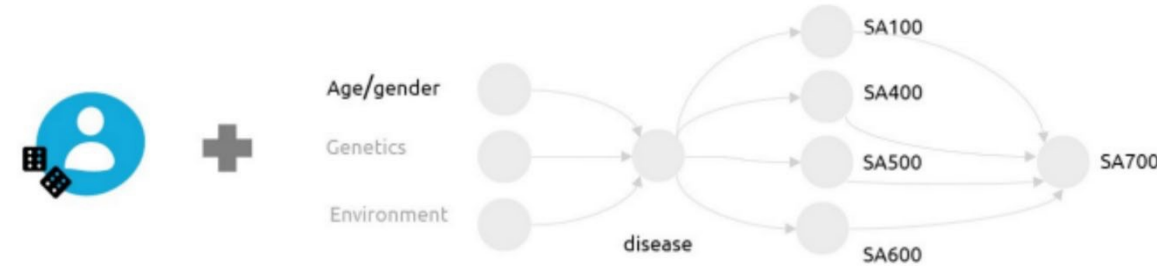- Design of data access processes
- European connectivity

Duration: November 2021- December 2024

Privacy & Utility Metrics

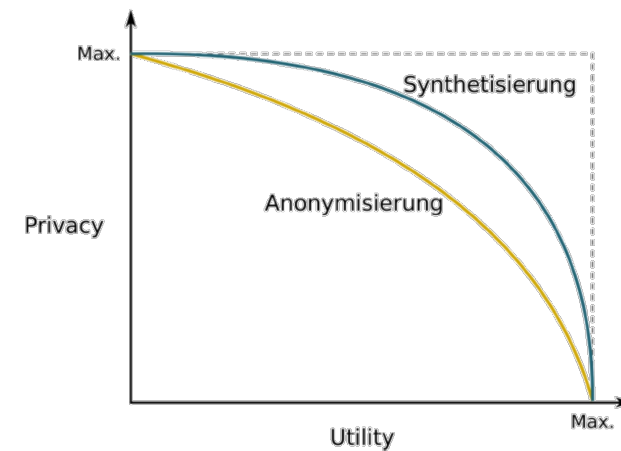Main Methods:

- Graphical Model (Bayesian Network)

- GANs

Privacy metrics:

- Differential privacy

- Shadow model attacks

Defined use cases for utility:

- Bleeding risk in for deep vein thrombosis on oral anticoagulation and concomitant antiplatelet therapy

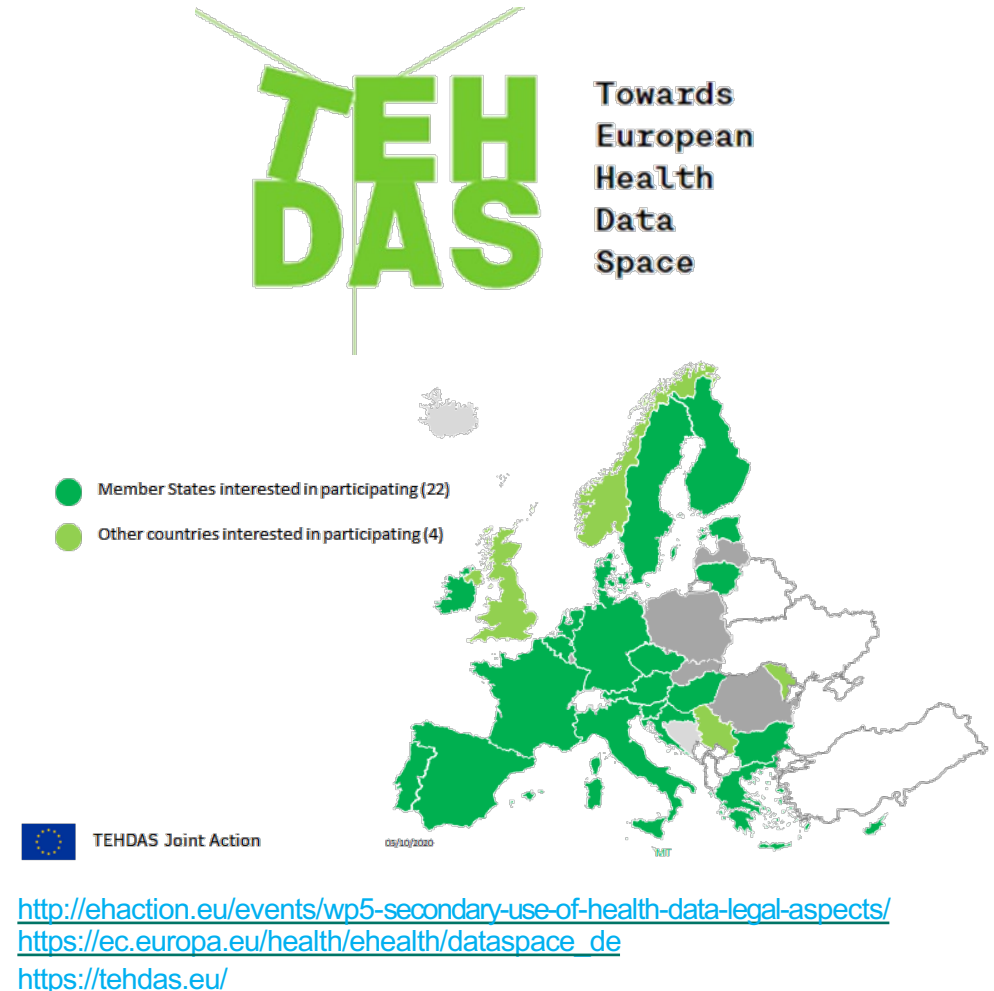- Prevalence & incidence of diabetes in Germany (2018 & 2019)

EHDS Initiatives TEHDAS -> TEHDAS2

Joint Action **T**owards the **E**uropean **H**ealth **Da**ta **S**pace with the following pillars:

- Reliable data governance system and principles for cross-border data use

- Data quality

- Secure infrastructure und interoperability

- HDL supports TEHDAS as part of a delegation coordinated by Federal Ministry of Health

Towards
European
Health
Data
Space

Member States interested in participating (22)

Other countries interested in participating (4)

TEHDAS Joint Action   05/10/2020

http://ehaction.eu/events/wp5-secondary-use-of-health-data-legal-aspects/
https://ec.europa.eu/health/ehealth/dataspace_de
https://tehdas.eu/

Federal Institute
for Drugs
and Medical Devices

HDL

Health Data@EU Pilot for a European Health Data Space on Secondary Use of Health Data


Thrombosis in COVID-19 patients

Foster a better understanding of the risk of thrombosis in COVID-19 patients

- Use case leader: European Medicines Agency

- Aims: Better understanding the risks of thrombosis in COVID-19 patients

- Data partners: Denmark, France, Croatia, Finland + DARWIN EU

- Health Data Lab is part of the research team

Federal Institute for Drugs and Medical Devices

HDL

https://ehds2pilot.eu/

Horizon Europe Project
Development, Optimisation & Implementation of AI-Methods for RWD Analyses in Regulatory Decision-Making & HTA along the Product Life-Cycle
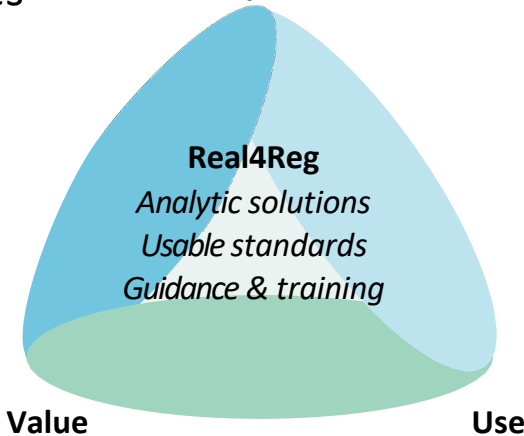
Duration: 2023-2026, ~ 7 Mio €
Consortium: 10 Partners, 6 EU countries, Lead: BfArM
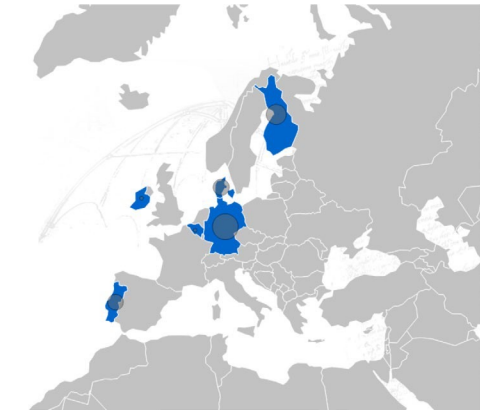
Objectives

Needs and Outcomes
- Unlocking potential of AI methods in the regulatory and HTA context
- Usable standards in RWD use
- Guidance and training in RWE use and RWD analyses for health regulatory and HTA bodies across all EU countries
- Acceptance and impact of RWD and synthetic data along the product lifecycle

Personal communication B. Hänisch

Thank you very much for your attention!

"Knowing is not enough; we must apply.

Willing is not enough; we must do."

Johann Wolfgang von Goethe. (and/or Bruce Lee)

Contact

Federal Institute for Drugs and Medical Devices
Health Data Lab
Kurt-Georg-Kiesinger-Allee 3
D-53175 Bonn

Contact person
Steffen Heß
fdz@bfarm.de
www.bfarm.de
www.healthdatalab.de
Phone +49 (0)228 99 307-4491