



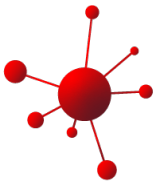
## The Swiss Personalized Health Network (SPHN) Federated Clinical Routine Dataset

### What is it?

- Answer:
  - ✓ The SPHN Federated Clinical Routine Dataset is datasets of the **clinical data collected in routine care** in SPHN format available Swiss healthcare institutions, which is described in the SPHN Metadata Catalog. The SPHN Metadata Catalog is not a direct data repository. More specifically, the catalogue contains metadata (i.e., structured descriptions) about these routinely collected datasets, including:
    - **What data is available** (e.g., diagnoses, laboratory results, procedures, medications, demographics)
    - **How the (meta)data is structured and coded**, the metadata follows the HealthDCAT standard (complying with the European Health Dataspace EHDS). It describes the data itself which uses standardized SPHN concepts and international terminologies (e.g. ICD-10, LOINC and SNOMED CT)
    - **Where the data is generated and stored** (which institution or data source): USZ, USB, INSEL, HUG, CHUV and KISPI
    - **Context and scope of the data**, such as patient populations or clinical domains: clinical routine data of patients which signed the general consent
    - **Data quality and governance information**, including documentation on how the data is managed, standardised and can be shared for research.

### Data access

- Questions:
  - How and who can have access to the data for research? Can non-SPHN affiliated Institutes also get Access?
  - How exactly can this data be requested, and who should we contact? What conditions must be met in order to receive the data?
- Answers:
  - ✓ **The catalogue entry (metadata) of the SPHN Federated Clinical Routine Dataset itself is publicly accessible online and can be explored by anyone** here: <https://fdp.dcc.sib.swiss/catalog/fc6e936d-ad4b-5b21-b8a4-8c7826bbec99>
  - ✓ This catalogue helps you identify what types of data are available as part of SPHN Federated Clinical Routine Dataset (e.g., demographics, diagnoses, procedures, lab results) and where they come from, but **it does not grant direct access to the underlying patient data itself**
  - ✓ Access to the underlying clinical data (beyond metadata) typically requires a formal research request submitted to the relevant data-holding institution (e.g., the university hospital that holds the data)
- Question:
  - How does the data access process look like (proposal? review? costs?) and how long does it take?
- Answer:
  - ✓ The data access follow the governance mechanism of the hospitals. In the catalog you find a contact persons and access conditions.
- Question:
  - Is the SPHN Metadata Catalog aligned with international initiatives?
- Answer:
  - ✓ The metadata catalog follows the HealthDCAT standard the standard for metadata in the European Health Dataspace.
  - ✓ What you can do with it:
    - ✓ The metadata catalog follows the HealthDCAT standard the standard for metadata in the European Health Dataspace.
    - ✓ What you can do with it:



- **Explore administrative metadata information:** temporal coverage of the data, contact point, data sources, data providers, inclusion & exclusion criteria, data use restriction...
- **Find detailed information on data content:** which SPHN 'concepts' are available in the dataset. An SPHN 'concept' is a precisely defined data item, with a clear meaning, format, and structure, often linked to international medical terminologies (like SNOMED CT or LOINC)
- **Explore in an interactive way the graph structure** (via Schema Scope)
- **Interact with the metadata catalogue using queries** (more technical)

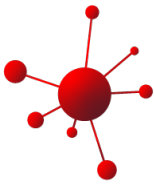
- Question:
  - Can I find out how many patient are female patient have diabetes?
- Answer:
  - ✓ Combination of criteria cannot be explored in the Metadata Catalog but will be possible in the SPHN Data Explorer
    - **An SPHN data explorer is under development** (giving total number of records, variable distribution and summary statistics)

### Type of data

- Question:
  - Is there any omics data available (e.g. Olink Protein panels, Metabolomics)?
- Answer:
  - ✓ No, currently the FedData only includes clinical routine data no omics data.
- Questions:
  - Is any **bio-modelling** aspects (like SBML) involved?
  - Are any **physiological signals**, such as ECG or EEG, available in the dataset?
- Answer:
  - ✓ **No**, there is no bio-modelling aspects involved.
  - ✓ The SPHN FedData includes **749,254 instances of Electrocardiographic Procedure**, availability of the different concepts can be explored here <https://fdp.dcc.sib.swiss/concept-availability/46db9d02-4441-5236-882d-9acd5296671f>.
- Questions:
  - Do they contain aspects related to **psychiatry or mental health**?
  - Does the federated dataset include psychiatric variables? If yes, does that include conditions such as major depression or dementia? Would it be possible to distinguish mid-life and late-life depression using age and longitudinal follow-up data?
  - Are psychiatric diagnoses (depression, dementia etc) or psychiatric medications (antidepressants, mood stabilisers) available and medication prescriptions timelines are accessible?
- Answer:
  - ✓ For the ICD-10 chapter "**Mental, Behavioral and Neurodevelopmental disorders F01-F99** », 433 different codes, with 1,465,571 instances are recorded in the FedData. For Medication N06A Antidepressants we have 26 different codes, 74249 instances. We will have all individual code level counts on the SPARQL endpoint soon, so that code-level queries can be done by the user itself.

### Data coding

- Questions:
  - How can we ensure that our own **documentation systems (register) are aligned with SPHN framework** and variable coding?
  - Clinical data are coded at the institutions (SNOMED; LOINC; ICD-10) is this done in the hospital's routine?
  - That is my question, too. Do the hospitals put the codes for each dataset, or do you SPHN have a way to put the code based on the dataset?
  - Do hospitals use the same units with a code such as LOINC?



- Answer:
  - ✓ The coding and structuring of the FedData is done according to the [SPHN Schema](#) and its recommended standards. The coding of the data, may happen on different levels,
    - 1. At the source (e.g. in the labs)
    - 2. At the level of the clinical data platform
    - 3. As part of the SPHN data generation pipeline.
  - ✓ While coding at the source, is the most accurate and sustainable it is not always possible, therefore level 2 and 3 are done. The units used for a LOINC code differ, unfortunately even within data from one hospital. Please note that LOINC codes may differ depending on the unit, e.g. molar mass versus weight mass.

### Joining / contributing to SPHN database

- Questions:
  - The expansion of the SPHN network to include five cantonal hospitals and the SCI has recently been announced (LUKS, KSA, KSB, HOCH, EOC). Are there already plans for the next round of expansion to include additional cantonal hospitals? What are the requirements for potential candidates?
  - Do you also plan onboard the non-university Hospitals into the Federated Routine Data Set Collection?
- Answer:
  - ✓ This is a very good question. In theory yes, but it will definitely take some time. We need to really work concretely on how to do it and how to find the resources for the hospitals to standardise all their data.

### SPHN tools and services in research

- Question:
  - How to ensure the biggest impact with HRO?
- Answer:
  - ✓ The biggest impact comes with the use of the dataset. Therefore we need to work on the structuring, medical coding of high quality data. On the other hand we need clear and aligned governance process for the secondary use of clinical routine data from hospitals.
- Question
  - How do you control for systematic bias across hospitals in the SPHN dataset? Which methods do you consider most robust for causal inference within SPHN? Are there examples where SPHN-based studies have already influenced clinical practice?
- Answer:
  - ✓ There are many biases when it comes to routine data from different hospitals (due to different clinical information system, different coding practices but also difference clinical processes), therefore researchers need to carefully evaluate this.
- Question
  - What will be up to you the future of SPHN, in particular with regard to DigiSanté? Also, a few years ago there was some talking about HospFair, will something similar be considered in the near future (easy access to routine data without the burden and costs of regular projects)?
- Answer:
  - ✓ HospFair was one of the programs we had with SPHN to really push this coding of LOINC and SNOMED CT. DigiSanté mostly focusing now on getting these standards in the healthcare itself, at the source. Research is the secondary use of it, both activities have to go hand in hand, many learnings of SPHN are now guiding DigiSanté. We are working close together to get these standards integrating at the source.

### Other questions

- Question
  - How often is the overall dataset regenerated/updated?
- Answer:
  - ✓ Under discussion with hospitals. It may not be the same for all hospitals (ideally at least monthly, but at the start maybe only twice a year)



- Question
  - How long it took to have this initial clinical routine dataset? With all the processes associated to it
- Answer:
  - ✓ It's really hard to say, everything we did in the 8-years of SPHN more or less contributed to it. The project and preparation however now took less than 1 year.
- Question
  - Thanks for having this presentation. I understood the ontologies that are being used. But I didn't understand what the data model for the data layer structure is. Are any open data standards being used or was a custom internal data structure built for this purpose? If the later, what is the reason to not use open data model standards?
- Answer:
  - ✓ We built everything on the open standards of W3C. Actually, the content (i.e. the concepts), which build the [SPHN Schema](#) was consolidated with the working group. The model was built on taking into account how the data structured and available in the hospitals.
- Question:
  - Since the system in each hospital has potentially seen an individual patient more than once during the 7 or so years, and the de-identification is done by your pipeline, are you keeping any information to allow linking observations to the same individual? Even if not now, in the future?
- Answer:
  - ✓ Which data belongs to the same patient is part of the hospital systems. The SPHN Connector can do de-identification and in this case would also store the mapping table, this information however is kept in the hospital, SPHN has no access to the installation of the SPHN Connector in the hospital neither the mapping tables. Most of the hospitals do the deidentification prior to using the SPHN Tooling. The hospitals could go back to the identifying information, not SPHN.