Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Data Management for Digital Health
Lecture Kickoff

Borchert, Dr. Schapranow

Data Management for Digital Health
Winter 2023

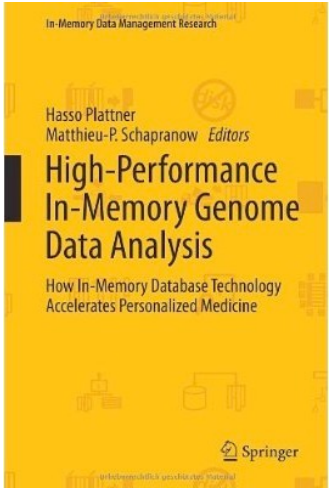
Lecture Organization
Administrative Contacts

- 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



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/

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)

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--cRElB5AZ--/1865smw5hbbt6jpg.jpg

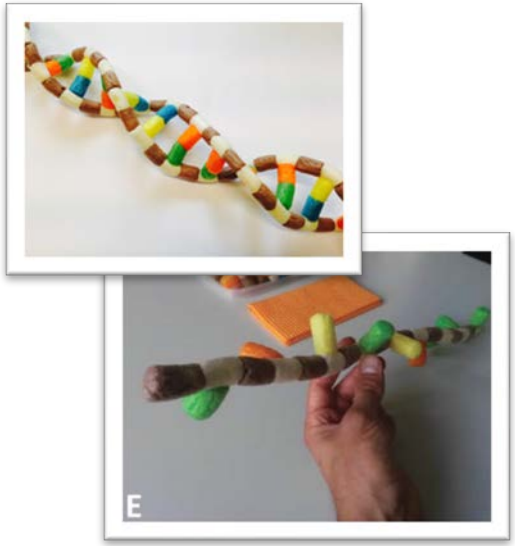
Lecture Organization
What we expect from you

- 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--cRElB5AZ--/1865smw5hbbt6jpg.jpg

Lecture Organization
Sneak Preview

- Hands-on exercises



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

- Experience the reality

Credit: Delaware State News/Dave Chambers

- Discuss with experts

Credit: Acuitus Medical

Agenda
Pillars of the Lecture

Medical Use Cases


Biology Recap


Oncology


Nephrology

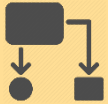

Infectious Diseases

Technology Foundation


Data Sources


Data Formats


Processing and Analysis


Software Architectures

Machine Learning

Data

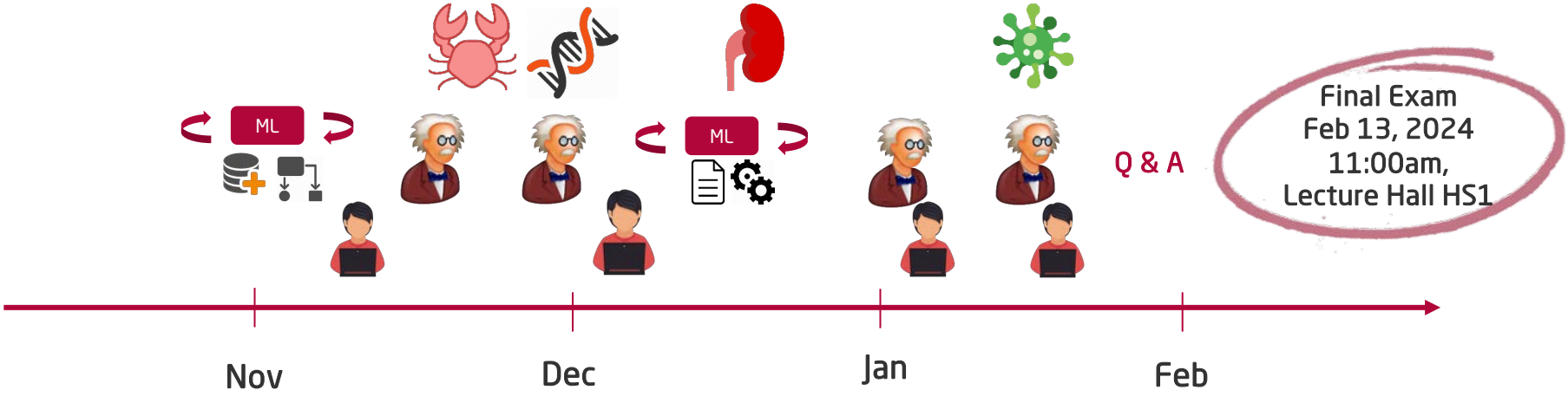
ML

Refine

Evaluate

Prediction + Probability

Lecture Schedule



- Lecture Kickoff
- Actors in Healthcare
- Digital Health Data

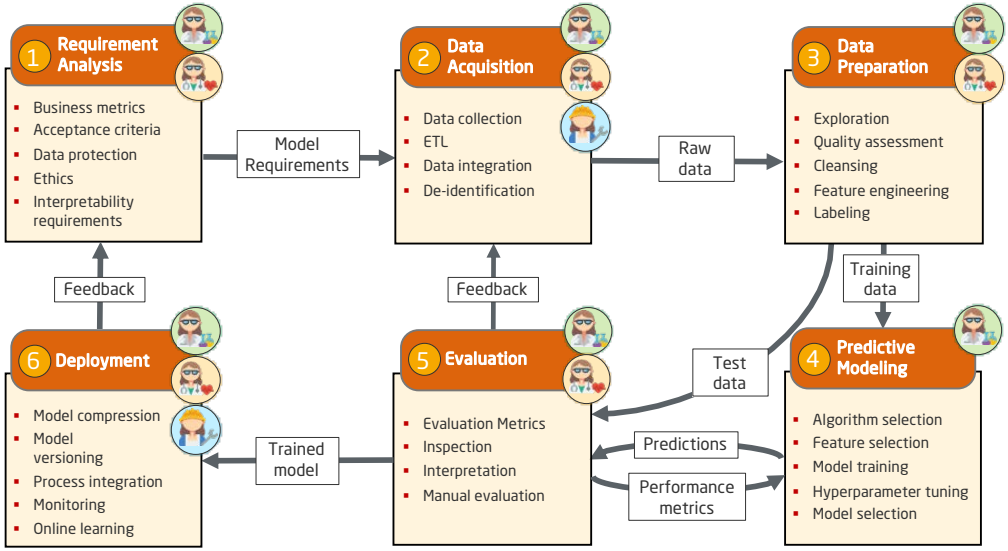
- Machine Learning (ML) Foundations
- Use Case Oncology
- Biology Recap

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

- Use Case Infectious Diseases
- Unsupervised ML

Lecture Contents
Machine Learning Foundations

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



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

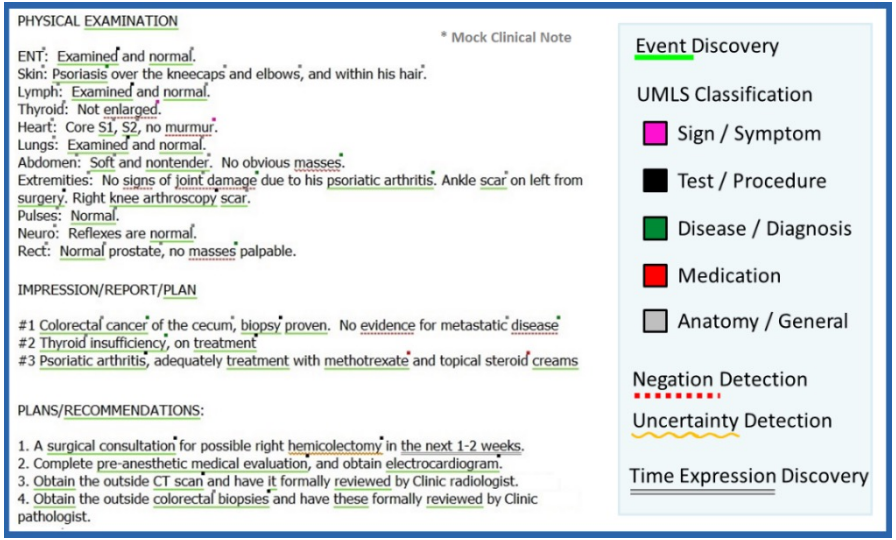
Lecture Contents
Biology Recap

- 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

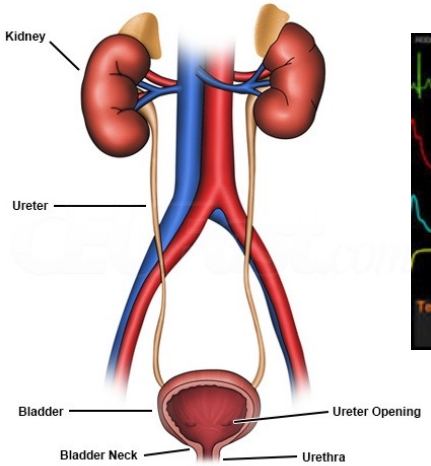
Lecture Contents
Processing of Medical Texts

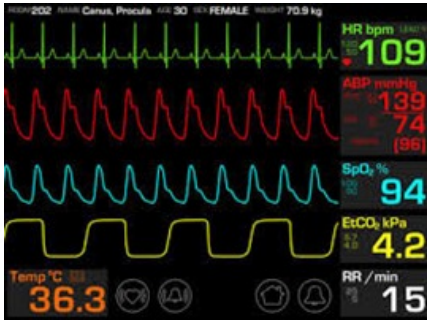
- Ontologies and knowledge engineering in digital health
- Methods for text mining and natural language processing
- Information extraction from unstructured text data
- Sequential machine learning models
- Real-world applications for text mining in oncology



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

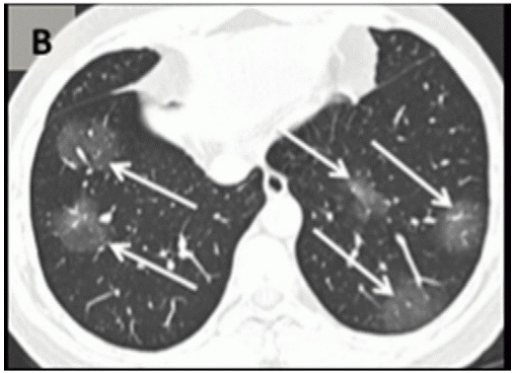




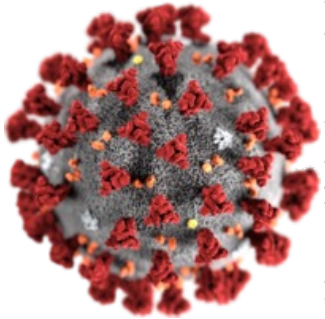
https://ceufast.com/course/urinary-tract-infections-the-unappreciated-giant ,http://www.appdropp.com/ios/ivital-signs/434101595

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



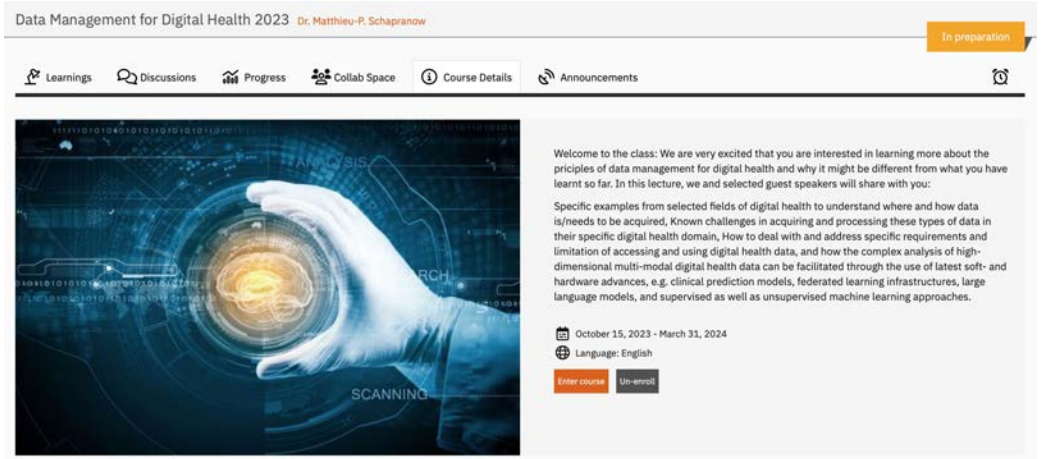
https://www.itnonline.com/content/coronavirus-ct-findings-linked-disease-time-course

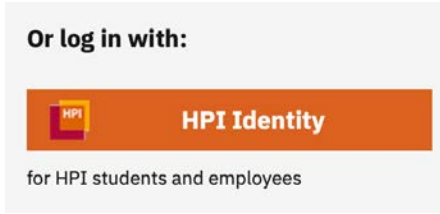


https://ceufast.com/course/urinary-tract-infections-the-unappreciated-giant ,http://www.appdropp.com/ios/ivital-signs/434101595

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





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 launch binder to start an Jupyter notebook server
  → no local installation required

README.md

Data Management for Digital Health 2022/23

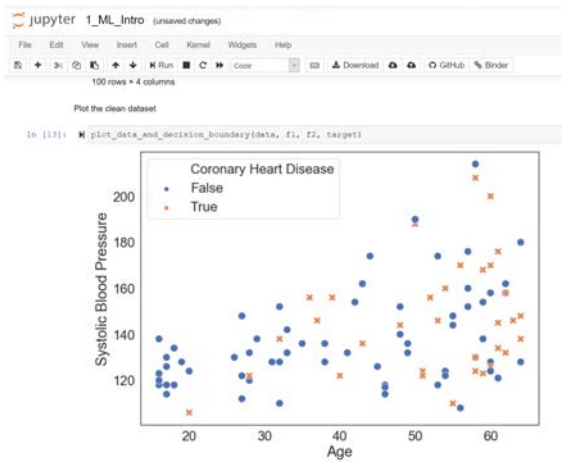
This repository contains the code to reproduce figures, metrics, and models for the 2022/23 version of the course.

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

launch binder

Contents:

- Week 2: Iterative ML Design Process



Open Questions?

https://msptelemarketing.com/wp-content/uploads/2017/09/19.jpg

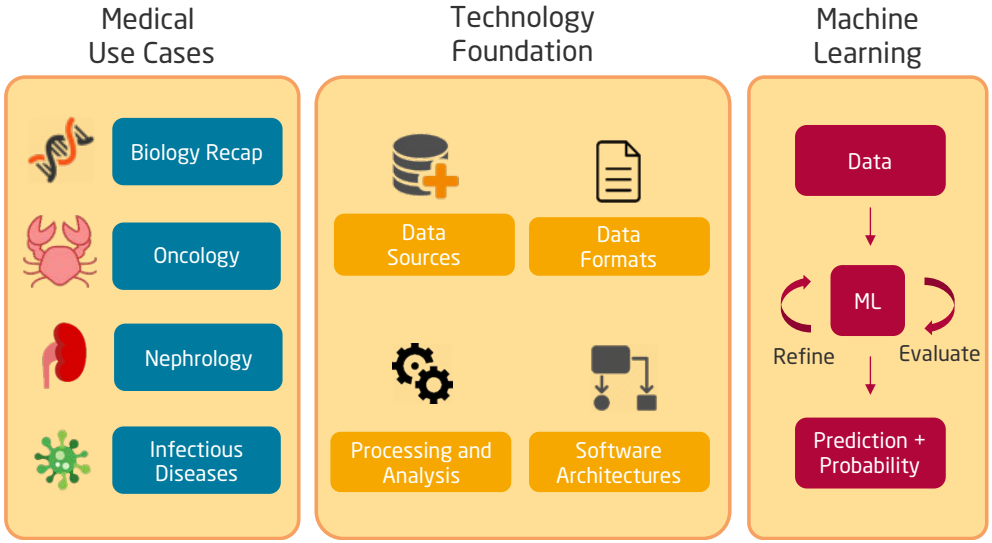
Interactive Poll
What are your favorite lecture topics?

A. Medical use cases

B. Technology foundation

C. Machine learning

D. All of them

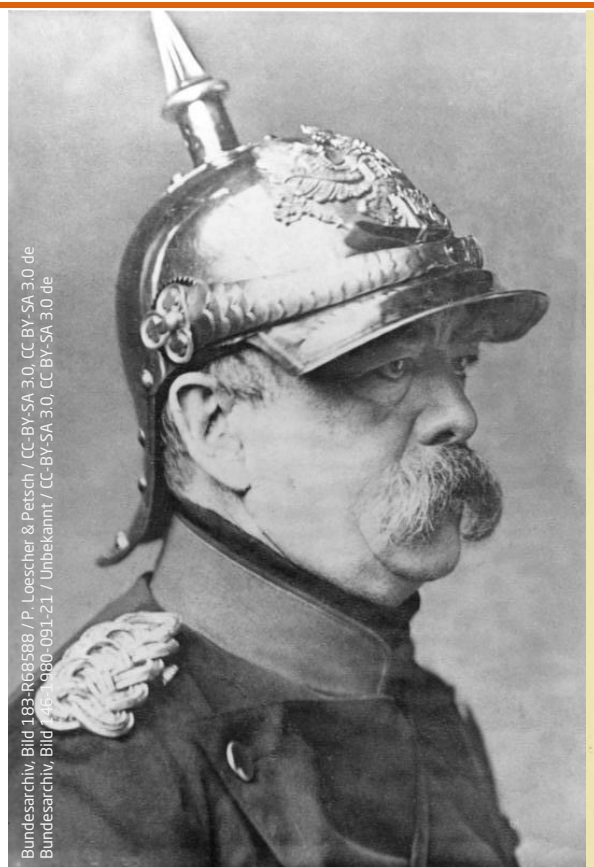


Interactive Poll
Who are the most important actors in healthcare?

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

https://msptelemarketing.com/wp-content/uploads/2017/09/19.jpg

History of the German Social Security Act



Introduction to Digital Health

Data Management for Digital Health, Winter 2023

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

German Social Security Triangle

Beneficiary
(Leistungsempfänger)

Service Provider
(Leistungserbringer)

Benefactor
(Leistungsträger)

https://www.familienratgeber.de/rechte-leistungen/kranken-pflegeversicherung/gkv-pv.php
Photo by HH E

German Social Security Triangle (SGB V)

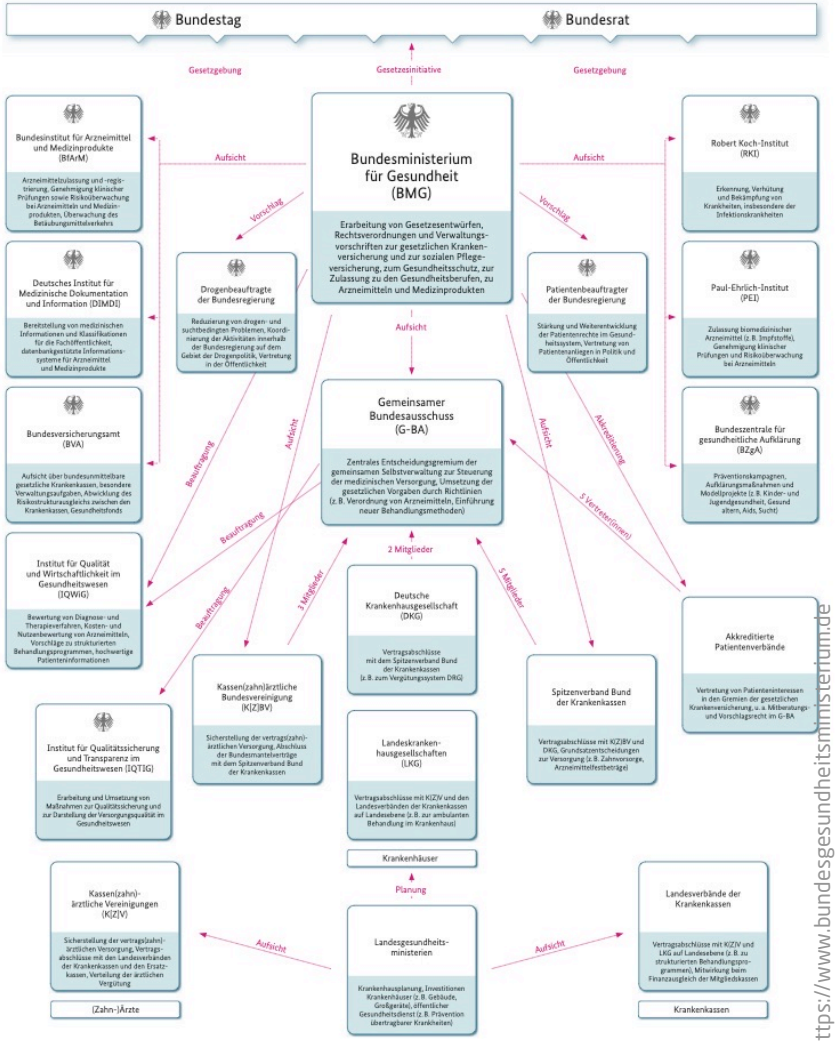
Service provider (Leistungserbringer)

- 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

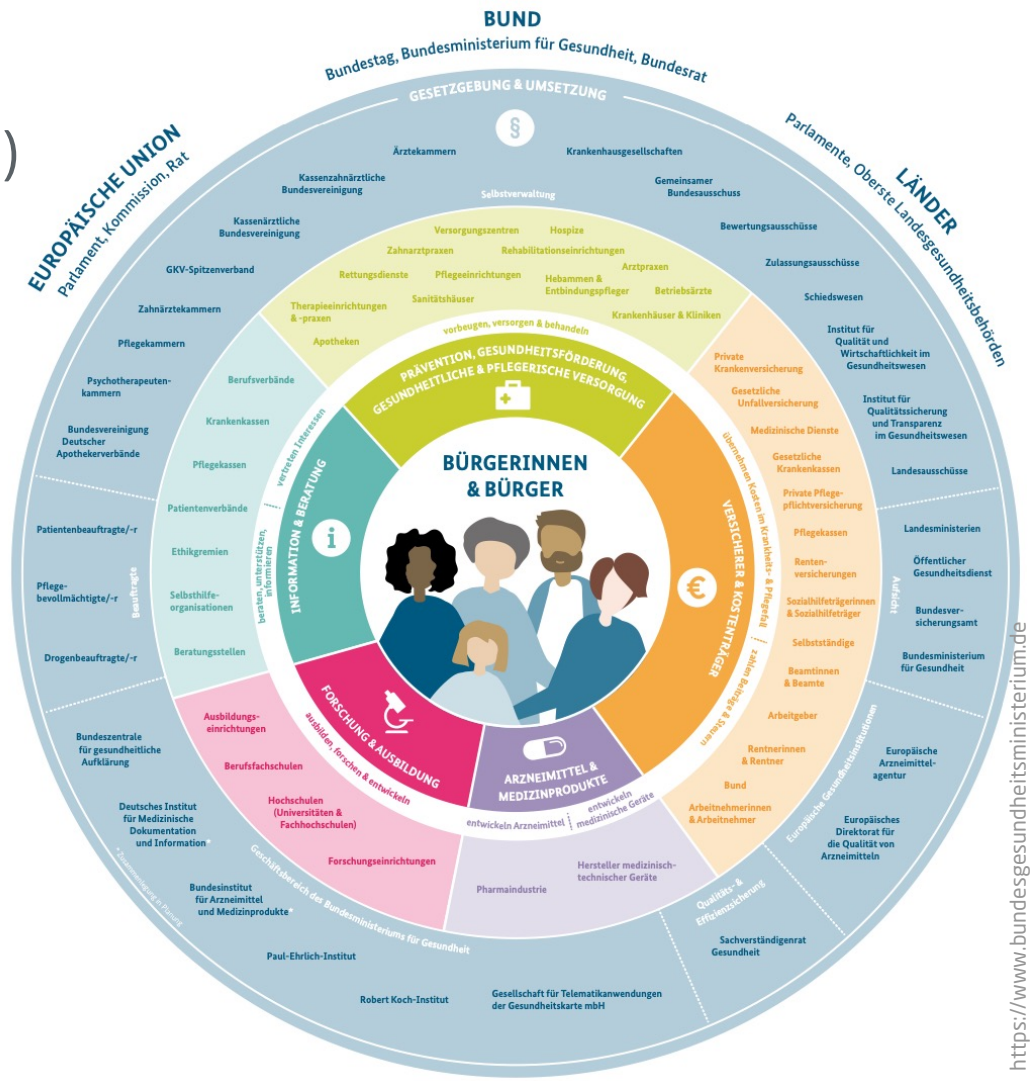
Who is missing?

Citizen



Actors within the German Healthcare System (cont'd)

- Updated version 2019



https://www.bundesgesundheitsministerium.de

Actors in Healthcare

- **Patients**

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

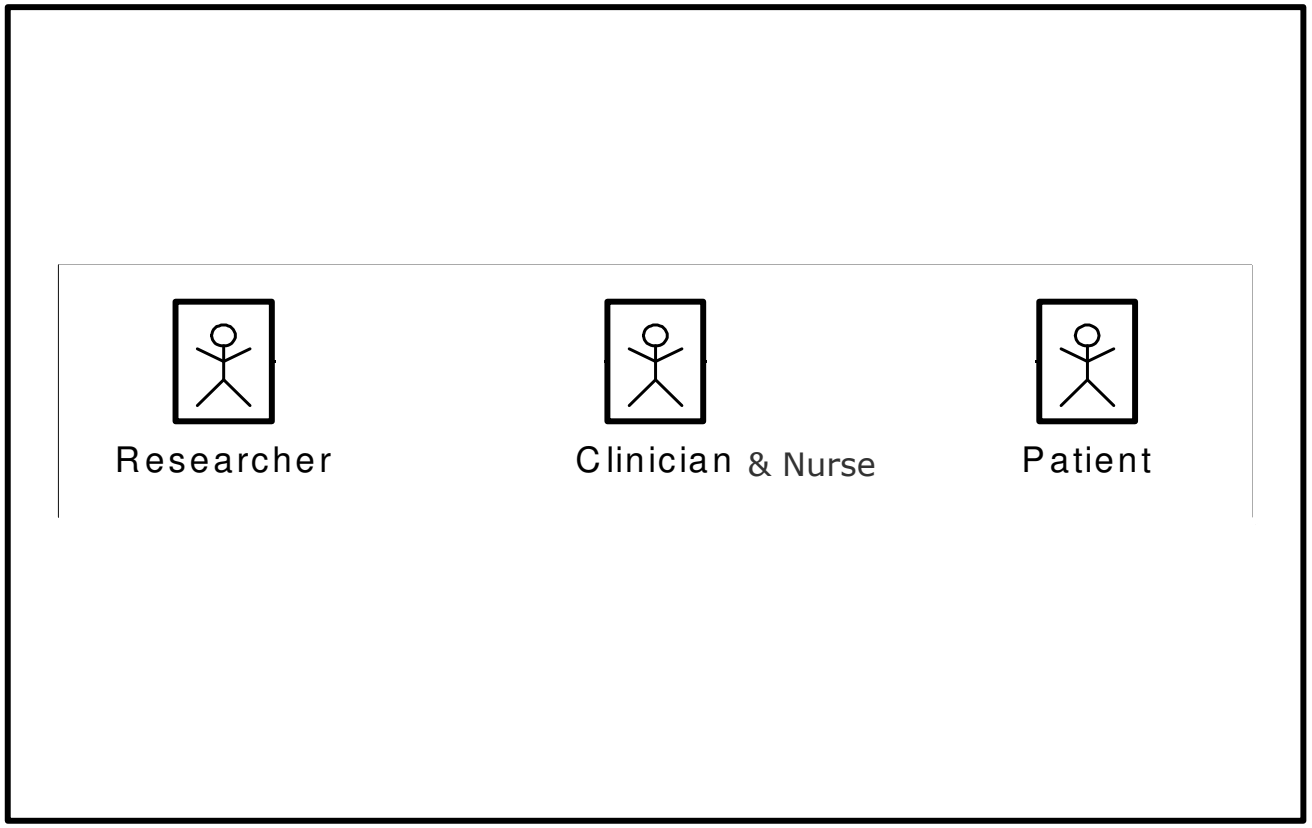
- **Clinicians and Nurses**

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

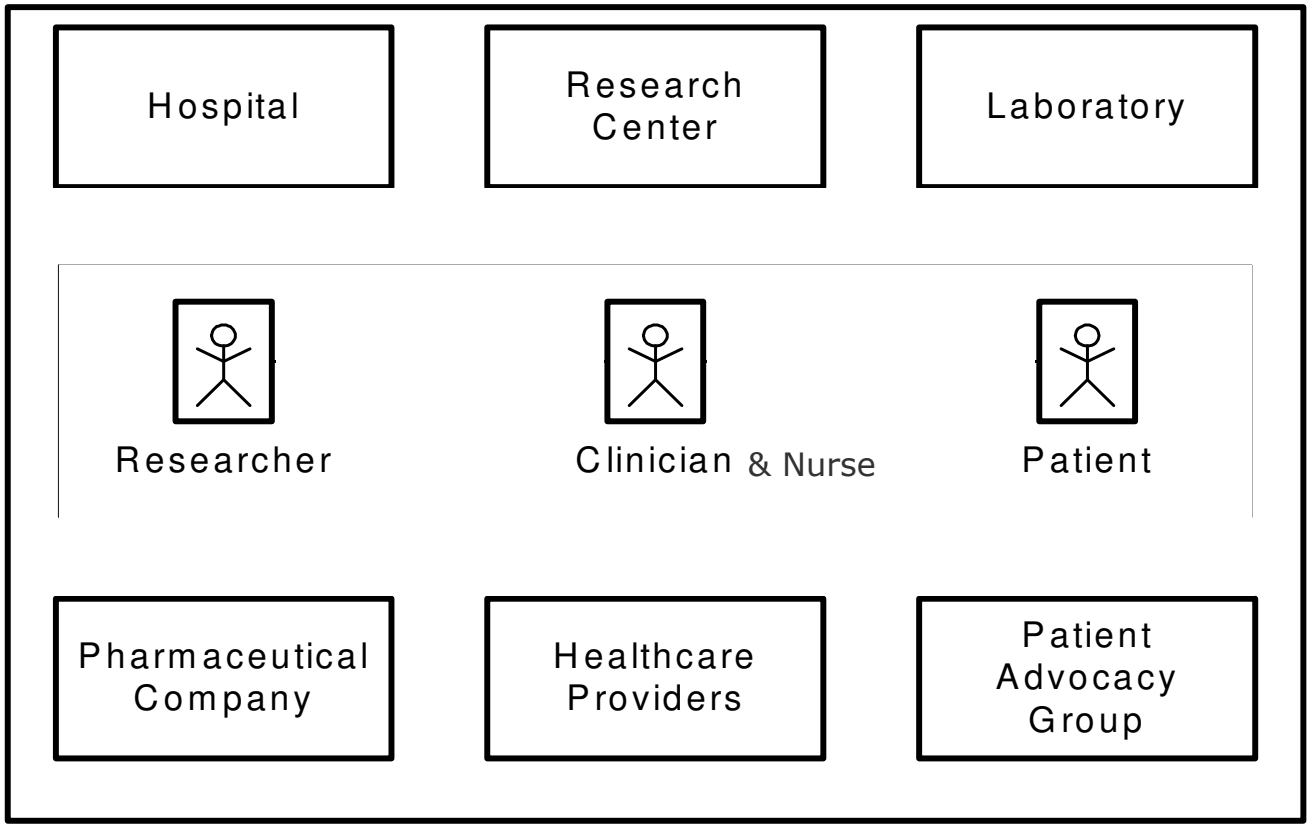
- **Researchers**

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

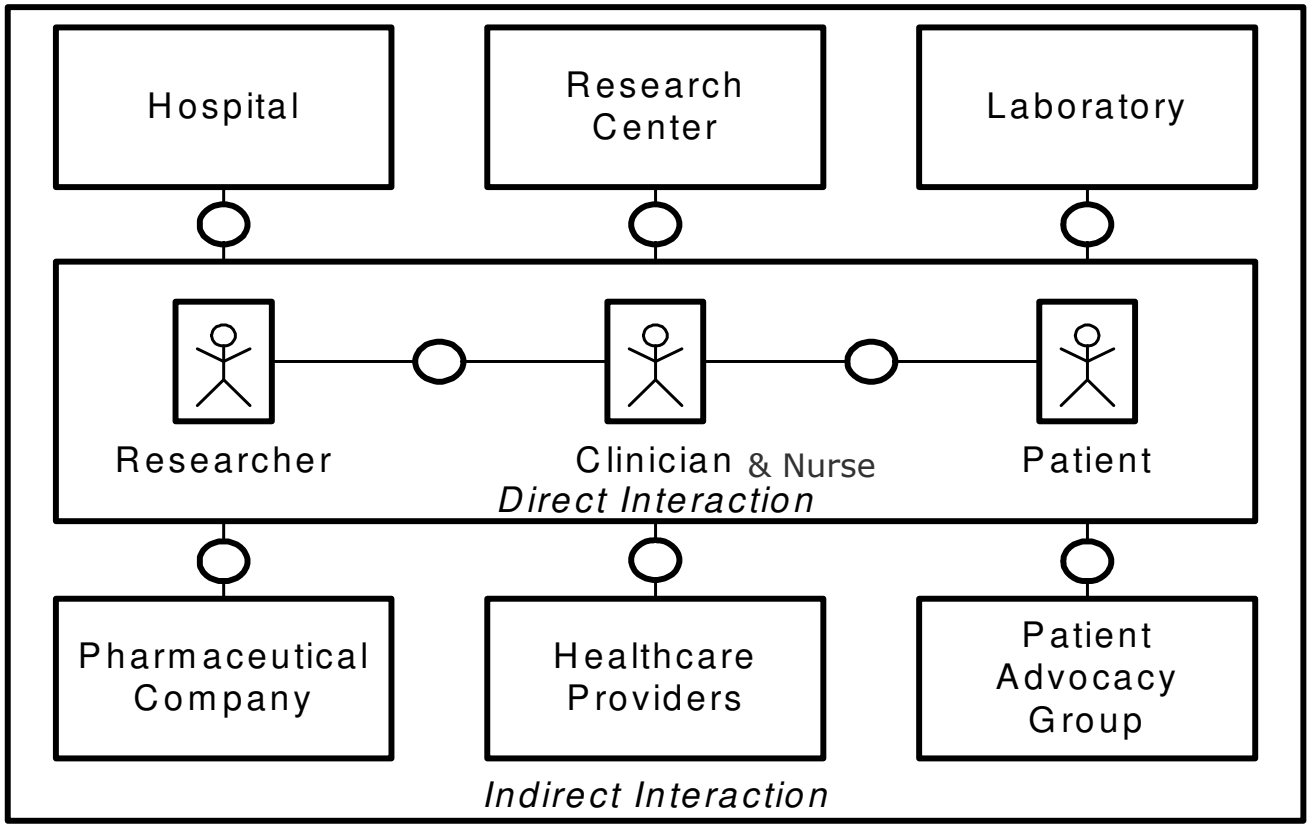
Intelligent Healthcare Networks in the 21st Century?



Intelligent Healthcare Networks in the 21st Century?



Intelligent Healthcare Networks in the 21st Century!

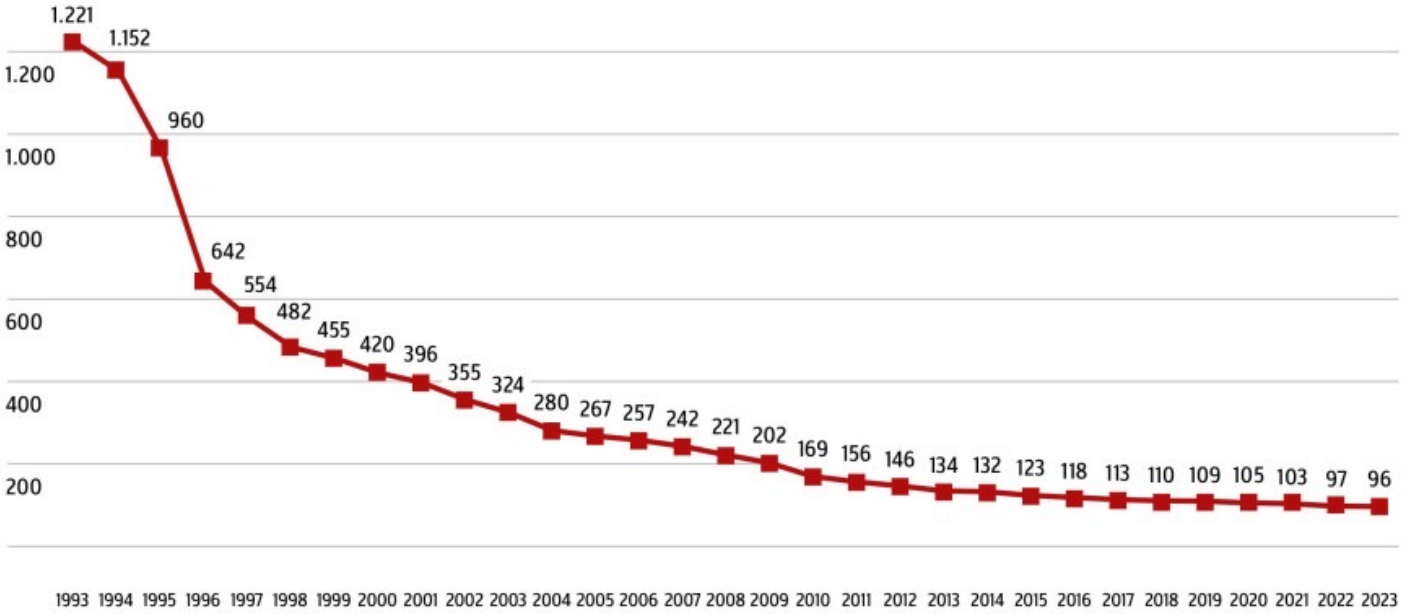


Key Aspects of the German Social Healthcare Systems

- 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

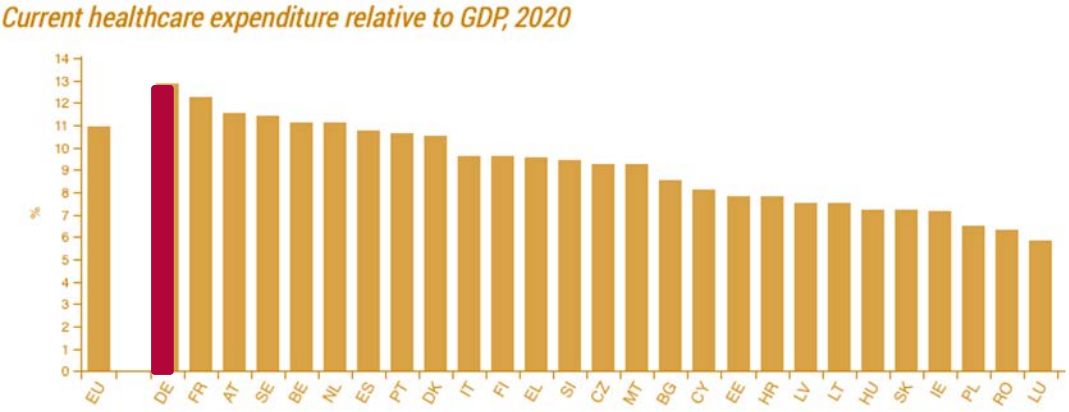
Photo by Thought Catalog

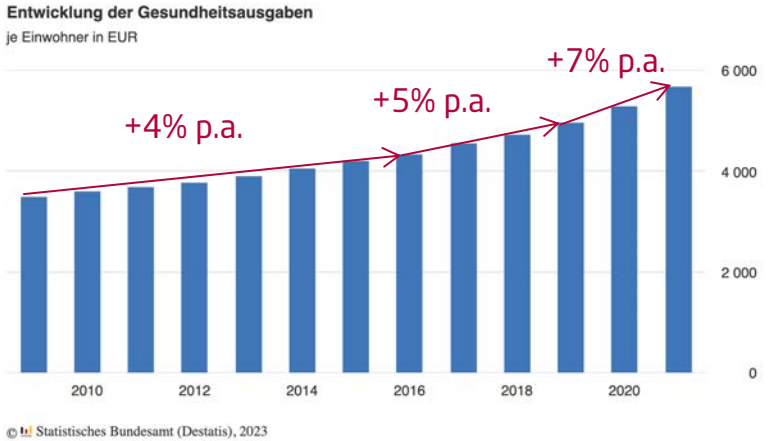
Number of Health Insurances in Germany



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

Facts You Should Know
Costs for Healthcare

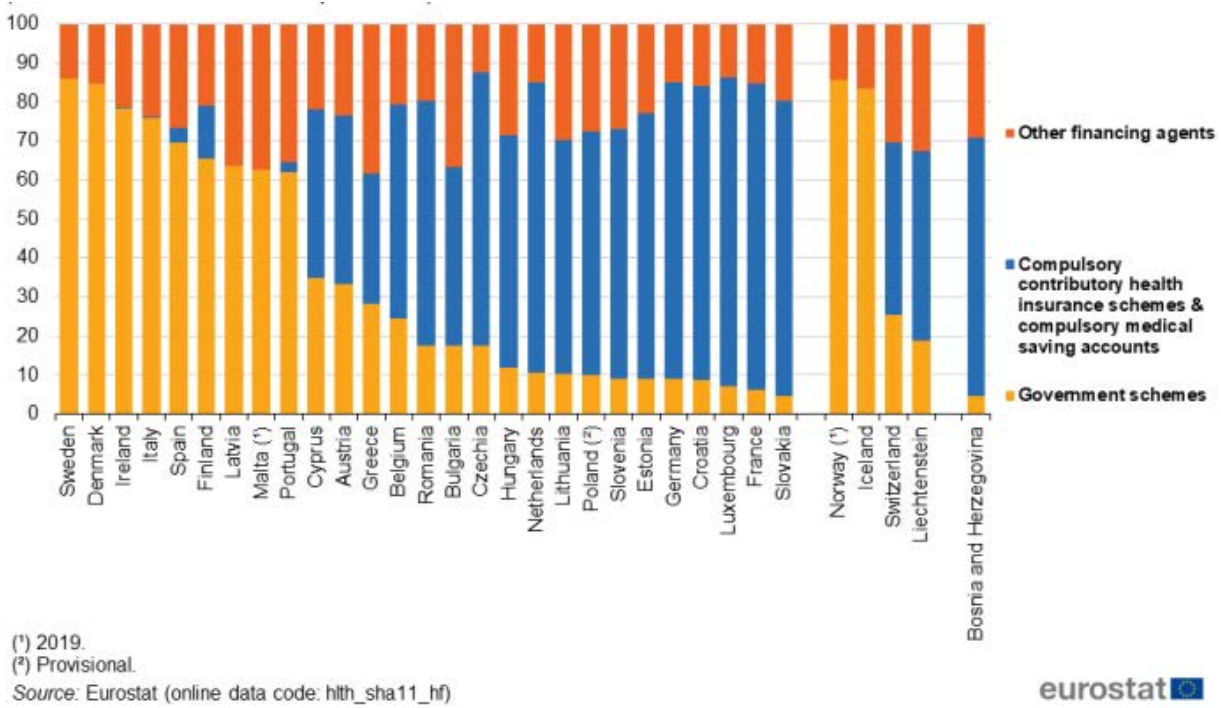




- Germany 5,699 EUR per capita resp. 13.2 % of GDP in 2021

OECD Health Statistics 2023, Eurostat Database, WHO, Global Health Expenditure Database, Destatis
https://ec.europa.eu/eurostat/statistics-explained/index.php/Healthcare_expenditure_statistics

Sources of Financing Healthcare Expenditures



(1) 2019.
(2) Provisional.
Source: Eurostat (online data code: hlth_sha11_hf)

eurostat

OECD Health Statistics 2020, Eurostat Database, WHO, Global Health Expenditure Database, Destatis
https://ec.europa.eu/eurostat/statistics-explained/index.php/Healthcare_expenditure_statistics

Federal Ministry of Health (BMG)

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

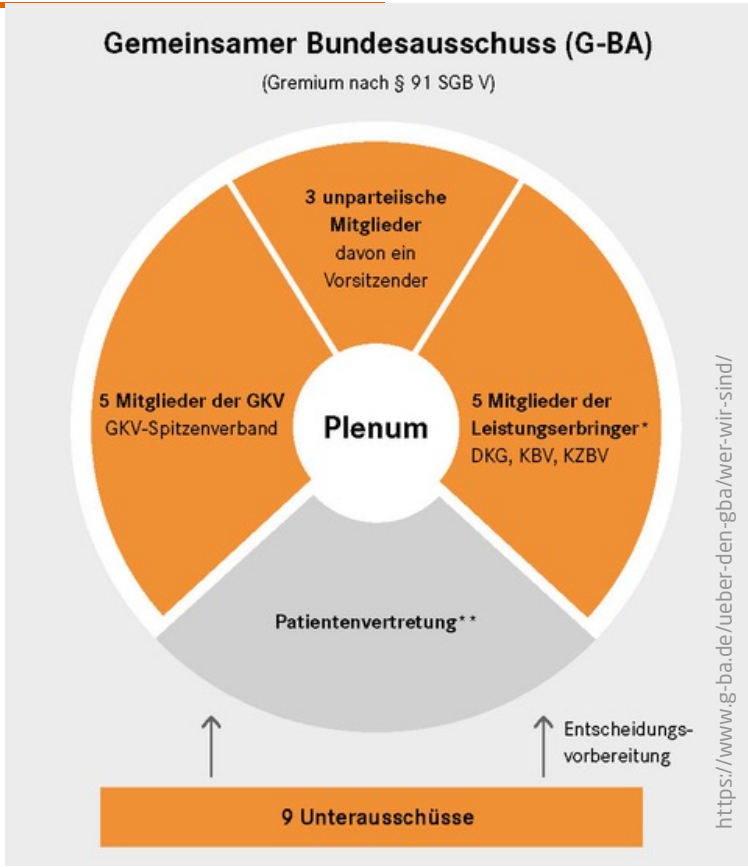
Ressortforschung des Bundesministeriums für Gesundheit

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

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.



Exchange of Healthcare Services

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

https://www.kite-uhn.com/labs

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)
- USA: Food and Drug Administration (FDA)

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 Kehlkopfoperierten 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.
- Lymphangioleiomyomatose, 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.

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

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

PubMed database
>23M articles

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

Cancer patient records
>160k records at NCT

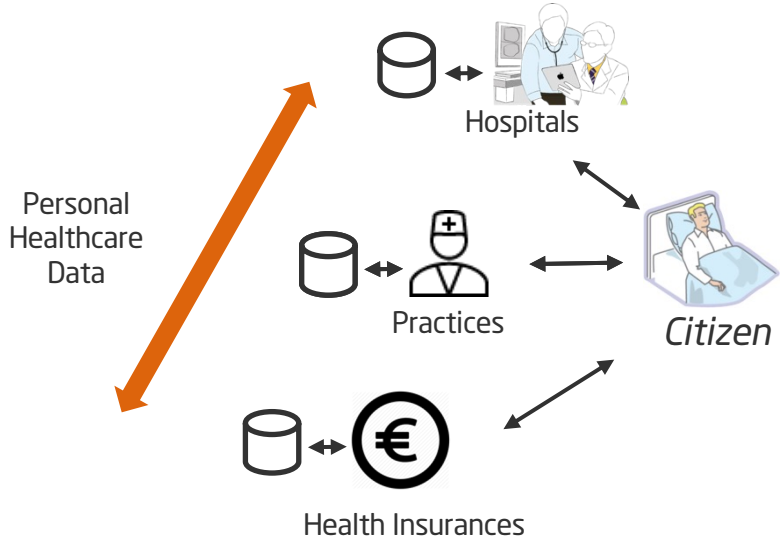
Prescription data
1.5B records from 10,000 doctors and
10M Patients (100 GB)

Clinical trials
Currently more than 30k
recruiting on ClinicalTrials.gov

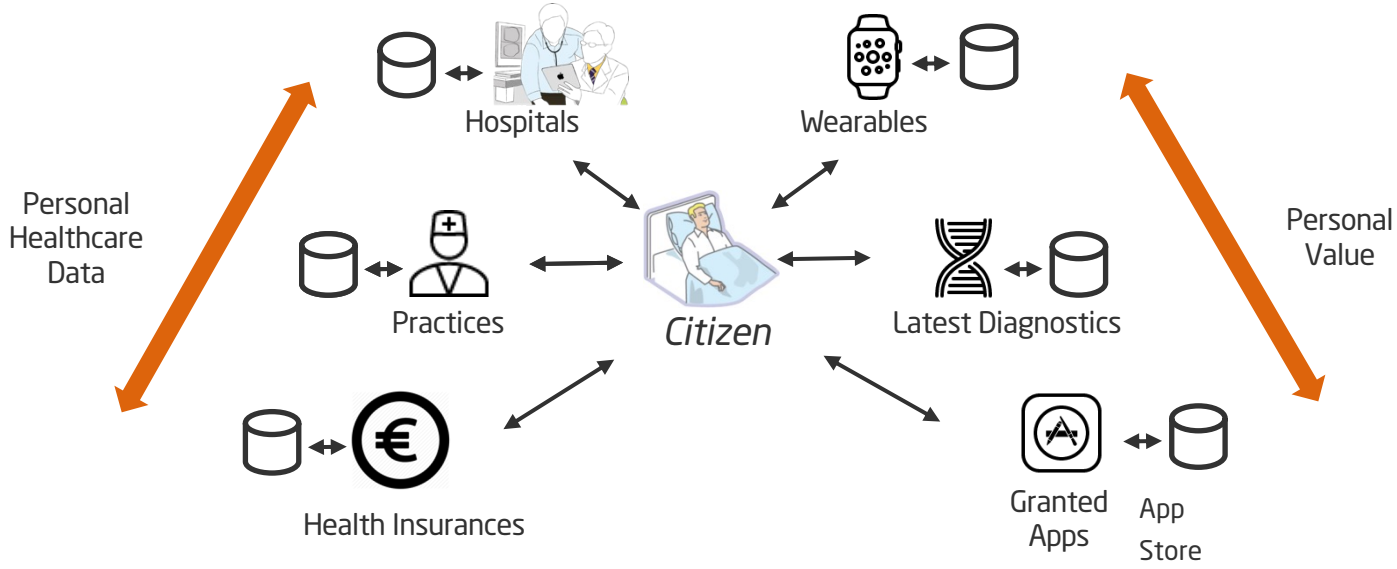
Citizens asks for Control of Personal Healthcare Data

Citizen

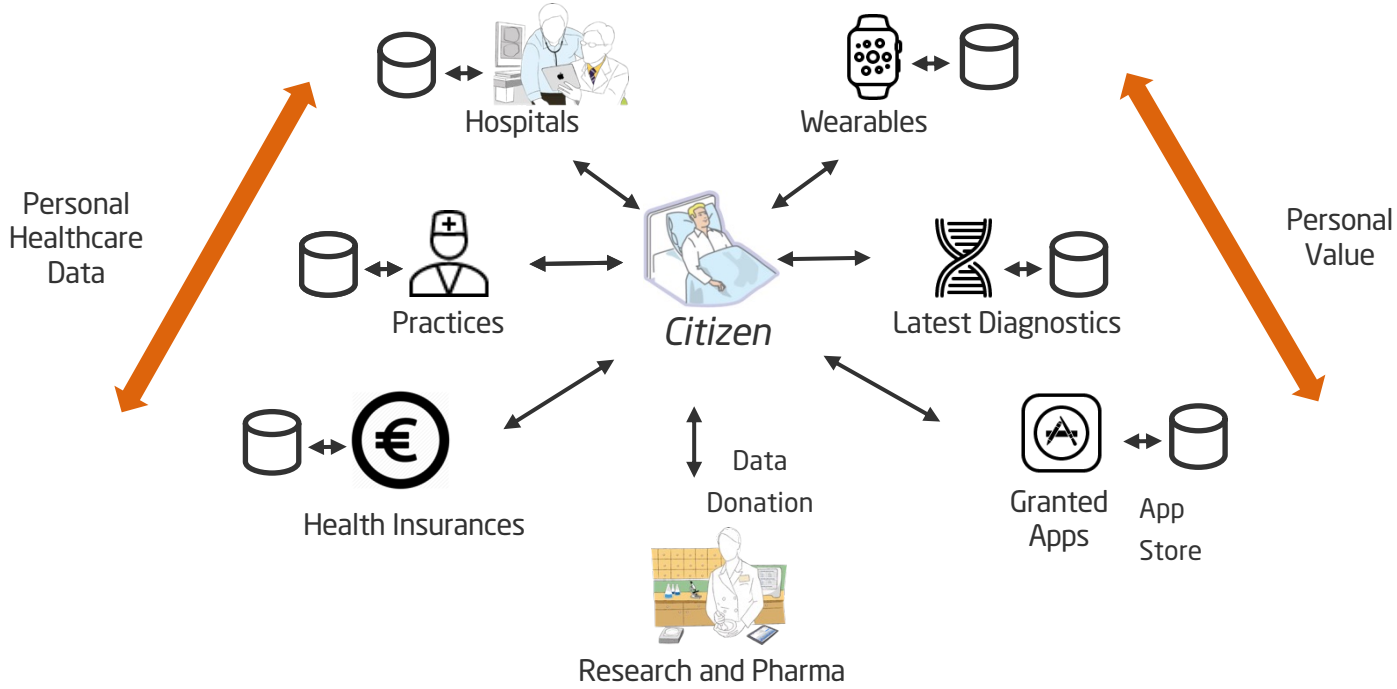
Integrated Control of Personal Healthcare Data



Personal Value of Citizens often through 3rd Party Vendors



Sovereign Control of Personal Healthcare Data



Our Vision
Real-time Access to Latest International Medical Knowledge

Do Not Forget to Enroll!

We want you!

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