Von klinischen (Routine-)Daten zu FAIRen Forschungsdaten: Der Beitrag des Swiss Personalized Health Network (SPHN) an das Schweizer Gesundheitsdaten-Ökosystem

Dr. Katrin Crameri
Direktorin Personalisierte Gesundheitsinformatik & SPHN Datenkoordinationszentrum
SIB Schweizerisches Institut für Bioinformatik

A project of SAMW ASSM
Schweizerische Akademie der Medizinischen Wissenschaften
Académie Suisse des Sciences Médicales
Accademia Svizzera delle Scienze Mediche
Swiss Academy of Medical Sciences

SIB Swiss Institute of Bioinformatics

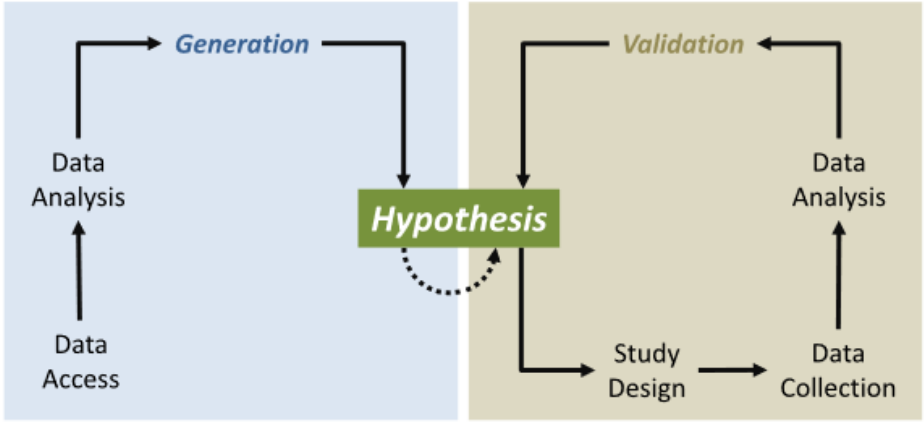
MEDIZIN INFORMATIK INITIATIVE

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

05.10.2022

Der Weg zur personalisierten Gesundheit

Angesichts der riesigen Datenmengen, die im Gesundheitswesen erfasst werden, sollten Entscheidungen in der Gesundheitsversorgung nicht mehr auf Bevölkerungsdurchschnittswerten beruhen, sondern individuelle Patientenmerkmale berücksichtigen.



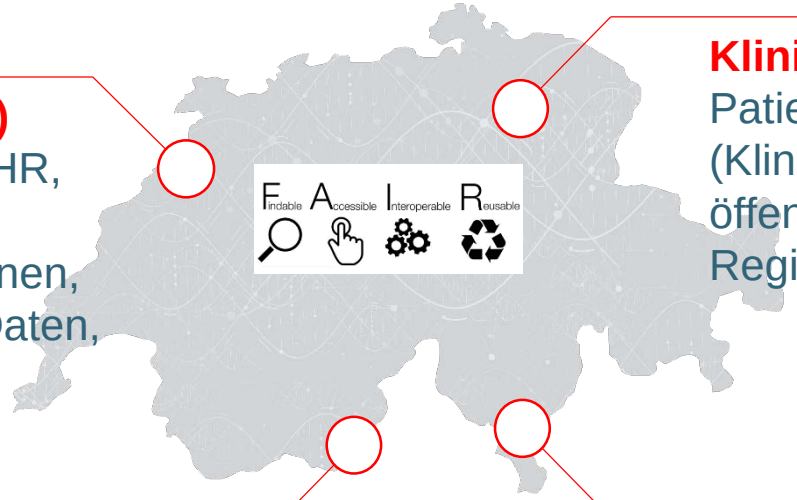
Big Data studies Controlled Clinical Trials

- Präzisere Diagnostik, Prävention, Prädiktion und Behandlung
- Verbesserte Versorgungsqualität
- Learning Health System

https://bmcmedresmethodol.biomedcentral.com/track/pdf/10.1186/s12874-019-0774-0.pdf

Datentypen und Datenquellen für die Forschung

Klinische (Routine) Daten (RWD)
Daten von Leistungserbringern (EHR, med. Register); e.g. Diagnose, Medikation, Vitalzeichen, Operationen, Laborwerte, Bilddaten, Outcome Daten, Kostendaten, etc.



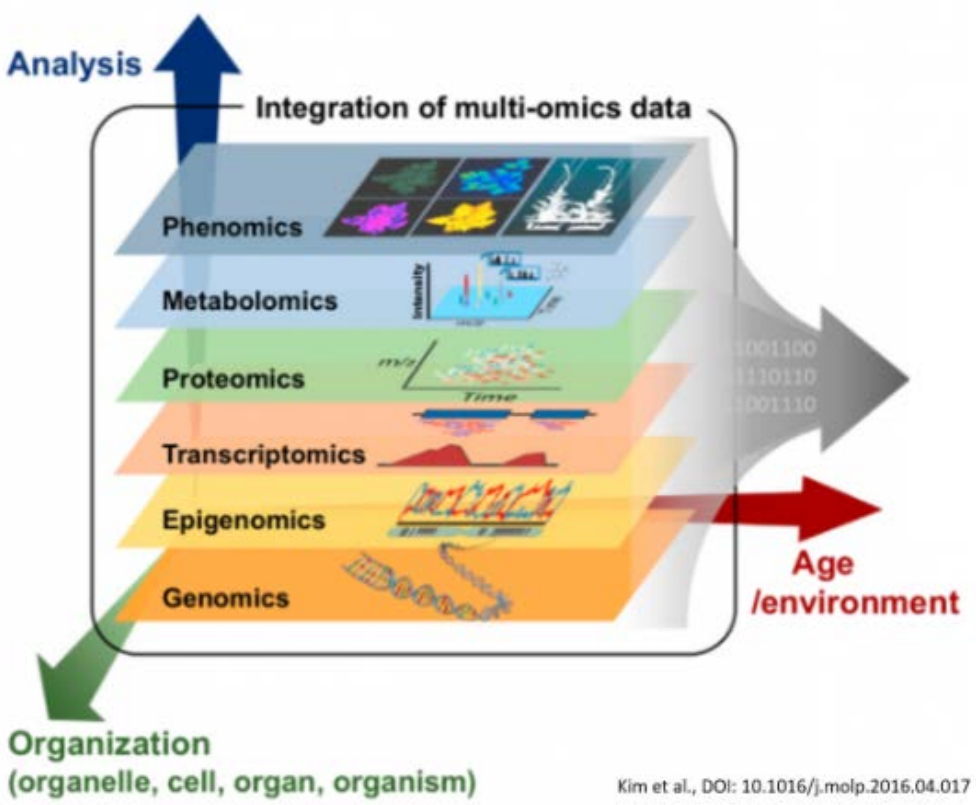
Klinische Forschungsdaten
Patientenorientierte Studiendaten (Klinische Studien, RCT, für die öffentliche Gesundheit relevante Register, longitudinale Kohorten usw.)

Molekulare und *omics Daten
(e.g. Genomics, epigenomics, transcriptomics, proteomics, metabolomics, Biomarker, etc.)

Healthy citizen Daten
Tracking Apps, Wearables, Sensors, Lebensstil-Daten, Daten aus sozialen Medien, etc.

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

Biologische Ebenen der personalisierten Gesundheit



Kim et al., DOI: 10.1016/j.molp.2016.04.017

Strategic Focus Area
Personalized Health and Related Technologies

Genomics	Proteotyping	Metabolomics & Lipidomics
CGAC Clinical Genomic Analysis Center	CPAC Clinical Proteotype Analysis Center	CMAC Clinical Metabolomics Analysis Center
Clinical Grade Sequencing (ISO 15189 accredited)	Quantitative Proteotyping	Small molecule analysis
Whole Genome Sequencing (WGS)	Proteotype analysis (DDA, DIA, PRM)	Targeted metabolomics
Whole Exome Sequencing (WES)	Post-translational modification analysis	Untargeted metabolomics
RNA Sequencing (RNASeq)	Spatial proteotype analysis	Lipidomics
Further information and pricing	Further information and pricing	Further information and pricing
Fast-track genomic data analysis	Fast-track proteomic data analysis	Fast-track metabolomic data analysis

Integrated Multi-Omic Data Interpretation

Integration, visualization, and analysis of omics data mapped onto biological networks and pathways

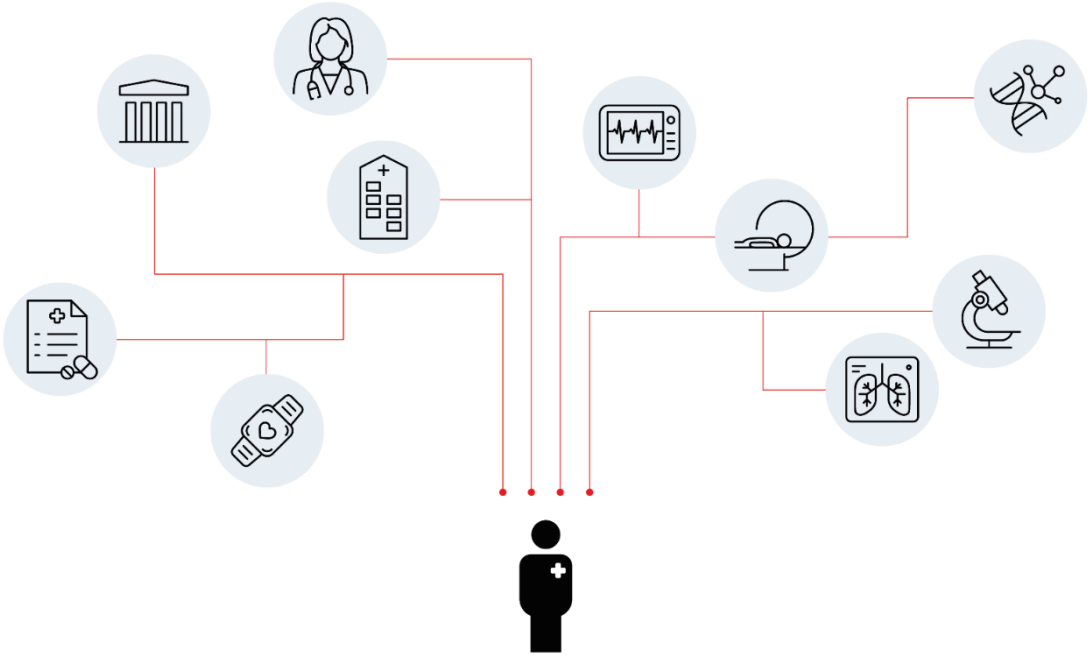
BioMedIT integration for data lineage tracking, data management, data sharing, secondary and tertiary analysis

ETH PHRT Swiss Multi-Omics Center: http://smoc.ethz.ch

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

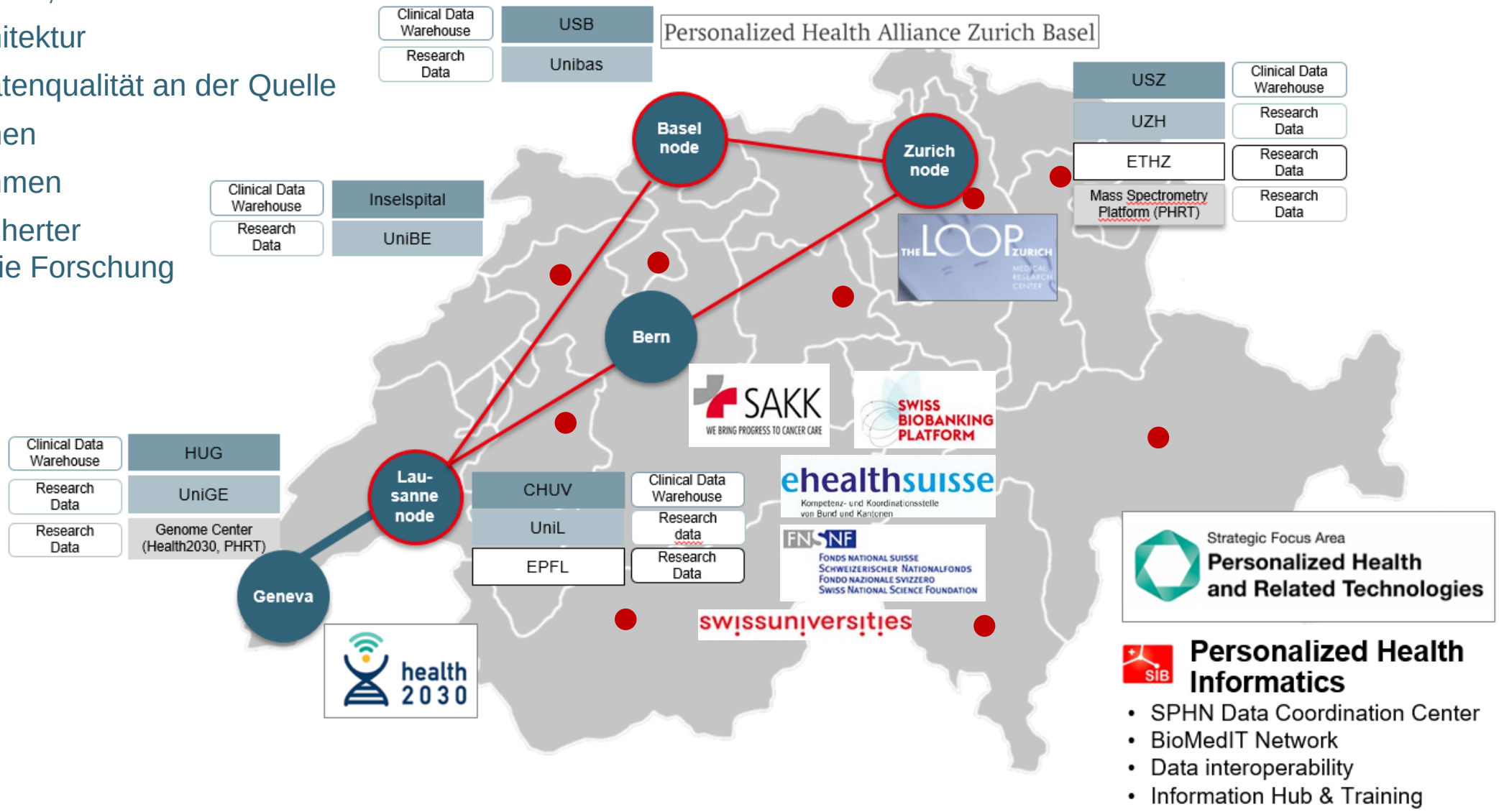
Personalisierte Gesundheit: Die 🇨🇭 Herausforderungen

- Nutzung von Routinedaten aus der Versorgung: fehlende Digitalisierung, fehlende Strukturierung und fehlende Standards
- Zusammenführung heterogener Datensätze aus diversen Quellen
- Datenqualität versus Patientenrealität
- Grosse, multizentrische Projekte mit vielen Akteuren
- Wie bringen wir sensible Daten zusammen und wo analysieren wir sie?



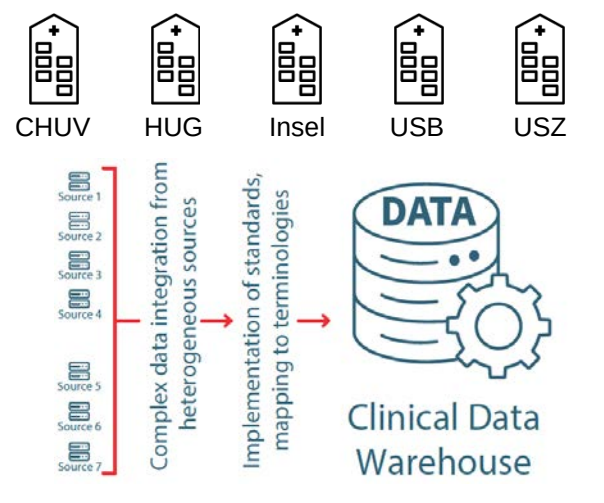
SPHN Setup und Prinzipien

- Forschungsinfrastruktur Initiative
- Laufzeit 2017-2024, CHF 135 Mio
- Dezentrale Architektur
- Investition in Datenqualität an der Quelle
- Ethischer Rahmen
- Rechtlicher Rahmen
- BioMedIT: gesicherter Datenraumfür die Forschung

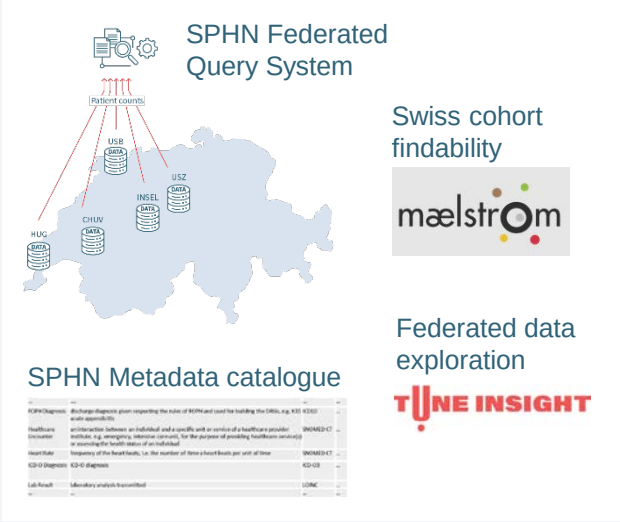


Erreichte Etappenziele von SPHN Ende 2022

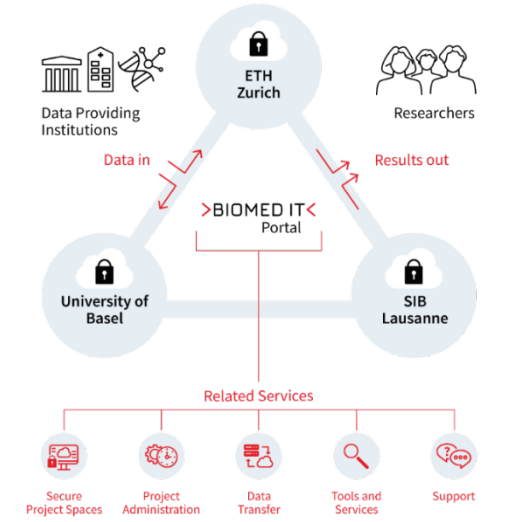
Klinische Datenplatformen an den Universitätsspitälern



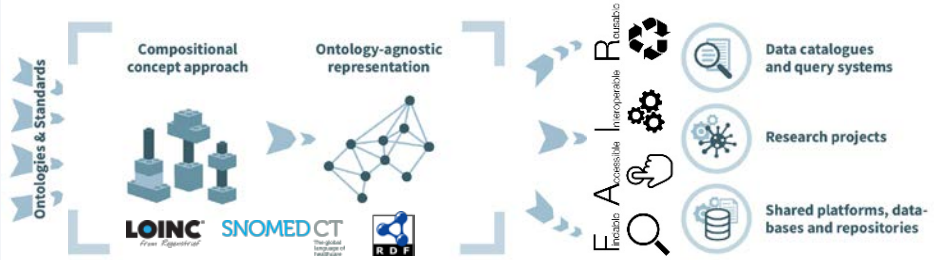
Auffindbarkeit von Daten



BioMedIT Network



SPHN Interoperabilitätsstrategie



Rechtl. Rahmen

Schulungen und Trainings

- Responsible Use of Health-Related Data for Research
- FAIR Health Data for Research
- Information Security
- BioMedIT: Research with Sensitive Data

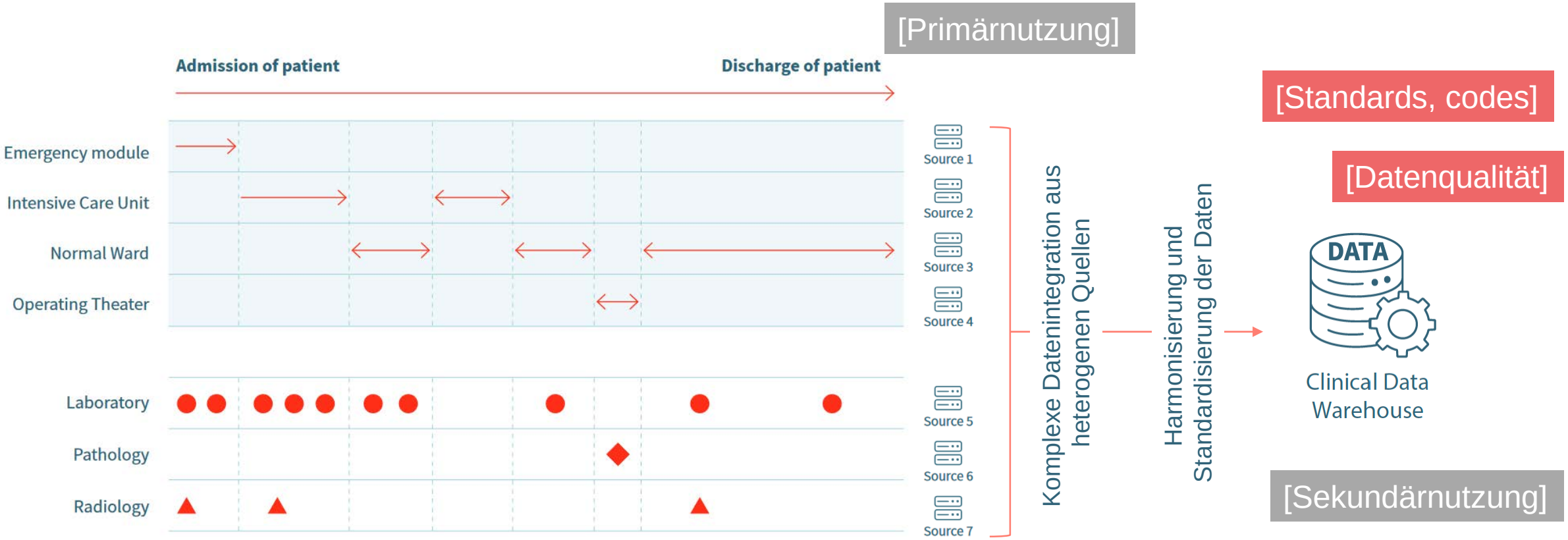
GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

Infrastrukturaufbau an den fünf Universitätsspitälern (UH)

A project of
SAMW ASSM
Schweizerische Akademie der Medizinischen Wissenschaften
Académie Suisse des Sciences Médicales
Accademia Svizzera delle Scienze Mediche
Swiss Academy of Medical Sciences
SIB Swiss Institute of Bioinformatics
MEDIZIN INFORMATIK INITIATIVE
GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

05.10.2022

Die Datenspur der Patient*innen im Spital



Sekundärnutzung von Gesundheitsdaten

Politische Entscheidungsträger

- Policy making

Regulierer / Behörden

- Public Health, Marktzulassung, Vigilanz
- Statistik, Monitoring
- uvm

Forschung

- Machbarkeitsabklärungen (Feasibility)
- Forschungsprojekte (z.B. Hypothesen-generierung, Biomarkersuche, ML/prediction, Referenz-kohorten, Arzneimittelwirkung/unerw. Wirkung, Prozess/Qualitäts-verbesserung, etc.)
- Rekrutierung von Patientenkohorten

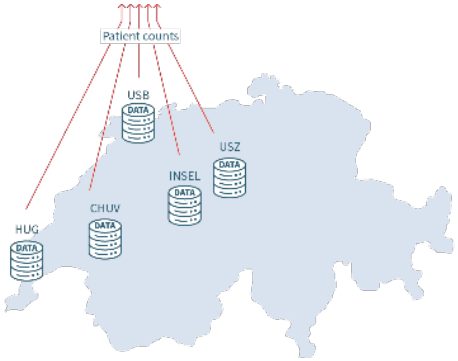
Leistungserbringer (operative Nutzung und Entwicklung)

- Datenauswertungen, Statistiken und Reports für Anwendungszwecke (z.B. Benchmarking, QS, Controlling, Simulationen oder prädiktive Einsatzplanung, etc.)
- Datenbereitstellung für Anwendungsfälle wie Tumorboard, Behandlungs-Dashboard, Gerätemanagement, Medikamenteninteraktionen, CIRS-Fälle, etc.
- Datenlieferungen für (medizinische) Register
- Datenaufbereitung und -übermittlung an Behörden (Meldepflicht), Patienten und Primärsysteme
- Betriebliches Monitoring, Qualitätsüberwachung, Life Cycle Management, Validierung der Datenqualität, Datenschutzverletzungen
- Process Mining, Forecasting, RPA
- Entwicklung Clinical-Decision-Support-Systeme

Mit freundlicher Genehmigung von Max Grimm

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

Auffindbarkeit von Daten



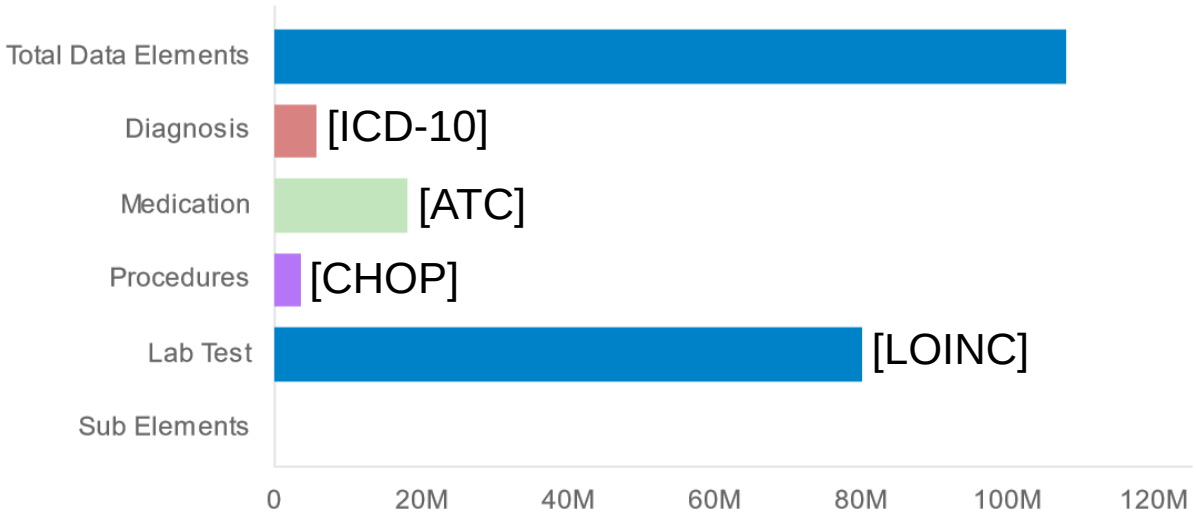
Föderiertes Abfragesystem: Machbarkeitsanalysen; über Daten aus allen 5 UH mit einem Tool durchsuchbar
> 0,5 Mio. Patienten (mit Generalkonsent); > 100 Mio. Datenelemente

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

SPHN Federated Query System

Daten der klinischen Datenplattformen aller fünf Universitätsspitäler (USB, USZ, Insel, CHUV, HUG)

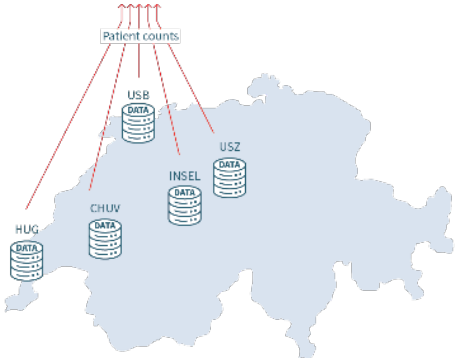
Insgesamt >100’000’000 Datenelemente



Über 500’000 Patient*innen (mit unterzeichnetem Generalkonsent)

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

Auffindbarkeit von Daten



Föderiertes Abfragesystem: Machbarkeitsanalysen über Daten aus allen 5 UH mit einem Tool durchsuchbar
> 0,5 Mio. Patienten (mit Generalkonsent); > 100 Mio. Datenelemente

Metadaten von 10 bevölkerungsbasierten und krankheitsspezifischen CH Kohorten im internationalen Maelstrom-Katalog; Vollständige Studienbeschreibungen durchsuchbar mit > 145'000 annotierten Variablen

Datenexploration und -analyse (Pilotphase):
TI4Health: Reichhaltiges, standardisiertes, strukturiertes Datenset von >1000 Melanompatient*innen

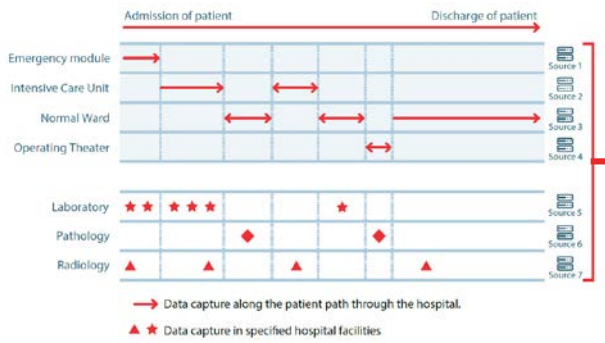
SPHN Interoperabilitäts-Strategie

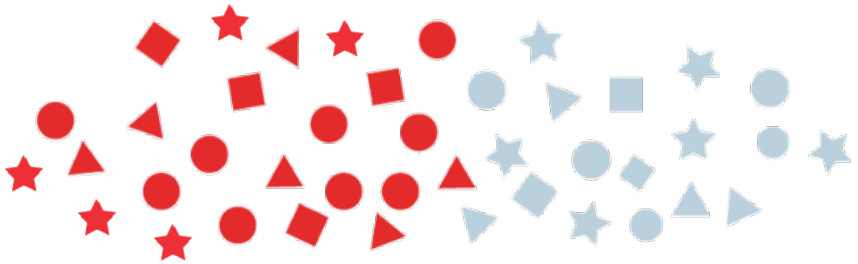
A project of
SAMW ASSM
Schweizerische Akademie der Medizinischen Wissenschaften
Académie Suisse des Sciences Médicales
Accademia Svizzera delle Scienze Mediche
Swiss Academy of Medical Sciences
SIB Swiss Institute of Bioinformatics
MEDIZIN INFORMATIK INITIATIVE
GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

05.10.2022

Die Notwendigkeit der Interoperabilität

Klinische Routinedaten



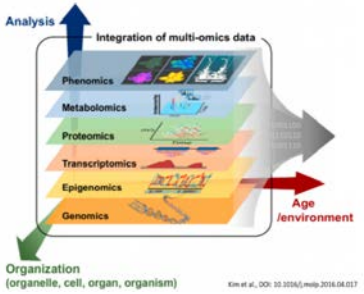


Andere gesundheitsbezogene Daten

ehealthsuisse
Kompetenz- und Koordinationsstelle von Bund und Kantonen

SWISS BIOBANKING PLATFORM

health2030 genome center

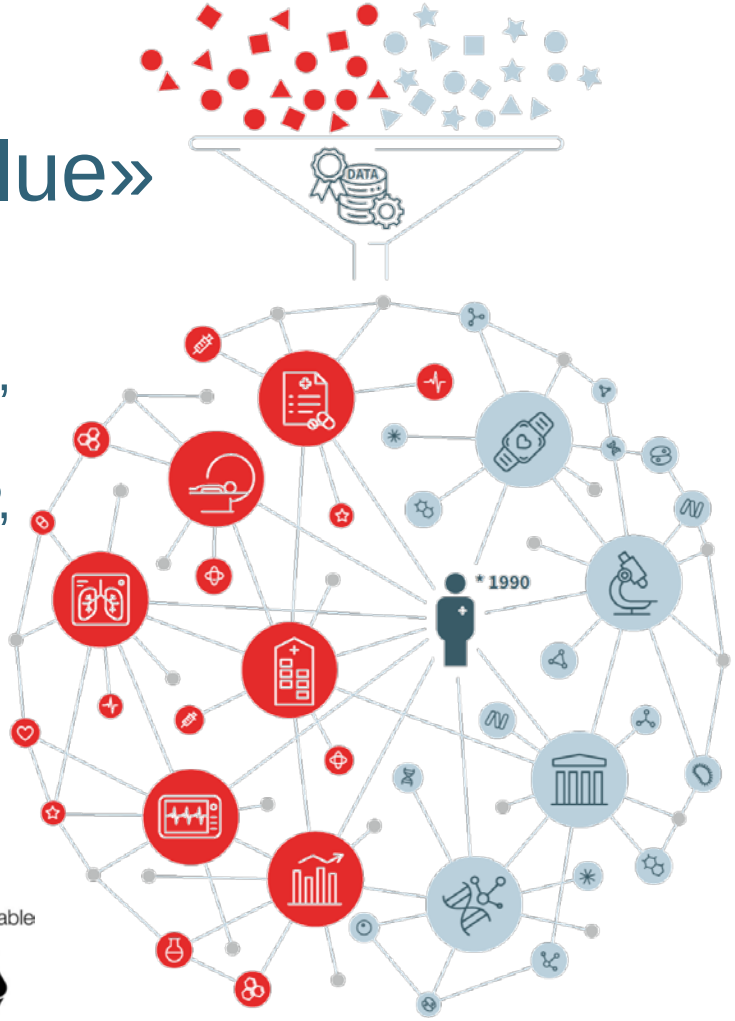


etc.

→ Die Bedeutung der Daten verstehen

→ Einfache Verknüpfung von Daten aus verschiedenen Quellen ermöglichen

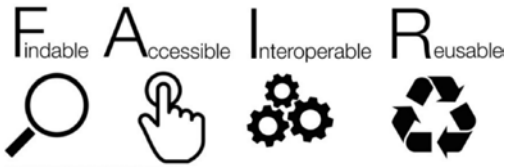
Der SPHN-Interoperabilitätsrahmen: «making sense of data and adding value»



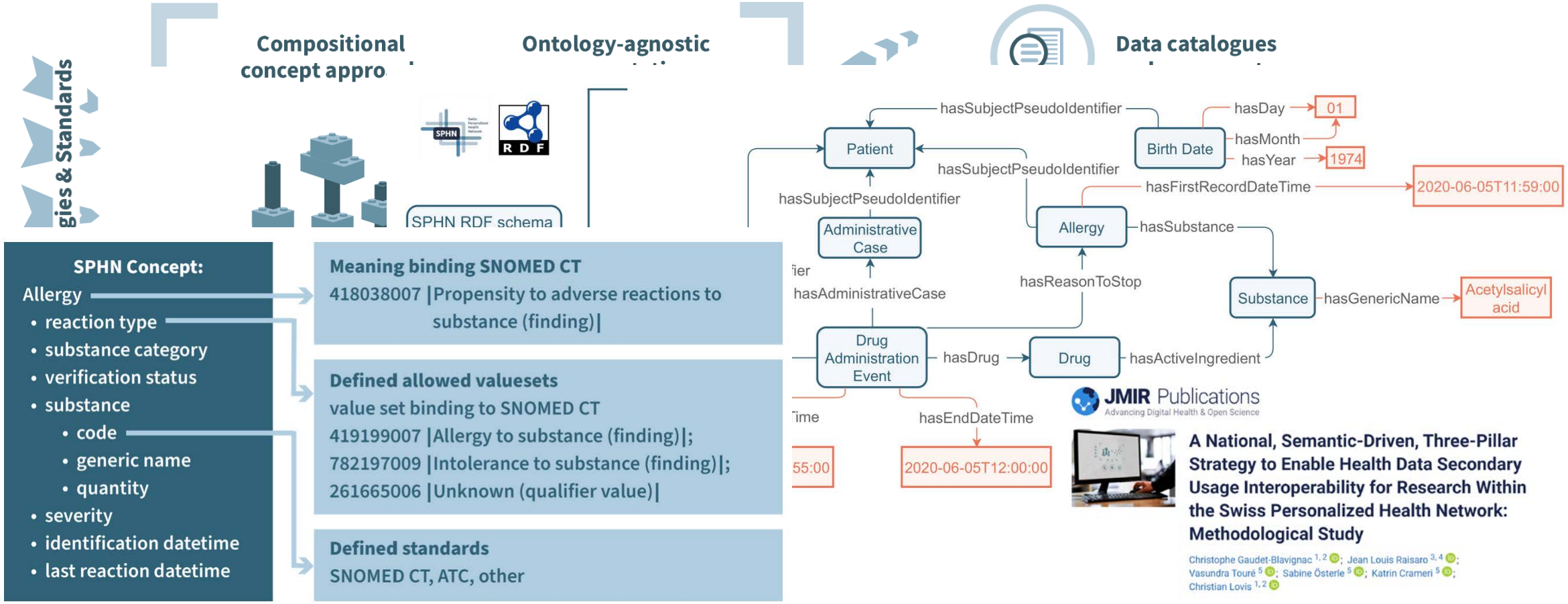
Semantik
- Verwendung eines kontrollierten Vokabulars für Konzepte, Valusets und Daten
- Daten-Standards: ATC, SNOMED CT, ICD-10-GM, CHOP, UCUM, LOINC, NANDA, ICD-O, MedDRA...
- Mehrstufige Kodierung und Mappings

Sprache für “knowledge representation”
- Semantic Web (RDF Schema)

Qualitätskontrolle
- Semantic Web (SHACL, SPARQL)



Die SPHN Interoperabilitäts-Strategie



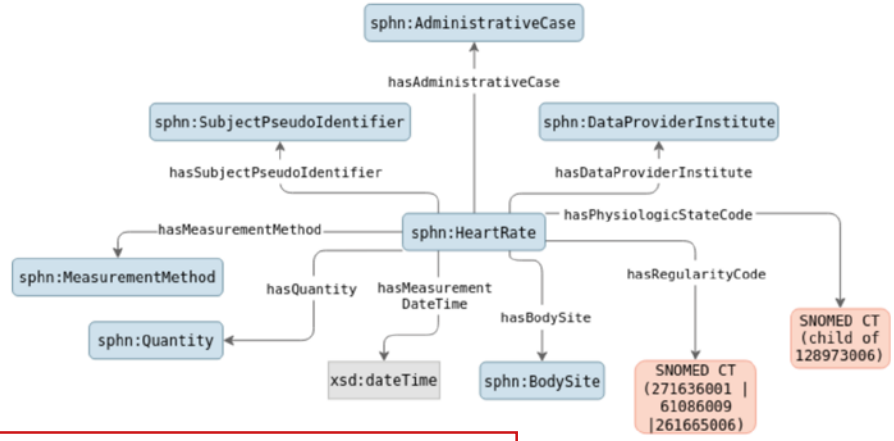
SNOMED CT meaning binding

Heart Rate [c]

URI https://biomedit.ch/rdf/sphn-ontology/sphn#HeartRate

Description frequency of the heart beats, i.e. the number of time a heart beats per unit of time

Schema representation



Meaning binding (Equivalent-classes)
LOINC 8867-4 | Heart rate | [c]
SNOMED 364075005 | Heart rate (observable entity) | [c]

Definition von SPHN Konzepten

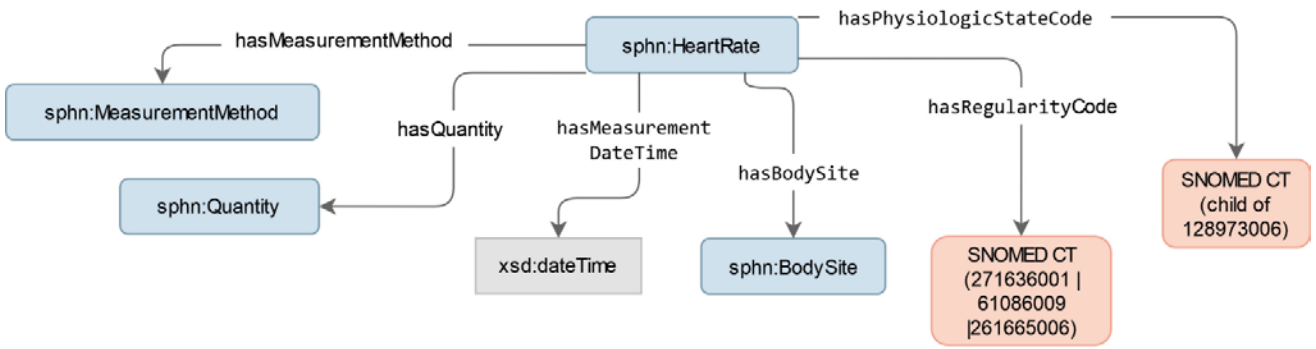
e.g., SPHN Concept Heart Rate – gebunden an SNOMED CT Konzep 364075005 |Heart rate (observable entity)|

→ Implementiert via owl:equivalentClass

→ 47/63 SPHN Konzepten hat ein meaning binding

→ Auch für LOINC angewandt wo immer möglich (LOINC: 8867-4 Heart rate).

SNOMED CT value set binding



Restrictions

▾ has physiologic state code [op] : children of the given classes are allowed

SNOMED 128973006 | Finding related physiologic patient state (finding) | [c]

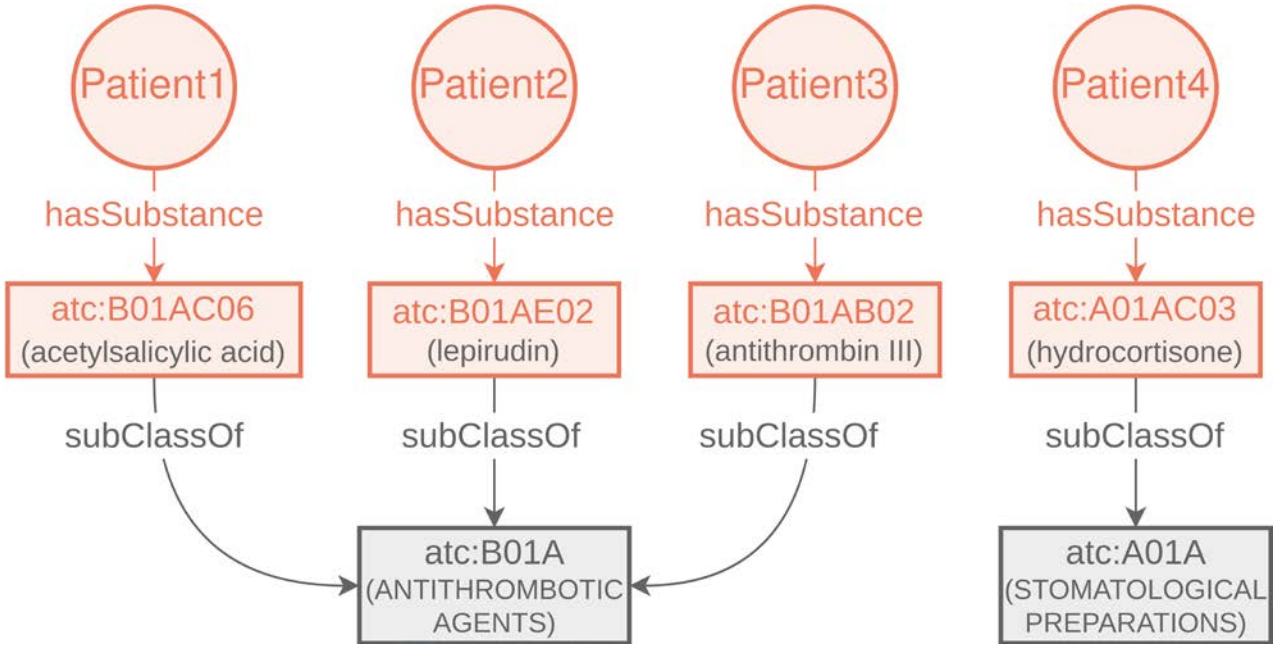
▾ has regularity code [op] : children of the given classes are not allowed

SNOMED 261665006 | Unknown (qualifier value) | [c] SNOMED 271636001 | Pulse regular (finding) | [c] SNOMED 61086009 | Pulse irregular (finding) | [c]

- Spezifische value sets: Eindeutige und ausgewählte Anzahl von SNOMED CT-Codes erlaubt
- Allgemeine value sets: Ganzer (oder mehrere) Teilgraph(en) als mögliche Werte erlaubt
- 37/52 value sets haben SNOMED CT value set binding

RDF-Inferenz durch die Verwendung von Terminologien

- Ontologien addieren Informationen zu den Daten (durch Klassen, hierarchische Strukturen usw.)
- Maschinen verstehen Codes
- Computer können schlussfolgern und Wissen ableiten, z.B. welche Patienten mit antithrombotischen Mitteln behandelt wurden



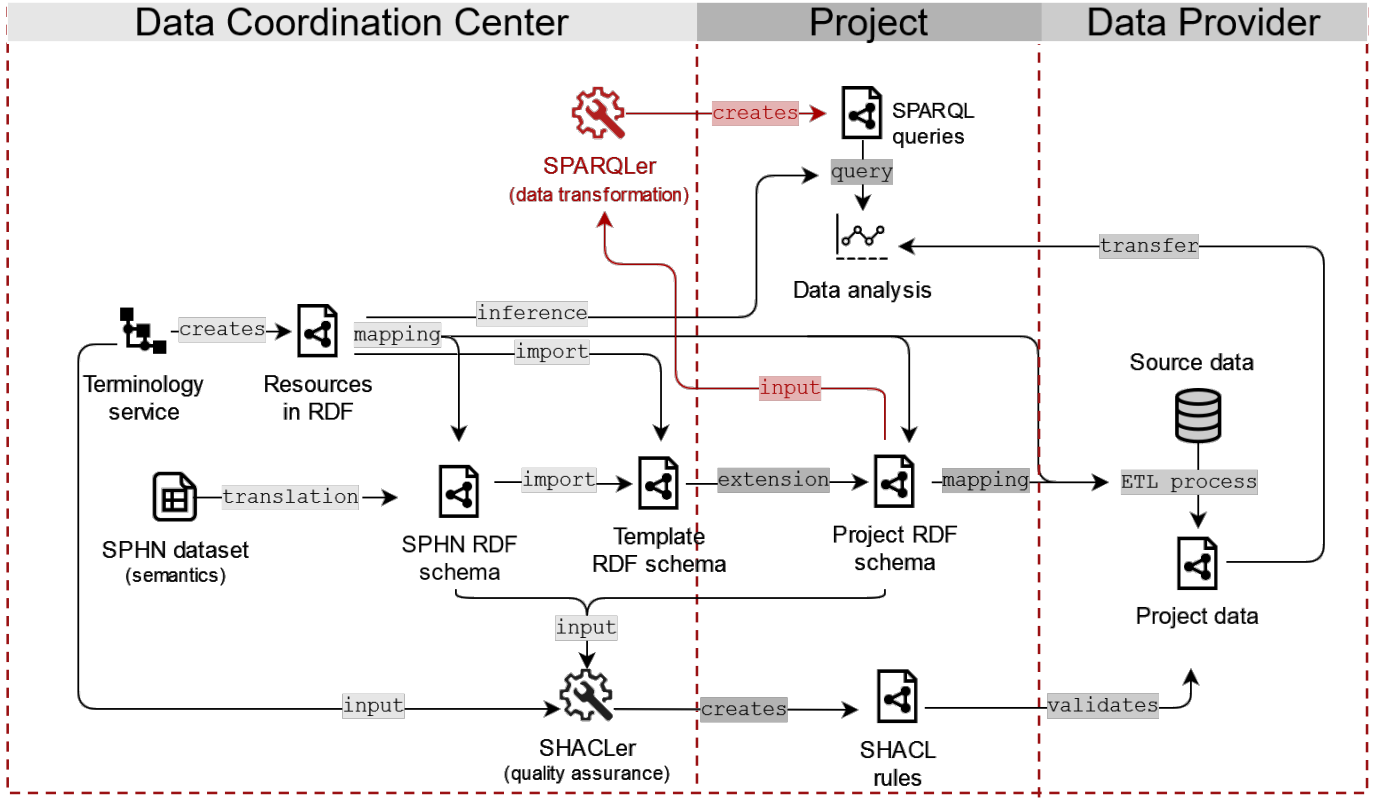
Vereinfachtes Beispiel: In ATC kodierte Stoffe enthalten hierarchische Informationen, die eine Schlussfolgerung aus den Daten ermöglichen. Acetylsalicylsäure, Lepirudin und Antithrombin III sind Antithrombotika, aber nicht Hydrocortison

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

Das «SPHN ecosystem towards FAIR data»

DCC Infrastrukturkomponenten und Services:

- SPHN Datenset (>80 Konzepte)
- SPHN Daten Schema (RDF)
- Vorlagen und Guidelines
- Terminologie Service
- Tools für Qualitätssicherung



preprints.org > medicine & pharmacology > other > doi: 10.20944/preprints202109.0505.v1

Preprint Technical Note Version 1 Preserved in Portico This version is not peer-reviewed

The SPHN Ecosystem Towards FAIR Data

Sabine Österle *, Vasundra Touré, Katrin Crameri

Version 1 : Received: 27 September 2021 / Approved: 29 September 2021 / Online: 29 September 2021 (17:32:40 CEST)

Open Access Article

DCC Terminology Service—An Automated CI/CD Pipeline for Converting Clinical and Biomedical Terminologies in Graph Format for the Swiss Personalized Health Network

by Philip Krauss [1], Vasundra Touré [2], Kristin Gnodtke [2], Katrin Crameri [2] and Sabine Österle [2,*]

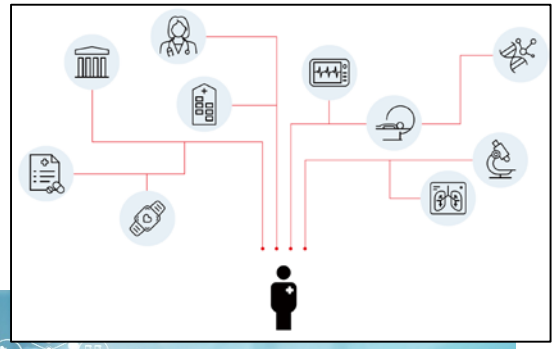
BioMedIT: Eine private Cloud für die Analyse von sensiblen Big Data

A project of

05.10.2022

Systematische Nutzung sensibler Daten in standortübergreifenden Forschungsprojekten: Was braucht es ?

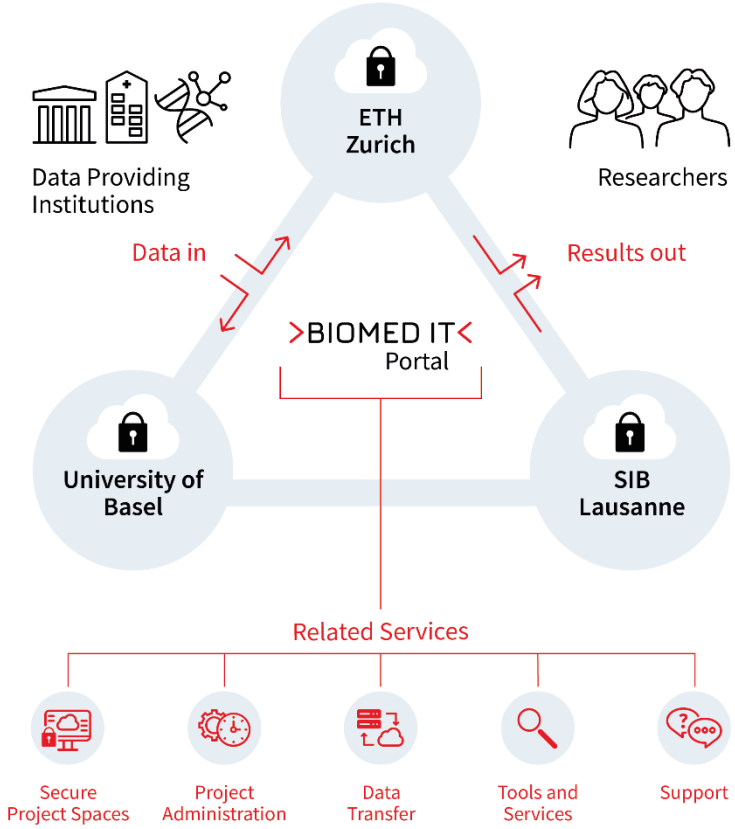
- Möglichkeiten zur sicheren Datenmobilisierung
- Leistungsstarke IT-Infrastrukturen für Big Data Computing und Speicherung (Hochleistungsrechenzentren)
- Sicherheitsmassnahmen für IKT-Systeme zum Schutz vertraulicher Informationen vor unbefugtem Zugriff, unbefugter Nutzung, usw.



GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

BioMedIT: Eine nationale, sichere IT Plattform

- Cloud-Technologie, aufgespannt zwischen ETHZ, Unibas und Unil/SIB, managed by SIB
- Daten- und Rechendienste (HPC), Fernzugriff, kollaborative Analysen
- Anbindung von >28 Datenanbieter, Ende-zu-Ende-verschlüsselte Datenübertragung
- Datenschutz und Datensicherheit «by design»
- Eine übergreifende Information Security Policy
- Daten verbleiben auf BioMedIT



GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

BioMedIT: Eine nationale, sichere IT Plattform

BioMedIT Nodes

sciCORE

Universität Basel

OpenStack
1.0 PB Storage
Compute Capacity:
•500 CPUs
•8 GPUs
•5 TB RAM
HPC Scheduler: SLURM
File System: CEPH, GlusterFS

openstack. slurm LabKey

SENSA

UNIL | Université de Lausanne · SIB Swiss Institute of Bioinformatics

OpenStack
400TB Storage
Compute Capacity:
•800 CPUs
•6 GPUs
•3 TB RAM
HPC Scheduler: SLURM
File System: Weka, CEPH, Scality

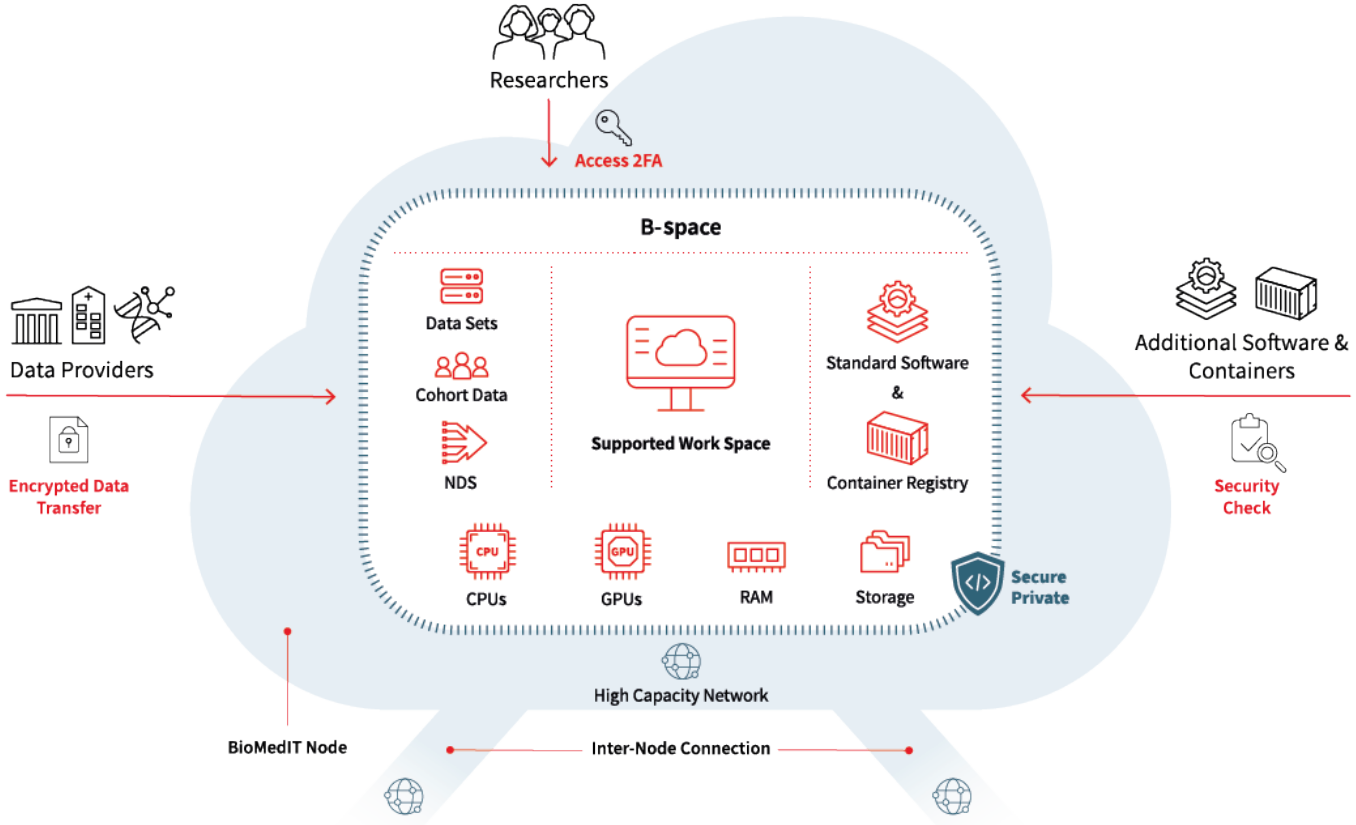
openstack. slurm WEKA

SIS

ETH zürich

OpenStack
2.0 PB Storage
Compute Capacity:
•4400 CPUs
•424 GPUs
•41 TB RAM
HPC Scheduler: SLURM
File System: Luster file storage

openstack. slurm openBIS



Anforderungen an die Datennutzung für Forschungsprojekte

A project of
SAMW ASSM
Schweizerische Akademie der Medizinischen Wissenschaften
Académie Suisse des Sciences Médicales
Accademia Svizzera delle Scienze Mediche
Swiss Academy of Medical Sciences
SIB Swiss Institute of Bioinformatics
MEDIZIN INFORMATIK INITIATIVE
GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

05.10.2022

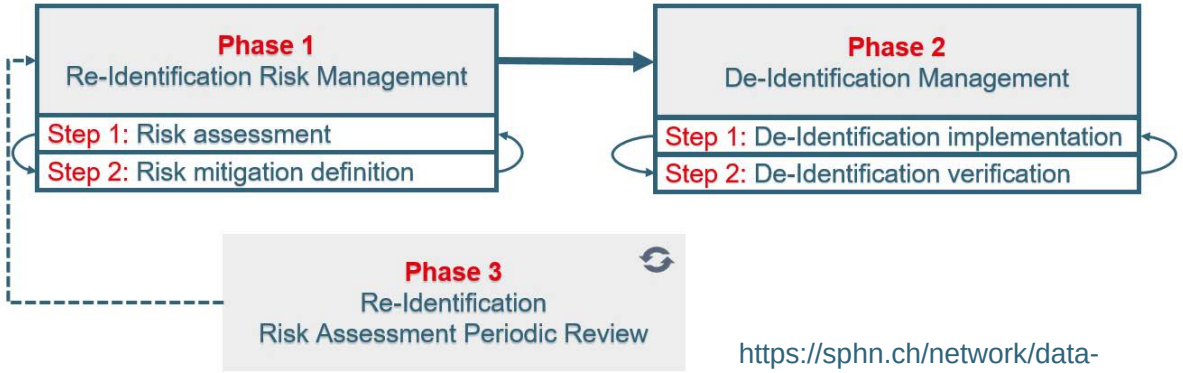
Anforderungen an die Datennutzung für Forschungsprojekte

- ✓ Projektplan
- ✓ Einverständniserklärung der Patienten (Konsent)
- ✓ Genehmigung der Ethikkommission
- ✓ Institutionelle Genehmigung (Data Governance Board)
- ✓ Technische Sicherheitsmassnahmen
- ✓ De-identifizierte Daten (pseudonymisiert/kodiert)
- ✓ Rechtliche Vereinbarung(en) zwischen den Projektpartnern (Konsortialvertrag/Datentransfer- und Nutzungsvertrag)

Individuelle Datenebene

De-Identifizierung gesundheitsbezogener Daten in Übereinstimmung mit den gesetzlichen Bestimmungen

- Rechtsgutachten als Auslegeordnung (risk-based vs rule-based)
- Praktische Anleitung für De-Identifizierung von Daten nach quantitativen und qualitativen risikobasierten Ansatz
- Pragmatischen Minderung des Re-Identifizierungsrisikos durch die Erstellung von De-Identifizierungsregeln



https://sphn.ch/network/data-coordination-center/de-identification/

Individuelle Datenebene

Pseudonymisierung, Anonymisierung

- Projektspezifische Patient*innen-IDs
- Veränderung von nicht ergebnisrelevanten Datenpunkten (z.B. Encounter date)

Verschleierung, Aggregierung, Verschlüsselung (bei Abfragesystemen in SPHN)

- Teilweise Aggregierung (z.B. Altersgruppenbildung)
- Verschleierung (*obfuscation*, z.B. Datenmengen <10 werden nicht angezeigt)
- K-Anonymität: Identifizierende Informationen jedes einzelnen Individuums von mindestens k-1 anderen Individuen ununterscheidbar
- Verschlüsselungen der lokalen Einzelergebnisse bei föderierten Abfragen

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

SPHN-Vorlagen für rechtliche Vereinbarungen

Overview of available legal agreement templates and their contents

The three types of agreement can be included in a single agreement (e.g.: CA+DTUA+CA; see templates below). When the parties enter into a CA for the purpose of exchanging data and wish to use a subcontractor for data processing the CA integrates a DTUA and DTPA.

- Consortium Agreement (CA)
- Consortium Agreement + DTUA + DTPA (multi node)
- Consortium Agreement + DTUA + DTPA (single node)

- DTUA + DTPA (multi node)
- DTUA + DTPA (single node)

- DTPA (multi node)
- DTPA (single node)

- DTUA without an external processor

www.sphn.ch/dtua

Verbleibende Herausforderungen

Herausforderungen und offene Fragen

- Eine Forschungsinfrastruktur-Initiative kann die digitale Transformation im Gesundheitswesen unmöglich ersetzen
- Jeder möchte mit klinischen Daten aus der Praxis arbeiten, aber niemand ist bereit, für die Standardisierung und Strukturierung der Daten zu zahlen
- «Standards don’t send you money now, they send you money later» (G. Grieve)
- FAIRe Daten nützen immer den anderen: Wie setzen wir die richtigen Anreize?
- Datenschutz und Datensicherheit machen Prozesse oft komplexer: Wie können wir Prozesse straffen?
- Datenschutz und Datensicherheit sind teuer: Wer kommt (langfristig) für die zusätzlichen Kosten im Rahmen der sicheren Datenräume auf?

«Take home» Botschaften

- Im lernenden Gesundheitssystem von morgen müssen Gesundheitsversorgung und Forschung Hand in Hand gehen
- Die Anforderungen an Daten sind immer gleich, egal ob Versorgungsqualität, Patientenakte, öffentliche Gesundheit, Patientensicherheit, etc.
- Es braucht eine nationale Datenstrategie, um nachhaltige, skalierbare Lösungen zu erreichen, damit:
  - die Schweiz den Anschluss nicht verliert,
  - die datengetriebene Forschung sich nicht nur auf ausländische Quellen stützt,
  - ein Gesundheitsdaten-Ökosystem mit breiter Nutzung entstehen kann

Acknowledgements

Personalized Health Informatics Group:

Owen Appleton, Jan Armida, Katrin Crameri, Patricia Fernandez Pinilla, Kristin Gnodtke, Simone Guzzi, Petar Horki, Shubham Kapoor, Julia Maurer, Michael Müller-Breckenridge, Sabine Österle, Christian Ribeaud, Vasundra Touré, Deepak Unni

The SPHN NSB and NAB, Task Forces & WGs

The BioMedIT Board and workforces @ ETHZ, Unibas, Unil/SIB; SIB Management

The SPHN Management Office: Thomas Geiger, Liselotte Selter, Sarah Vermij

@CrameriKatrin
@SPHN_ch

Katrin.Crameri@sib.swiss
dcc@sib.swiss I info@sphn.ch

www.sphn.ch I www.sib.swiss/phi
www.BioMedIT.ch