



Data Management for Digital Health Lecture Kickoff

Borchert, Dr. Schapranow
Data Management for Digital Health
Winter 2023

Lecture Organization

Administrative Contacts

- Florian Borchert
- Dr. Matthieu-P. Schapranow

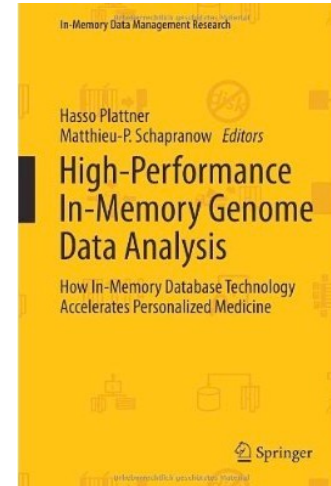
✉ <First Name>.<Last Name>@hpi.de



Digital Health Cluster
Hasso Plattner Institute Campus III
Rudolf-Breitscheid-Str. 187
14482 Potsdam, Germany

we.analyzegenomes.com

 @AnalyzeGenomes



Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
2

Lecture Organization

Time and Dates

- Location: HPI Campus II, L-1.06
- Tuesdays 11.00am-12.30pm (s.t.)
- Thursdays 01.30pm-03.00pm (s.t.)

- Enroll for the lecture until **Oct 31, 2023** (firm deadline)
- Lecture website: <https://we.analyzegenomes.com/dm4dh2023/>



HPI Hasso Plattner Institut

Home

Scope of the lecture



Welcome to the online class: we are very excited that you are interested in learning more about the foundations data management for digital health. A very relevant topic not only in times of worldwide COVID-19 pandemic. In this lecture, we will provide you specific examples from the field of digital health to understand where and how data is acquired, what are the challenges with these specific types of data, and how to handle them with latest technology advances. We will link to latest worldwide developments in fighting the COVID-19 pandemics and provide you with a better understanding of the latest decisions and developments, where applicable.

After participating in the course, you will be equipped with the ability to:

- > assess requirements of selected real-world use cases from the **medical** field,
- > select latest **technology** building blocks to create viable healthcare software solutions, and
- > analyze requirements for data analysis and processing, e.g. for **machine learning**.

In the course, we will have invited guest speakers sharing their real-world experience with you in a brief presentation. You will also have the chance to raise your questions and discuss with them in the course of the lecture.

Further details about the structure of the lecture will be shared in the first course of the lecture with you.

Lecture Organization Grading

- Credit points: 6 graded ECTS
- Final grading will be determined by final exam at the end of the course (100%)
- Bear in mind: You have to pass all intermediate exercises prior to participate the final exam (Prüfungsvorleistung)



Introduction to Digital Health

Data Management for
Digital Health, Winter
2023

Lecture Organization

What you can expect from us

- Broaden your horizons in the fields of
 - Digital health,
 - Life sciences, and
 - Data challenges and opportunities
- Experience real-world use cases
- Hands-on experiments with selected tools
- Invited talks by key experts in the field
- Get experience in collaborative project work



<http://i.kinja-img.com/gawker-media/image/upload/s--cREIB5AZ--/1865smw5hbtt6jpg.jpg>

Introduction to Digital Health

Data Management for
Digital Health, Winter
2023

5

Lecture Organization

What we expect from you

- Commitment to the lecture and exercises
- Attend lectures regularly
- Participate in group discussions and expert talks
- Perform autonomous research to dig deeper into topics
- Support through your expertise also your fellow students
- Update us as supervisors on any issues you might encounter



<http://i.kinja-img.com/gawker-media/image/upload/s--cREIB5AZ--/1865smw5hbbt6jpg.jpg>

Introduction to Digital Health

Data Management for
Digital Health, Winter
2023

6

Lecture Organization Sneak Preview

- Hands-on exercises



Credit: Caine et al. *A 3D-DNA Molecule Made of PlayMats*. 2015

- Experience the reality



Credit: Delaware State News/Dave Chambers

- Discuss with experts



Credit: Acuitus Medical

**Introduction to Digital
Health**

Data Management for
Digital Health, Winter
2023

7

Agenda

Pillars of the Lecture

Medical Use Cases



Biology Recap



Oncology



Nephrology



Infectious
Diseases

Technology Foundation



Data
Sources



Data
Formats



Processing and
Analysis



Software
Architectures

Machine Learning

Data



Refine

Evaluate



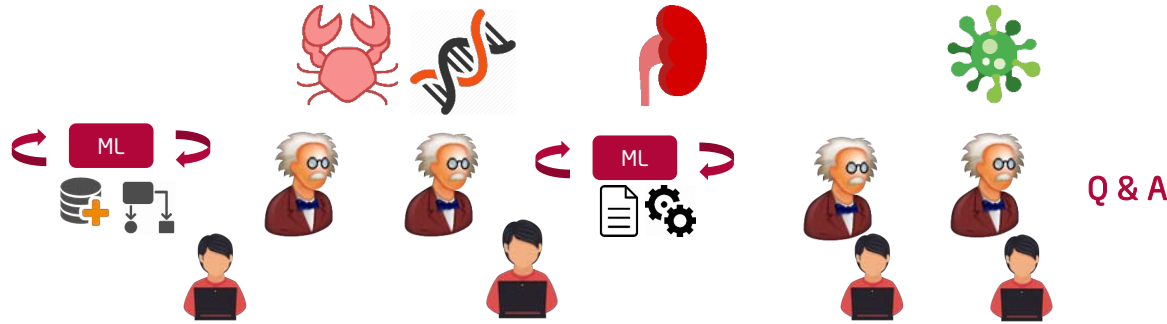
Prediction +
Probability

**Introduction to Digital
Health**

Data Management for
Digital Health, Winter
2023

8

Lecture Schedule



Final Exam
Feb 13, 2024
11:00am,
Lecture Hall HS1

Nov

Dec

Jan

Feb

- Lecture Kickoff
- Actors in Healthcare
- Digital Health Data

- Machine Learning (ML) Foundations
- Use Case Oncology
- Biology Recap

- Natural Language Processing
- Use Case Nephrology & Intensive Care
- Supervised ML & Deep Learning

- Use Case Infectious Diseases
- Unsupervised ML

Introduction to Digital Health

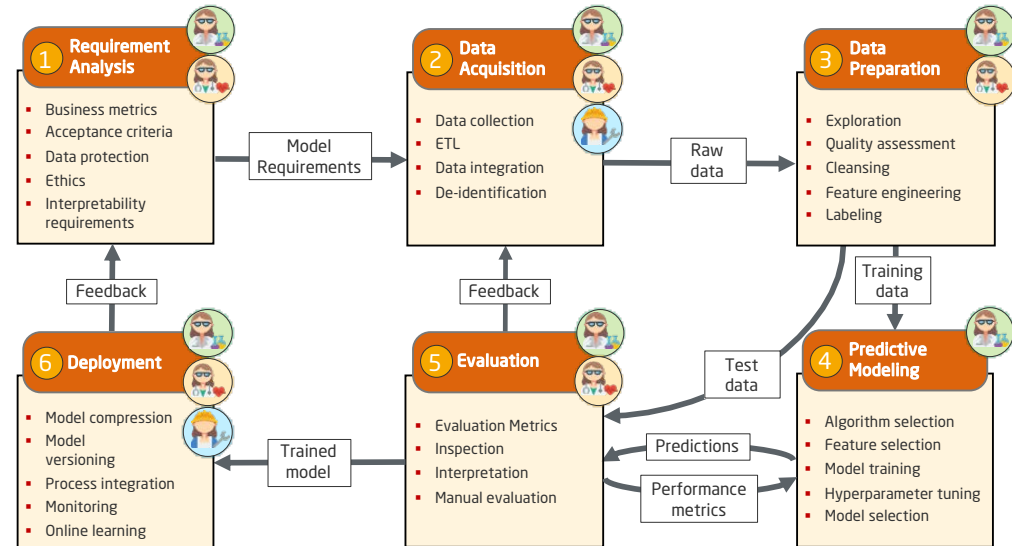
Data Management for
Digital Health, Winter
2023

Lecture Contents

Machine Learning Foundations



- What makes ML special in digital health?
- How does a process model for ML in digital health look like?
- What external factors and frameworks need to be incorporated for digital health?
- Why is digital health data special?



Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
10

Lecture Contents

Use Case Oncology



- Important aspects and how to distinguish them precisely, e.g. personalized, stratified, and precision medicine
- Clinical oncology process, from diagnosis to treatment
- Special data formats in oncology, e.g. *omics and imaging
- Computer-assisted detection and diagnosis, e.g. application of convolutional neural networks
- Distributed, scalable computing and processing workflows
- Real-world application examples, e.g. Medical Knowledge Cockpit, Medical Image Data Analysis
- Expert talks on oncology



White House, President Obama speaks on the Precision Medicine Initiative, Jan 30, 2015

Introduction to Digital Health

Data Management for
Digital Health, Winter
2023

Lecture Contents

Biology Recap



- Discovery of cells and their functional components
- Cells and their life cycle
- Structure of DNA
- The human genome and its medical use
- Sequencing technologies



HPI/Matthieu-P. Schapranow

Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
12

Lecture Contents

Processing of Medical Texts



- Ontologies and knowledge engineering in digital health
- Methods for text mining and natural language processing
- Information extraction from unstructured text data
- Sequential machine learning models
- Real-world applications for text mining in oncology

PHYSICAL EXAMINATION * Mock Clinical Note

ENT: Examined and normal.
Skin: Psoriasis over the kneecaps and elbows, and within his hair.
Lymph: Examined and normal.
Thyroid: Not enlarged.
Heart: Core S1, S2, no murmur.
Lungs: Examined and normal.
Abdomen: Soft and nontender. No obvious masses.
Extremities: No signs of joint damage due to his psoriatic arthritis. Ankle scar on left from surgery. Right knee arthroscopy scar.
Pulses: Normal.
Neuro: Reflexes are normal.
Rect: Normal prostate, no masses palpable.

IMPRESSION/REPORT/PLAN

#1 Colorectal cancer of the cecum, biopsy proven. No evidence for metastatic disease.
#2 Thyroid insufficiency, on treatment
#3 Psoriatic arthritis, adequately treatment with methotrexate and topical steroid creams

PLANS/RECOMMENDATIONS:

1. A surgical consultation for possible right hemicolectomy in the next 1-2 weeks.
2. Complete pre-anesthetic medical evaluation, and obtain electrocardiogram.
3. Obtain the outside CT scan and have it formally reviewed by Clinic radiologist.
4. Obtain the outside colorectal biopsies and have these formally reviewed by Clinic pathologist.

Event Discovery

UMLS Classification

- Sign / Symptom
- Test / Procedure
- Disease / Diagnosis
- Medication
- Anatomy / General

Negation Detection

Uncertainty Detection

Time Expression Discovery

Introduction to Digital Health

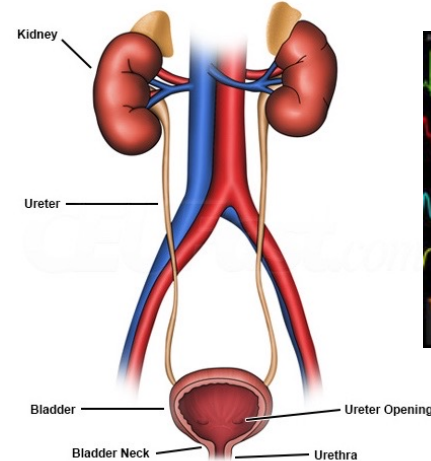
Data Management for
Digital Health, Winter
2023
13

Lecture Contents

Use Case Nephrology



- Nephrology: why do you have two kidneys?
- How to predict severe post-transplant patient risks, e.g. clinical predictive modeling using supervised machine learning
- Reasons for admission to an Intensive Care Unit (ICU), common risks and challenges during ICU stays
- How to detect severe event, e.g. using ICU sensors



<https://ceufast.com/course/urinary-tract-infections-the-unappreciated-giant>
<http://www.appdropp.com/ios/ivital-signs/434101595>

Introduction to Digital Health

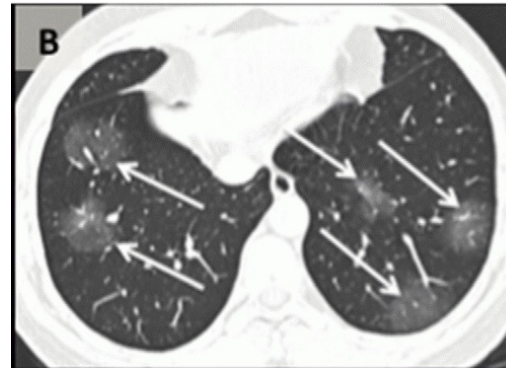
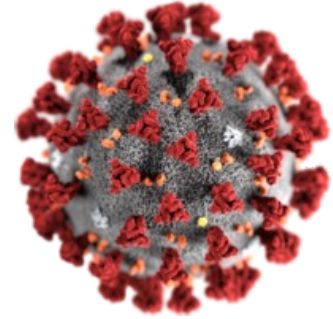
Data Management for
Digital Health, Winter
2023
14

Lecture Contents

Use Case Infectious Diseases



- How to group infectious diseases
- Overview and important facts you should know about infectious diseases
- How to prevent and detect infections
- Epidemics, pandemics and ways to contain the spread of infections
- Vaccinations, history and categories, latest development and innovative approaches
- Special focus topic: COVID-19 and latest status and findings




Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
15

Online Tools and Methods: Student Community for Exercises and Async. Exchange

- What: openHPI provided by HPI
- Where: <https://we.analyzegenomes.com/dm4dh2023-community/>
- Requirements: One-time online enrollment for the course: you must use the HPI identity provider to enroll to the openHPI course
- Purpose: We will use it for exercises and as community platform for all students to exchange and discuss content of the lecture

Or log in with:




HPI Identity

for HPI students and employees

Data Management for Digital Health 2023 Dr. Matthieu-P. Schapranow In preparation

Learnings Discussions Progress Collab Space Course Details Announcements



Welcome to the class: We are very excited that you are interested in learning more about the principles of data management for digital health and why it might be different from what you have learnt so far. In this lecture, we and selected guest speakers will share with you:

Specific examples from selected fields of digital health to understand where and how data is/needs to be acquired. Known challenges in acquiring and processing these types of data in their specific digital health domain. How to deal with and address specific requirements and limitation of accessing and using digital health data, and how the complex analysis of high-dimensional multi-modal digital health data can be facilitated through the use of latest soft- and hardware advances, e.g. clinical prediction models, federated learning infrastructures, large language models, and supervised as well as unsupervised machine learning approaches.

October 15, 2023 - March 31, 2024
Language: English


[Enter course](#) [Un-enroll](#)

Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
16

Online Tools and Methods:

Optional: Deep Dive Machine Learning


- What: Jupyter notebooks providing deep dive into ML code
- Where: <https://we.analyzegenomes.com/dm4dh2023-code/>
- Requirements: Use GitHub or  to start an Jupyter notebook server
→ no local installation required

☰ README.md

🔗 Data Management for Digital Health 2022/23

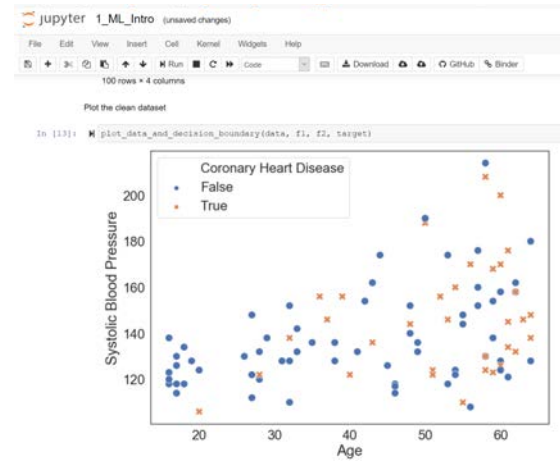
This repository contains the code to reproduce figures, metrics, and models for the 2022/23 version of the [course](#).

To run all notebooks interactively with MyBinder, click here (and wait for a few seconds):



Contents:

- Week 2: [Iterative ML Design Process](#)



Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
17

Open Questions?



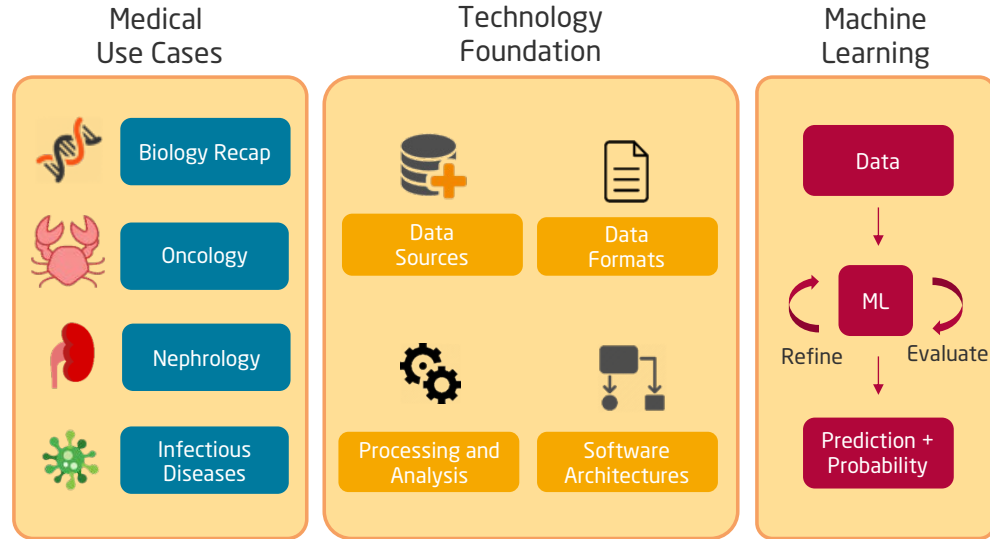
Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
18

Interactive Poll

What are your favorite lecture topics?

- A. Medical use cases
- B. Technology foundation
- C. Machine learning
- D. All of them



Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
19

Interactive Poll

Who are the most important actors in healthcare?

- A. Medical doctors
- B. Nurses and care professionals
- C. Pharma and researchers
- D. Insurance companies
- E. Citizens



Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
20

History of the German Social Security Act



Bundesarchiv, Bild 183-RE6588 / P. Loeschner & Petsch / CC-BY-SA 3.0, CC-BY-SA 3.0 de
Bundesarchiv, Bild 183-RE6588-091-21 / Umbekannt / CC-BY-SA 3.0, CC-BY-SA 3.0 de



Introduction to Digital Health

Data Management for Digital Health, Winter 2023
21

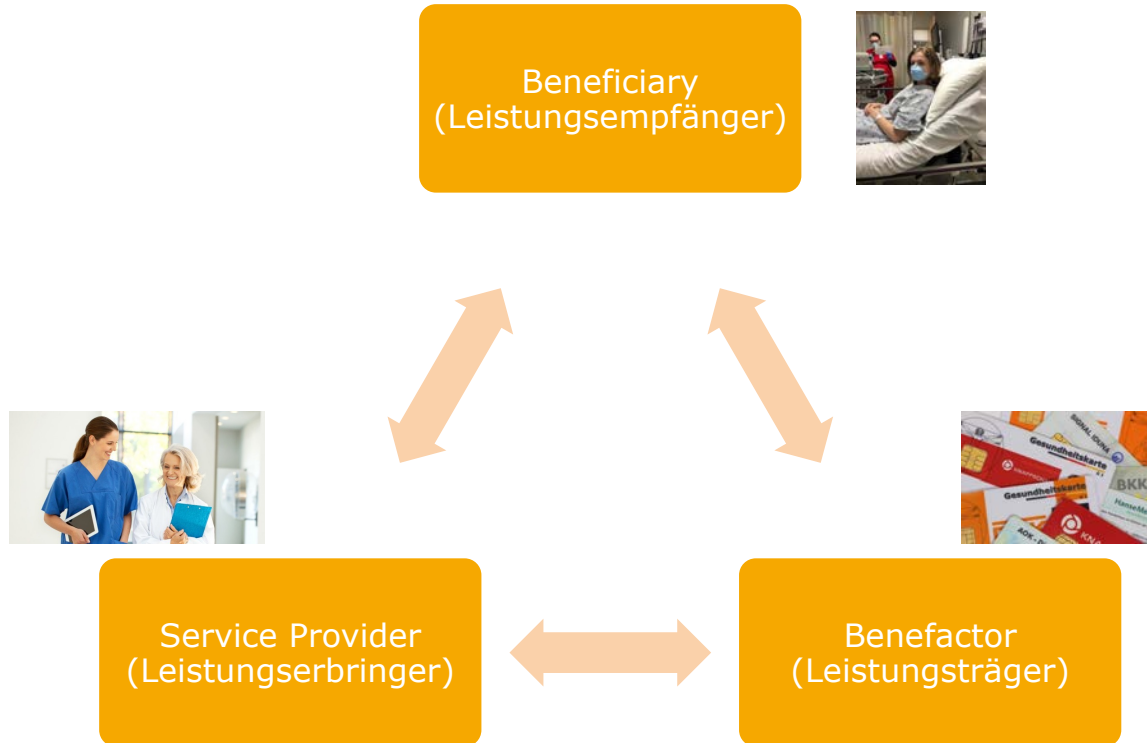
Legal Framework of the German Social Security Act

Book	Title
SGB I	General part
SGB II	Basic security benefits for job seekers
SGB III	Employment promotion
SGB IV	Common regulations for the social security act
SGB V	Statutory health insurance
SGB VI	Statutory pension fund
SGB VII	Statutory accident insurance
SGB VIII	Child and youth services
SGB IX	Rehabilitation and participation of handicapped persons
SGB X	Social administration tools and data security
SGB XI	Social nursing care
SGB XII	Social welfare
SGB XIV	Social compensation

Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
22

German Social Security Triangle



Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
23

German Social Security Triangle (SGB V)

Service provider (Leistungserbringer)

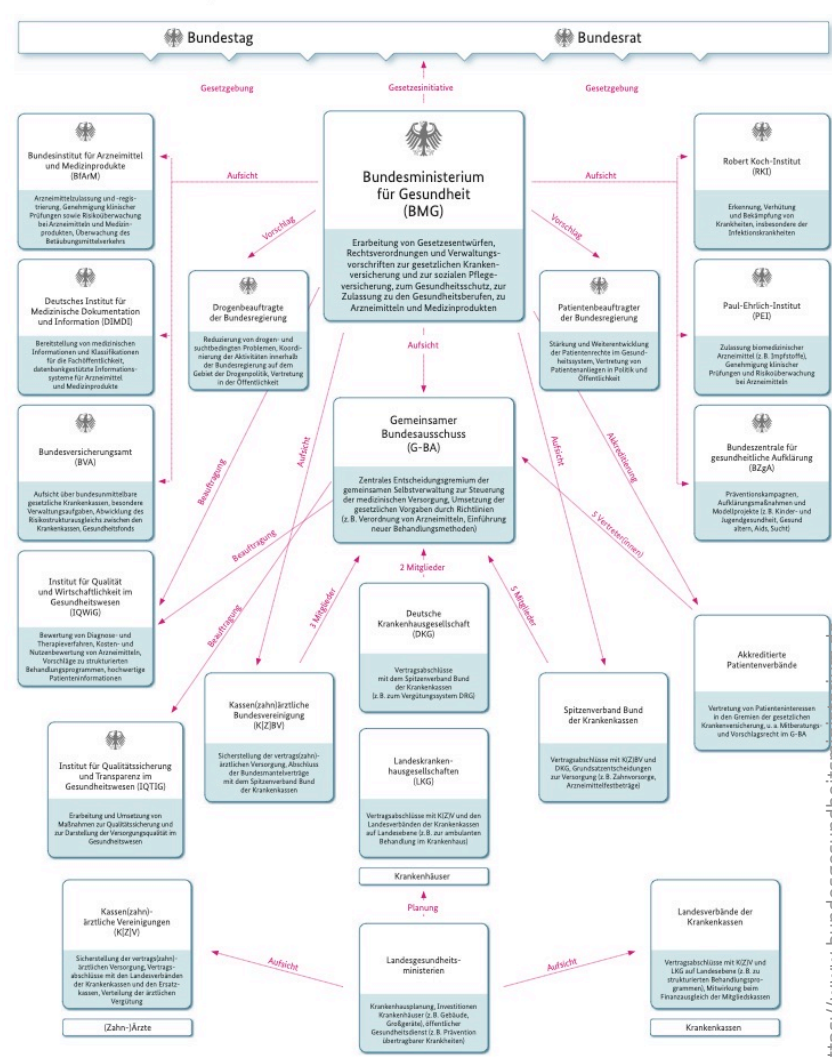
- Medical doctors of all professions incl. dentists
- Pharmacies
- Hospitals
- Therapists
- Experts for pharmaceuticals and medical appliances
- Ambulance and emergency services
- Care provider

Actors of the German Healthcare System

Who is missing?

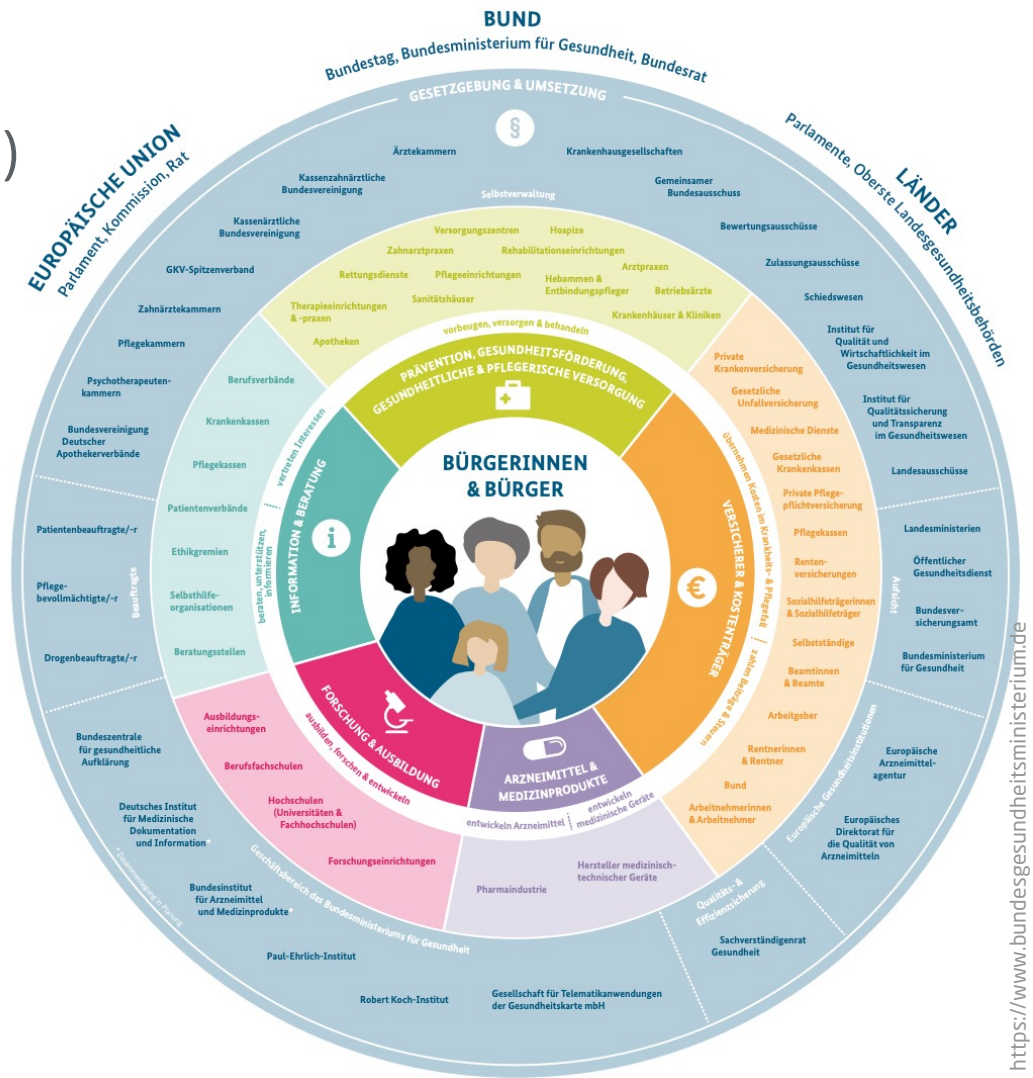


Citizen



Actors within the German Healthcare System (cont'd)

■ Updated version 2019



■ Patients



- Individual anamnesis, family history, and background
- Require fast access to individualized therapy

■ Clinicians and Nurses



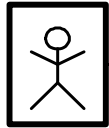
- Identify root and extent of disease using laboratory tests
- Evaluate therapy alternatives, adapt existing therapy

■ Researchers

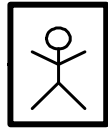


- Conduct laboratory work, e.g. analyze patient samples
- Create new research findings and come-up with treatment alternatives

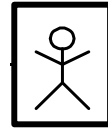
Intelligent Healthcare Networks in the 21st Century?



Researcher



Clinician & Nurse

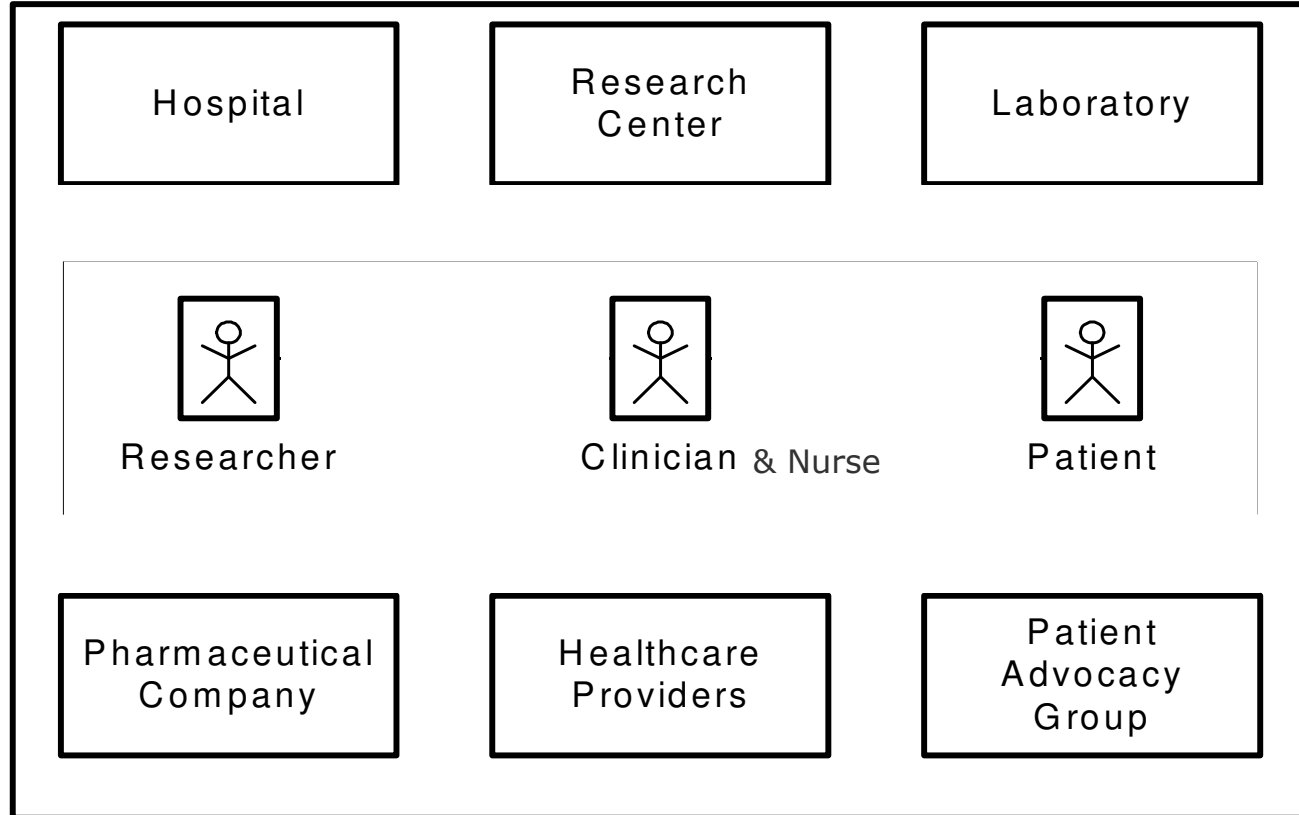


Patient

**Introduction to Digital
Health**

Data Management for
Digital Health, Winter
2023
28

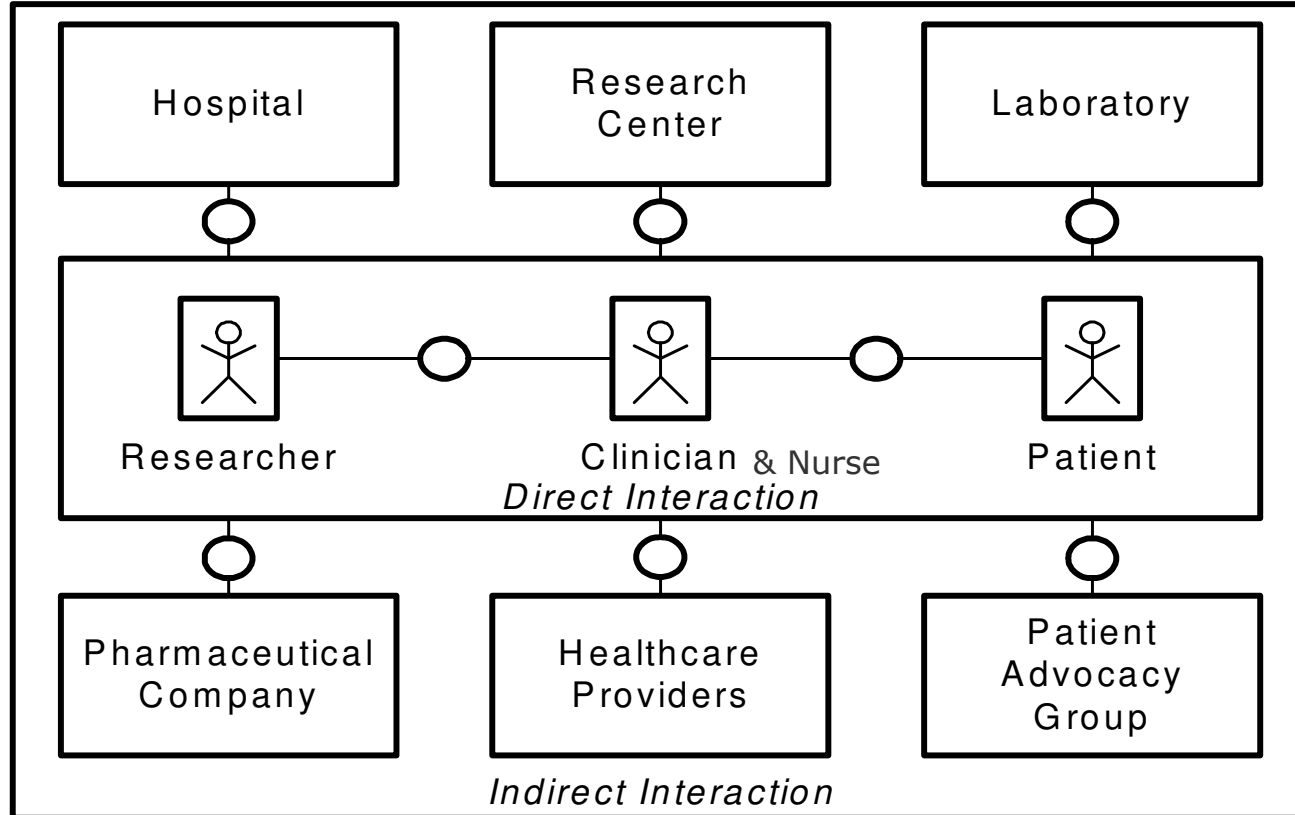
Intelligent Healthcare Networks in the 21st Century?



Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
29

Intelligent Healthcare Networks in the 21st Century!

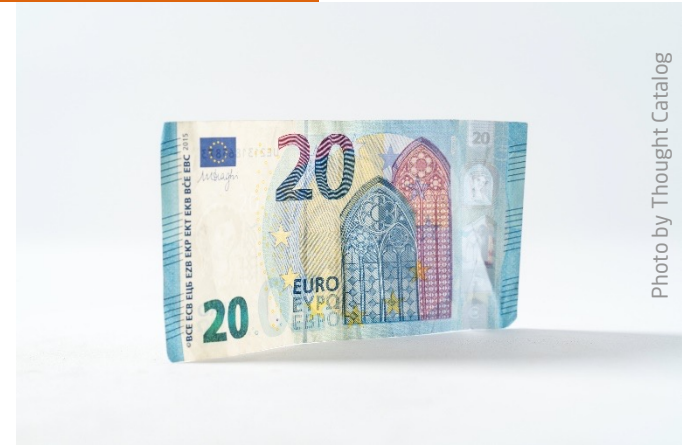


Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
30

Key Aspects of the German Social Healthcare Systems

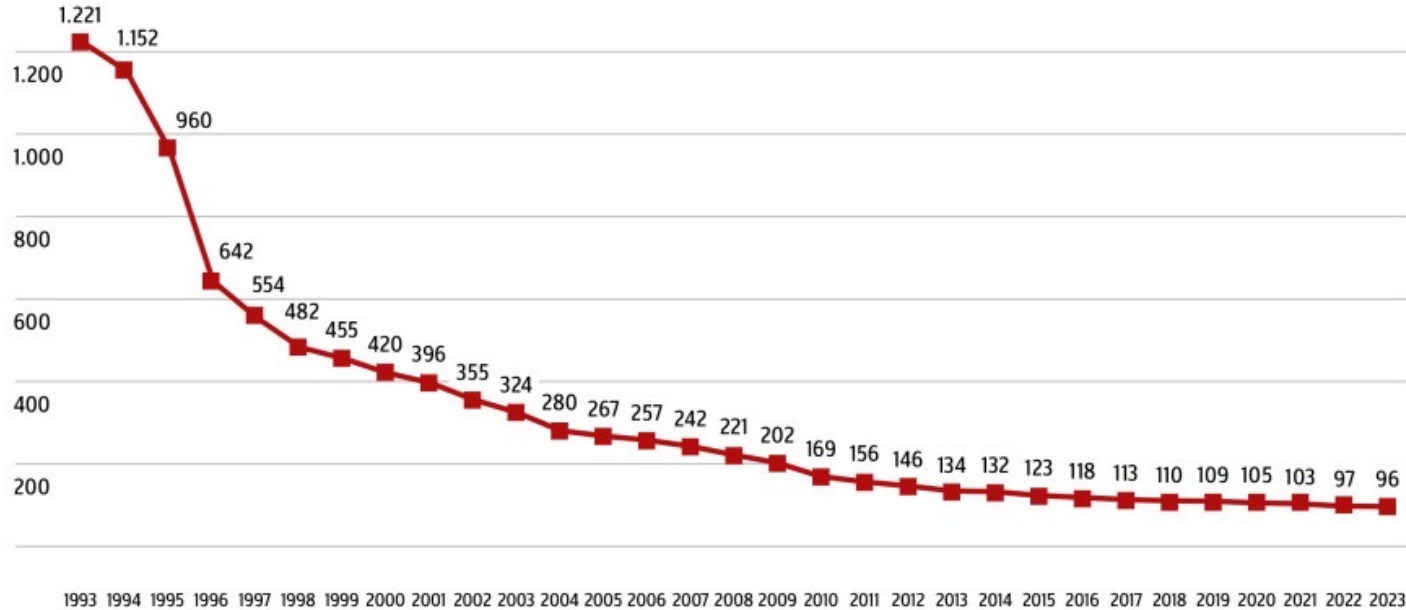
- Statutory insurance (covers approx. 70M members)
 - Private system also exists
 - Financial fees of insured people, employers, and tax
- Principle of solidarity
 - Everyone has access to same quality of healthcare services
 - Fees are based on monthly income
- Self administration
 - Legal framework determined by federal laws
 - Operational and implementational aspects are jointly agreed on, e.g. G-BA



Introduction to Digital Health

Data Management for
Digital Health, Winter
2023

Number of Health Insurances in Germany



Quelle und Darstellung: GKV-Spitzenverband; Stichtag: 1. Januar

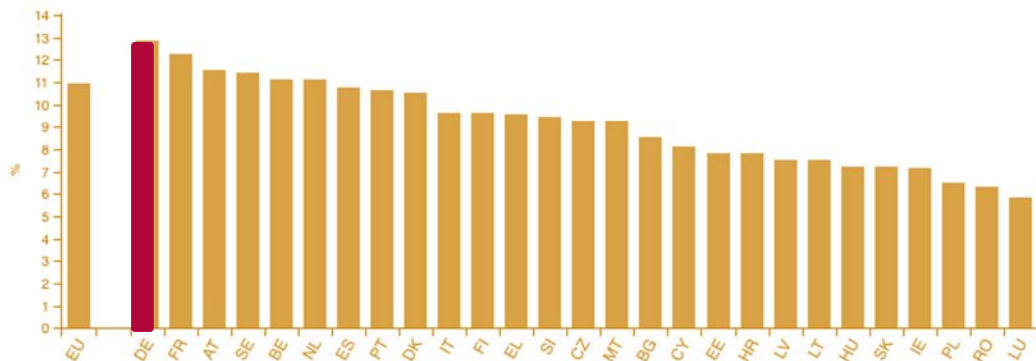
Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
32

Facts You Should Know Costs for Healthcare

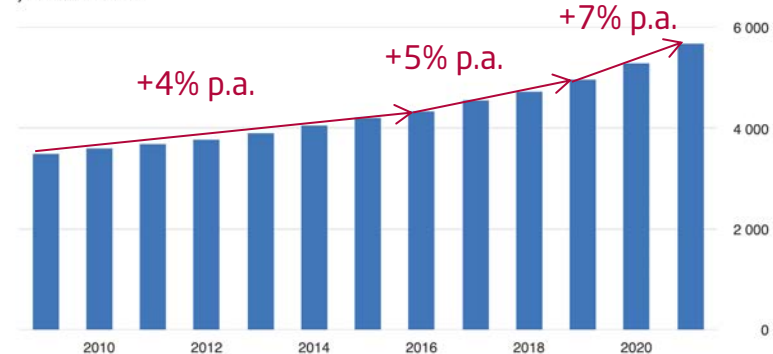


Current healthcare expenditure relative to GDP, 2020



■ Germany 5,699 EUR per capita resp. 13.2 % of GDP in 2021

Entwicklung der Gesundheitsausgaben
je Einwohner in EUR



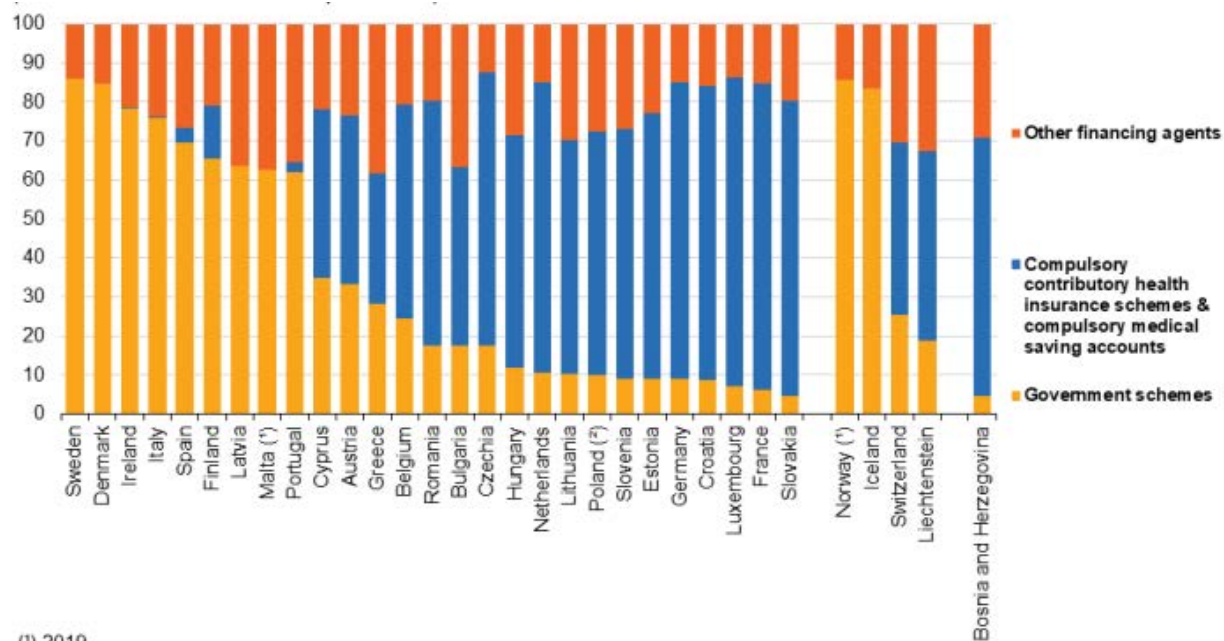
© Statistisches Bundesamt (Destatis), 2023

Introduction to Digital Health

Data Management for
Digital Health, Winter
2023

33

Sources of Financing Healthcare Expenditures



(*) 2019.

(?) Provisional.

Source: Eurostat (online data code: hith_sha11_hf)

- Development of laws, legal frameworks, and regulations
- Supervision of:
 - Statutory health insurances,
 - Nursing care insurance,
 - Health protection,
 - Health professions, and
 - Pharmaceuticals & medical devices.



Ressortforschung des Bundesministeriums für Gesundheit

Das Bundesministerium für Gesundheit (BMG) bereitet regelmäßig weitreichende politische Entscheidungen vor, die einen direkten Einfluss auf das alltägliche Leben der Bürgerinnen und Bürger haben.

► [weiterlesen](#)

Introduction to Digital Health

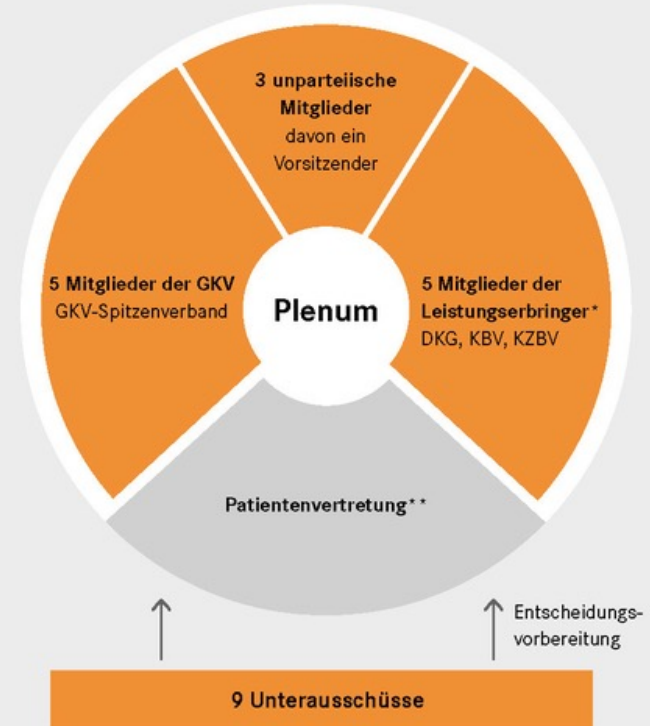
Data Management for
Digital Health, Winter
2023
35

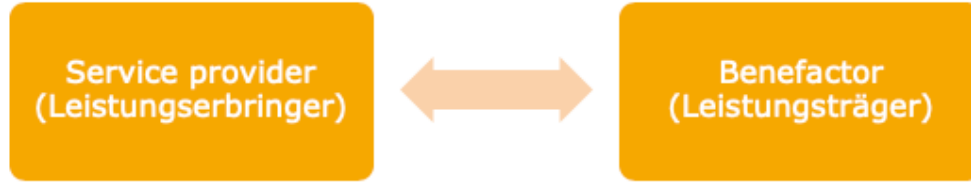
Joint Federal Committee (G-BA)

- Self administration of the German federal healthcare system
- Tasks are defined in SGB V
- Consists of representatives of:
 - Medical doctors, dentists, psychotherapist,
 - Statutory health insurance,
 - Hospitals, and
 - Patients.
- Aims:
 - Define health services covered by public health insurances,
 - Control implementation of legal frameworks, and
 - Assurance constant quality.

Gemeinsamer Bundesausschuss (G-BA)

(Gremium nach § 91 SGB V)





Association of statutory health insurance physicians on state level

- Members: Health care providers
- Every party, who invoices to health insurance companies

Health insurance companies

- Members: Health care receivers
- Manage contribution fees of members and negotiate general agreements

Types of Care

- Ambulant care services:
 - Provided by local medical experts at home or in their practices
- Stationary care:
 - Typically delivered as long-term hospital stay
- Rehabilitation:
 - Focuses on recovery, follows long-term care



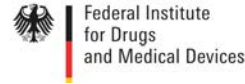
Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
38

Who decides about new Pharmaceuticals and Devices?

■ Germany

- Federal Institute for Drugs and Medical Devices
- Paul Ehrlich Institute (PEI) for vaccinations



■ Europe: European Medicines Agency (EMA)



EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH



■ USA: Food and Drug Administration (FDA)



Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
39

Patient Advocacy Organizations (PAO)

- Organized as non-for-profit organizations
- Specialized per medical indication
- Support members on multiple levels, e.g. legal, rights, contact to world-wide experts
- Patient- and caregiver-oriented education
- Represent patient group in multiple committees and bodies, e.g. legal as well as research

- BSL Bundesverband Selbsthilfe Lungenkrebs e.V.
- Bundesverband Niere e.V.
- Landesvereinigung Selbsthilfe Berlin e.V.,
- Bundesarbeitsgemeinschaft Werkstätten für behinderte Menschen e.V.
- Achse - Allianz Chronisch seltener Erkrankungen Bundesverband Niere e.V.
- Deutsche Stiftung chronisch Kranke
- Bundesverbandes der Kehlkopferierten e. V.,
- Deutsche Alzheimer Gesellschaft e.V.
- Leukämie und Lymphom SHG Halle
- Gesundheitsladen Bielefeld (Paritätische NRW)
- Verbraucherzentrale Bundesverband e.V.
- Frauenselbsthilfe nach Krebs e.V.
- PatV im Koordinierungsausschusses des G-BA
- diabetesDE - Deutsche Diabetes Hilfe
- Deutsche ILCO e.V. - Selbsthilfeverein für Darmkrebsbetroffene und Stomaträger*innen
- Alzheimer Ethik gem.e.V.
- Lymphangioliomyomatose, LAM
- Schlaganfall-Ring Schleswig-Holstein e.V.
- BNeV – Prävention, Zystennieren
- Hämochromatose-Vereinigung Deutschland e.V.
- BRCA-Netzwerk – Hilfe bei familiärem Brust- und Eierstockkrebs e.V.

Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
40

Digital Health Hype of Hope?

- Health data acquisition is no longer the bottle neck, but its analysis
- Health data remains scattered across silos, with limited benefits for individuals and the community
- Citizens ask for access to their personal healthcare
- No. of health apps growing, but holistic view on the citizen is still missing
- Advances in hardware and software, e.g. machine learning, in-memory databases, enable high-throughput data processing



Introduction to Digital Health

Data Management for
Digital Health, Winter
2023

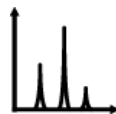
The Challenge

Distributed Heterogeneous Data Sources



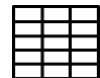
Human genome/biological data

600GB per full genome
15PB+ in databases of leading institutes



Human proteome

160M data points (2.4GB) per sample
>3TB raw proteome data in ProteomicsDB



Hospital information systems

Often more than 50GB



Cancer patient records

>160k records at NCT



PubMed database

>23M articles



Medical sensor data

Scan of a single organ in 1s
creates 10GB of raw data



Prescription data

1.5B records from 10,000 doctors and
10M Patients (100 GB)



Clinical trials

Currently more than 30k
recruiting on ClinicalTrials.gov

Introduction to Digital Health

Data Management for
Digital Health, Winter
2023

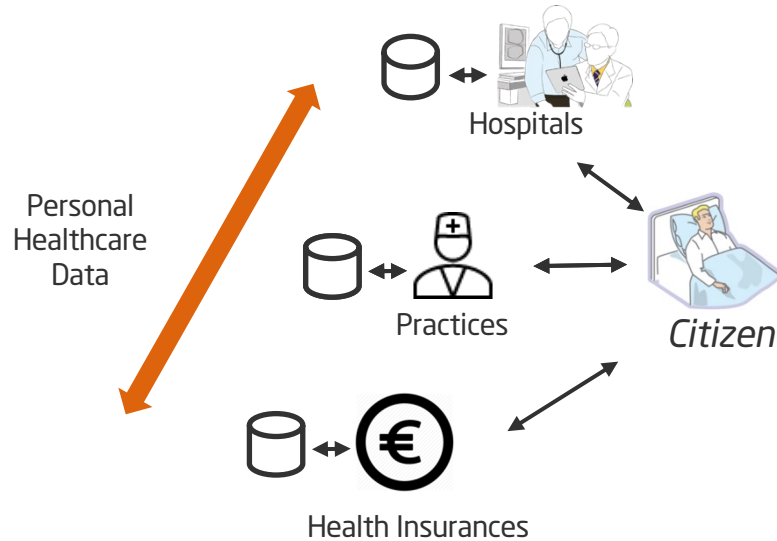
42

Citizens asks for Control of Personal Healthcare Data



Citizen

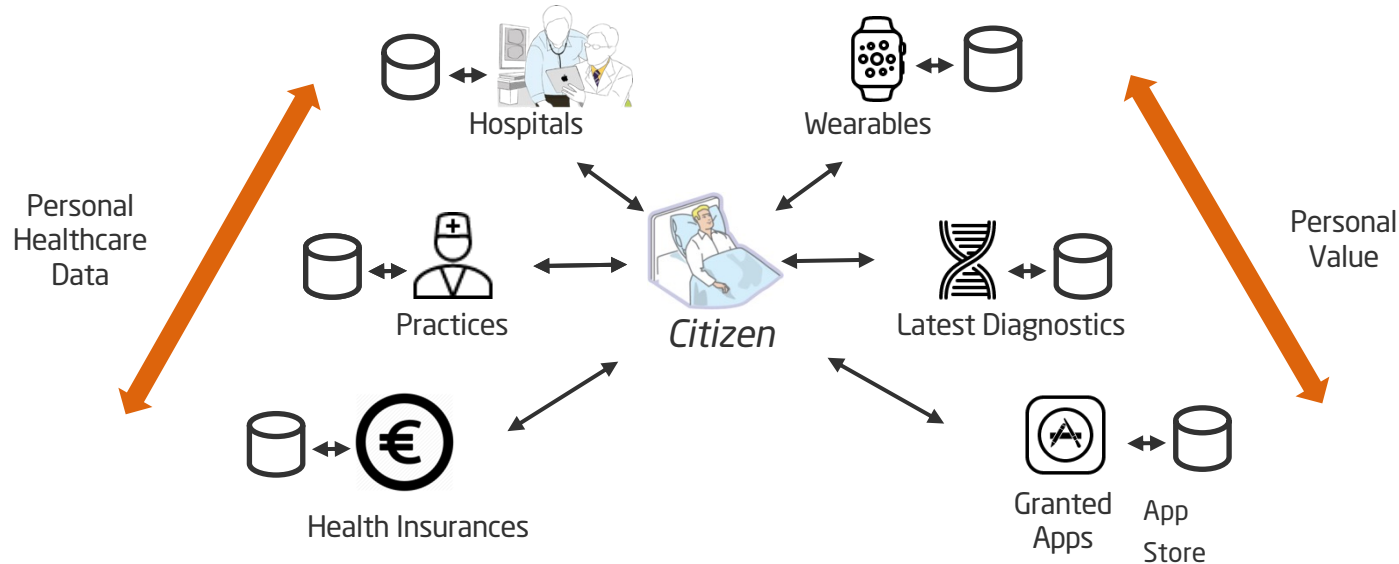
Integrated Control of Personal Healthcare Data



Introduction to Digital Health

Data Management for
Digital Health, Winter
2023

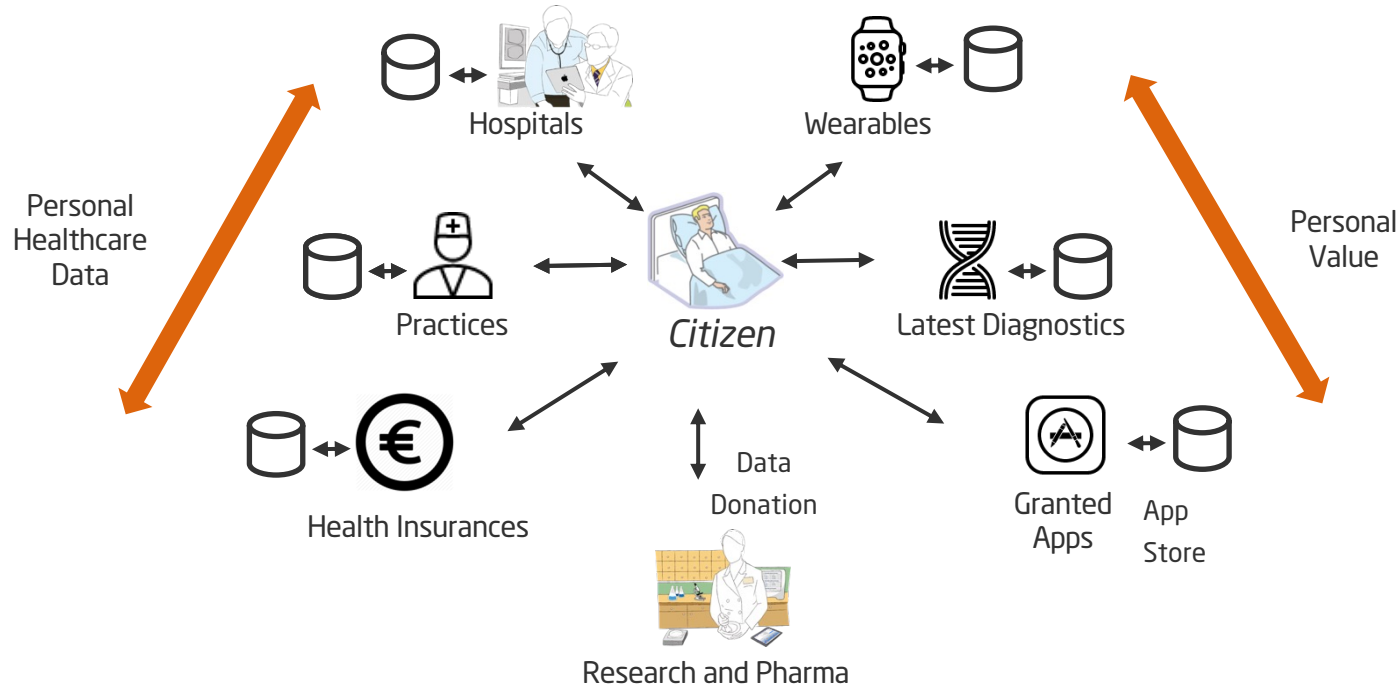
Personal Value of Citizens often through 3rd Party Vendors



Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
45

Sovereign Control of Personal Healthcare Data



Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
46

Our Vision

Real-time Access to Latest International Medical Knowledge



DOCTOR



Do Not Forget to Enroll!



We want you!



**Introduction to Digital
Health**

Data Management for
Digital Health, Winter
2023

48

Lecture Organization

Administrative Contacts

- Florian Borchert
- Dr. Matthieu-P. Schapranow

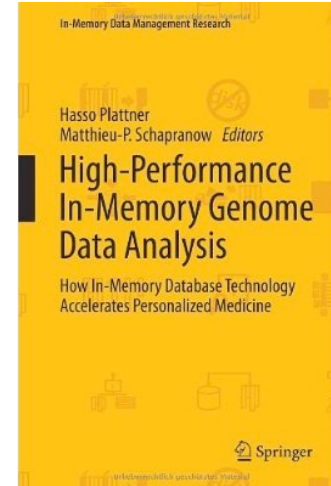
✉ <First Name>.<Last Name>@hpi.de



Digital Health Cluster
Hasso Plattner Institute Campus III
Rudolf-Breitscheid-Str. 187
14482 Potsdam, Germany

we.analyzegenomes.com

 @AnalyzeGenomes



Introduction to Digital Health

Data Management for
Digital Health, Winter
2023
49