



# Data & Computing Services for Personalized Health: a Paradigm Shift

Dr. Diana Coman Schmid  
Personalized Health Data Services Manager  
Scientific IT Services, ETHZ  
[diana.coman@id.ethz.ch](mailto:diana.coman@id.ethz.ch)

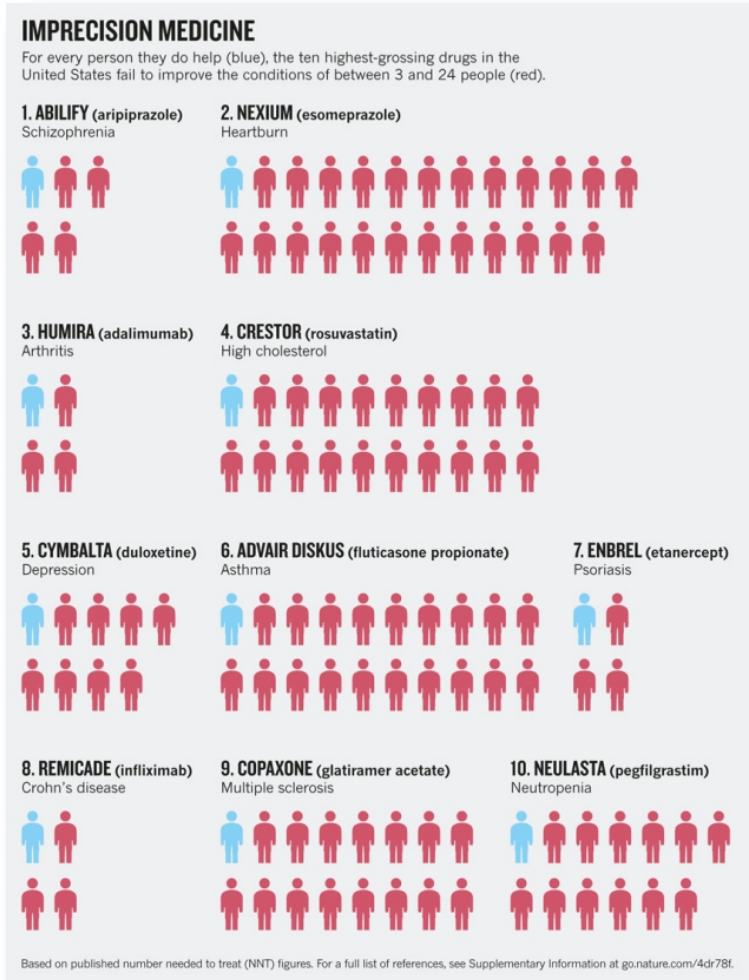
**hpc-ch forum**  
May 17, 2018

# Outline

- **Personalized Health (PH) > Data Driven**
- PH Services > a Paradigm Shift
- Data Driven PH, as we speak – worldwide and in Switzerland

# Data Driven Personalized Health > the Need

## Imprecision Medicine



Schork, 2015 *Nature*

## Precision Medicine

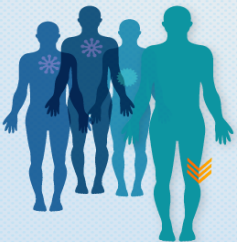
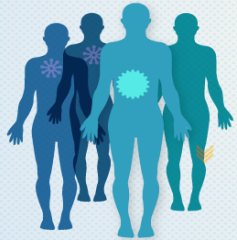
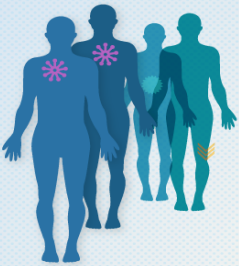


[www.cancer.gov](http://www.cancer.gov)

# Data Driven Personalized Health > the Need

## Precision Medicine

Discovering unique therapies that treat an individual's cancer based on the specific genetic abnormalities of that person's tumor.



www.cancer.gov

## Personalized Health



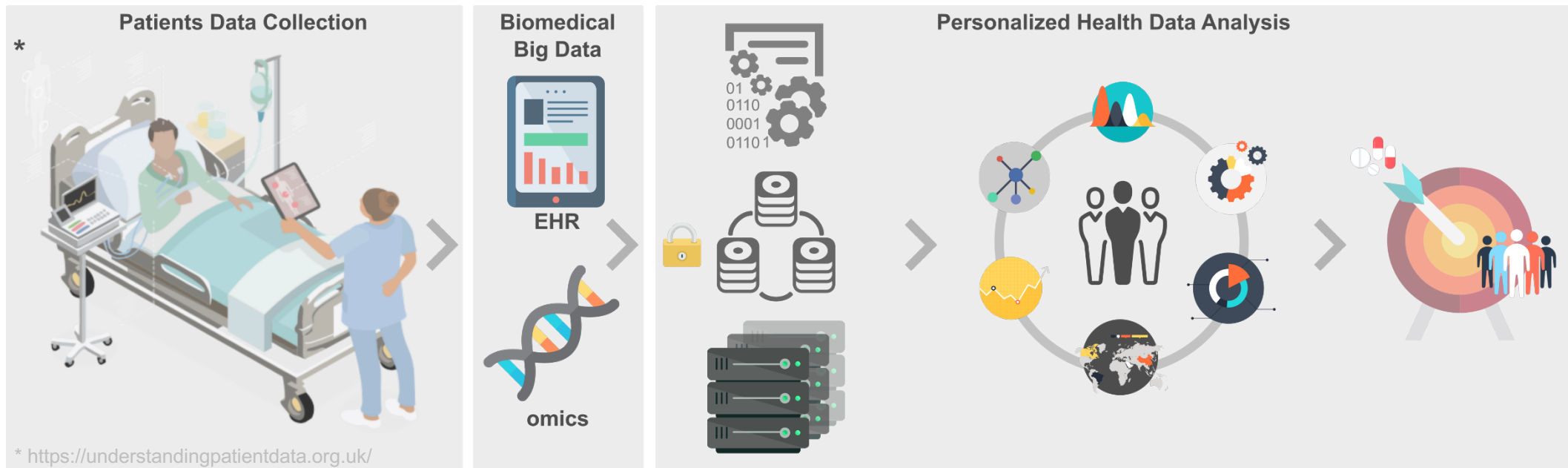
<https://blogs.cdc.gov/genomics/2015/03/02/precision-public/>

- beyond individualized treatment of sick individuals
- promotes **Health at Population** level



# Personalized Health > Data Driven

The goal is to provide the right treatment, at the right moment to the right patients (precision medicine) and in the same time to ensure as many people as possible stay healthy (prevention; personalized health).



# Personalized Health (PH) > Data Driven

What is new is **Data** Driven Personalized treatment and prevention >> **PH**

**Data**: large volumes (unstructured) that exceed **Storage** and **Computation** capacity of standard PCs

## Rare Disease

- **< 1 in 2,000** people (Europe)
- **7,000** rare diseases (+5/week)
- **80 %** have a genetic component
- *100,000 Genomes Project* (UK)  
**20-25 %** are “potentially actionable” diagnoses

## Cancer

## Complex Disease

# Personalized Health > it helped them

What is new is **Data** Driven Personalized treatment and prevention >> **PH**

**Data**: large volumes (unstructured) that exceed **Storage** and **Computation** capacity of standard PCs

## Rare Disease

- **< 1 in 2,000** people (Europe)
- **7,000** rare diseases (+5/week)
- **80 %** have a genetic component
- *100,000 Genomes Project* (UK)  
**20-25 %** are “potentially actionable” diagnoses

## Jessica suffers from a rare condition that was diagnosed through DNA analysis



<https://www.ft.com/content/d2e21cea-d684-11e6-944b-e7eb37a6aa8e>

- Jessica: four-year-old
- epilepsy, poorly coordinated movement, slow mental development
- **mutation in the gene *SLC2A1***
- 1 month on a **special diet**: improvement on her speech, energy levels and general steadiness (*100K Genomes*, UK)

# Personalized Health > it helped them

What is new is **Data** Driven Personalized treatment and prevention >> **PH**

**Data**: large volumes (unstructured) that exceed **Storage** and **Computation** capacity of standard PCs

## Rare Disease

- **< 1 in 2,000** people (Europe)
- **7,000** rare diseases (+5/week)
- **80 %** have a genetic component
- *100,000 Genomes Project* (UK)  
**20-25 %** are “potentially actionable” diagnoses

## Cancer

- **39–95 %** patients: “potentially actionable” diagnoses
- 570 studies, 32,149 patients<sup>1</sup>  
**extended overall survival** (median 13.7 months versus 8.9 months;  $P = 0.0001$ )

Carrasco-Ramiro et al., 2017 *Nature*

## Complex Disease

- e.g., type 2 diabetes
- **multifactorial** (e.g., life style)
- **heterogenous**
- **>150** common genetic variants
- gut **microbiome** content
- complex **health record data**

Zeevi et al., *Cell*, 2015



# Artificial Intelligence (AI) in Personalized Health > because of Data

What is new is **Data Driven Personalized** treatment and prevention >> PH

**Data**: large volumes (unstructured) that exceed **Storage** and **Computation** capacity of standard PCs >> **AI**

## Google DeepMind

- 2018: predicts **heart disease** by looking at your eyes<sup>2</sup>
- **AI** ( $n=297,360$ ; retina scans, general medical data)
- **Accurate** predictions (70 % vs. 72 % standard blood tests)
- **Quicker** (no blood tests)
- **Needs clinical testing**

<sup>2</sup> Poplin et al., 2018 *Nature*

## IBM Watson Health

- **90 % concordance** with tumor board recommendation for breast cancer<sup>3</sup>
- AI **40 sec.** vs. Doctor **12 min.**
- one patient (glioblastoma), whole genome and RNA sequencing<sup>4</sup>
- AI **10 min.** vs. Bioinformatician **160 h.**

<sup>3</sup> <http://www.ascopost.com/News/44214>

<sup>4</sup> Wrzeszczynski et al., 2017 *Neurology Genetics*

## Keep an eye on<sup>5</sup>:

- **Careskore (US)**: Zeus
- **Zephyr Health (US, UK, IN)**: Illuminate
- **Oncora Medical (US)**: Unity, Focus, Acumen
- **Atomwise (US)**: AtomNet
- **Sentrian (UK,US)**: Remote Patient Intelligence

<sup>5</sup> <http://medicalfuturist.com/top-artificial-intelligence-companies-in-healthcare/>

# PH can only happen in presence of AI > because of Data

*“We are drowning in information but starved for knowledge” John Naisbitt (1982)*

- omics (gen-, epigen-, transcript-, etc.) >> **Bioinformatics**: domain specific and/or integrative analyses
- eHealth (electronic health records, histopathological and radiological images, etc.) >> **AI**: machine learning
- citizen data (digital footprints, mobile apps) >> **AI**: machine learning

## Knowledge

> **6,000** Mendelian disorders studied at genetic level

**Most** with unknown role in health and disease

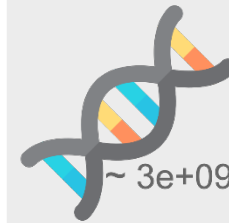
## Data

> **9,797,389,622,108,464** total bases (SRA@NIH, 9e+15 in 2017)

**16,439,169,142,390,369** (2018)

**2025** prediction: 1e+21

## 1x Human Genome Sequencing



~ 3e+09 bases



100 gigabytes

## High Performance Computing (HPC)



**80 % Unstructured Data**  
20 % Structured Data

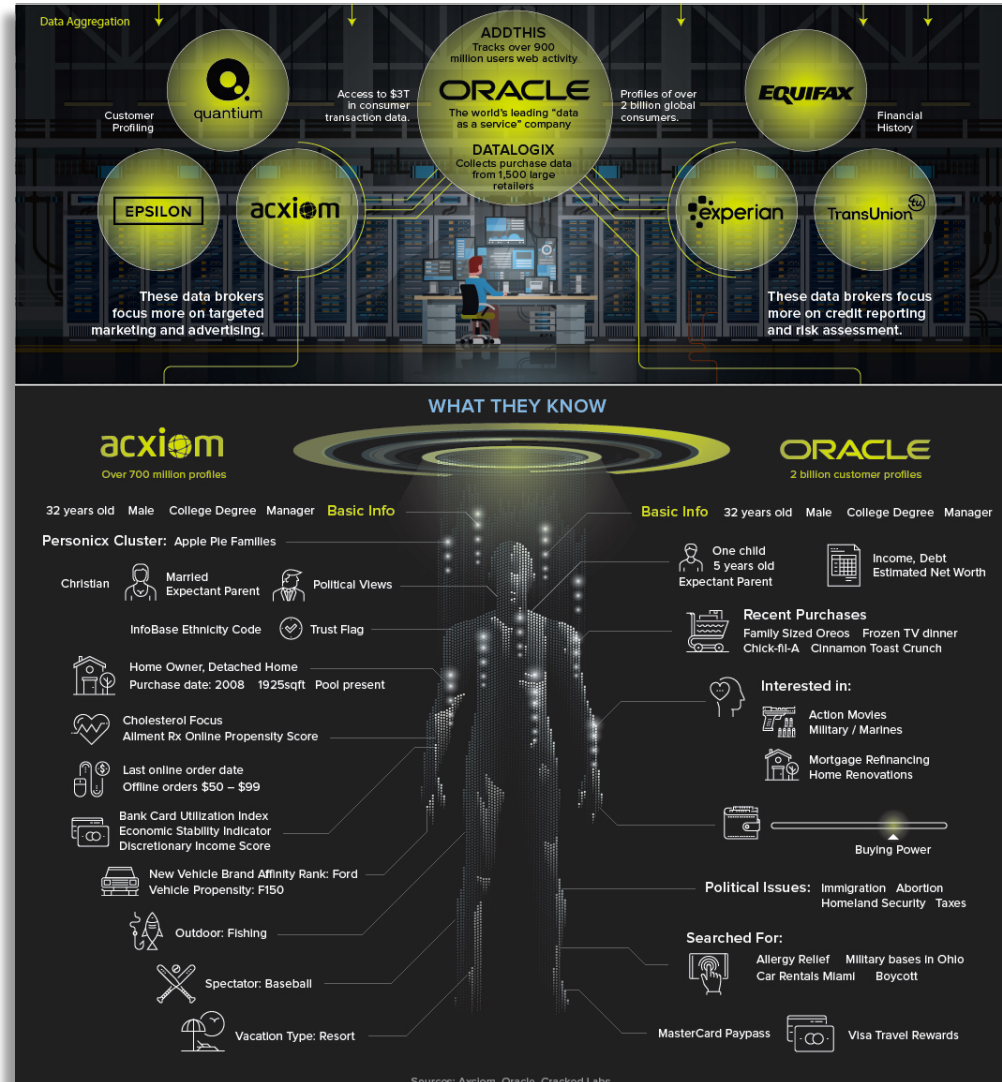
# Outline

- Personalized Health (PH) > Data Driven
- **PH Services > a Paradigm Shift**
- Data Driven PH, as we speak – worldwide and in Switzerland

# Digital Life > do You have the control?

## Personal Data Footprint :

- Social media
- **Banking**
- **Health**
- ...



## Personal Data Governance :

- **Ownership (CH !EU)**
- **Security & Privacy**
- **Storage, Processing & Sharing**
- *Is my gut bacteria data (microbiome) Patient or Bacterial data?*
- ...



# Patient Informed Consent > my own, October 2017

## Aufklärungsblatt zur Weiterverwendung von biologischem Material und gesundheitsbezogenen Personendaten für die Forschung

Version 2.0 vom 5.12.2016

### Sehr geehrte Patientin, sehr geehrter Patient

Im Verlauf Ihres Aufenthalts am UniversitätsSpital Zürich (USZ) werden gesundheitsbezogene Daten von Ihnen erhoben und möglicherweise auch biologische Proben Ihres Körpers gesammelt (Blut und andere Körperflüssigkeiten, Gewebeproben). Dieses biologische Material in Verbindung mit den Daten ist für die medizinische Forschung sehr wertvoll. Wir bitten Sie deshalb um Ihre Einwilligung, dieses Material und Ihre Daten für die Forschung zu nutzen.

### Ihre Einwilligung für die Forschung

Mit Ihrer Einwilligung können Forschende Ihre Daten wissenschaftlich auswerten und Untersuchungen mit den biologischen Proben durchführen. Von der Einwilligung betroffen sind Daten aus Ihrer Krankengeschichte, zum Beispiel Resultate von klinischen, bildgebenden und laborchemischen Untersuchungen, genetische Daten (Ergebnisse aus Untersuchungen des Erbguts), aber auch Angaben zu Ihrer Person (Alter, Geschlecht), zu durchgeführten Therapien und zum Krankheitsverlauf. Bei den Proben handelt es sich um von Ihnen stammendes, bereits vorhandenes biologisches Material (Blut, Urin oder Gewebeproben), das beispielsweise für Ihre Diagnose oder Behandlung entnommen wurde und nicht weiter benötigt wird. Die Verwendung der Daten und Proben und die Weitergabe an Forschende im In- und Ausland unterliegen strengen Vorschriften.

Es ist möglich, dass Sie im Verlauf Ihres Aufenthalts an unserem Spital ein weiteres Mal um eine Einwilligung für Forschungszwecke angefragt werden. Dies kann zum Beispiel dann der Fall sein, wenn die für Sie zuständige Klinik zusätzliche Proben von Ihnen für eine Biobank entnehmen möchte oder eine spezielle Fragestellung wissenschaftlich untersucht werden soll. Ihr Behandlungsteam wird Sie gegebenenfalls darüber informieren.

### Schutz Ihrer Daten und Proben

Nur sehr wenige Personen sind berechtigt, die Daten Ihrer Krankengeschichte einzusehen. Diese Personen sind für Ihre Behandlung zuständig oder haben im Rahmen eines bewilligten Forschungsprojekts die Erlaubnis, Einblick in Ihre Daten zu nehmen.

**Daten**, die für Forschungszwecke verwendet werden, müssen so bald wie möglich **verschlüsselt** werden, das heisst, alle Angaben, die auf Ihre Identität hinweisen – wie Name, Geburtsdatum, Versicherungsnummer usw. – werden durch einen Code ersetzt. Nur wer Zugang zum **Schlüssel** hat (Dokument, das Code und Name verbindet), kann also Rückschluss auf Ihre Person nehmen.

Die **biologischen Proben** werden in einer Biobank sicher aufbewahrt. Eine Biobank ist eine systematische Sammlung von Proben und verknüpfbaren Daten unter klar festgelegten Bedingungen. Biologisches Material

## Aufklärungsblatt zur Weiterverwendung von biologischem Material und gesundheitsbezogenen Personendaten für die Forschung

Version 2.0 vom 5.12.2016

und genetische Daten dürfen ausschliesslich **verschlüsselt oder anonymisiert** an Forschende weitergegeben werden. Anonym bedeutet, dass sämtliche identifizierenden Angaben unkenntlich gemacht oder gelöscht werden, sodass ein Rückschluss auf Ihre Person nicht mehr möglich ist.

Werden Daten und Proben in verschlüsselter Form an Forschende **ausserhalb** des UniversitätsSpital Zürich weitergegeben, verbleibt der Schlüssel am USZ, wo er von einer nicht am Forschungsprojekt beteiligten Stelle sicher verwahrt wird. Bei Forschung im Ausland muss gewährleistet sein, dass mindestens dieselben Anforderungen an den Datenschutz eingehalten werden wie in der Schweiz.

Jedes Forschungsprojekt muss zudem von der zuständigen Ethikkommission bewilligt werden. Diese überprüft, ob das Projekt und dessen Durchführung wissenschaftlich und ethisch vertretbar ist und ob die rechtlichen Bedingungen, insbesondere der Datenschutz, eingehalten werden.

### Forschungsergebnisse

Die Erkenntnisse von Forschungsprojekten mit Daten und Proben tragen meist erst für zukünftige Patientinnen und Patienten zu einer verbesserten medizinischen Betreuung bei. Sollte aber ein Ergebnis für Ihre persönliche Gesundheit relevant sein, würden Sie nach Möglichkeit informiert werden (dies ist bei Forschung mit anonymisierten Proben nicht möglich). Solche Situationen treten aber sehr selten auf.

Durch die Überlassung Ihrer Daten und Proben für die Forschung entsteht kein Anspruch auf eine Beteiligung an möglichen Gewinnen, die aus den Erkenntnissen entstehen könnten. Ihnen oder Ihrer Krankenversicherung entstehen durch die Forschungsprojekte **keine Kosten**.

### Ihre Rechte

Ihre Einwilligung ist freiwillig und gilt grundsätzlich unbegrenzt. Sie haben aber jederzeit das Recht, Ihre Einwilligung ohne Begründung zurückzuziehen (**Widerruf**). Wenden Sie sich hierfür bitte an Ihre behandelnde Klinik. Bei einem Widerruf werden Ihre Daten und Proben nicht mehr für Forschungsprojekte ausgehändigt.

Ob Sie sich für oder gegen eine Einwilligung entscheiden oder Ihre Einwilligung widerrufen, hat keinen Einfluss auf Ihre medizinische Betreuung.

### Weitere Informationen

Falls Sie noch Fragen zur Weiterverwendung von Daten und Material für die Forschung haben, wenden Sie sich an Ihre behandelnde Ärztin / Ihren behandelnden Arzt oder besuchen Sie unsere Website [www.usz.ch/forschung](http://www.usz.ch/forschung)

## Einwilligungserklärung

Patientenetikette

zur Weiterverwendung von gesundheitsbezogenen Personendaten und biologischem Material für die Forschung.

Name und Vorname der Patientin / des Patienten:

COMAN SCHMID DIANA ELENA

Geburtsdatum:

Ich bestätige, dass

- ich das zu dieser Einwilligungserklärung gehörende Aufklärungsblatt (Version 2.0 vom 5.12.2016) erhalten habe und mich hinreichend informiert fühle.

Ich willige ein, dass

- meine gesundheitsbezogenen Daten (inkl. genetischer Daten) und biologischen Proben in der oben beschriebenen Form zu Forschungszwecken verwendet werden dürfen.

 Ja

 Nein

Mit der Überlassung Ihrer gesundheitsbezogenen Daten und Ihres Materials leisten Sie einen wertvollen Beitrag für die biomedizinische Forschung.

Wir danken Ihnen dafür herzlich.

Zürich  
Ort

Datum 03.10.2017

Kausen

Unterschrift Patientin/Patient

Name/Vorname/Unterschrift vertretungsberechtigte Person

Generalkonsent, Version 2.0 vom 5.12.2016



## Patient Informed Consent > my own, October 2017

Mit Ihrer Einwilligung können Forschende Ihre Daten wissenschaftlich auswerten und Untersuchungen mit den biologischen Proben durchführen. Von der Einwilligung betroffen sind Daten aus Ihrer Krankengeschichte, zum Beispiel Resultate von klinischen, bildgebenden und laborchemischen Untersuchungen, **genetische Daten** (Ergebnisse aus Untersuchungen des Erbguts), aber auch Angaben zu Ihrer Person (Alter, Geschlecht), zu durchgeführten Therapien und zum Krankheitsverlauf. Bei den Proben handelt es sich um

**Daten**, die für Forschungszwecke verwendet werden, müssen so bald wie möglich **verschlüsselt** werden, das heisst, alle Angaben, die auf Ihre Identität hinweisen – wie Name, Geburtsdatum, Versichertennummer usw. – werden **durch einen Code ersetzt**. Nur wer Zugang zum **Schlüssel** hat (Dokument, das Code und Name verbindet), kann also Rückschluss auf Ihre Person nehmen.

und genetische Daten dürfen ausschliesslich **verschlüsselt oder anonymisiert** an Forschende weitergegeben werden. **Anonym bedeutet**, dass sämtliche identifizierenden Angaben unkenntlich gemacht oder gelöscht werden, sodass ein Rückschluss auf Ihre Person nicht mehr möglich ist.

nen und Patienten zu einer verbesserten medizinischen Betreuung bei. Sollte aber ein **Ergebnis für Ihre persönliche Gesundheit relevant sein**, würden Sie nach Möglichkeit informiert werden (dies ist bei Forschung mit anonymisierten Proben nicht möglich). Solche Situationen treten aber sehr selten auf.



# PH, Biomed. Research & Services > Sensitive Personal Data

## Federal Act on Data Protection



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

### - Art. 3 Definitions

The following definitions apply:

- a. *personal data (data)*: all information relating to an identified or identifiable person;
- b. *data subjects*: natural or legal persons whose data is processed;
- c. *sensitive personal data*: data on:
  1. religious, ideological, political or trade union-related views or activities,
  2. **health**, the intimate sphere or the racial origin,
- e. *processing*: any operation with personal data, irrespective



# PH, Biomed. Research & Services > Sensitive Personal Data

Standard de-identification and anonymization are ineffective with Omics data



How to **define** anonymization in the context of current data & analytics

- CH-FADP, HRA, HRO, EU-GDPR, HIPAA
- Omics

Anonymization & **Usability**:

- Clinical trial: control & treatment
- Regression: age vs. age categories
- Machine learning: gender, family history, etc.

US: About 4 reported medical data breaches (affecting 500+ people) every week

[https://ocrportal.hhs.gov/ocr/breach/breach\\_report.jsf](https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf)





# PH, Biomed. Research & Services > **Paradigm Shift**

## Secure & Powerful Data & Computing Infrastructure & Services

### Legal & Ethics

- Sensitive personal data
- Data types exceed the legal specification > Ethics
- Data sharing is prerequisite for PH

### Infrastructure & Services

- Distributed and large data lakes
- Data centralization/federation
- Interoperability
- Data management (controlled data access)
- Computing power (AI)

### Usage

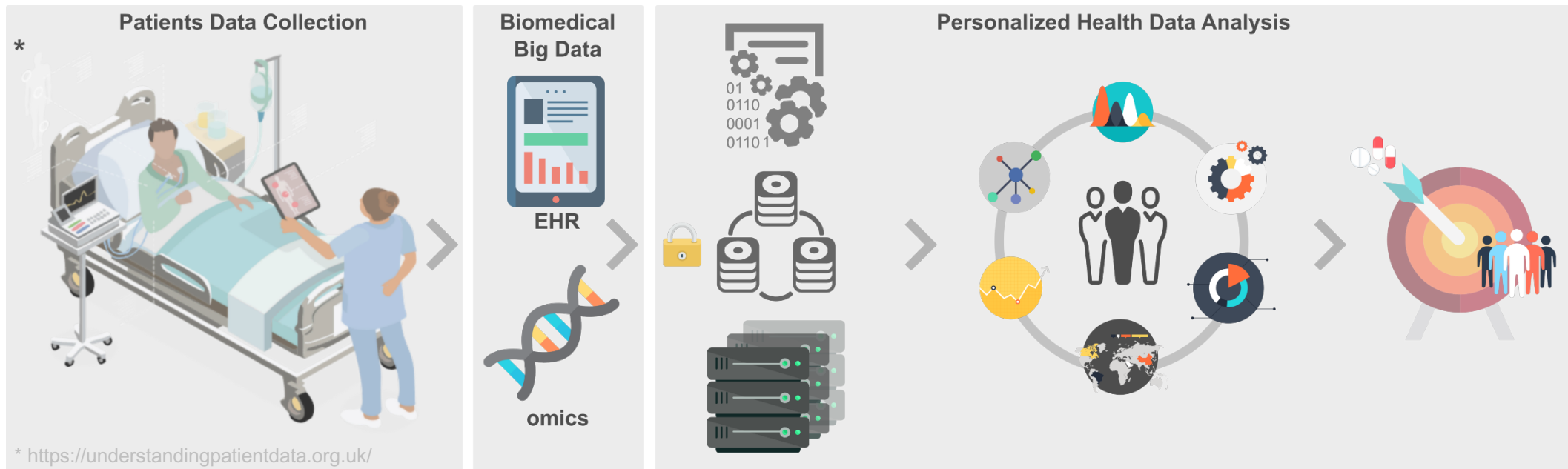
- Researchers: security & usability
- Clinicians: UX
- Citizens

# Research & Services in **Personalized Health** > need **Hybrids**

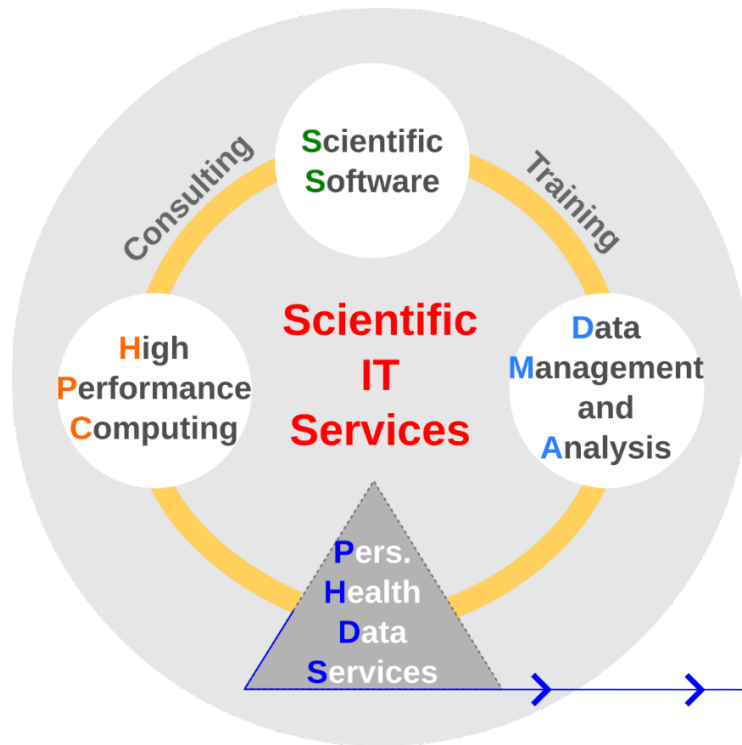
**Why:** prevent, diagnose and treat unfavorable health conditions more precisely

**Who:** hybrid ecosystem: medical, IT, computational and statistical experts

**How:** security; “Big Data”: storage, management and analysis; data exchange & interoperability



# Personalized Health Data Services > at the SIS, ETHZ



Leonhard Med



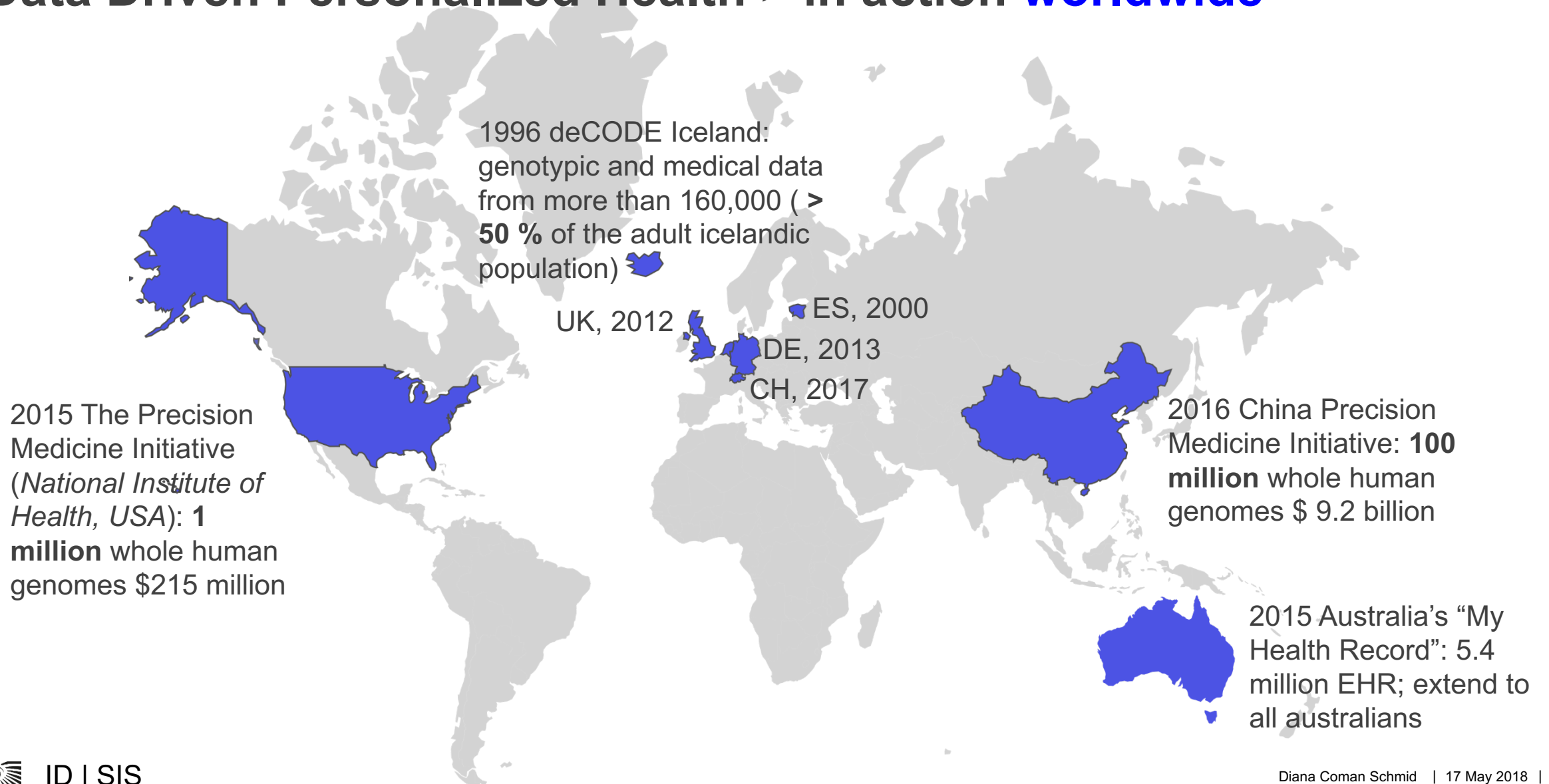
**Leonhard Med:** dedicated computing platform for secure data driven biomedical research

- Legal: ETH; ISO/IEC 27001; Federal Act on Research involving Human Being; Federal Act on Data Protection; EU-GDPR; HIPAA
- Multi-factor authentication
- Data encryption
- Controlled in/out traffic
- Auditable user activity (logged)

# Outline

- Personalized Health (PH) > Data Driven
- PH Services > a Paradigm Shift
- **Data Driven PH, as we speak – worldwide and in Switzerland**

# Data Driven Personalized Health > in action **worldwide**





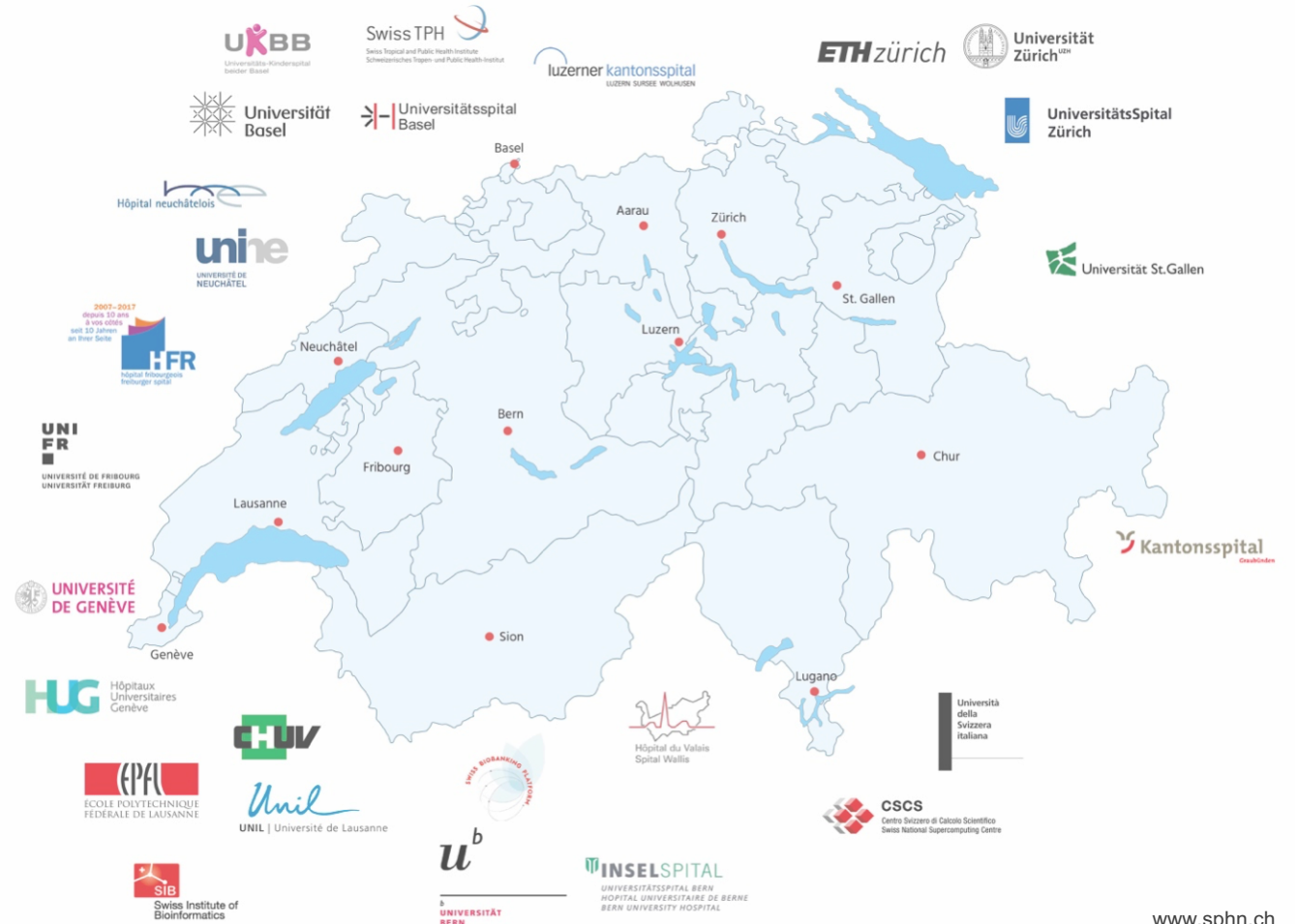
# Data Driven Personalized Health as we speak > in Switzerland



15 projects  
15.66 Mio.CHF



27 projects  
24.55 Mio.CHF



# SPHN &/ PHRT > ongoing projects examples

**Infrastructure > implementation :: development :: driver**

## SVIP-O

- **single place** for genetic mutations in **cancer**
- **Swiss hospitals** patients
- **curated** database
- clinicians/researchers could **base their diagnoses** on the same harmonized high-quality information

## PSSS

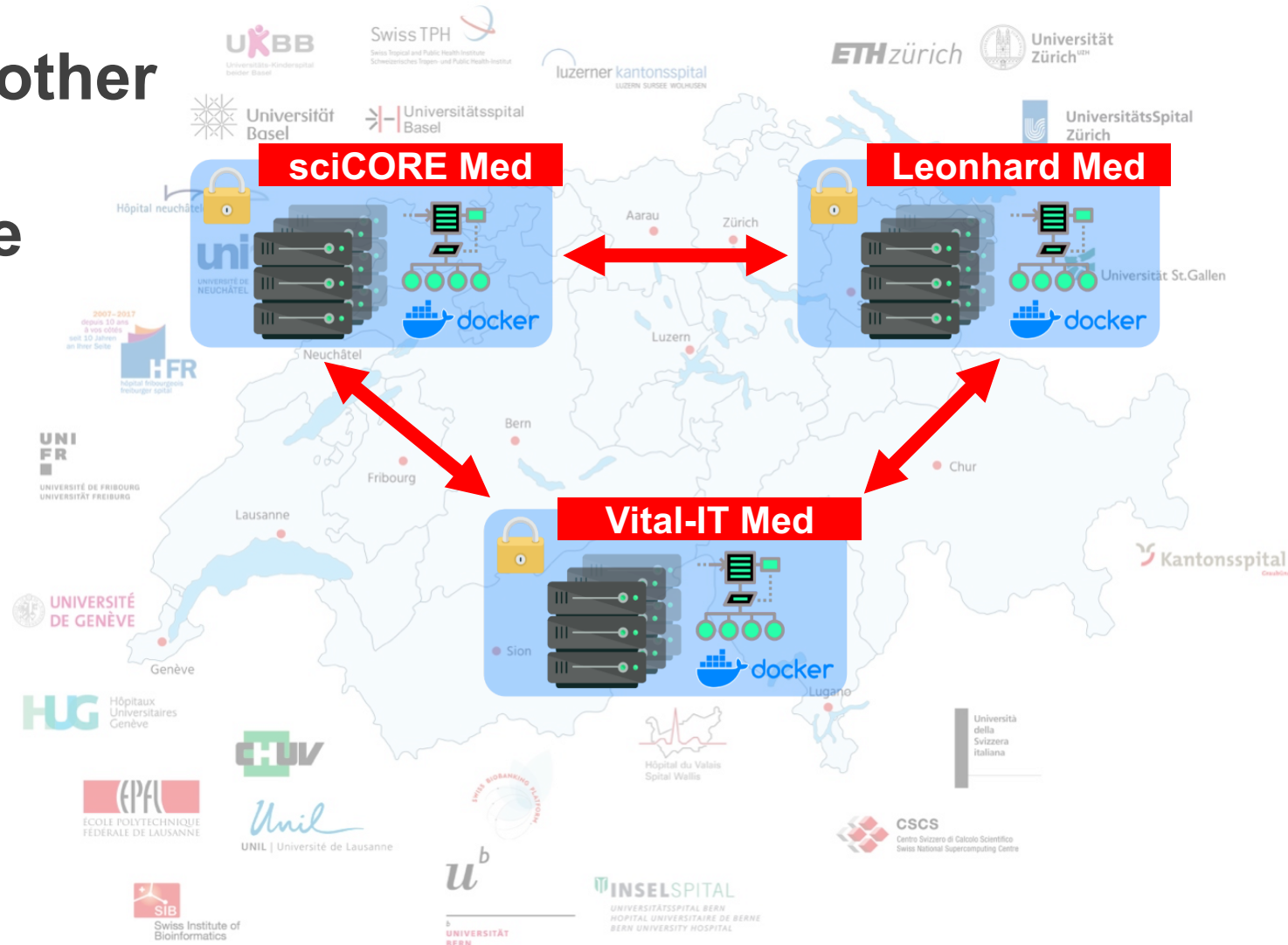
- **sepsis** is associated with high morbidity and mortality.
- **all Swiss** University Hospitals, Universities and ETH Zürich
- **ICU & omics** based biomarkers
- **Machine Learning** to recognize a bacterial sepsis **earlier** and to **predict** its course

## MIDATA.coop

- Platform for Ethical and Fair **Citizen/Patient** Participation in Personalized Health Research
- **Copies of personal data** are stored in **citizen-owned** and **citizen-controlled** MIDATA cooperative/s
- **Not for profit**

# Data Driven Personalized Health as we speak > in Switzerland

SPHN/PHRT/other  
Secure  
Infrastructure  
BioMedIT



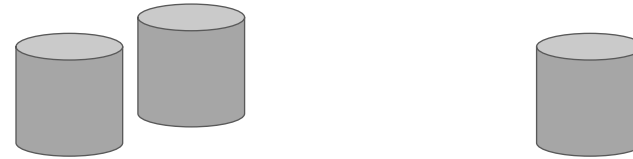


## Leonhard Med > Secure Data and Computing Platform

- **Data Driven Biomedical Research** > Individual Research Groups
- **Harmonized, Interoperable & Collaborative Swiss Biomed. Res.** > SPHN/PHRT
- **Standardized Clinical Research Analytics** > Clinical Research Service (NEXUS ETHZ)
- **Citizen Data** > MIDATA.coop

# Active Data Management > transfer, storage, computing, sharing

Data Sources  
(distinct)



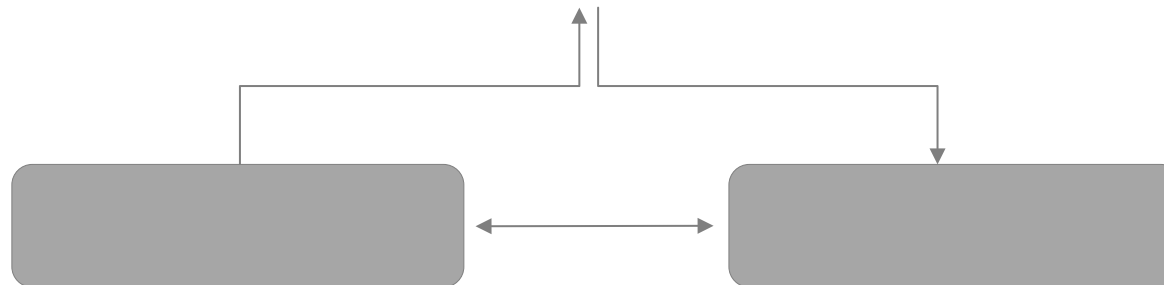
Data Storage & Analysis  
(central)



Backup



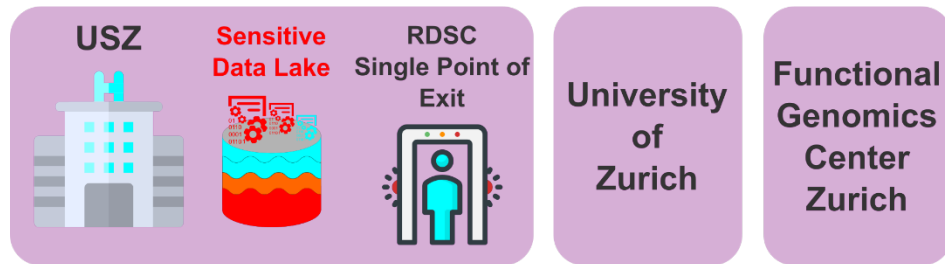
Data, Workflows  
Sharing & Interoperability



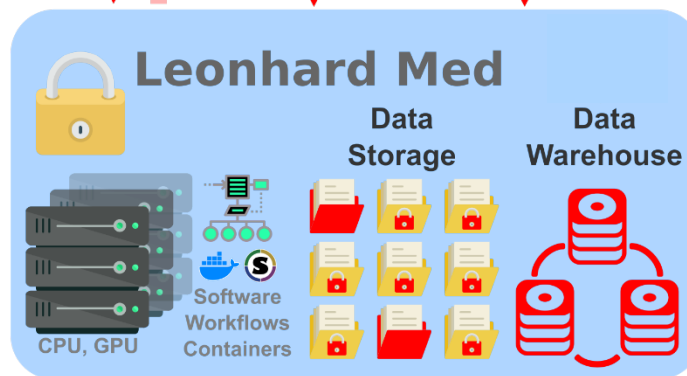


# SPHN/PHRT Project Z ■ > lead SIS ETHZ

Data Sources  
(distinct)



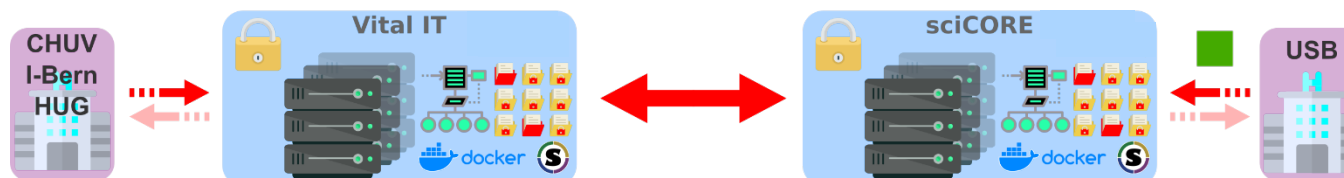
Data Storage & Analysis  
(central; \*federation)



Data, Workflows  
Sharing & Interoperability

BioMedIT

Data Sources  
(distinct)



# Leonhard Med for Personalized Health

## Hub Platform:

- to access, process and share biomedical data **securely**
- for **powerful** computing in a HPC environment
- where the intensive computing and **state-of-the-art software** solutions are readily available
- **dynamically evolving** to match the projects and user needs
- where the **user experience is a central focus, providing balance in terms of security and usability**

## Concluding remarks

- Data driven PH Research is happening worldwide and in Switzerland, as we speak
- Paradigm shift for PH Services: secure and powerful infrastructure is instrumental but not sufficient
- Hybrid expertise is paramount
- Data Security Awareness should be/come the *modus operandi*

# Acknowledgements

hpc-ch forum > Christian Bolliger

## SIS colleagues:

Thomas Wüst

Bernd Rinn

Christian Bolliger

Michal Okoniewski

Balazs Laurenczy

Caterina Barillari

Alex Upton

Allen Neeser

# Thank you !

Scientific IT Services: <https://sis.id.ethz.ch>

Leonhard HPC: <https://scicomp.ethz.ch/wiki/Leonhard>

Personalized Health Data Services: [diana.coman@id.ethz.ch](mailto:diana.coman@id.ethz.ch)