GDSD 2025

DATA-DRIVEN MEDICINE

PROF. ANNE SCHWERK

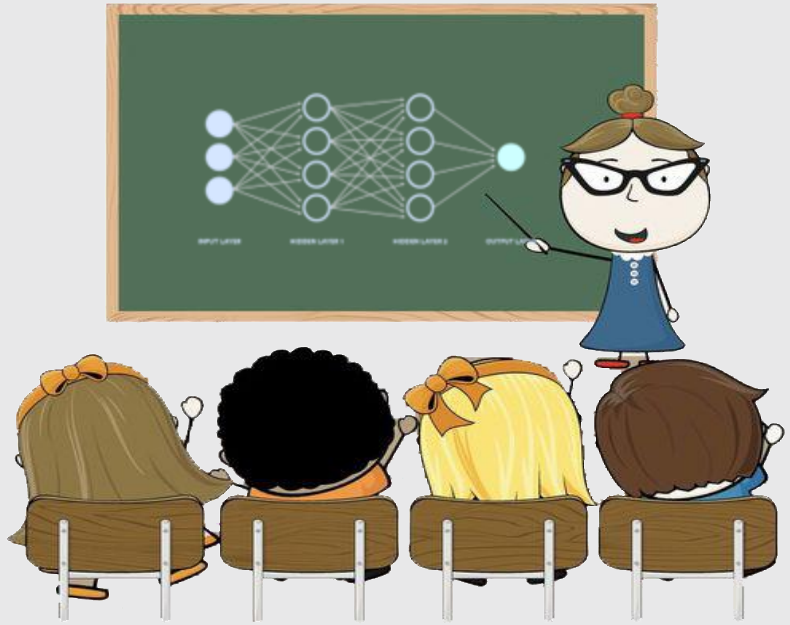
- INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES
- CHARITÉ / BERLIN INSTITUTE OF HEALTH

AI-DRIVEN MEDICINE

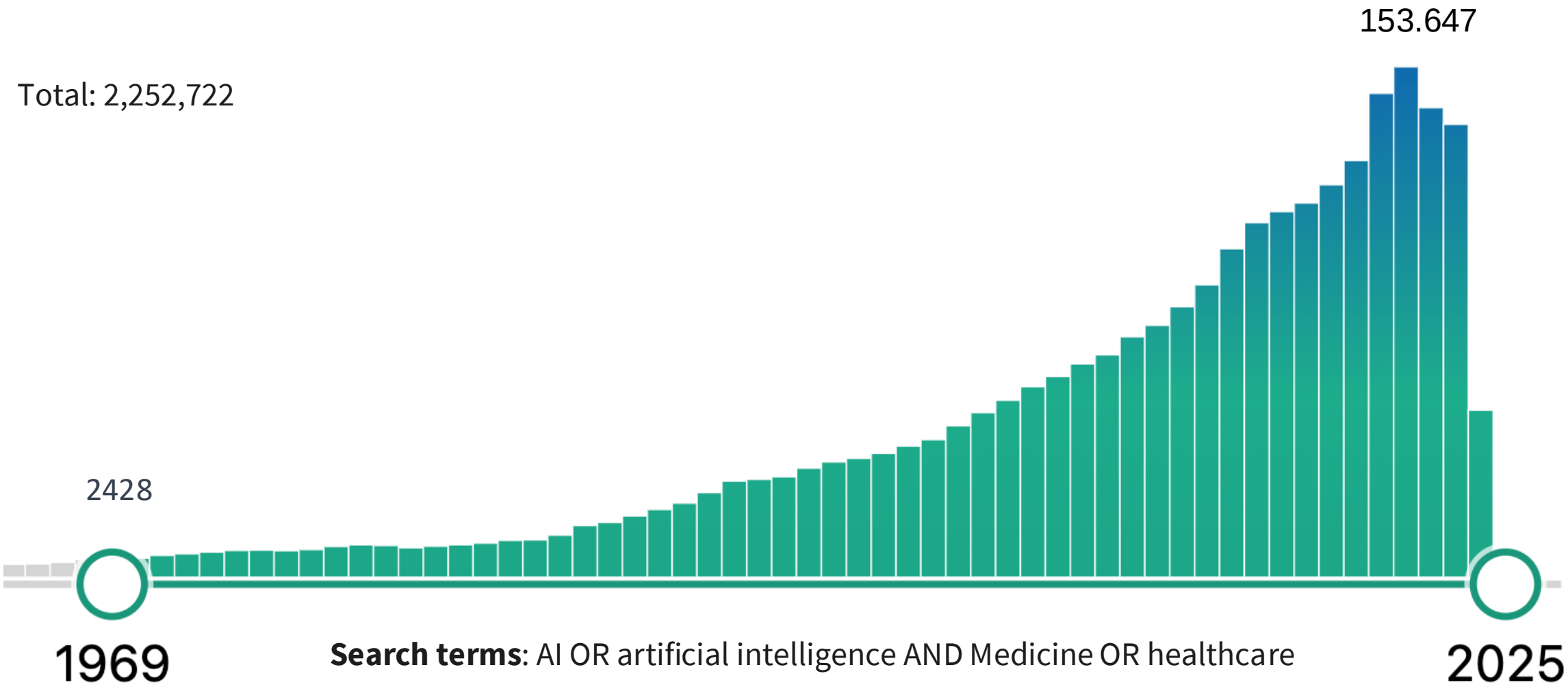
INTERNATIONALE HOCHSCHULE

#1

Overview: AI and Healthcare Data



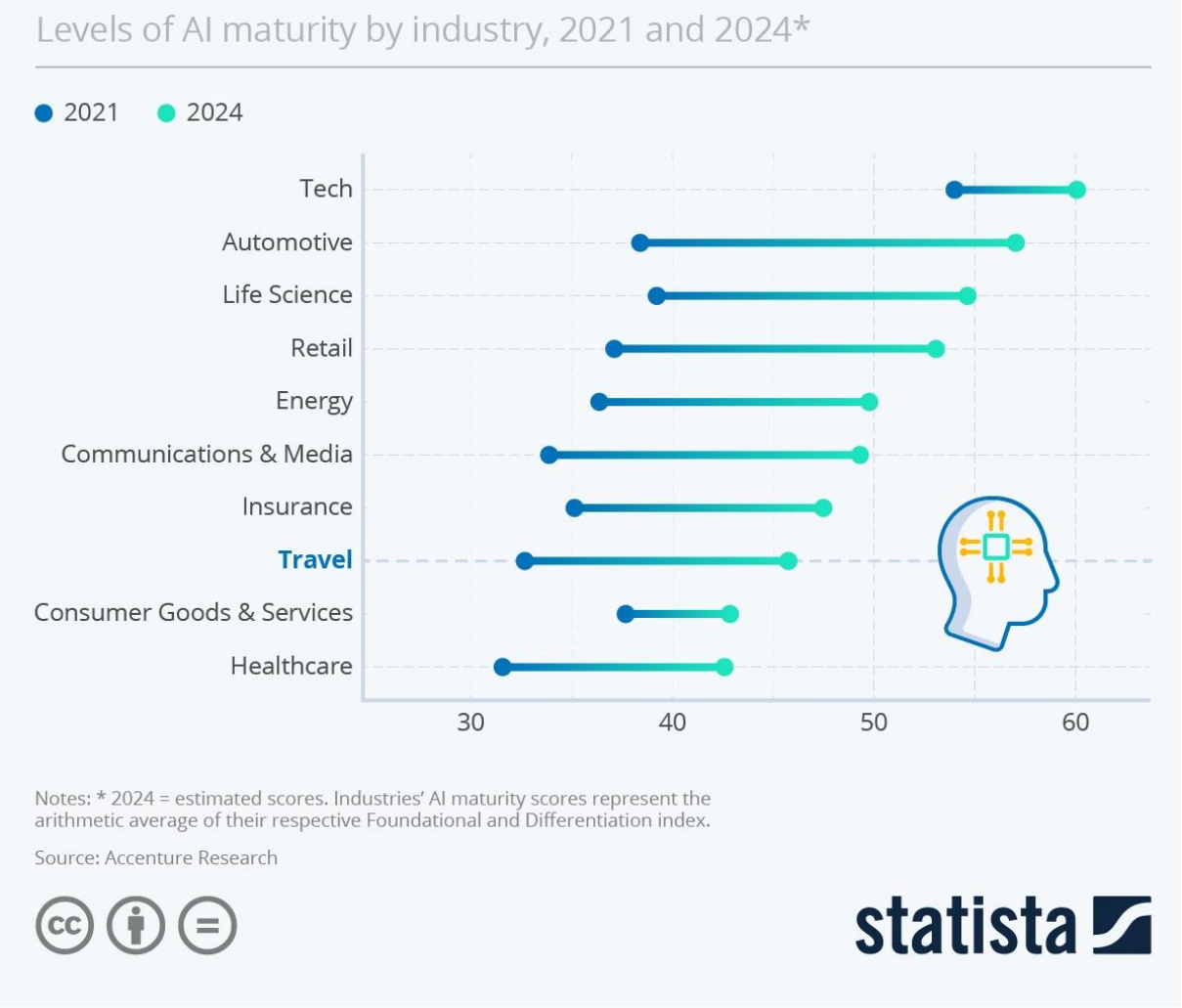
INCREASED RELEVANCE OF MEDICAL AI
PUBLICATIONS



Search terms: AI OR artificial intelligence AND Medicine OR healthcare

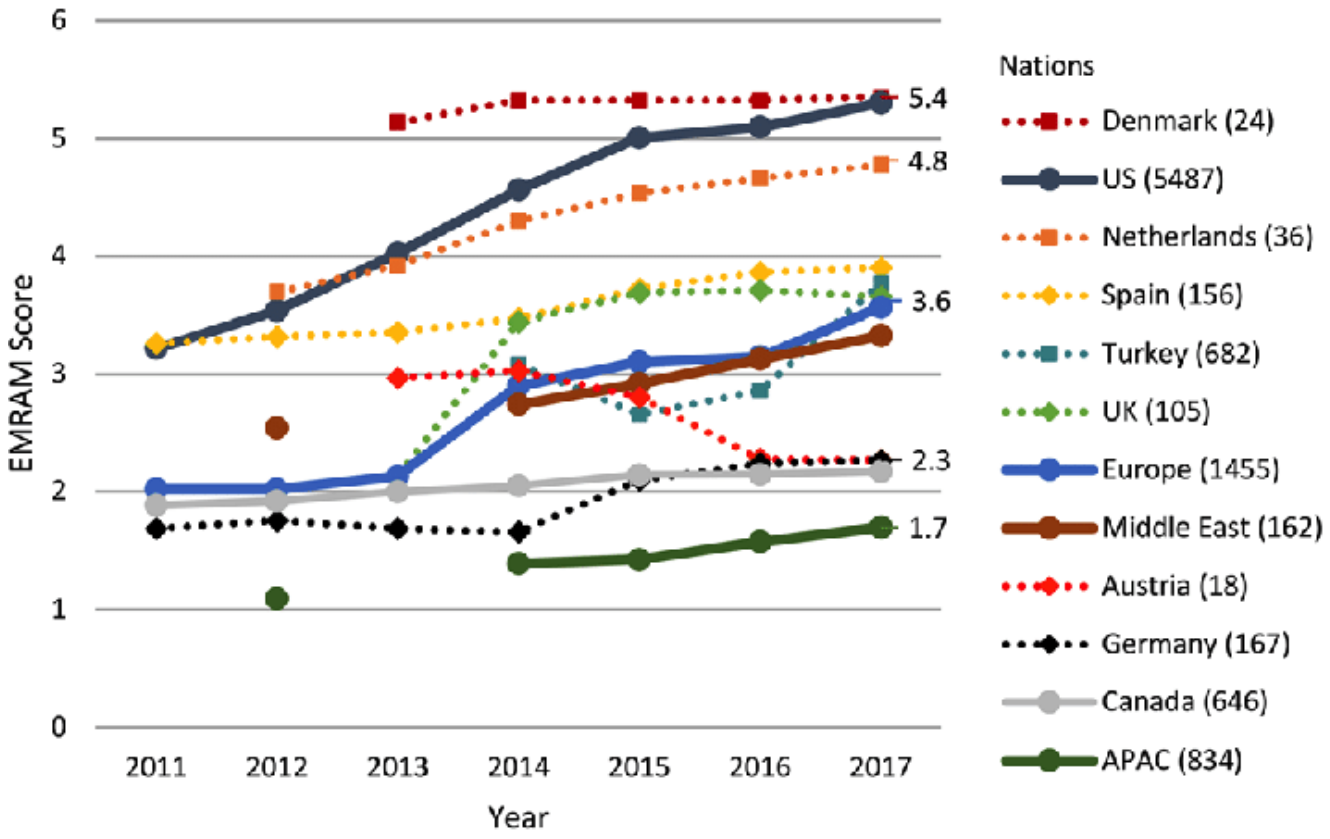
Source of the image: PubMed

ADOPTION OF AI PER INDUSTRY



Source: https://indatalabs.com/blog/ai-adoption-by-industry

LACK OF DIGITALISATION

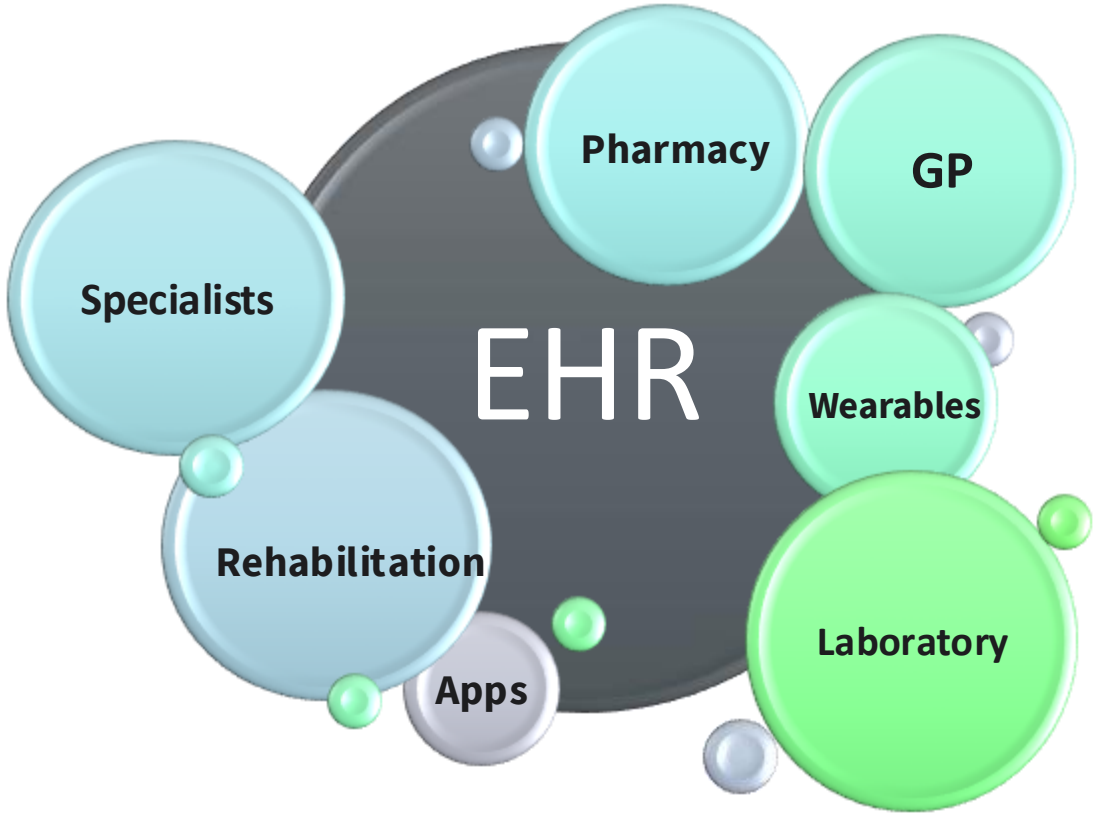


STAGE	HIMSS Analytics EMRAM EMR Adoption Model Cumulative Capabilities
7	Complete EMR; External HIE; Data Analytics, Governance, Disaster Recovery, Privacy and Security
6	Technology Enabled Medication, Blood Products, and Human Milk Administration; Risk Reporting; Full CDS
5	Physician documentation using structured templates; Intrusion/Device Protection
4	CPOE with CDS; Nursing and Allied Health Documentation; Basic Business Continuity
3	Nursing and Allied Health Documentation; eMAR; Role-Based Security
2	CDR; Internal Interoperability; Basic Security
1	Ancillaries - Laboratory, Pharmacy, and Radiology/Cardiology information systems; PACS; Digital non-DICOM image management
0	All three ancillaries not installed

Bukowski et al. (2020); https://www.nexus-marabu.de/nachricht/neue-himss-emram-kriterien

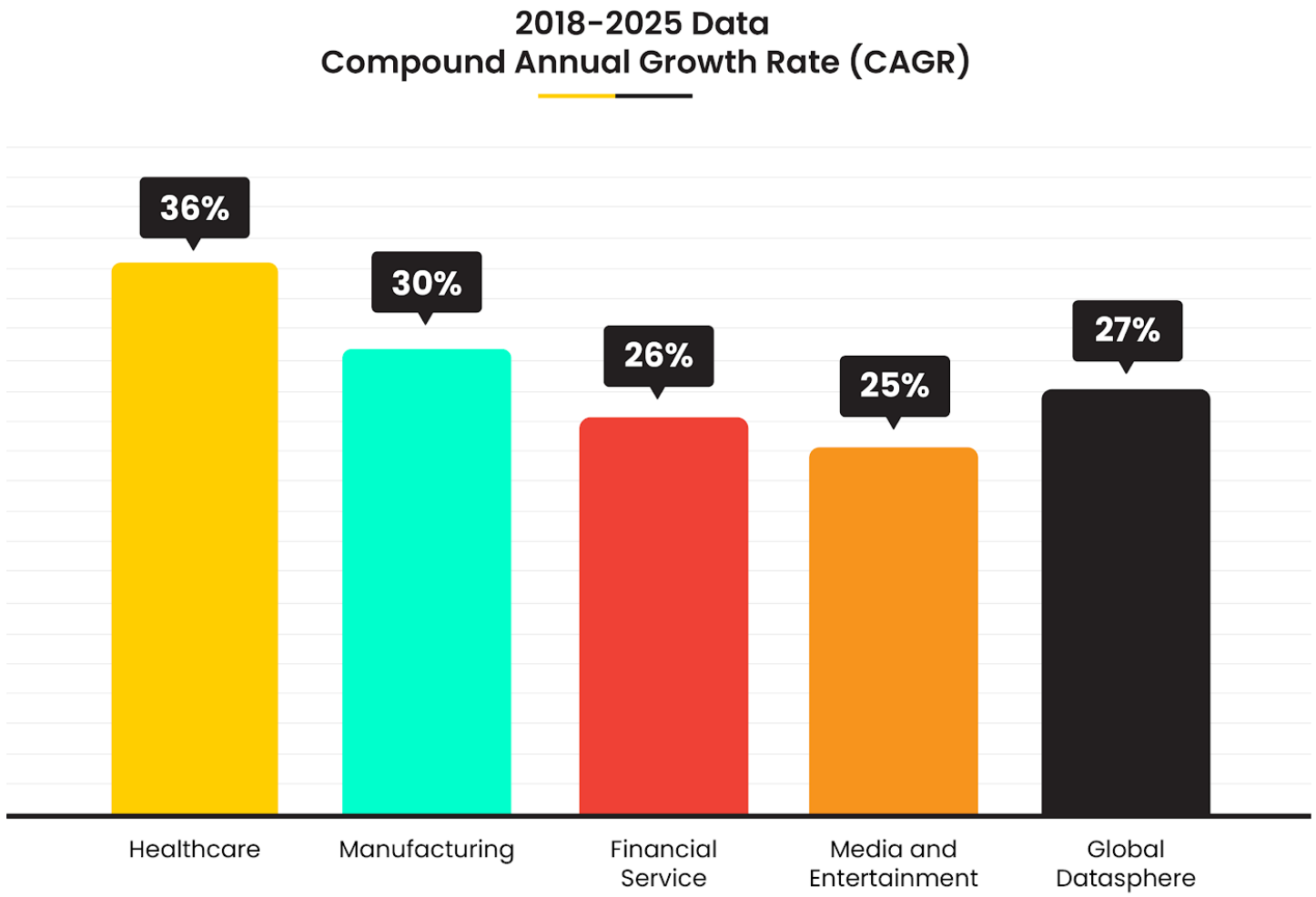
CHALLENGES: HEALTHCARE DATA

- Lack of data & inhibited sharing:
  - Labeled data
  - Benchmark data sets & open data
  - Data silos
  - Lack of cross-validation options
  - Lack of standardization/ interoperability
- Data quality:
  - Imbalanced data
  - Missing data
  - Incomplete data
  - Standardization



Source of the image: Schwerk (2023)

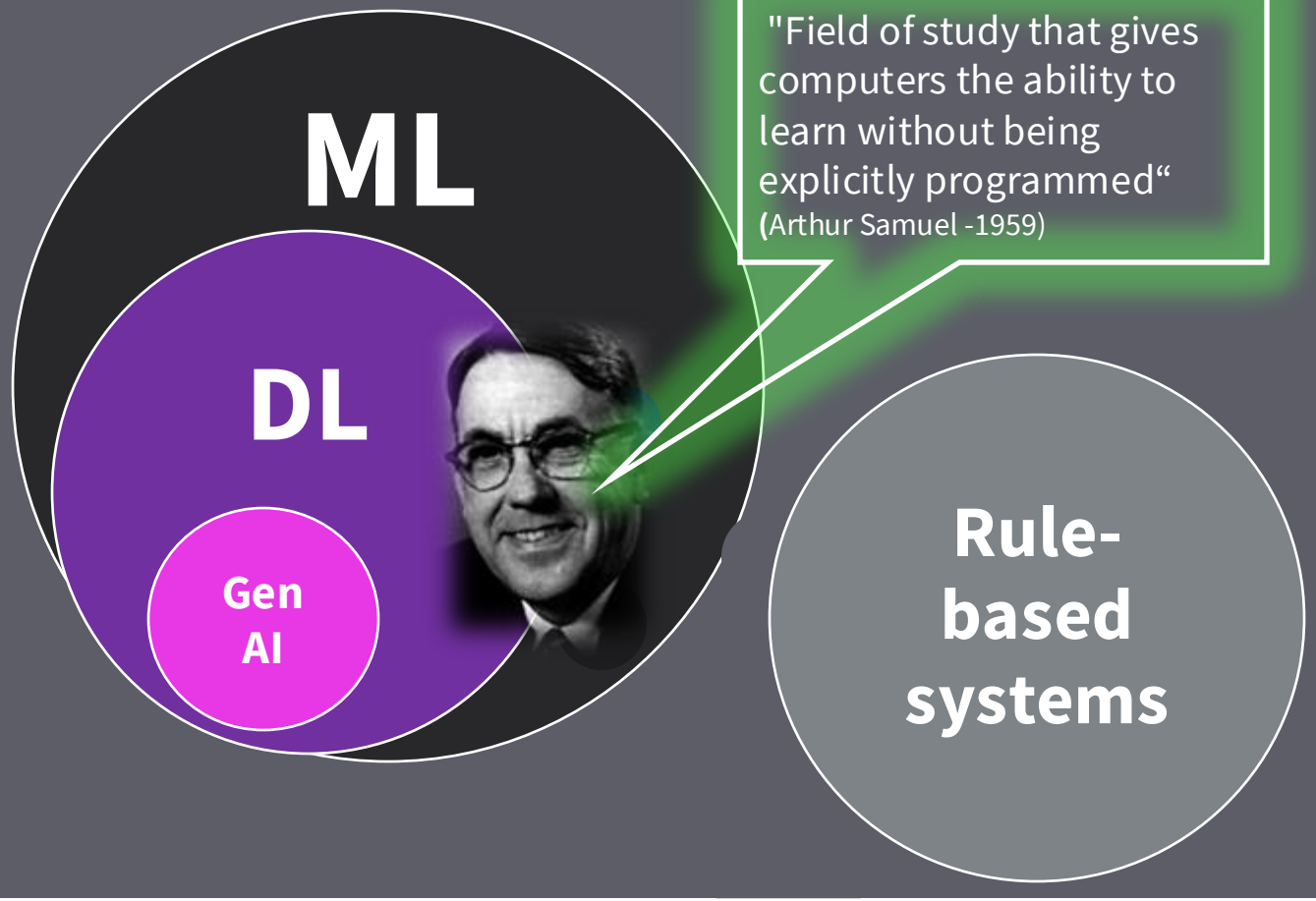
HEALTHCARE DATA INCREASE



Source: https://www.rbccm.com/en/gib/healthcare/episode/the_healthcare_data_explosion

DEFINITION: AI

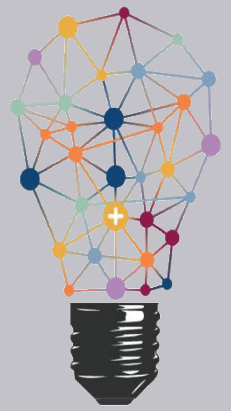
Artificial Intelligence: IT systems with human-like behavior – based on statistical & mathematical models



THE END OF THEORY - DEDUCTION VERSUS INDUCTION

INTERNATIONALE HOCHSCHULE

Deduction

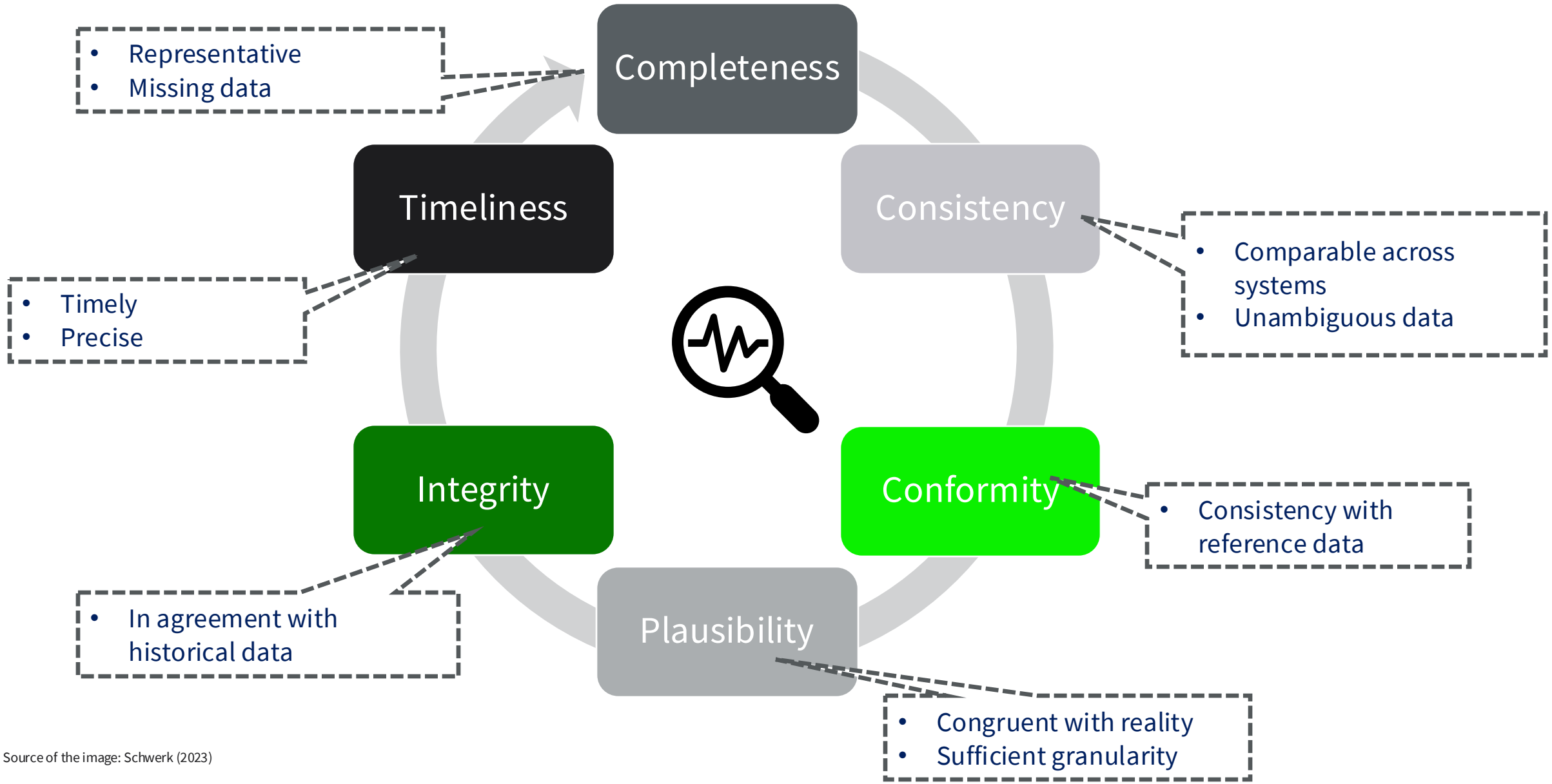


1. Hypothesis
2. Experiment
3. Data collection
4. Data analysis
5. Validation

Induction

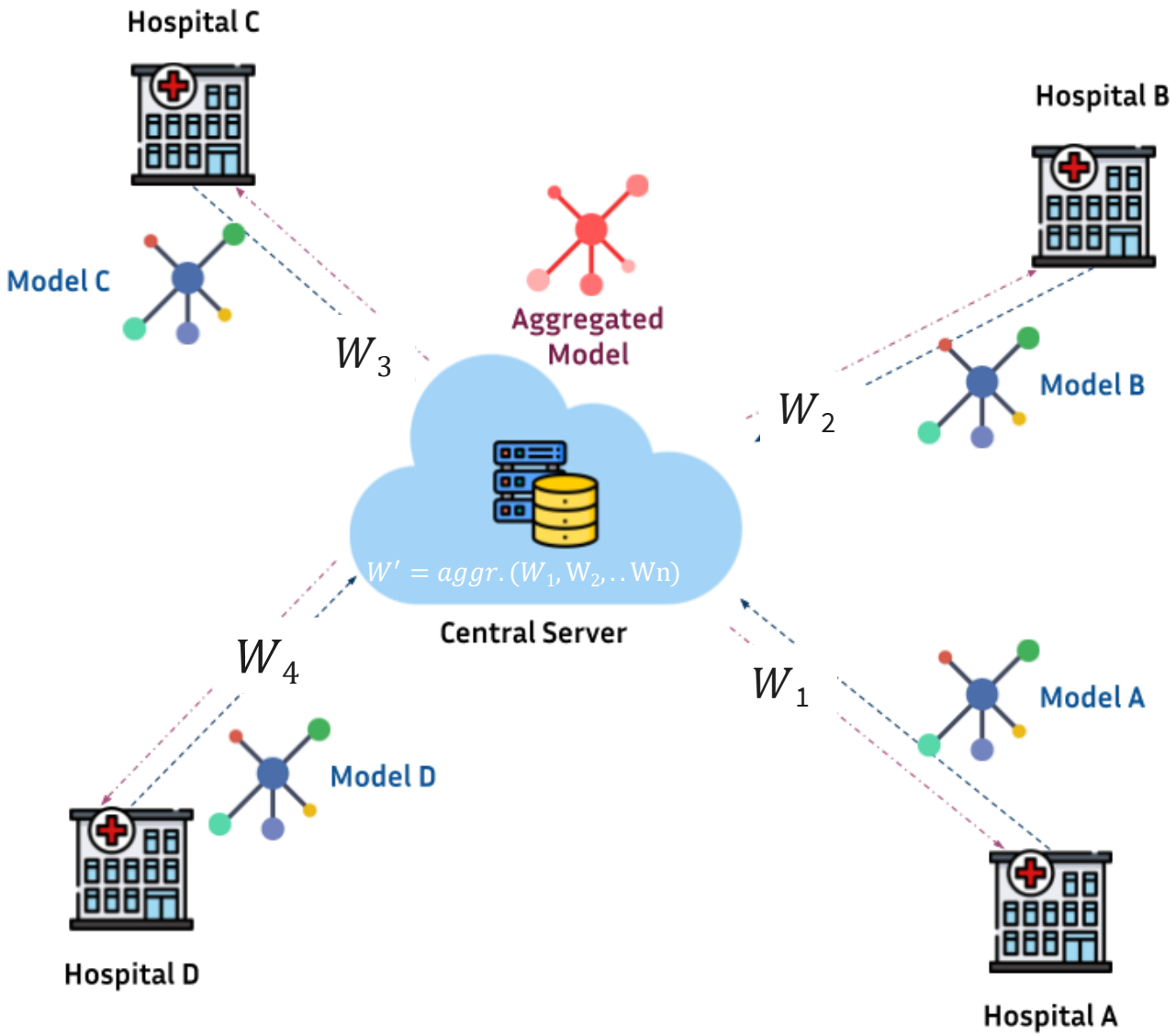
1. Big Data Integration
2. Data mining
3. Pattern recognition
4. Hypothesis generation
5. Validation

IMPORTANCE OF DATA: DATA QUALITY



Source of the image: Schwerk (2023)

IMPROVED DATA QUALITY: FEDERATED ML

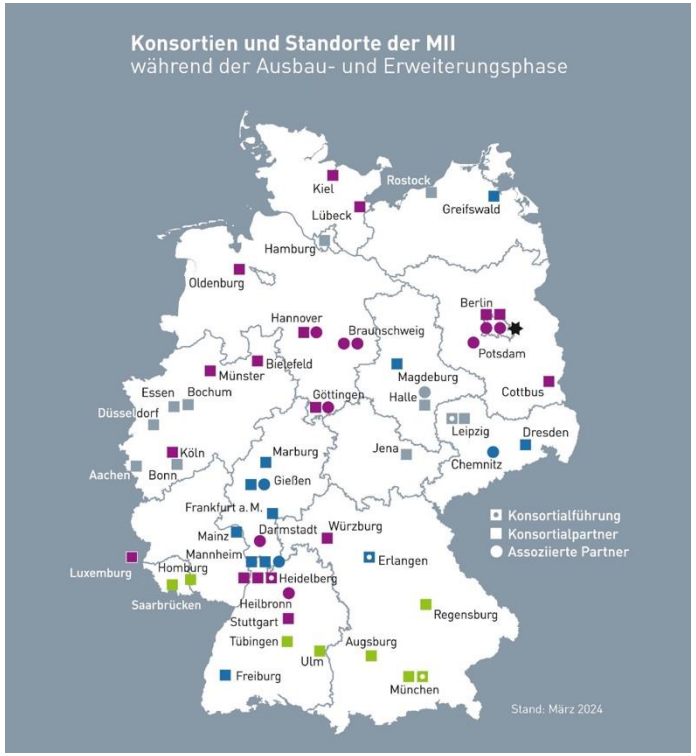


Source: https://fedbiomed.org/

- **Data protection:** Data remains at the original location - only model parameters are shared
- **Decentralization:** Hospitals train local models and send encrypted parameters to central coordinators
- **Iterative process:** The coordinator aggregates local models into a global model and shares it. The process is repeated until the model converges

RESEARCH DATA PORTAL FOR HEALTH (FDPG)

- Initiated by the Medical Informatics Initiative (MII) launched in 2016 by the BMBF
- A federated network for centralized and decentralized data access
- Core data set from primary IT systems of the universities →local data integration centers
- Standardization to MII format
- Depending on the basis of use:
  - Distributed evaluations
  - Central evaluations (with broad consent)



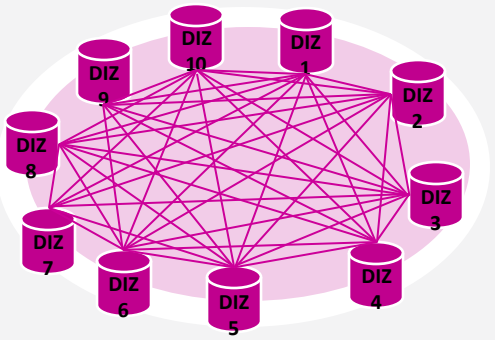
- 15 M EHRs
- 160 M Diagnoses
- 1.5 Billion Lab values

→ **2-3 M uncoded rare diseases**

THE FDPG

Open data and collaborative networks

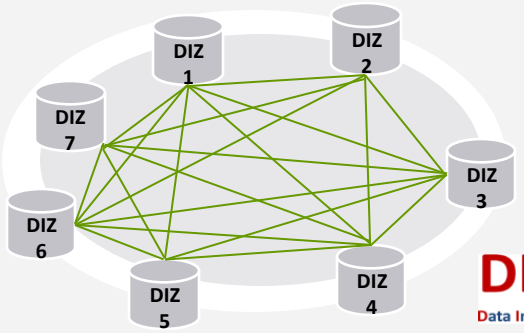
Data analytics and ML for personalized medicine.



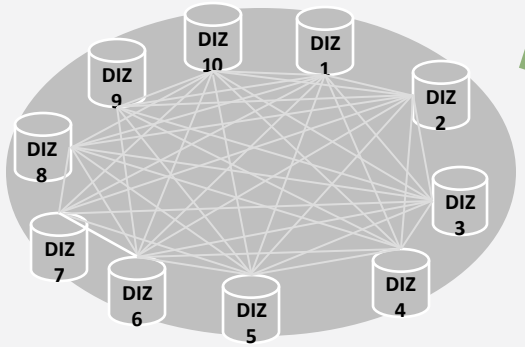
HiGHmed Medical Informatics

Researcher

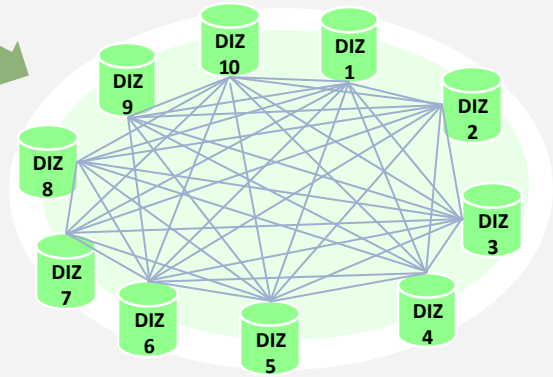
FORSCHUNGS-DATENPORTAL FÜR GESUNDHEIT



DIFUTURE Data Integration for Future Medicine



SMITH Smart Medical Information Technology for Healthcare



miracum Medical Informatics in Research and Care in University Medicine

Real-world data for medical innovations

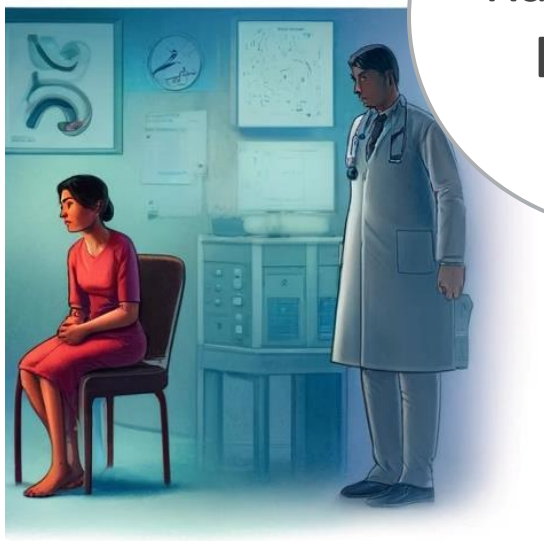
Data integration and analysis

INTERNATIONALE HOCHSCHULE

#2

Examples

RARE DISEASES: DIAGNOSTIC ODYSSEY



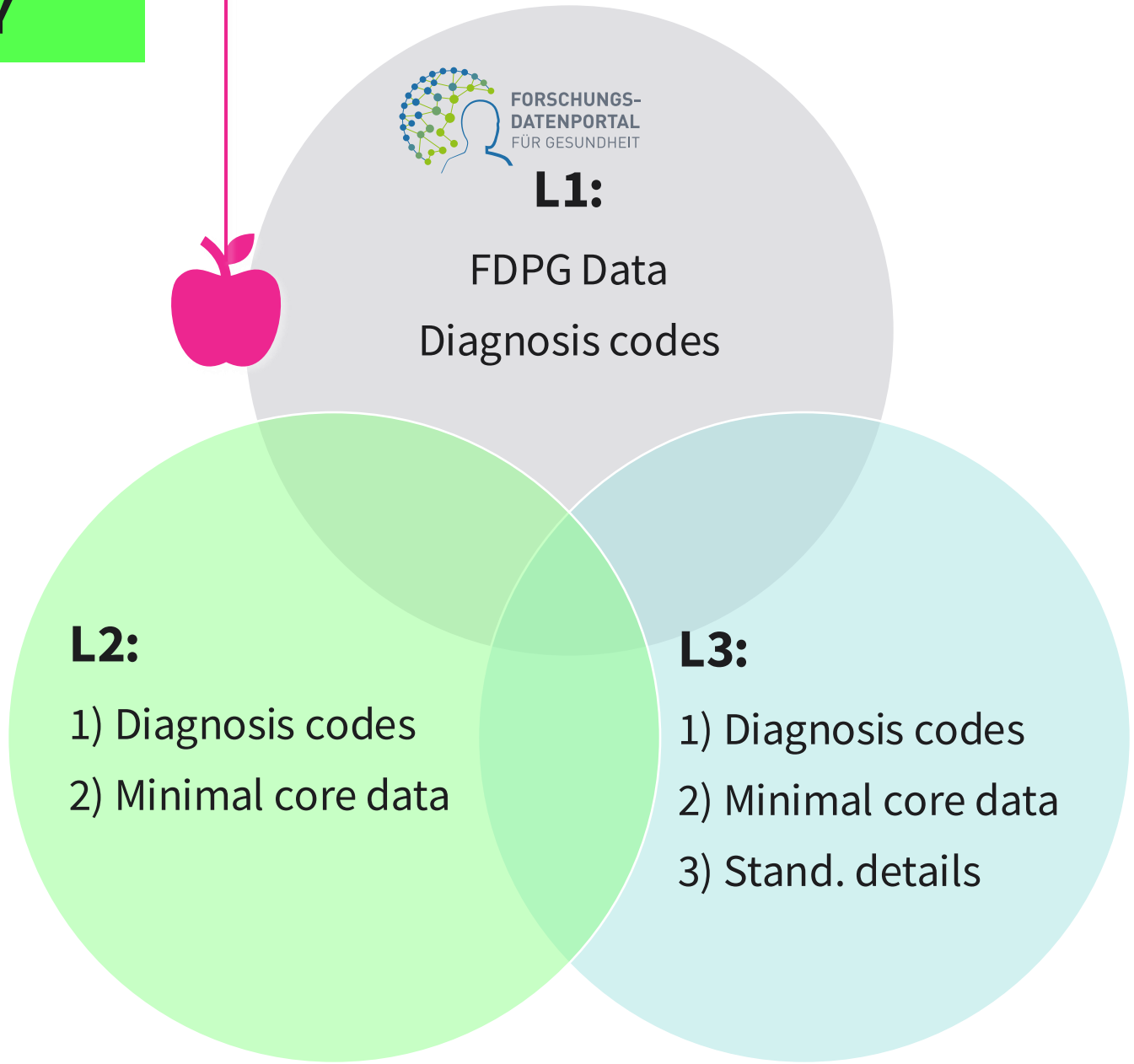
Rare Disease Dilemma

- Most rare diseases are genetic — 72%
- Most rare diseases affect children — 75%
- Many are misdiagnosed — 70%
- On average it takes 8 years to diagnose — 8 years

Source: www.Eurordis.org

DATA AVAILABILITY

- FDPG Data
- Orpha / ICD Codes



- Addition through clinics
- **Goal**: Europe-wide analysis

- Special data:
  - Progress data
  - Intervention data
  - PROMs
  - Omics
- European Registers

ML FOR RARE DISEASE (RD) DIAGNOSTICS

INTERNATIONALE HOCHSCHULE

Morbus Osler

- Undected RDs
- EHR-based Phenotypes
- **Example**: Morbus Osler

- Combined markers
- Cheaper Diagn.
- **Novel Diagnoses**:
  - Combined: Genetic (PRS) Epigenetic

Hypophosphatasia

RD Identification

Diagnostic Precision

Subgroup Identification & therapy

Novel Biomarker

Parkinson's disease

- Precision medicine
- Treatment

Monitoring
- Interventions
- Progression

- Diagnostics

Gaucher disease

L1: ML FOR HYPOPHOSPHATASIA DIAGNOSIS

- Hereditary disease: ALPL gene mutation
- 400 + disease-causing ALPL variants
- Symptoms:
  - **Severe**: Bone demineralization, respiratory failure, seizures
  - **Mild**: Tooth loss, periodontal disease
- Diagnosis: Ø 5.7 years delayed
  - Frequent misdiagnoses
  - ALP value + symptoms + genetics
  - Specific orpha code
- Incorrect treatment: bisphosphonates → Bone damage

→ Early diagnosis is crucial

L1 Analysis: Use ALP (+PLP) biomarkers + phenotypes to determine more specific biomarker thresholds
→ identify new patients and improve diagnostic precision

– Khan et al. 2024

L1: RBS FOR MORBUS OSLER DIAGNOSTIK

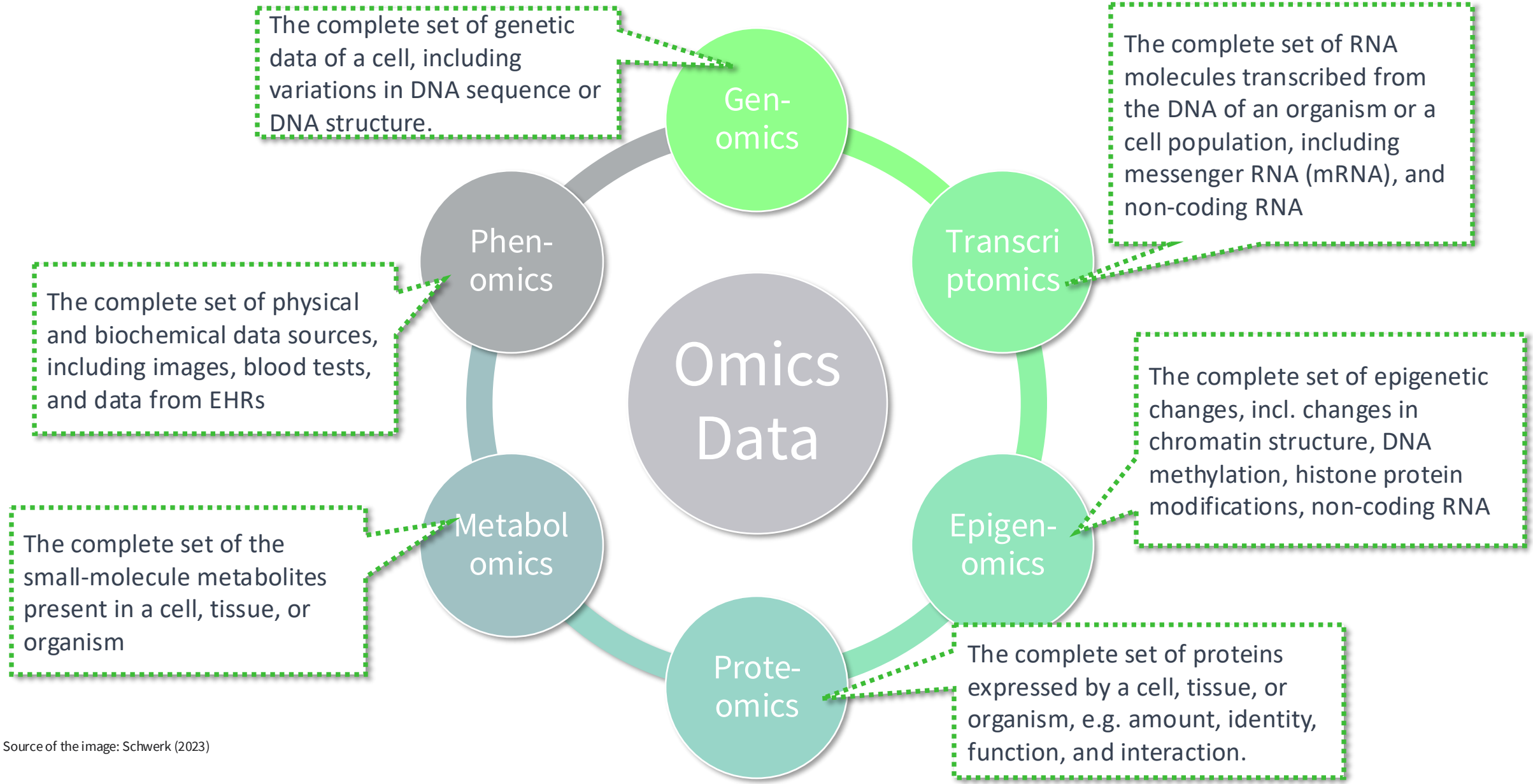
- Hereditary disease: : ENG, ACVRL1, or SMAD4 gene mutation
- 600 + disease-causing variants
- Symptoms:
  - **Severe:** Arteriovenous malformations → Bleeding, strokes, cardiac stress
  - **Mild:** Nosebleeds (epistaxis), telangiectasias on skin and mucous membranes
- Diagnosis: Ø 26 years delayed
  - Frequent misdiagnoses
  - Curacao criteria + genetics
- Incorrect treatment: anticoagulants → Bleeding

→ Early diagnosis is crucial

L1 Analysis: Using the ICD-10 code for identification & extended phenotyping
→ RBS on Curacao Criteria → New identifications

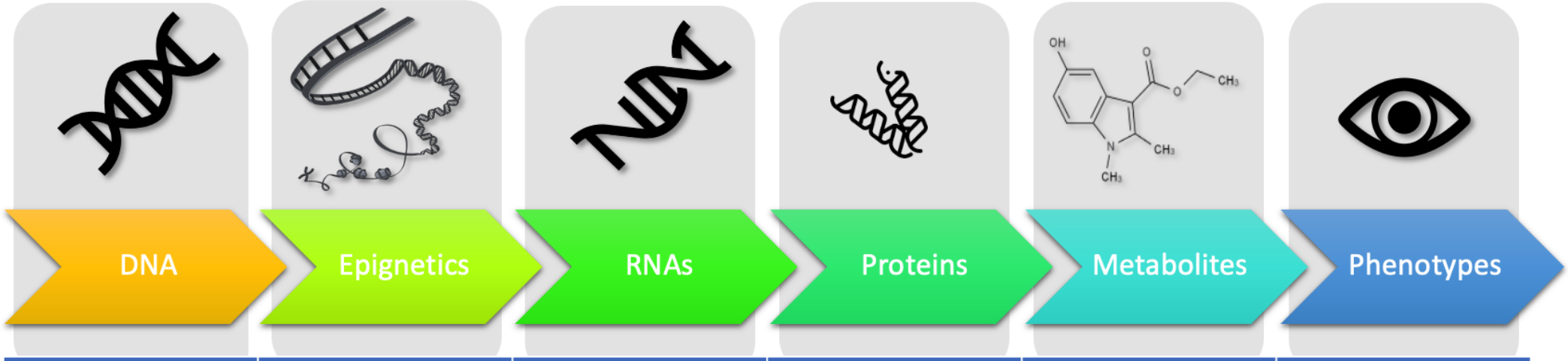
– Pierucci et al. 2012

L3 DATA: OMICS



Source of the image: Schwerk (2023)

OMICS DATA AND ANALYSES



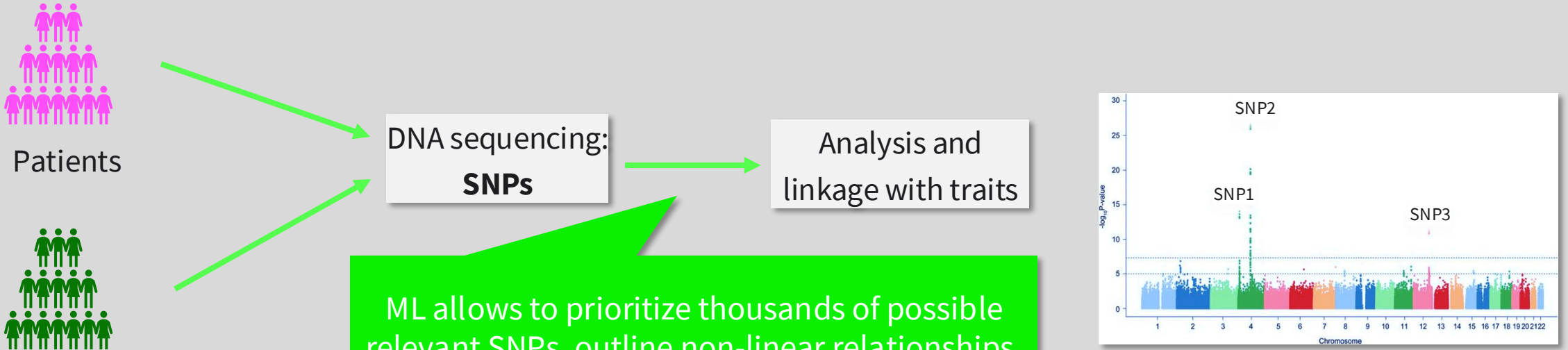
Omics	Genomics	Epigenomics	Transcriptomics	Proteomics	Metaolomics	Phenomics
Analysis	• eQTL • GWAS • PRS • Gene prioritization • Epistasis • Gene-based stratification	• Gene prioritization • Tumor heterogeneity • Drug Response Prediction	• Differential gene expression • Batch effect removal • Gene prioritization	• pQTL • Biomarker • Spatial proteomics: segmentation & protein mis-localizations	• Preprocessing • mQTL • Biomarker • Disease profiles • Metabolite identity/ pathway prediction	• PheWAS • Integrated omics • Image-based biomarker • EHR biomarker/ subgrouping

— Source of the image: Schwerk (2023)

SUBGROUP IDENTIFICATION: GENOME WIDE ASSOCIATION STUDIES

GWAS:

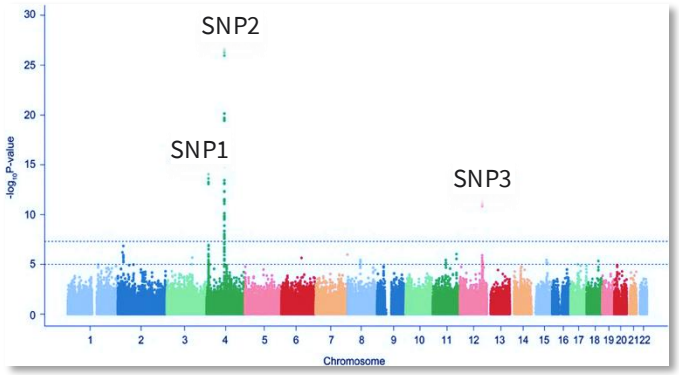
- Analyzes the genomes of large groups of people with and without a disease / trait
- Identifyies genomic variants (SNPs) - by assessing linkage disequilibrium



Source of the image : Schwerk (2023) and adapted from Matsuo et al. (2016;)

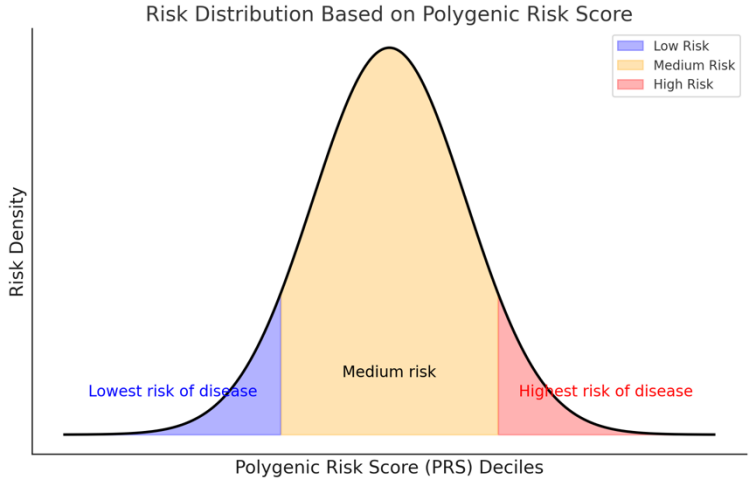
DIAGNOSTIC PRECISION: POLYGENIC RISK SCORE CALCULATION

- PRS sum weighted effect sizes of risk variants from GWAS to estimate disease susceptibility
- Linear approach: does not account for complex genetic interactions (e.g., epistasis)
- **ML helps to:**
  - Account for non-linear interactions
  - Combine multiple existing PRS



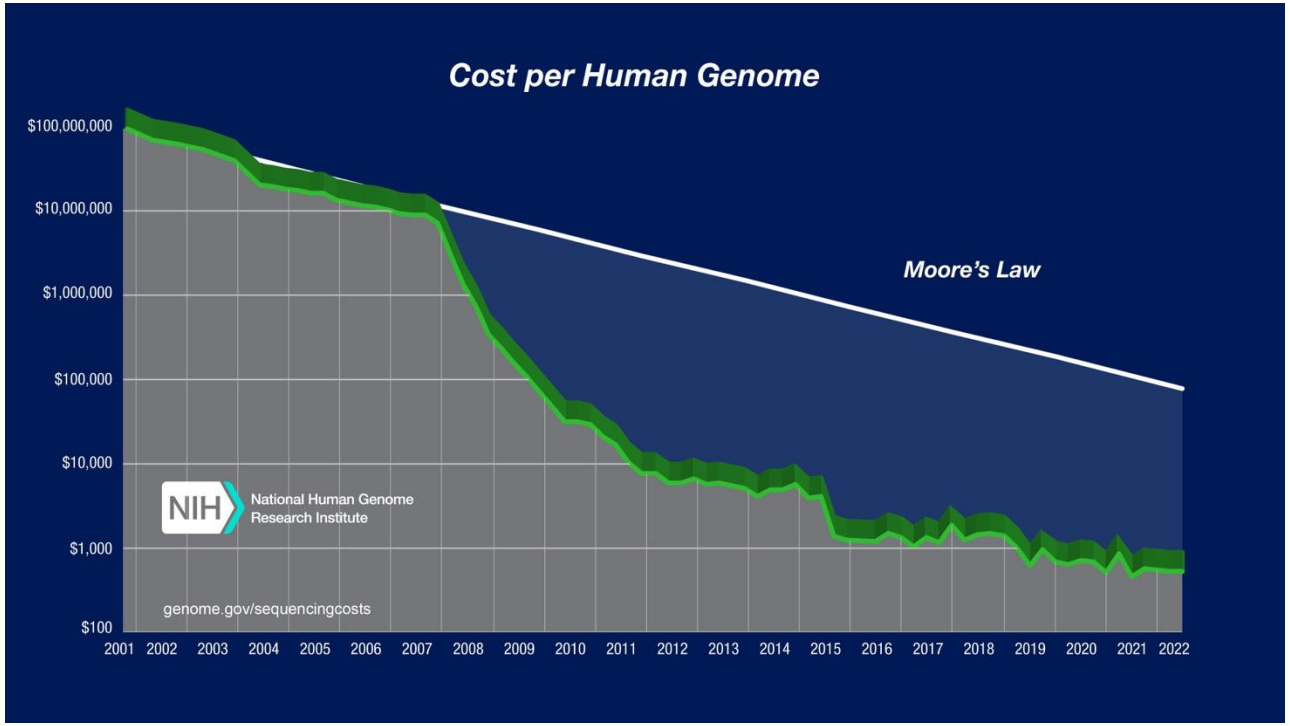
GWAS

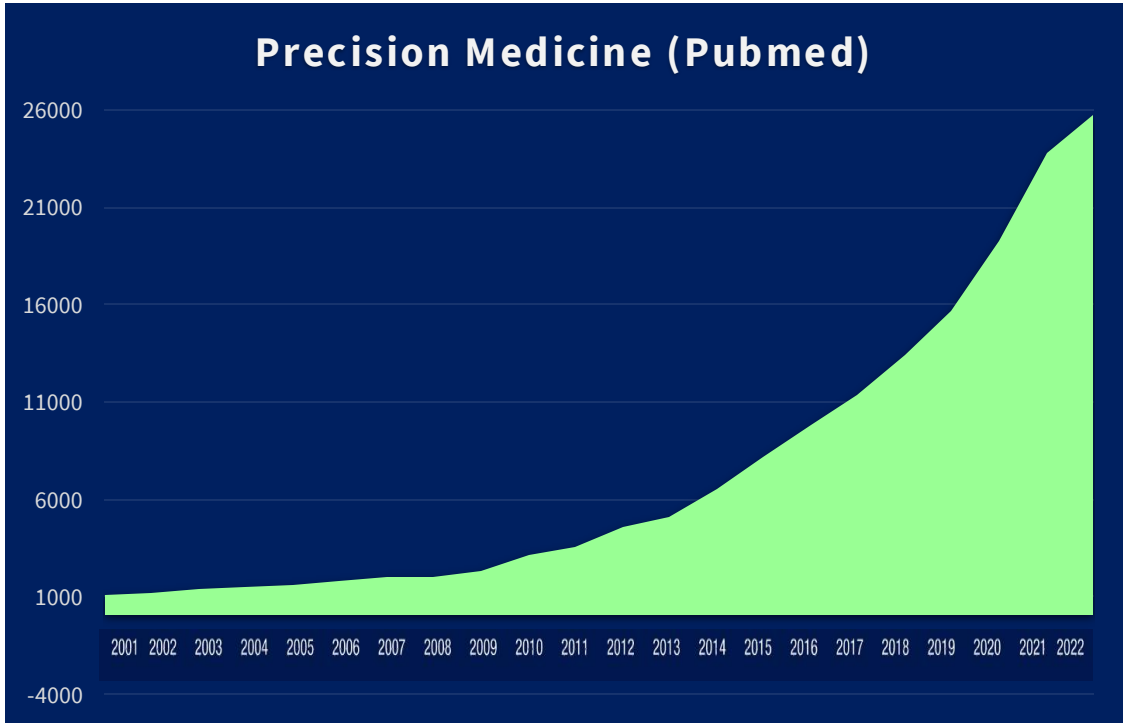
Top genetic variants
e.g. SNPs
(largest effect size)



Risk scoring

COST PER GENOME AND PRECISION MEDICINE PUBLICATIONS





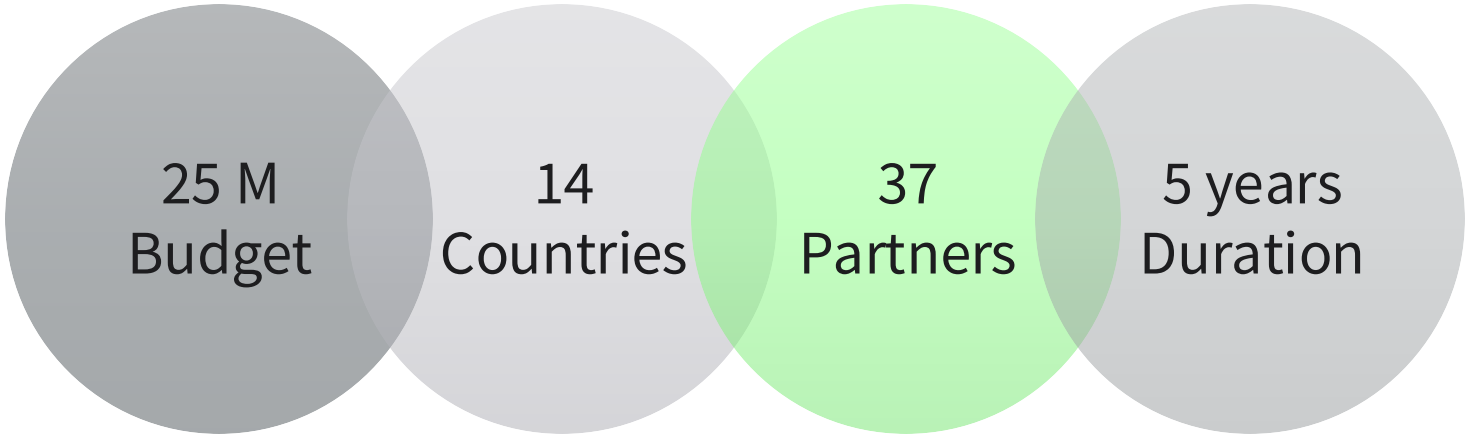
Source of the image: NIH

SCREEN4CARE

SCREEN 4CARE

INTERNATIONALE HOCHSCHULE

- **Genetic newborn screening:**
  - Early diagnosis of genetic RDs
- **Digital tools:**
  - Symptom checker
  - EHR algorithms
- **Infrastructure**:
  - Federated ML



Improved the accuracy and speed of diagnosis

https://www.screen4care.eu/

#3

Take-Home-Message & Discussion

- **AI as a medical game changer:**
  - Access to digital data
  - High quality data

- **Expected AI revolution:**
  - Fastest growing healthcare data source
  - Text data and LLMs

INTERNATIONALE HOCHSCHULE

THANK YOU

Prof. Dr. Anne Schwerk
anne.schwerk@iu.org