Next steps for nationwide data networks

The Swiss Personalized Health Network (SPHN) and BioMedIT: Enable use and exchange of interoperable health data for research

Katrin Crameri, PhD, MPH
Director Personalized Health Informatics, SIB
Needs to drive data-intensive infectious disease research and practice

Expand shared open-access datasets and establish standards for data integration:

Recommendations

- Utilize new technologies and best practices to support innovative data collection methods that lower the barriers for researchers to collect, integrate, and share data.
- Provide incentives for researchers to harmonize data and metadata, and promote transparent and reproducible research.
- Establish environments that provide mechanisms of aggregating and harmonizing independent data sets and make high quality data available more quickly and in more useful formats for access by scientists and machines.

National Institute of Allergy and Infectious Diseases, National Institutes of Health 2018
Data-streams for ID research

Medical encounter data
EHR, insurance claims, hospital discharge records, death certificates, etc.
→ information on disease status, symptoms, confirmed infections, drug or vaccine-related adverse events
→ individual level or aggregated

Participatory syndromic data
Crowd-sourced data in which volunteers self-report an array of symptoms.
→ no provision of confirmed infection status for specific pathogens but
→ provide individual-level health data in near real time

Molecular or *omics data
Data of host, pathogen, and vector: Genomics, transcriptomics, proteomics, metabolomics, DNA methylation, etc.

Non-health digital data
Derive from use of Internet search engines, social media, or mobile phones.
→ provide information on health-related behaviour, including contact and travel patterns, vaccine status and sentiments
→ key for understanding and modelling disease transmission

Findability
Accessibility
Interoperability
Reusability
The clinical research setting
The health data research setting

Different:
- Data types
- Data standards
- Data formats

Distributed High capacity

Findability? Accessibility?

Data sources

Data governance regulations
Legal framework (DTUA)

Re-use:
- Data-sharing
- combination of DS
- Secure Transport
- Long-term storage
  ...

Petr Holub et al., Biopreserv Biobank. 2018 Apr;16(2):97-105.
Nationwide data networks need **FAIR data**

**Findability:** Data in silos, small total numbers, missing catalogues, lacking meta-data

**Accessibility:** Who to contact? Data governance, lock-in privilege, willingness to share?

**Interoperability:** Heterogeneity across data sources; formal, accessible, shared, and broadly applicable language for knowledge representation = hard work

**Reusability:** Proper description, legal and ethical restraints (consent, DTUA, ethics approval), long-term storage strategies?
Government funded initiative, mandated to SAMS and SIB:

- **2017-2020**: focus on university hospitals, bottom-up projects, and BioMedIT (CHF 68 Mio)
- **2021-2024**: consolidate and expand the network (CHF ?)

Promote the development, implementation, and validation of coordinated infrastructures to make health data interoperable and shareable for research in CH
SPHN: a research infrastructure initiative

- **A decentralized approach:** SPHN does not build a central data lake. Data stays with the data providers and is used in a project-specific manner.

- **Getting data out of silos:** Working towards a FAIR use of health data for research, and a more democratic access to data.

- **Building, connecting and aligning systems:** Establishing a national infrastructure network, consisting of various modules and components, harmonizing systems: improve data quality through structuring, standardisation.

- **Reaching interoperability of data:** semantic and technical interoperability – within projects, across projects, across systems, over time, internationally aligned.
SPHN Infrastructure components (examples)

Development of research data management systems at the 5 Swiss university hospitals with a connection to Biobanks / sample information;

Deployment of a distributed federated query system over all university hospitals to enable feasibility studies

New technologies and innovative solutions for consent management, data governance, data collection, de-identification, etc.; Establishment and maintenance of shared platforms, cohorts/registries and data-bases for research and health care

Creation of a national infrastructure resource to securely transfer, store, manage and process confidential research data (BioMedIT) that can jointly be used
Data interoperability – The landscape of health data standards and data models

-> There is no easy or one-size-fits-all solution

-> Every model or standard has its strength in a certain domain

Thanks to Kees van Bochove, Founder, The Hyve @keesvanbochove
SPHN Clinical data semantic strategy

Making care system data available for research:
→ Align orthogonal visions and diverse use cases
→ How to know today what will be important for tomorrow’s research?
Compositional approach of the semantic framework

1. Concepts are generalizable building blocks which can be used in different contexts
2. Each concept contains all information necessary to understand it
3. Concepts can be combined to composed concepts, which again can be combined into more complex compositions

---

<table>
<thead>
<tr>
<th>concept or attribute</th>
<th>concept name</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>concept</td>
<td>Biosample</td>
<td>a specific quantity of biological material</td>
</tr>
<tr>
<td>attribute body site</td>
<td>Body site</td>
<td>body site of biosample collection</td>
</tr>
<tr>
<td>attribute datetime</td>
<td>Date time</td>
<td>datetime of biosample collection</td>
</tr>
<tr>
<td>attribute material</td>
<td>Material</td>
<td>the type of biological material</td>
</tr>
<tr>
<td>concept</td>
<td>Biobanksample</td>
<td>biosample stored in a biobank</td>
</tr>
<tr>
<td>attribute sample ID</td>
<td>Sample ID</td>
<td>unique identifier of sample</td>
</tr>
<tr>
<td>attribute biosample</td>
<td>Biosample</td>
<td>a specific quantity of biological material</td>
</tr>
<tr>
<td>concept</td>
<td>Lab Result</td>
<td>laboratory analysis transmitted</td>
</tr>
<tr>
<td>attribute structured lab result</td>
<td></td>
<td>code of information transmitted</td>
</tr>
<tr>
<td>attribute biosample</td>
<td>Biosample</td>
<td>a specific quantity of biological material</td>
</tr>
<tr>
<td>attribute analysis date time</td>
<td></td>
<td>the analysis takes place</td>
</tr>
<tr>
<td>attribute value</td>
<td>Result</td>
<td>result of the laboratory</td>
</tr>
<tr>
<td>attribute unit</td>
<td>Unit</td>
<td>unit of the result</td>
</tr>
</tbody>
</table>
Concepts have recommended standards

Value sets or terminology standards are define for each concept, wherever possible national and international standards are applied

Lab Result

→ LOINC

Code

→ SPHN value set e.g. taken from eHealth Suisse, Federal Office of Statistics, FHIR …
Description formalism for transport and storage

- High degree of flexibility
- Data model-agnostic
- Concepts need to be expressible without changing the semantic meaning

→ Resource Description Framework (RDF)
- Flexible triple store
- Easy queries with SPARQL
- Could be a good solution also outside SPHN
Data models

Numerous data models

- CDISC mandatory model for FDA
- HL7 - Healthcare interoperability
- CDA, RIM, FHIR, ...
- OMOP – Public health & R/D community
- systematic analysis of disparate observational databases

→ SPHN must be able to speak with everybody!
Data-analysis: leverage the potential of health-related data

- Data needs to be interoperable, of high quality, and available to researchers in various disciplines (under controlled access)
- Strong capabilities in clinical bioinformatics, computational biology and computational service infrastructure are required
- Big data analyses and machine learning require high-performance IT infrastructures for computing and storage
- Security measures for ICT systems are necessary to protect confidential information from unauthorized use, modification, loss or release
The BioMedIT network

BioMedIT provides researchers with access to a secure and protected computing environment for analysis of sensitive data without compromising data privacy.

*DCC: Data Coordination Center
The BioMedIT network

Routine data
Omics data
Cohort data
Imaging data

Harmonized data standards (semantic, data formats, metadata)

Regulatory framework
- Consent
- Ethical approval
- Data transfer and use agreement (DTUA)
The BioMedIT network

**Encrypted file**

**Routine data**

**Omics data**

**Cohort data**

**Imaging data**

**Harmonized data standards** (semantic, data formats, metadata)

**Regulatory framework**
- Consent
- Ethical approval
- Data transfer and use agreement (DTUA)

**End to end encryption**
The BioMedIT network

- Regulatory framework
  - Consent
  - Ethical approval
  - Data transfer and use agreement (DTUA)

- Remote accessible
  - Single sign on: SWITCH edu-ID

- End to end encryption
- Technical interoperability

- Harmonized data standards
  (semantic, data formats, metadata)

- Routine data
- Omics data
- Cohort data
- Imaging data

- Encrypted file

- One Information Security Policy

- HOSPITALS, UNIVERSITIES, PLATFORMS
The BioMedIT network

Regulatory framework
- Consent
- Ethical approval
- Data transfer and use agreement (DTUA)

Harmonized data standards
(semantic, data formats, metadata)

Routine data
Omics data
Cohort data
Imaging data

End to end encryption

One Information Security Policy

Remoteely accessible
Single sign on: SWITCH edu-ID

Technical interoperability

BioMedIT - isolated project space

(i)CORE Basel
(ii)SIS Zurich

Cutting-edge technology:
- Configurable software stack
- Storage and compute capacity
- Backup and archiving

Encryptd file

Compute Environment

HOSPITALS, UNIVERSITIES, PLATFORMS
Conclusion

Key success factors of SPHN:

- the need for a **common understanding** of the vision and scope
- commitment of all players to **share health data** for the benefit of society (citizens, patients, hospitals, research institutes, etc.)
- a **transdisciplinary approach** between clinicians, researchers, bioinformaticians, machine learning experts, etc.
- **FAIR** data and **nationwide interoperability** of health data
- nationally coordinated and supported **data infrastructures**
- process innovation in research and health care.
Conclusion

- Healthcare and research must go hand in hand in the digital learning healthcare system of tomorrow
- Higher quality of data at the moment of capture saves a lot of money on the back-end
- Biggest challenges: #Scalability #Sustainability #CulturalShift
www.sib.swiss/phi
www.sphn.ch
dcc@sib.swiss
@CrameriKatrin
@SPHN_ch
@PHRT_CH
The health data research setting: Navigating Health Data Ethics
University Hospitals: logical architecture

User groups
Service and user interaction Layer
Data Processing and Storage Layer
Data Source Systems

Hospital internal
Medical Informatics
SPHN
Other external clients

Request Workflow Engine
- Text search tool, data exploration, frontends, links to analysis environment, etc.

Data Processing & Modelling
- Key/Cohort Mgmt, Pseudonymisation, Normalisation, Semantic Interoperability

University Hospital internal Data Lake

- Structured data
- semi-structured data
- unstructured data

- Patient data with all types of consent status, typically identifying data

Data Integration
- Technical quality assurance, referencing, etc.

Data Source Systems

- Clinical Routine Data
- Biosample Data
- Pathology Report Data
- Media Data
- Neonat/KISpi Data

- Lab Data
- Oncology Data
- IntensiveCare Data
- Study/Registry Data
- *omics Data

Standardisation according to SPHN requirements, structuring (NLP)
University Hospitals: logical architecture

Service and user interaction Layer

Data Processing and Storage Layer

User groups

Hospital internal
Medical Informatics
SPHN
Other external clients

Request Workflow Engine
Text search tool, data exploration, frontends, links to analysis environment, etc.

Data Processing & Modelling
Key/Cohort Mgmt, Pseudonymisation, Normalisation, Semantic Interoperability

University Hospital internal Data Lake

Structured data
semi-structured data
unstructured data

Patient data with all types of consent status, typically identifying data

Data Integration
Technical quality assurance, referencing, etc.

Data Source Systems

Clinical Routine Data
Biosample Data
Pathology Report Data
Media Data
Neonat/KISpi Data

Lab Data
Oncology Data
IntensiveCare Data
Study/Registry Data
*omicsData

Data collection / data capture in a more structured and standardized way